Hierarchical Multimodal Deep Learning for Fuel Model Classification Using LiDAR and Sentinel-2*

Andreu Bonet-Pavia[1], Fernando Martínez-Plumed[1], Pablo Crespo-Peremarch[2], Cèsar Ferri[1], and Luis Ángel Ruiz[2]

[1] Valencian Research Institute for Artificial Intelligence (VRAIN), Valencia, Spain
abonpav@etsinf.upv.es
{fmartinez,cferri}@dsic.upv.es
[2] Geo-Environmental Cartography and Remote Sensing Group (CGAT), Valencia, Spain
{pabcrepe,laruiz}@cgf.upv.es

Abstract. Accurate regional-scale fuel type mapping is a cornerstone of proactive wildfire management and ecosystem preservation, especially within highly heterogeneous Mediterranean landscapes. Traditional workflows often rely on labor-intensive manual photointerpretation or tabular machine learning pipelines that depend heavily on handcrafted features. This paper introduces an end-to-end multimodal deep learning framework for regional fuel model classification across the Valencian Community (Spain), using the Scott and Burgan taxonomic system. The proposed framework natively fuses unstructured, low-point-density Airborne Laser Scanning (ALS) point clouds with multi-temporal Sentinel-2 spectral sequences. To exploit the nested structure of vegetation classifications, we introduce FuelTree: a hierarchical neural network head that factorizes joint probabilities into coarse superclasses and conditional fine-grained fuel models, guaranteeing taxonomic consistency. We benchmark several 3D spatial encoders (PointNet, PointNet++, DGCNN, Point Transformer) and temporal modules (GRU, LSTM) against state-of-the-art tabular Gradient Boosting Decision Tree (GBDT) ensembles (XGBoost, LightGBM, CatBoost). Our results demonstrate that deep learning models consistently outperform tabular baselines; the top-performing DGCNN-GRU model achieved a fine-grained macro-F_1 score of 0.776, outperforming the best GBDT baseline (XGBoost, macro-F_1 of 0.747). These findings highlight the advantages of end-to-end representation learning for capturing high-dimensional interactions between 3D canopy structures and seasonal phenology.

Keywords: Wildfire Management · Multimodal Fusion · Deep Learning · Airborne Laser Scanning (ALS) · Sentinel-2 Time Series · Hierarchical Classification.

---

* Work done as a student researcher at the ENIA-UPV Chair
Code: https://github.com/andreucs

1 Introduction

Accurate and up-to-date fuel type mapping is vital for effective wildfire management [28] and ecosystem planning [29], particularly in the Mediterranean basin where recurrent fires heavily alter forest structure and composition [30]. While climate change accelerates the frequency and severity of extreme wildfires [5], traditional ground surveys and manual photointerpretation remain too costly, slow, and limited in coverage to provide timely updates [6]. Artificial Intelligence (AI) has emerged as a powerful paradigm to extract insights from massive geospatial datasets [35]. However, the explosive growth of heterogeneous Remote Sensing (RS) data—combining multi-temporal satellite imagery and 3D sensor modalities—strains conventional processing approaches, which struggle to fully exploit such high-dimensional informational richness [20].

To ingest these complex RS datasets, traditional workflows often rely on manual feature engineering to transform spatial and spectral data into tabular formats [18,13]. Within this tabular machine learning domain, Gradient Boosting Decision Trees (GBDTs) [15] offer competitive advantages in terms of accuracy and interpretability, yet they remain heavily underutilized in forestry compared to standard Random Forests [3]. Conversely, deep learning models allow for end-to-end representation learning but require careful architectural design. Despite recent advancements in Mediterranean fuel mapping [9,13], a critical research gap persists in efficiently fusing medium-resolution multispectral time series with low-point-density Airborne Laser Scanning (ALS) data for regional-scale modeling under complex taxonomic structures.

To address these limitations, this paper proposes a multimodal framework for regional fuel model classification across the Valencian Community. The primary contributions of this work are threefold:

- The integration of multimodal deep learning techniques to fuse low-point-density ALS data with Sentinel-2 multi-temporal spectral sequences.
- The design of a hierarchical neural network architecture that explicitly exploits the nested taxonomic structure of the Scott and Burgan fuel classification system.
- A comprehensive comparative study benchmarking state-of-the-art tabular GBDT ensembles against deep learning models, evaluating the trade-offs between manual feature engineering and end-to-end representation learning.

This paper is structured as follows. Section 2 details the study area, ground truth reference, and the preprocessing workflows for both ALS and Sentinel-2 datasets. Section 3 outlines the spatial sampling pipeline and describes the tabular ensembles and multi-modal deep learning architectures evaluated. Section 4 presents and discusses the experimental results and benchmarking trade-offs. Finally, Section 5 draws conclusions and outlines avenues for future work.

Financiado por la Unión Europea NextGenerationEU

2 Study Area and Data

2.1 Study Area

The study was conducted in the Valencian Community, located on the eastern Iberian Peninsula along the Mediterranean coast. The region comprises coastal plains, agricultural mosaics, shrublands, and mountain systems, resulting in highly heterogeneous vegetation structures and fuel arrangements. Sampling was performed over spatially distributed ALS tiles covering the territory; tile selection was specifically designed to preserve representation across burnable fuel classes rather than to sample a single contiguous block.

2.2 Ground Truth Reference

The reference product is the Fuel Model Map of the Valencian Community, generated at 10×10 m spatial resolution following the 2018 modernization of regional forest fuel cartography. The map contains 18 local classes: 14 forest or burnable fuel categories and four non-burnable categories associated with urban areas, agriculture, water, and bare ground. In accordance with the Scott and Burgan Standard Fire Behavior Fuel Models (SFBFMs) [40], this work focuses on $m = 13$ of these models, explicitly excluding the non-burnable classes and the `SB3` class due to its scarcity (see Figure 1). These labels are grouped into coarser superclasses $\mathcal{C} = \{\mathrm{GR}, \mathrm{SH}, \mathrm{TU}\}$ using the first two characters of the fuel code.

Table 1: Taxonomic descriptions of the Scott and Burgan Standard Fire Behavior Fuel Models (SFBFMs) evaluated in this study.

Fuel Code	**Description**
`GR2`	Low Load, Dry Climate Grass
`GR4`	Moderate Load, Dry Climate Grass
`GR7`	High Load, Dry Climate Grass
`GR8`	High Load, Very Coarse, Humid Climate Grass
`SH1`	Low Load, Dry Climate Shrub
`SH3`	Moderate Load, Humid Climate Shrub
`SH4`	Low Load, Humid Climate Timber–Shrub
`SH5`	High Load, Dry Climate Shrub
`SH9`	Very High Load, Humid Climate Shrub
`TU1`	Low Load, Dry Climate Timber–Grass–Shrub
`TU2`	Moderate Load, Humid Climate Timber–Shrub
`TU3`	Moderate Load, Humid Climate Timber–Grass–Shrub
`TU5`	Very High Load, Dry Climate Timber–Shrub

[3] Photographic Key Source

Financiado por la Unión Europea NextGenerationEU · Gobierno de España · Ministerio de Asuntos Económicos y Transformación Digital · ENIA · España | digital 2026 · Cátedra ENIA UPV · UPV VRAIN · nunsys GROUP

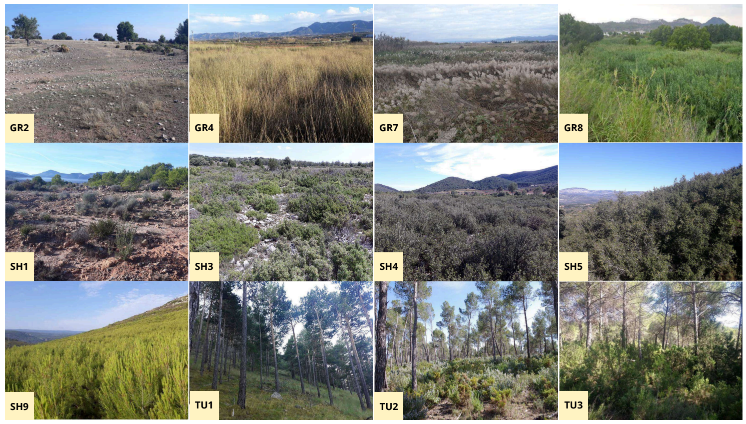

Fig. 1: Photographic Key of selected fuel models across the study area (detailed taxonomic descriptions are provided in Table 1). Note: A photographic reference for model `TU5` is unavailable.[3]

2.3 Airborne Laser Scanning (ALS) Data and Processing

ALS data were obtained from the Spanish National Plan for Aerial Orthophotography (PNOA) between 2015 and 2021, with a point density ranging from 0.5 to 1 point $\cdot$ m^{-2}. Point clouds were preprocessed using PDAL [11]. The pipeline retained ASPRS classes 1–5, removed statistical outliers using $k = 8$ nearest neighbors and a distance multiplier of 3.0 [38], estimated height above ground via Delaunay interpolation [12], and filtered points to retain normalized heights $Z \in [-2, 50]$ m. Negative heights were subsequently capped at zero. Since visual inspection indicated that certain flight lines contained horizontal or vertical offsets, overlapping returns were removed prior to sampling. For each sampled location, a square ALS crop with a side length of 12 m was extracted, corresponding to the 10 m reference pixel plus a 1 m buffer on each side.

The tabular ALS representation comprises 84 predictors, including: point count, height statistics ($z_{\max}$, $\bar{z}$, s_z, skewness, kurtosis, entropy, percentiles $q_p(z)$ for $p \in \{5, 10, \ldots, 95\}$, cumulative height-bin percentages, and the percentage of returns above $\bar{z}$); intensity statistics and percentiles; return-number proportions for returns 1–5; ground-return and vegetation-class proportions; eigenvalue-based shape descriptors [37,45]; the Rumple index [21]; and the vertical complexity index (VCI) [14].

2.4 Sentinel Data and Processing

Sentinel-2 data were extracted via Google Earth Engine for each sampled geometry and organized into annual time series. Each series was smoothed using a Savitzky–Golay filter [39] and interpolated to a regular 15-day grid, yielding $T = 24$ observations per sample. The temporal window was matched to the ALS

Financiado por la Unión Europea NextGenerationEU

acquisition year when imagery was available; otherwise, the closest complete annual sequence was utilized to avoid sparse or incomplete time series.

The deep learning models utilize eight Sentinel-2 bands {`B3`, `B4`, `B5`, `B6`, `B7`, `B8`, `B11`, `B12`} alongside four spectral indices as temporal inputs: the Normalized Difference Vegetation Index (NDVI) [36], the Normalized Difference Moisture Index (NDMI) [16], the Modified Soil Adjusted Vegetation Index 2 (MSAVI2) [33], and the Normalized Burn Ratio Index (NBRI) [23]. Similarly, the tabular models leverage handcrafted metrics derived from the multi-temporal sequences of these same four spectral indices. For each index sequence $x = (x_1, \ldots, x_T)$, with $T = 24$ and $\Delta t = 15$ days, the handcrafted metrics include the mean, median, standard deviation, interquartile range, minimum, maximum, time of maximum encoded as $(\sin(2\pi t^*/T), \cos(2\pi t^*/T))$, and the area under the curve $\sum_t x_t \Delta t$ approximated by trapezoidal integration.

3 Methods

3.1 Spatial Sampling

The spatial sampling pipeline initializes by clipping the 10 m reference fuel model raster to the boundaries of the PNOA ALS tiles. Candidate tiles are filtered according to three criteria: a minimum spatial dimension of 50×50 pixels, zero unclassified (*nodata*) pixels, and a minimum burnable-pixel fraction (ρ_b) defined as:

$$\rho_b = \frac{n_b}{n_b + n_{nb}} \geq 0.2,$$

where n_b and n_{nb} denote the respective counts of burnable and non-burnable pixels within the tile. To mitigate edge-effect label noise near taxonomic boundaries, a 5×5 morphological kernel is applied to the raster. A pixel p is considered pure, and thus eligible for sampling, if and only if it satisfies:

$$\min_{u \in \mathcal{N}_5(p)} y_u = \max_{u \in \mathcal{N}_5(p)} y_u = y_p,$$

provided that the local label y_p belongs to a valid burnable fuel class. To ensure optimal regional representation, a genetic algorithm (GA) [19] is employed to select a subset of $k = 80$ raster tiles from the filtered candidates. For any candidate solution S yielding a global class histogram $h_c(S) = \sum_{i \in S} n_{ic}$, the fitness function is formulated as the harmonic mean:

$$F(S) = \frac{m}{\sum_{c \in \mathcal{Y}} (h_c(S) + \epsilon)^{-1}},$$

where m represents the number of target classes and ϵ is a small smoothing parameter (see Figure 2). The GA optimization process, including hyperparameter tuning, was executed via Optuna [2] using 120 trials across four distinct random seeds, subject to a 60-second execution deadline per run and a median pruner strategy.

Financiado por la Unión Europea NextGenerationEU · Gobierno de España · ENIA · España digital 2026 · Cátedra ENIA UPV · VRAIN · nunsys GROUP

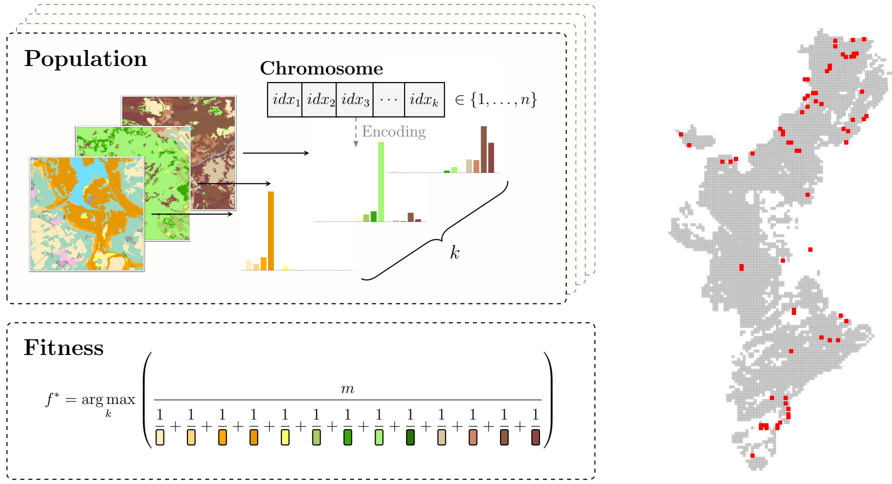

Fig. 2: Overview of the tile selection mechanism using a genetic algorithm (GA) to optimize class balance. The fitness function evaluates candidate chromosome populations to maximize the harmonic mean of fuel model representations, ensuring an even regional distribution of the $k = 80$ selected tiles across the Valencian Community (red markers).

Within the optimization-selected tiles, final sampling locations are drawn exclusively from the validated pure burnable pixels. To prevent spatial autocorrelation and oversampling, a class-wise greedy Poisson-disk [4] approximation enforces a minimum planar separation threshold of 25 m. This spatial constraint is computationally optimized utilizing a KD-tree structure [27] to accelerate nearest-neighbor distance queries. Following a random shuffling of candidate coordinates, any point within a pair falling below this Euclidean threshold is discarded. Each retained coordinate uniquely defines the center of an ALS spatial crop and establishes a corresponding Sentinel-2 plot identifier (see Figure 3).

3.2 Ensembles

The tabular machine learning baseline comprises four state-of-the-art tree-based architectures: Random Forest [3], XGBoost [8], LightGBM [22], and CatBoost [31]. Each classifier ingests a concatenated flat feature vector:

$$\mathbf{x}_i = [\phi_{\mathrm{ALS}}(P_i), \phi_{\mathrm{S2}}(T_i)],$$

where $\phi_{\mathrm{ALS}}(P_i)$ represents the 84 handcrafted statistical descriptors extracted from the cropped point cloud P_i, and $\phi_{\mathrm{S2}}(T_i)$ represents the 40 time-series metrics derived from the Sentinel-2 sequence T_i. Hyperparameter tuning was auto-

Financiado por la Unión Europea NextGenerationEU · España | digital · ENIA UPV · VRAIN · nunsys GROUP

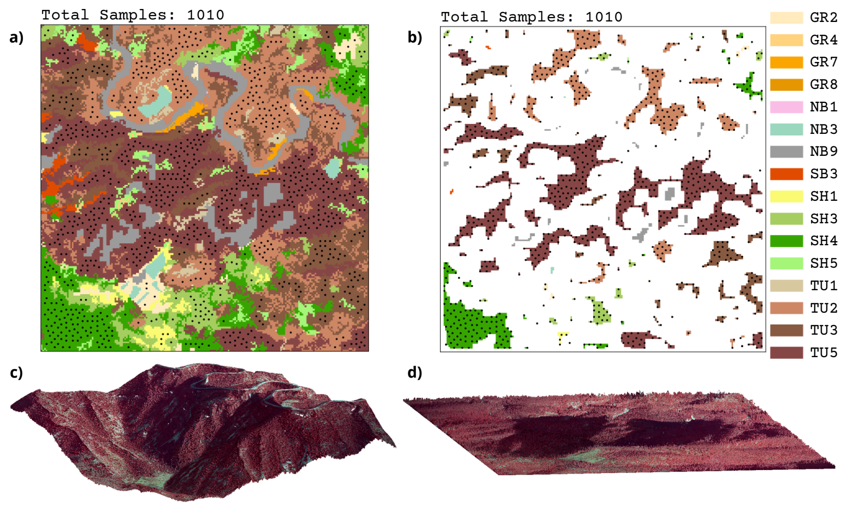

Fig. 3: Visual representation of the spatial sampling. (a) Distribution of the final sampled coordinates (black dots) overlaid on the candidate pure pixels of the reference fuel map. The text annotation "Total Samples: 1010" denotes the local sample density contained exclusively within this specific candidate tile. (b) Inspection of the same raster tile showcasing the morphological kernel constraint, illustrating that samples are restricted strictly to validated pure pixel interiors. (c–d) CloudCompare 3D visualizations of the identical ALS tile, contrasting (c) the raw point cloud return against (d) the fully preprocessed point cloud after outlier removal, height normalization, and flight-line offset correction.

mated via Optuna, utilizing a Tree-structured Parzen Estimator (TPE) [2] sampler paired with a median pruner to optimize the macro-F_1 score over a 5-fold stratified cross-validation scheme. Following optimization, the top-performing hyperparameter configuration for each model was utilized to retrain the architecture on the complete training partition prior to final validation on the held-out test set.

3.3 Deep Learning

The deep learning models process raw, unstructured point clouds and multi-spectral time series natively, omitting manual feature engineering. Each LiDAR return is represented as a five-dimensional vector (x, y, z, i, r), where the spatial coordinates are centered and scaled relative to a bounding unit sphere, and the intensity (i) and return number (r) are normalized to the range $[0, 1]$. Farthest Point Sampling (FPS) [7] downsamples each point cloud to a fixed size of $N = 128$ points; for crops falling below this density, points are randomly duplicated to maintain tensor consistency. Data augmentation during training includes random

Financiado por la Unión Europea NextGenerationEU · España digital · Cátedra ENIA UPV · VRAIN · nunsys GROUP

rotation around the vertical axis, scaling within the interval $[0.95, 1.05]$, Gaussian jitter ($\sigma = 0.005$), and random translation noise ($\sigma = 0.01$) [7].

The spatial encoding module f_P is evaluated across four distinct structural paradigms: PointNet [7], PointNet++ [32], DGCNN [44], and Point Transformer [50]. These architectures introduce contrasting inductive biases, ranging from global symmetric aggregation and hierarchical local grouping to dynamic graph neighborhoods and self-attention mechanisms [43]. Concurrently, the temporal encoder f_T—implemented as either a Gated Recurrent Unit (GRU) [10] or a Long Short-Term Memory (LSTM) network [17]—processes the Sentinel-2 sequence $T_i \in \mathbb{R}^{24 \times d_T}$. Multi-modal fusion is achieved through an additive combination after projecting both modalities into a unified latent space:

$$\mathbf{z}_i = f_P(P_i) + f_T(T_i).$$

To leverage the taxonomic structure of the Scott and Burgan classification system, a hierarchical classification head is integrated into the architecture (see Figure 4). This component adapts the foundational WordTree concept introduced in YOLO9000 [34] to implement our specialized FuelTree framework. Structurally, the classification heads incorporate a dropout [41] layer with a rate of 0.3 for regularization, followed by LeakyReLU [47] activation functions. Let $c \in \mathcal{C}$ denote a coarse fuel category (superclass) and $y \in \mathcal{Y}_c$ represent a fine-grained target class nested within that category. The network outputs concurrent logits for both taxonomic tiers. The joint probability of a fine-grained class is thus decomposed conditionally:

$$p(y \mid \mathbf{z}) = p(c(y) \mid \mathbf{z})\, p(y \mid c(y), \mathbf{z}),$$

where structurally invalid fine classes (i.e., those outside the active superclass $\mathcal{Y}_c$) are masked out dynamically by setting their logit values to $-\infty$. The joint network is optimized via a hierarchical negative log-likelihood loss:

$$\mathcal{L}_{\text{tree}} = -\log p(c_i \mid \mathbf{z}_i) - \log p(y_i \mid c_i, \mathbf{z}_i),$$

augmented with label smoothing [42]. Conversely, the baseline configuration utilizes a standard flat classification head that predicts fine-grained classes directly.

3.4 Experimental Setup

To ensure robust evaluation, a spatially aware data split grouped by individual LiDAR tile was enforced on the total dataset ($N = 78,078$): 15% of the total sample pool (12,792 samples) was reserved exclusively for the independent test partition (see Table 2). The remaining pool of 65,286 samples was partitioned such that 10% formed the validation set for neural network early stopping, while the remaining 90% was used for model training. Model performance is reported using overall accuracy and macro-F_1 scores at the fine and coarse levels.

Neural architectures were optimized using the AdamW [26] solver with an initial learning rate of 3×10^{-4}, a weight decay coefficient of 10^{-4}, and a cosine

Financiado por la Unión Europea NextGenerationEU · España digital 2026 · Cátedra ENIA UPV · VRAIN · nunsys GROUP

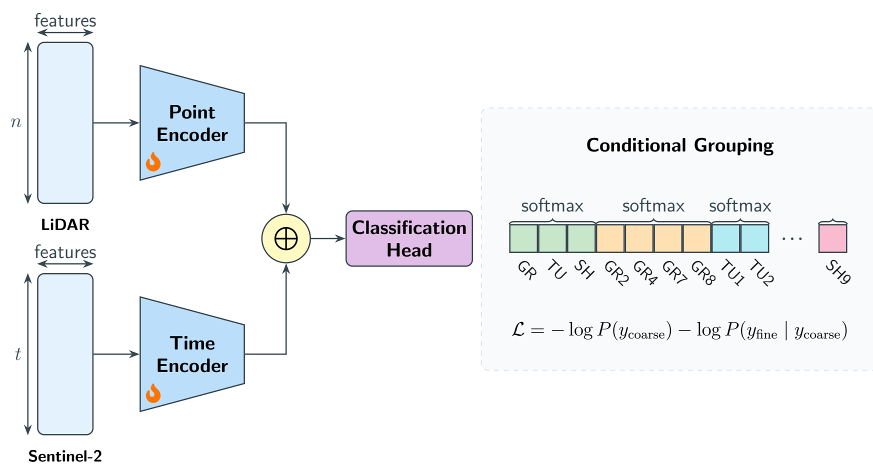

Fig. 4: Architecture of the multi-modal deep learning framework featuring conditional taxonomic grouping. Raw LiDAR crops and Sentinel-2 time series are processed by independent spatial and temporal encoders. Their latent embeddings are fused additively ($\oplus$) and passed to a hierarchical classification head that factorizes probabilities into coarse superclasses ($\mathcal{C}$) and conditional fine-grained fuel models.

learning rate scheduler [25] decaying to a minimum learning rate of 10^{-6}. Models were trained for up to 80 epochs with early stopping using a patience threshold of 15 epochs, a batch size of 256, and label smoothing [42] with a factor of 0.05. All deep learning experiments were executed using an NVIDIA RTX A5000 GPU and an AMD Ryzen 9 9950X 16-core processor.

4 Results and Discussion

4.1 Deep Learning Performance and Architectural Trade-offs

The quantitative evaluation of the multimodal deep learning models (Table 3) reveals clear trends regarding spatial and temporal encoding paradigms. Models equipped with a Gated Recurrent Unit (GRU) consistently outperform their Long Short-Term Memory (LSTM) counterparts across almost all spatial backbones, while maintaining a lower parameter footprint. Among the spatial encoders, the Dynamic Graph Convolutional Neural Network (`DGCNN`) paired with a GRU achieved the highest overall performance, yielding a fine-grained F_1-macro score of 0.776 and a coarse-grained accuracy of 0.913. PointNet++ (`PN2`) demonstrated competitive robustness, marking it as a strong candidate for hierarchical feature extraction. Conversely, the Point Transformer (`PT`) underperformed despite its significantly larger parameter capacity (4.4M parameters).

Financiado por la Unión Europea NextGenerationEU

Table 2: Summary of sample distributions across the fine-grained Scott and Burgan fuel models and coarse superclasses for the training and independent test partitions.

Superclass (%)	**Codes**	**Train**	**Test**	**Total**
Grasslands (GR) (18.70%)	`GR2`	8,149	1,244	9,393
	`GR4`	1,574	295	1,869
	`GR7`	1,992	293	2,285
	`GR8`	946	111	1,057
Shrublands (SH) (48.33%)	`SH1`	2,696	1,201	3,897
	`SH3`	6,454	744	7,198
	`SH4`	9,312	1,858	11,170
	`SH5`	5,987	448	6,435
	`SH9`	7,144	1,894	9,038
Timber-Understory (TU) (32.96%)	`TU1`	1,707	264	1,971
	`TU2`	6,741	2,270	9,011
	`TU3`	8,253	1,500	9,753
	`TU5`	4,331	670	5,001
Total	—	**65,286**	**12,792**	**78,078**

This indicates that heavy self-attention mechanisms may suffer from overfitting when digesting low-point-density ALS data over complex taxonomic structures.

Table 3: Performance metrics of multimodal deep learning architectures. Results are reported as (Fine / Coarse) taxonomic levels. The asterisk (*) denotes the flat baseline model trained with a standard multiclass softmax head.

	LSTM [17]	GRU [10]	**Params**	F_1**-macro** ↑	**Acc.** ↑
PN [7]	✓		900K	0.754/0.884	0.821/0.906
PN2 [32]	✓		1M	0.758/0.889	0.825/0.908
DGCNN [44]	✓		900K	0.759/0.887	0.825/0.905
PT [50]	✓		4.4M	0.755/0.886	0.818/0.903
PN [7]		✓	850K	0.765/0.894	0.831/0.912
PN2 [32]		✓	950K	0.764/0.898	0.835/0.914
DGCNN [44]		✓	850K	0.776/0.894	0.837/0.913
PT [50]		✓	4.4M	0.753/0.877	0.818/0.901
PN* [7]		✓	850K	0.767/0.891	0.830/0.909

Financiado por la Unión Europea NextGenerationEU · Gobierno de España · Ministerio de Asuntos Económicos y Transformación Digital · ENIA · España | digital 2026 · Cátedra ENIA UPV · UPV VRAIN · nunsys GROUP

4.2 Impact of the Hierarchical FuelTree Classification Head

Comparing the hierarchical PointNet configuration (PN + GRU) against the flat baseline (PN* + GRU) reveals virtually identical performance with minimal variations. Formal statistical testing would be required to establish any actual distinction between the two approaches. Consequently, the value of the FuelTree framework is purely structural rather than empirical: it mathematically guarantees taxonomic consistency by eliminating impossible fine-class predictions.

4.3 Deep Learning vs. Tabular Ensembles

The benchmarking results for the tree-based tabular ensembles are summarized in Table 4. Among these baselines, XGBoost delivered the top performance with a fine-grained F_1-macro of 0.747 and an accuracy of 0.818, closely followed by LightGBM and CatBoost. Random Forest lagged significantly behind across all metrics.

When contrasted with the deep learning framework, even the simplest neural (PN + LSTM) matches or exceeds the best-performing GBDT ensemble (XGBoost). This gap underscores the superiority of end-to-end representation learning, which successfully captures high-dimensional multi-temporal and 3D geometric interactions that manual feature engineering fails to encapsulate.

Table 4: Performance metrics of tabular machine learning ensemble baselines utilizing 124 handcrafted statistics. Results are reported as (Fine / Coarse) taxonomic levels.

Model	**F_1-macro ↑**	**Acc. ↑**
XGBoost [8]	0.747/0.878	0.818/0.878
LightGBM [22]	0.745/0.878	0.818/0.893
CatBoost [31]	0.743/0.875	0.811/0.891
Random Forest [3]	0.728/0.862	0.790/0.879

4.4 Modality Ablation Study

An ablation study was conducted to isolate the individual contributions of 3D structural data and multi-temporal spectral sequences (Table 5). Models restricted exclusively to ALS data exhibit a severe performance bottleneck at the fine-grained level, with F_1-macro scores limited to 0.49–0.51 across all architectures. Although low-density LiDAR (0.5–1 point $\cdot$ m^{-2}) provides sufficient geometric context for coarse ecosystem classification (up to 0.835 accuracy), it lacks the taxonomic resolution required to separate closely related fuel models.

In contrast, Sentinel-2 spectral-only models yield a substantial performance leap. The deep learning temporal encoders (`GRU` and `LSTM`) achieve a fine F_1-macro of 0.709, outperforming the tree-based baselines. Ultimately, fusing both modalities triggers a strong cross-modal synergy. Integrating structural LiDAR as a geometric refiner elevates the top-performing paradigm (`DGCNN` + `GRU`) to its peak fine F_1-macro of 0.776, proving that robust regional-scale mapping requires joint, end-to-end representation learning.

Table 5: Ablation study evaluating the impact of data modalities across different architectural paradigms. Results are reported as (Fine / Coarse) taxonomic levels.

	LiDAR	**Sentinel-2**	**Params**	**F_1-macro ↑**	**Acc. ↑**
PN2 [32]	✓		700K	0.494/0.798	0.613/0.835
DGCNN [44]	✓		600K	0.494/0.784	0.618/0.827
GRU [10]		✓	650K	0.709/0.859	0.789/0.876
LSTM [17]		✓	950K	0.709/0.857	0.783/0.875
Random Forest [3]	✓		-	0.489/0.769	0.620/0.816
		✓	-	0.641/0.800	0.703/0.800
LightGBM [22]	✓		-	0.518/0.771	0.648/0.823
		✓	-	0.648/0.818	0.727/0.821

5 Conclusions and Future Work

This paper presented a multimodal deep learning framework for regional-scale fuel model classification across the Valencian Community, successfully fusing low-point-density ALS data with Sentinel-2 multi-temporal spectral sequences. The experimental benchmarking demonstrates that deep learning architectures consistently outperform state-of-the-art tabular GBDT ensembles such as XGBoost and LightGBM. This outcome underscores the advantage of end-to-end

Financiado por la Unión Europea NextGenerationEU · España | digital 2026 · Cátedra ENIA UPV · VRAIN · nunsys GROUP

representation learning over traditional handcrafted feature engineering, as the neural networks effectively capture high-dimensional interactions between 3D canopy structures and seasonal phenological variations.

Several promising avenues remain open for future research. First, the framework could integrate pretrained, self-supervised foundation models [46,48,49] to leverage large volumes of unlabeled geospatial data and enhance model adaptability across heterogeneous Mediterranean regions. Within this context, exploring the Joint Embedding Predictive Architecture (JEPA) [1] paradigm represents a viable direction to learn robust, noise-resilient representations of complex ecosystems without relying excessively on hand-engineered data augmentations.

Second, future workflows could transition from isolated point-wise sampling to contiguous, grid-like spatial structures. Utilizing Convolutional Neural Networks (CNNs) to process these continuous regular grids would allow the pipeline to explicitly model spatial context and horizontal topological relationships between neighboring samples. Finally, incorporating probabilistic methods to systematically quantify both epistemic and aleatoric uncertainties [24] is crucial to enhance the framework's operational reliability, providing explicit confidence metrics for risk assessment and planning.

References

1. Assran, M., Duval, Q., Misra, I., Bojanowski, P., Vincent, P., Rabbat, M., LeCun, Y., Ballas, N.: Self-supervised learning from images with a joint-embedding predictive architecture (2023), https://arxiv.org/abs/2301.08243
2. Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B.: Algorithms for hyper-parameter optimization. In: Advances in Neural Information Processing Systems. vol. 24, pp. 2546–2554. Curran Associates, Inc. (2011)
3. Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001)
4. Bridson, R.: Fast poisson disk sampling in arbitrary dimensions. In: ACM SIGGRAPH 2007 Sketches. p. 22–es. SIGGRAPH '07, Association for Computing Machinery, New York, NY, USA (2007). https://doi.org/10.1145/1278780.1278807
5. Camps-Valls, G., Fernández-Torres, M., Cohrs, K.H., Höhl, A., Castelletti, A., Pacal, A., Robin, C., Martinuzzi, F., et al.: Artificial intelligence for modeling and understanding extreme weather and climate events. Nature Communications **16**(1), 1919 (2025)
6. Carbone, A., Spiller, D., Laneve, G.: Fuel type mapping using a cnn-based remote sensing approach: A case study in sardinia. Fire **6**(10) (2023)
7. Charles, R.Q., Su, H., Kaichun, M., Guibas, L.J.: PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 77–85. IEEE (2017)
8. Chen, T., Guestrin, C.: XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794. KDD '16, Association for Computing Machinery (2016)
9. Chrysafis, I., Damianidis, C., Giannakopoulos, V., Mitsopoulos, I., Dokas, I.M., Mallinis, G.: Vegetation Fuel Mapping at Regional Scale Using Sentinel-1, Sentinel-2, and DEM Derivatives—The Case of the Region of East Macedonia and Thrace, Greece. Remote Sensing **15**(4) (2023)

Financiado por la Unión Europea NextGenerationEU

10. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling (2014), https://arxiv.org/abs/1412.3555
11. Contributors, P.: Pdal point data abstraction library (Aug 2025). https://doi.org/10.5281/zenodo.10884408, https://doi.org/10.5281/zenodo.10884408
12. Delaunay, B.: Sur la sphère vide. Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et naturelles **1934**(6), 793–800 (1934)
13. Domingo, D., de la Riva, J., Lamelas, M.T., García-Martín, A., Ibarra, P., Echeverría, M., Hoffrén, R.: Fuel Type Classification Using Airborne Laser Scanning and Sentinel 2 Data in Mediterranean Forest Affected by Wildfires. Remote Sensing **12**(21) (2020)
14. van Ewijk, K.Y., Treitz, P.M., Scott, N.A.: Characterizing forest succession in Central Ontario using LiDAR-derived indices. Photogrammetric Engineering & Remote Sensing **77**(3), 261–269 (2011)
15. Friedman, J.H.: Greedy function approximation: A gradient boosting machine. The Annals of Statistics **29**(5), 1189 – 1232 (2001)
16. Gao, B.C.: Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space. Remote sensing of environment **58**(3), 257–266 (1996)
17. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. Neural Comput. **9**(8), 1735–1780 (1997)
18. Hoffrén, R., Lamelas, M.T., de la Riva, J.: Classification and Mapping of Fuels in Mediterranean Forest Landscapes Using a UAV-LiDAR System and Integration Possibilities with Handheld Mobile Laser Scanner Systems. Remote Sensing **16**(18) (2024)
19. Holland, J.H.: Genetic algorithms. Scientific american **267**(1), 66–73 (1992)
20. Hong, D., Li, C., Li, X., Camps-Valls, G., Chanussot, J.: Foundation Models in Remote Sensing: Evolving from unimodality to multimodality. IEEE Geoscience and Remote Sensing Magazine (2026)
21. Jenness, J.S.: Calculating landscape surface area from digital elevation models. Wildlife Society Bulletin **32**(3), 829–839 (2004)
22. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y.: LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in neural information processing systems **30** (2017)
23. Key, C.H., Benson, N.C.: Landscape assessment: Ground measure of severity, the composite burn index; and remote sensing of severity, the normalized burn ratio. Tech. Rep. RMRS-GTR-164-CD, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO (2006)
24. Kondylatos, S., Bountos, N.I., Prapas, I., Zavras, A., Camps-Valls, G., Papoutsis, I.: Probabilistic machine learning for noisy labels in earth observation. Scientific Reports **15**(1), 35890 (2025)
25. Loshchilov, I., Hutter, F.: SGDR: Stochastic Gradient Descent with Warm Restarts (2017), https://arxiv.org/abs/1608.03983
26. Loshchilov, I., Hutter, F.: Decoupled Weight Decay Regularization. In: International Conference on Learning Representations (2019)
27. Maneewongvatana, S., Mount, D.: Analysis of approximate nearest neighbor searching with clustered point sets, pp. 105–123. American Mathematical Society (12 2002). https://doi.org/10.1090/dimacs/059/06
28. McCaffrey, S.: Thinking of wildfire as a natural hazard. Society and Natural Resources **17**(6), 509–516 (2004)

Financiado por la Unión Europea NextGenerationEU

29. Pausas, J.G., Keeley, J.E.: A burning story: the role of fire in the history of life. BioScience **59**(7), 593–601 (2009)
30. Pausas, J.G., Llovet, J., Rodrigo, A., Vallejo, R.: Are wildfires a disaster in the mediterranean basin?–a review. International Journal of wildland fire **17**(6), 713–723 (2008)
31. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A.: CatBoost: unbiased boosting with categorical features. Advances in neural information processing systems **31** (2018)
32. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017)
33. Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S.: A modified soil adjusted vegetation index. Remote sensing of environment **48**(2), 119–126 (1994)
34. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7263–7271 (2017)
35. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat: Deep learning and process understanding for data-driven earth system science. Nature **566**(7743), 195–204 (2019)
36. Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W.: Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. NASA/GSFC Type III Final Report **371**, 1–130 (1973)
37. Rusu, R.B.: Semantic 3D object maps for everyday manipulation in human living environments. KI-Künstliche Intelligenz **24**(4), 345–348 (2010)
38. Rusu, R.B., Marton, Z.C., Blodow, N., Dolha, M., Beetz, M.: Towards 3D Point Cloud Based Object Maps for Household Environments. Robotics and Autonomous Systems **56**(11), 927–941 (2008), semantic Knowledge in Robotics
39. Savitzky, A., Golay, M.J.: Smoothing and differentiation of data by simplified least squares procedures. Analytical chemistry **36**(8), 1627–1639 (1964)
40. Scott, J.H., Burgan, R.E.: Standard fire behavior fuel models: a comprehensive set for use with Rothermel's surface fire spread model. Tech. Rep. RMRS-GTR-153, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station (2005)
41. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15**(1), 1929–1958 (2014)
42. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2818–2826. IEEE (2016)
43. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention Is All You Need. Advances in neural information processing systems **30** (2017)
44. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic Graph CNN for Learning on Point Clouds. ACM Trans. Graph. **38**(5), 146:1–146:12 (2019)
45. West, K.F., Webb, B.N., Lersch, J.R., Pothier, S., Triscari, J.M., Iverson, A.E.: Context-driven automated target detection in 3D data. In: Automatic Target Recognition XIV. vol. 5426, pp. 133–143. SPIE (2004)
46. Wu, X., DeTone, D., Frost, D., Shen, T., Xie, C., Yang, N., Engel, J., Newcombe, R., Zhao, H., Straub, J.: Sonata: Self-supervised learning of reliable point rep-

resentations. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 22193–22204 (2025)
47. Xu, B., Wang, N., Chen, T., Li, M.: Empirical Evaluation of Rectified Activations in Convolutional Network (2015), https://arxiv.org/abs/1505.00853
48. Zhang, Y., Wu, X., Lao, Y., Wang, C., Tian, Z., Wang, N., Zhao, H.: Concerto: Joint 2D-3D Self-Supervised Learning Emerges Spatial Representations (2026), https://arxiv.org/abs/2510.23607
49. Zhang, Y., Wu, X., Yang, Y., Fan, X., Li, H., Zhang, Y., Huang, Z., Wang, N., Zhao, H.: Utonia: Toward One Encoder for All Point Clouds (2026), https://arxiv.org/abs/2603.03283
50. Zhao, H., Jiang, L., Jia, J., Torr, P., Koltun, V.: Point Transformer. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 16239–16248. IEEE (2021)

A Figures

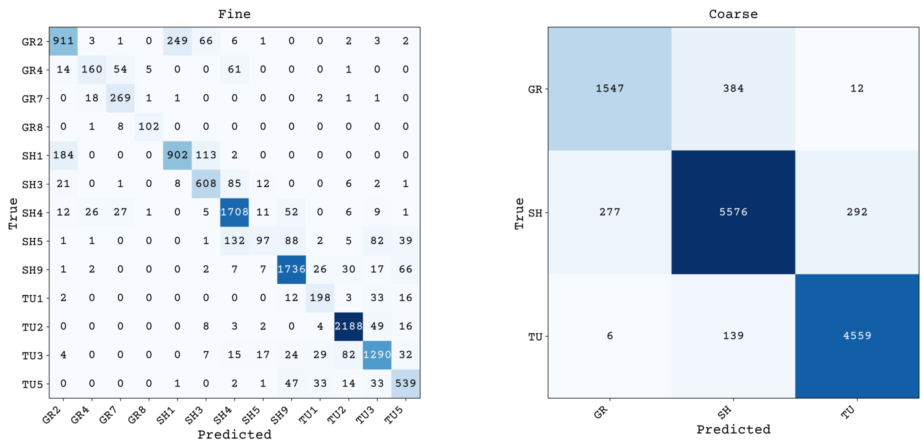

Fig. 5: Confusion matrices for the (DGCNN + GRU) architecture at the fine-grained (left) and coarse-grained (right) taxonomic levels.

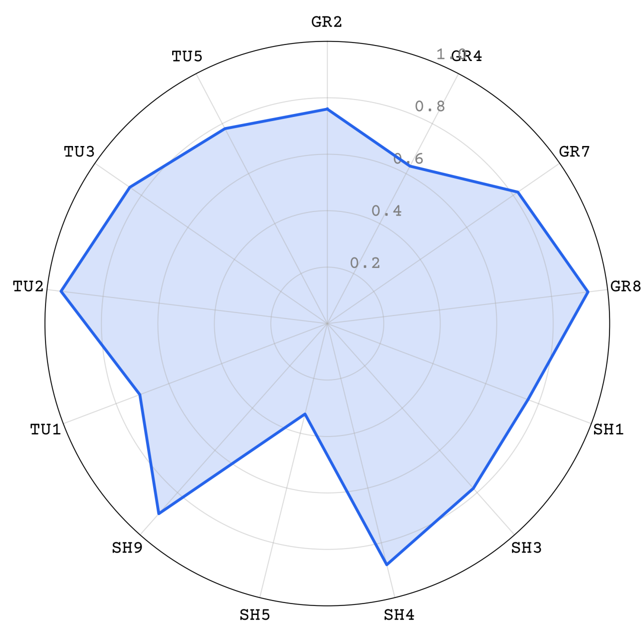

Fig. 6: Per-class F_1-score performance profile for the (DGCNN + GRU) model displayed as a radar chart.

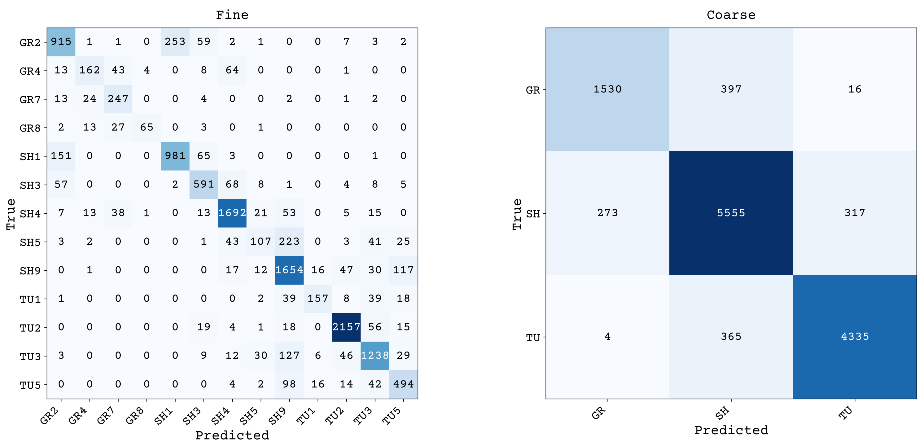

Fig. 7: Confusion matrices for LightGBM at the fine-grained (left) and coarse-grained (right) taxonomic levels

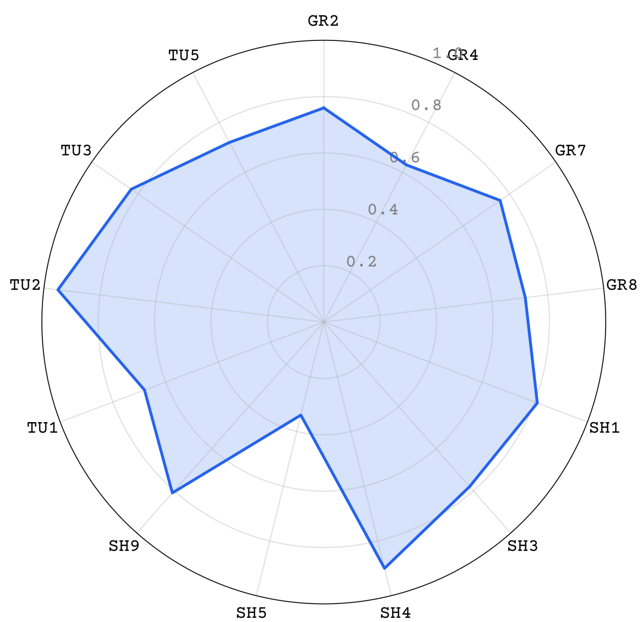

Fig. 8: Per-class F_1-score performance profile for the LightGBM model displayed as a radar chart.

Financiado por la Unión Europea NextGenerationEU