Toward Generalized Multi-Typological Classification of Cultural Heritage: A Random Forest Approach

Giuseppe Antuono [1], Valeria Cera[2], Daniela Ciarlo[3]

[1] Dept. of Civil, Building and Environmental Engineering, University of Naples Federico II, Naples, Italy - giuseppe.antuono@unina.it
[2] Dept. of Architecture, University of Naples Federico II, Naples, Italy - valeria.cera@unina.it
[3] Dept. of Engineering for Innovation, University of Salento, Lecce, Monteroni, Italy - danielaciarlo.dc@gmail.com

Keywords: AI, semantic segmentation, Random Forest, hyperparameter optimization, benchmarking, point cloud classification.

Abstract

The increasing adoption of point clouds in the digital documentation of Cultural Heritage (CH) has made three-dimensional semantic segmentation a key step for data interpretation and analysis. In this context, Deep Learning (DL) approaches have demonstrated high performance, albeit at the cost of substantial computational requirements and the need for large annotated datasets. Within this framework, the present study investigates the potential of leveraging a traditional supervised Machine Learning (ML) approach - Random Forest (RF) - through targeted optimization of training and validation procedures. To this end, the RF_CHC (Random Forest for Cultural Heritage Classification) model is proposed. Aimed at improving accuracy and, in particular, generalization capability in the semantic classification of architectural CH point clouds, RF_CHC integrates statistical hyperparameter calibration through the adoption of cross-validation procedures. The performance of RF_CHC was evaluated and compared with literature models (RF4PCC and optimized RF4PCC), demonstrating improved classification consistency and greater robustness across heterogeneous datasets, while highlighting the potential of an optimized ML-based approach as a competitive or complementary solution to currently prevalent DL models in the CH domain.

1. Introduction

The establishment of point clouds as a privileged form of digital representation of Cultural Heritage (CH) has driven a methodological shift, steering research toward Artificial Intelligence (AI) approaches - particularly Machine Learning (ML) and Deep Learning (DL) techniques - capable of automatically inferring semantic structures and hierarchical relationships from complex architectural datasets (Gîrbacia, 2024; Matrone et al., 2020a; Grilli et al., 2019).

Until the first two decades of the twenty-first century, supervised ML approaches dominated the field of Three-dimensional Point Cloud Semantic Segmentation (3DPCSS), with a prevalent use of the Random Forest (RF) algorithm, which progressively established itself as one of the most widely adopted methods due to its robustness compared to other approaches and its ability to operate effectively even on relatively small datasets.

Over the past 6 years, a progressive consolidation of DL techniques has been observed, driven by their superior performance in the segmentation of multiscale 3D spatial information at different levels of granularity, from the scene to the individual element (Betsas et al., 2025a; Bello et al., 2020; Liu et al., 2019). The challenges that initially accompanied the diffusion of DL - such as the geometric complexity of historical architectural elements, often difficult to standardize, together with the scarcity of large annotated datasets (Terruggi et al., 2020) - now appear increasingly manageable, also thanks to the adoption of synthetic data (Morbidoni et al., 2020; Wang et al., 2019).

Within this scientific context, characterized by an operational positioning centered on DL, the present research is situated. Motivated by a spirit of exploratory curiosity, it aims to investigate whether and how a more traditional ML approach, such as RF, can be optimized for 3DPCSS in order to enhance its competitiveness in light of the capabilities demonstrated by DL. Specifically, in the first phase of the research presented here, the focus is placed on improving RF performance in the classification of CH point clouds in terms of consistency in recognizing architectural classes conventionally acknowledged in the State of the Art (“floor”, “wall”, “column”, “arch”, “vault”, “stair”, “window/door”, “molding”, “roof”, “other”), as well as on its generalization capability, while leveraging the intrinsic advantages of ML approaches (the ability to operate on limited annotated datasets, low computational cost, and a reduced setup phase).

Starting from the consideration that RF can achieve good performance provided that feature selection and definition are carefully calibrated (Buldo et al., 2024), several studies have shown that the integration of new features, together with the adoption of automated feature selection procedures and class balancing analyses, contributes to improving segmentation accuracy (Gil and Arayici, 2025; Moyano et al., 2024).

Despite these advances, a critical challenge remains the limited generalization capability of RF models when applied to CH datasets, with the associated risk of overfitting and excessive dependence on the specific characteristics of the training data. In this regard, careful hyperparameter selection - although representing a fundamental step in balancing model complexity and predictive performance on heterogeneous architectural datasets (Feurer et al., 2015; Bergstra and Bengio, 2012) - has not yet been fully optimized, as the use of cross-validation techniques, which are instrumental for this purpose, is still not systematically established in the literature (Figure 1).

In this direction, the present study introduces the RF_CHC (Random Forest for Cultural Heritage Classification) model, aimed at promoting a more generalized classification of CH point cloud models by integrating statistical hyperparameter calibration through cross-validation and architectural class balancing strategies (Lin and Nguyen, 2020).

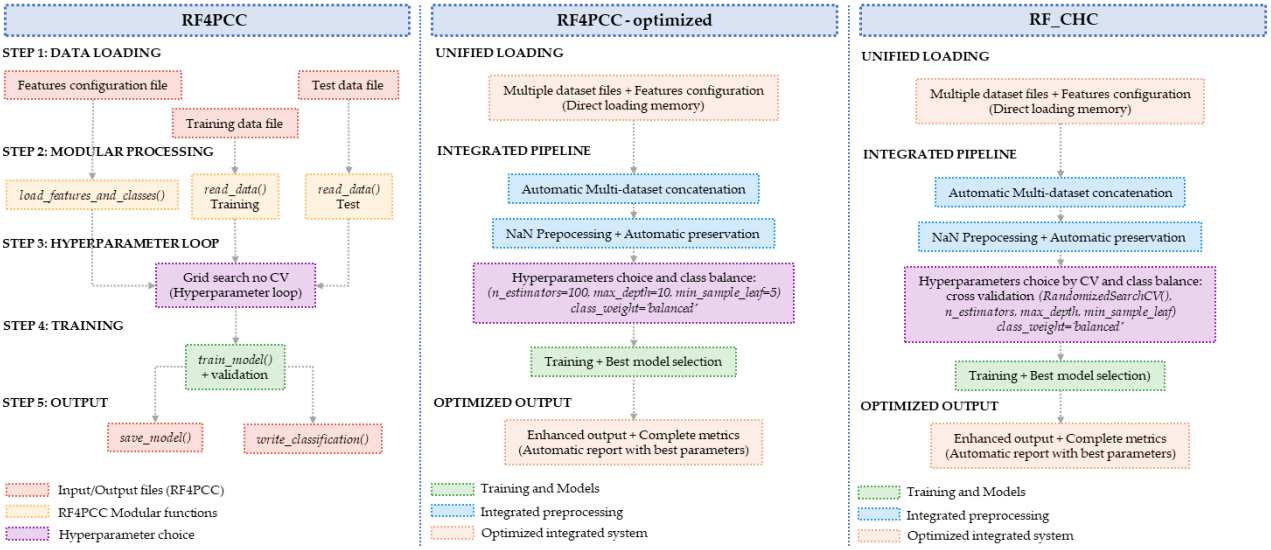


Figure 1. Comparison between RF4PCC model – RF4PCC optimized model – RF_CHC model.

To assess its effectiveness and generalization capability, the new model was compared with the RF4PCC (Grilli and Remondino, 2020) and RF4PCC-optimized (Antuono et al., 2025) models, evaluating key metrics such as accuracy and F1-score. Finally, the obtained results were related to those reported in the literature for the most common DL algorithms currently in use and/or under development, to better understand the potential and future evolution of the RF-based approach.

2. State of art

Studies that have addressed supervised ML approaches for 3DPCSS in the CH domain have highlighted the robustness of RF which, within the formal differences inherent to the ML and DL paradigms (Cao et al., 2022), has been shown to deliver satisfactory performance (Yang et al., 2023).

Although recent advances in the DL domain have progressively steered research toward the increasingly widespread adoption of these techniques, opportunities for further improving RF remain, suggesting the relevance of re-examining this approach with the aim of achieving increasingly robust results while simultaneously benefiting from the intrinsic advantages of ML-based methods.

A central operational issue of RF models applied to CH lies in their limited generalization capability, with the associated risk of overfitting and excessive dependence on the specific characteristics of the training dataset. Indeed, such models tend to over-adapt to the patterns of the case studies used for training, thereby compromising performance when applied to different architectural typologies or to morphologically and stylistically diverse contexts. At the same time, the literature still lacks a truly systematic approach to hyperparameter optimization through cross-validation techniques, which represent a critical element for balancing model complexity and predictive capability across heterogeneous and non-homogeneous datasets. As is well known, hyperparameters are model configuration parameters that are not learned from the data during training but must be set in advance (such as, for example, the number of trees in the forest or the maximum tree depth). Their proper calibration has a significant impact on classifier performance. Cross-validation is a validation technique that iteratively partitions the data into training and validation subsets, allowing model robustness to be assessed across different dataset splits and reducing the variance of performance estimates. Among hyperparameter optimization methods based on cross-validation, *GridSearchCV* and *RandomizedSearchCV* represent the most established and widely adopted approaches (Bischl et al., 2023). In the present study, these two methods were compared in order to identify the most appropriate strategy. *GridSearchCV* systematically explores all possible combinations of hyperparameters within a predefined discrete search space, ensuring identification of the optimal solution but with exponentially increasing computational complexity. *RandomizedSearchCV* randomly samples a predefined number of configurations from the hyperparameter space, offering greater computational efficiency and proving particularly advantageous for large search spaces, while achieving solutions of comparable quality. In previous studies, such as the RF4PCC model, a validation strategy was adopted in which the test set was directly used for hyperparameter selection, leading to potential contamination of performance estimates and overfitting to the specific test partition. Similarly, in RF4PCC-optimized, hyperparameters were selected empirically without a systematic validation procedure. These methodological limitations reduce the reliability of performance evaluations and the generalization capability of the models (Figure 1). Moreover, the results achieved in recent years and the emergence of new approaches have increasingly driven the adoption of DL methods for the automatic recognition of historical architectural elements. These include point-wise MLP-based methods such as PointNet (Qi et al., 2017) and PointNet++ (Qi et al., 2017); point-based convolutions (Thomas et al., 2019); recurrent neural networks (Huang et al., 2018); attention mechanisms and Transformer-based architectures, such as Point Transformer v1 (PTv1) (Zhao et al., 2021) and Point Transformer v2 (PTv2) (Wu et al., 2022); and finally discretization-based methods, which convert point clouds into discrete or multidimensional representations and subsequently apply convolutions for 3D semantic segmentation, as in the case of OACNNs (Peng et al., 2024). Within this context, the present work proposes the development of a generalized RF model for the classification of different architectural typologies of cultural heritage (RF_CHC), integrating systematic hyperparameter tuning methodologies and validating its generalization capability against pre-existing models (RF4PCC and RF4PCC-optimized). The objective is to reduce overfitting and enhance the classifier's adaptability to the morphological and stylistic diversity of historical buildings, while simultaneously assessing performance improvements through comparison with the main DL approaches currently employed for the automatic recognition of historical architectural elements.

3. Methodology

This contribution focuses on improving the generalization capability of the RF classifier to enable the classification of different architectural typologies within Cultural Heritage, extending the application of the RF4PCC-optimized framework beyond the specific context of monastic case studies. The newly developed model is compared with the previously validated RF4PCC model to assess improvements in terms of generalization capability.

The experimental process was structured into several phases:

(i) Implementation via cross-validation of the RF4PCC-optimized model: integration of cross-validation techniques during the training phase for automatic hyperparameter selection, with the aim of improving model stability and predictive performance. The model was tested on a point cloud of a monastic cloister, consistent with the training dataset.

(ii) Application of the RF4PCC-optimized model to new architectural typologies (type A – porticoes of religious chapels; type B – urban porticoes): assessment of generalization capability.

(iii) Development and validation of a specialized framework, RF_CHC: implementation of the RF4PCC-optimized dataset, automatic hyperparameter selection, and class balancing strategies aimed at achieving greater generalization across heterogeneous architectural typologies. The framework was tested on type A and B cases to verify effective cross-typological generalization capability.

To further verify the validity of the proposed approach, the RF_CHC model was compared with the baseline RF4PCC model, which has already been validated in the literature. Both models were tested on a point cloud external to the training data, representing the façade of the Church of the Santissima Pietà in Teggiano (SA), in order to objectively assess improvements in generalization capability. Furthermore, to evaluate the performance of the proposed ML-based approach with respect to DL techniques, the RF_CHC model was applied to the point cloud of the portico of the Via Crucis at the Sacro Monte of the SS. Trinità di Ghiffa (A_SMG_portico), included in the ArCH Dataset. The obtained results were compared with those reported in Betsas et al. (2025b), who evaluated several DL algorithms on the same point cloud, allowing a direct comparison in terms of accuracy and generalization capability.

3.1 Dataset

To test the proposed methodology, improve the model's predictive performance through cross-validation techniques, and assess its applicability to the recognition of elements belonging to different architectural typologies, several datasets were considered. These included both datasets specifically designed for the experimentation and pre-existing datasets representative of different architectural contexts. For the implementation of the RF4PCC-optimized model, a dataset designed for cloister typologies was used. The corresponding point clouds were acquired using a Terrestrial Laser Scanner (TLS), within the framework of a collaboration with the Religious Provinces of the Franciscan Order in the Campania region, and were manually annotated (Cera et al., 2025). The dataset includes eight point clouds of monastic cloisters from the convents of San Francesco in Montella (AV), S.S. Pietà in Teggiano (SA), San Francesco in Padula (SA), San Lorenzo Maggiore in Naples (NA), Sant' Andrea in Nocera Superiore (SA), Sant' Antonio in Nocera Inferiore (SA), San Francesco in Solofra (AV), and San Francesco in Benevento (BN).

To test the generalization capability of the RF4PCC-optimized model on new architectural typologies, the external ArCH – Architectural Cultural Heritage dataset (Matrone et al., 2020b) was used, selecting several of the seventeen annotated scenes. These include the Chapel of Gesù al Tribunale di Anna in the Special Reserve of the Sacro Monte of Varallo (VC) and a building with porticoes in Piazza Santo Stefano in Bologna (BO). Although belonging to different architectural contexts and characterized by distinct functions and formal articulations, these case studies share recurring elements - such as arches, columns, and vaults - which allow the evaluation of the recognition of common components of historical architecture and the verification of the model's ability to generalize beyond the original dataset.

To further strengthen generalization capability, the RF_CHC model dataset was expanded by including selected scenes from the ArCH dataset. Specifically, the Room of Columns of the Valentino Castle in Turin (3_VAL_room), three chapels of the Sacro Monte of Varallo (5_SMV_chapel_1, 6_SMV_chapel_2to4, 8_SMV_chapel_28), and two portions extracted from the portico of Santo Stefano Street in Bologna (11_SStefano_portico_2) were integrated. The inclusion of these environments - characterized by heterogeneous architectural elements, materials, and lighting conditions - allowed the dataset to be enriched with greater morphological and stylistic variability.

For further validation on more complex architectural typologies and forms, and to assess improvements achieved by the RF_CHC training model, a dataset related to religious buildings with articulated plan-volumetric structures was constructed, such as the Church of the SS. Pietà in Teggiano (SA). Moreover, the complexity of the spatial and volumetric organization of these buildings made it possible to further test the generalization capability of the new RF_CHC training model on a typology also used for comparison with DL-based segmentation approaches, namely the portico of the Via Crucis at the Sacro Monte of the SS. Trinità of Ghiffa.

To optimize processing times, modular portions of the point clouds were considered for all typological models analyzed, each representative of all architectural classes. This choice reduces computational load without compromising the validity of the results, thanks to the modularity of the selected segments and their ability to provide comprehensive information for class recognition.

3.2 Implementation via cross-validation of RF4PCC – optimized model (i)

The first phase, related to the optimization of the RF4PCC-optimized model, focuses on three hyperparameters, each explored within a range of values selected to balance performance and computational cost while avoiding overfitting:

- *n_estimators* [50, 100, 200, 300, 500]: controls the number of trees in the forest; excessively high values increase computational costs without guaranteeing proportional performance improvements.
- *max_depth* [5, 10, 15, 20, None]: regulates the maximum depth of the decision trees; greater depths may lead to overfitting on training-set-specific patterns, thereby compromising generalization capability.
- *min_samples_leaf* [1, 3, 5, 10, 15]: defines the minimum number of samples required to form a leaf; higher values improve generalization, being particularly effective for minority classes.

To identify the most effective approach for optimal hyperparameter selection, two tuning methodologies were compared: *RandomizedSearchCV*, which randomly explores the parameter space, and *GridSearchCV*, which evaluates all possible parameter combinations.

The issue of class imbalance - typical of architectural point cloud datasets, where structural elements such as walls and floors numerically dominate over decorative or detailed components - was addressed by keeping the *class_weight="balanced"* configuration constant throughout all training and optimization phases.

Class	Top Features	Best Radii
Arch	Verticality; Sphericity; Surface variation	1.0m (Ve); 0.75m (Ve); 1.0m (Sp); 0.25m (Ve); 0.75m (Su)
Column	Eigenentropy; Planarity; Verticality	0.75m (Ei); 1.0m (Pl); 1.0m (Ve)
Molding	Verticality	1.0m (Ve)
Floor	Verticality; Omnivariance; Sphericity	0.5m (Ve); 0.75m (Ve); 1.0m (Ve); 0.25m (Ve); 0.5m (Om); 0.1m (Ve); 1.0m (Sp); 1.0m (Om)
Window/Door	Verticality	1.0m (Ve)
Wall	Verticality; Omnivariance; Anisotropy	0.25m (Ve); 0.75m (Ve); 0.5m (Ve); 1.0m (Om); 0.1m (An)
Stairs	Verticality	1.0m (Ve); 0.75m (Ve)
Vault	Verticality; Anisotropy	0.75m (Ve); 1.0m (Ve); 0.25m (Ve); 0.1m (An)
Roof	Verticality	0.5m (Ve)
Other	Sphericity; Anisotropy	1.0m (Sp); 0.1m (An)

Table 1. Features selected by RF_CHC model

This strategy automatically assigns weights inversely proportional to class frequency, enabling robust performance estimation in the presence of strong geometric heterogeneity and uneven class distributions, conditions that are typical of architectural heritage applications.
To verify the effectiveness of the implementation, the modified model was initially tested on a portion of a cloister point cloud, selected for its geometric homogeneity with respect to the training dataset. This condition made it possible to evaluate the impact of hyperparameter optimization in a controlled and comparable context, reducing variables related to morphological complexity before proceeding with validation on more diverse architectural typologies.

3.3 Application of the implemented RF4PCC-optimized model on new architectural typologies (ii)

The RF4PCC-optimized predictive model, developed in a previous study (Antuono et al., 2025) from a heterogeneous dataset related to the architectural typology of monastic cloisters, was subsequently applied to two point clouds external to the training dataset and characterized by different architectural typologies, to evaluate its generalization capability. The two selected point clouds, extracted from the ArCH dataset, represent respectively a portico of a religious chapel (7_SMV_chapel_24, Sacro Monte of Varallo) and an urban portico (10_SStefano_portico_1, Portico of Santo Stefano, Bologna).
This approach made it possible to verify whether the model - although trained on a heterogeneous dataset confined to a specific architectural typology - was able to effectively recognize the same classes of architectural elements (“floor”, “wall”, “column”, “arch”, “vault”, “stair”, “window/door”, “molding”, “roof”, “other”) in morphologically and functionally different contexts, such as the porticoes of religious buildings and urban porticoes. This phase represented a fundamental step in understanding the generalization limits of the optimized model and in motivating the subsequent development of the RF_CHC framework.

3.4 Development and validation of RF_CHC model (iii)

To further enhance the model’s generalization capability across different architectural typologies of Cultural Heritage, a new framework named RF_CHC (Random Forest for Cultural Heritage Classification) was developed, synergistically integrating the most effective aspects of the pre-existing RF4PCC and RF4PCC-optimized models. In particular, from the RF4PCC model, the strategy of expanding the typological diversity of the training dataset was adopted, acknowledging the importance of exposing the model to heterogeneous architectural contexts in order to strengthen its robustness. From the RF4PCC-optimized model, the automatic feature selection methodology based on p-value analysis and the use of triangular matrices for dimensional optimization of the feature space were incorporated, with all processing and model implementation carried out in Python (Table 1). This procedure enables the identification of the most informative features while eliminating redundant or non-significant ones, thereby reducing the risk of overfitting and improving the model’s computational efficiency. In addition, the class balancing strategy (*class_weight="balanced")* was retained to effectively handle the uneven distribution of architectural categories within the dataset.
During the training phase, cross-validation was implemented through *RandomizedSearchCV*, selected for its ability to efficiently explore the hyperparameter space (*n_estimators*, *max_depth*, *min_samples_leaf*) at a significantly lower computational cost than *GridSearchCV*, while maintaining comparable performance, as demonstrated in the preliminary phase of the study. The training dataset was then expanded with respect to that used for the RF4PCC-optimized model, which included seven point clouds of monastic cloisters. To increase typological diversity, six additional point clouds from the ArCH dataset were integrated, selected to represent morphologically distinct architectural contexts: 3_VAL_room, 5_SMV_chapel_1, 6_SMV_chapel_2to4, 8_SMV_chapel_28, and two portions extracted from 11_SStefano_portico_2.
This dataset composition allowed the RF_CHC model to learn from a more representative set of data reflecting the complexity and variability of cultural architectural heritage, thereby promoting improved generalization capability to architectural typologies not present in the initial training set.

3.5 Comparison with RF4PCC model and DL algorithms

To verify the actual generalization capability of the new RF_CHC framework on architectural contexts external to the training dataset, a comparative validation was conducted with the RF4PCC model, which has already been validated and consolidated in the literature as a reference for the classification of Cultural Heritage (CH) architectural elements. For this validation phase, a point cloud representing a portion of the façade of the Church of the Santissima Pietà in Teggiano (SA) was selected; this dataset was not included in any previous training set (Figure 2). The introduction of a new point cloud was necessary because the point clouds previously used for generalization testing (7_SMV_chapel_24 and 10_SStefano_portico_1), although external to the RF_CHC training dataset, were already present in the original RF4PCC dataset, as they also originated from the ArCH dataset.

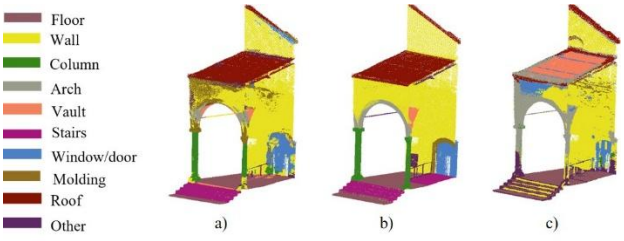


Figure 2. Point cloud classification: RF_CHC model (a), ground truth (b), RF4PCC model (c).

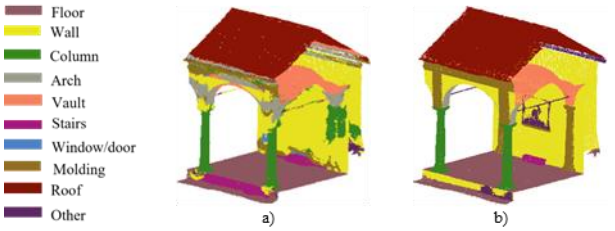


Figure 3. "A_SMG_portico" cloud classification: RF_CHC model (a), ground truth (b).

The use of a completely novel case study for both models thus enabled a fair and objective comparison, eliminating potential biases arising from prior data exposure and providing a rigorous assessment of improvements in terms of accuracy, F1-score, and the ability to recognize different architectural classes in morphologically complex and non-standardized historical architectures.

To evaluate the performance of the proposed ML-based approach with respect to DL techniques, the RF_CHC model was applied to the point cloud of the portico of the Via Crucis at the Sacro Monte of the SS. Trinità di Ghiffa (A_SMG_portico), included in the ArCH Dataset (Figure 3). This case study was selected because it was used as an experimental dataset in the study by Betsas et al. (2025b), in which several DL algorithms were evaluated on the same point cloud. The use of the same test dataset enables a direct and methodologically comparison between the two approaches.

4. Results and discussion

The results of each of the previous methodological phases were evaluated as they progressed to guide and understand the performance of the implementations carried out.

4.1 RF4PCC evaluation – optimized model implementation

In the first phase of the study, the effectiveness of the hyperparameter tuning methodologies implemented within the RF4PCC-optimized framework was demonstrated.

The comparison between methodologies shows that *GridSearchCV* and *RandomizedSearchCV* achieved substantially equivalent performance, with weighted F1-scores of 0,694 and 0,689, respectively (Figure 4). *RandomizedSearchCV* offers advantages in terms of computational efficiency while maintaining optimal results. The comparative analysis between the implemented model and the previous RF4PCC-optimized study shows that automated hyperparameter optimization led to a significant increase in overall accuracy from 0,58 to 0,70 and in weighted F1-score from 0,61 to 0,69. Based on these results, *RandomizedSearchCV* was adopted in subsequent phases for hyperparameter selection, as it ensures a good balance between performance and processing time (3-fold stratified cross-validation; optimal parameters: *n_estimators* = 200, *max_depth* = 20, *min_samples_leaf* = 5).

The cross-domain validation of the RF4PCC-optimized model, tested on two specific point clouds from the ArCH dataset - the Chapel of Santa Maria delle Virtù (7_SMV_chapel_24) and the Portico of Santo Stefano (10_SStefano_portico_1) - representative of architectural contexts different from the monastic training setting, revealed discrete generalization capability. The overall accuracy values were 0,63 and 0,77, respectively, with weighted F1-scores of 0,60 and 0,75 (Figure 5). Some categories exhibited particularly strong performance: the class "column" achieved an F1-score of 0,83 with 10.468 correctly classified points in the chapel and 0,76 with 19.034 points in the portico, while "vault" achieved 16.730 correct points (F1-score: 0,71) in the chapel and 39.329 (F1-score: 0,87) in the portico. Conversely, critical issues persist in distinguishing between "arch" and "vault", with very low F1-score values (0,32 and 0,33), as well as in separating "molding" from "wall" and "window/door" from "wall". These results highlight the need to develop a model with greater generalization capability, able to operate effectively across different architectural typologies of cultural heritage.

4.2 Performance analysis of RF_CHC model

To address the requirement of greater generalization capability, the RF_CHC model was created by enriching the training dataset with six additional point clouds from the ArCH dataset, representative of heterogeneous architectural contexts. The new model was subsequently subjected to an automatic feature selection process aimed at optimizing classification performance across diversified architectural scenarios.

The results obtained with the RF_CHC model demonstrate substantial improvements over the previous approach. In the case of the Chapel of Santa Maria delle Virtù, performance increased significantly: the class "arch" showed an increase in F1-score from 0,32 to 0,66, the class "molding" from 0,062 to 0,23, and the class "window/door" from 0,19 to 0,83.

Overall, accuracy increased from 0,63 to 0,70 and weighted F1-score from 0,66 to 0,70. Even more pronounced results were obtained for the Portico of Santo Stefano, where the class "arch" reached an F1-score of 0,71 (from 0,33), "molding" reached 0,70 (from 0,48), and "window/door" reached 0,46 (from 0,12), with overall accuracy increasing from 0,77 to 0,86 and weighted F1-score from 0,75 to 0,86 (Figure 6).

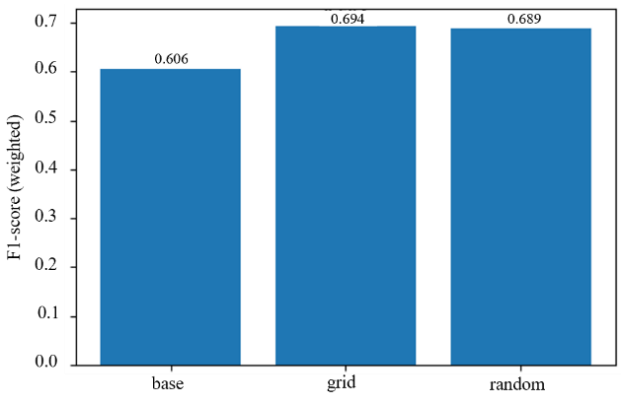
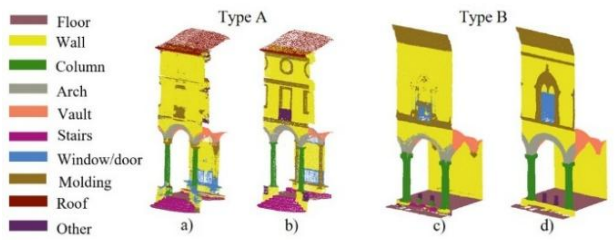


Figure 4. Comparison between F1-scores of the RF4PCC-optimized model – *GridSearchCV* – *RandomizedSearchCV*.



Accuracy	0,63	0,77
F1-score	0,60	0,75

Figure 5. Type A – B classification: implemented RF4PCC-optimized model (a) (c), ground truth (b) (d).

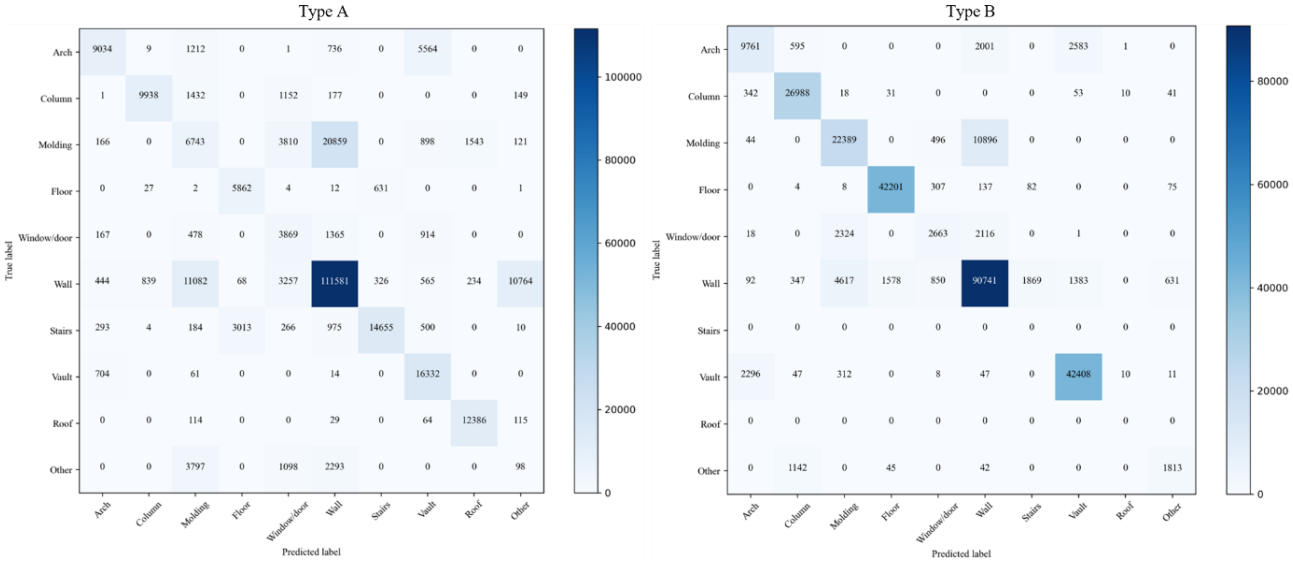


Figure 6. Confusion matrix – Type A and B classification by RF_CHC model.

RF_CHC	0,83	0,79	0,82	0,32	0,23	0,65	0,57	0,023	0,85	0,19
RF4PCC	0,79	0,73	0,065	0,32	0,00	0,00	0,40	0,00	0,17	0,044
	Floor	Wall	Column	Arch	Vault	Stairs	Window/ door	Molding	Roof	Other

Table 2. Class-wise F1-score of the RF_CHC and RF4PCC models.

To further validate the generalization capability of the new model, a direct comparison with the RF4PCC model - already consolidated in the literature - was carried out by applying both approaches to a new point cloud not used in previous phases. The results highlight a clear improvement achieved by the RF_CHC model (Table 2): while RF4PCC obtained an accuracy of 0,44 and a weighted F1-score of 0,43, the RF_CHC model reached 0,69 and 0,70, respectively, corresponding to an increase of approximately 57% in accuracy and 63% in weighted F1-score.

4.3 Evaluation of the results of the comparison between RF_CHC model and DL algorithms

The performance of the proposed ML-based approach with respect to DL techniques was evaluated by applying the RF_CHC model to the point cloud "A_SMG_portico" included in the ArCH Dataset and comparing the results with those obtained by several DL algorithms analyzed in the State of the Art (Betsas et al., 2025b) (Figure 7). The comparison of metrics revealed differentiated performance dynamics between the two families of approaches. The RF_CHC model achieved an mIoU (mean Intersection over Union, a metric measuring the average overlap between predictions and ground-truth labels across all classes) of 0,493, an accuracy of 0,612, and a weighted F1-score of 0,637.
An analysis of the results of the DL algorithms tested by Betsas et al. (2025b) on the same point cloud shows that the most performant algorithms, such as DGCNN Mod+3Dfeat and PTv1, achieved higher accuracy values (0,914 and 0,963, respectively) and high weighted F1-scores (0,915 and 0,736), demonstrating the ability of advanced DL models to capture complex patterns in architectural data. However, it is noteworthy that not all approaches ensured superior performance: algorithms such as OACNNs and PTv2, despite achieving apparently high accuracy values (0,889 and 0,909), exhibited severe performance drops in other metrics, with weighted F1-scores of 0,379 and 0,275 and mIoU values of 0,237 and 0,199, respectively lower than those obtained by RF_CHC. DGCNN also showed limited performance with an mIoU of 0,376, lower than that of the proposed model. In particular, the RF_CHC model occupies an intermediate position in terms of mIoU (0,49), outperforming three of the five tested DL algorithms and positioning itself between DGCNN and more advanced models such as DGCNN Mod+3Dfeat (0,599) and PTv1 (0,556) (Table 3). The comparative analysis highlights how the adopted generalization strategy - based on a heterogeneous training dataset and optimized feature selection - enables the RF_CHC model to achieve competitive performance with respect to DL approaches, outperforming some advanced algorithms on key metrics while maintaining performance consistency.

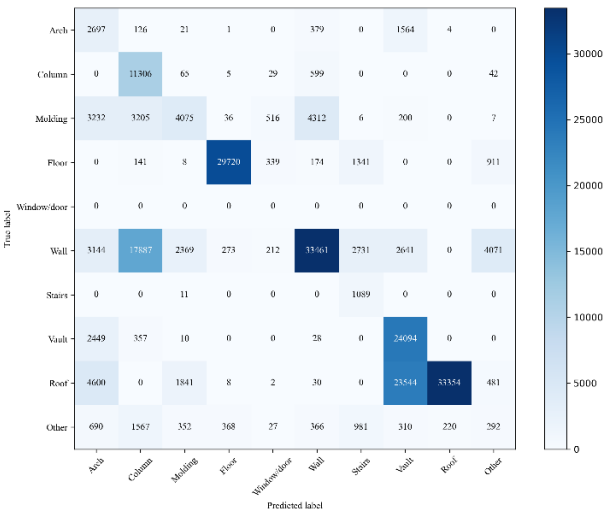
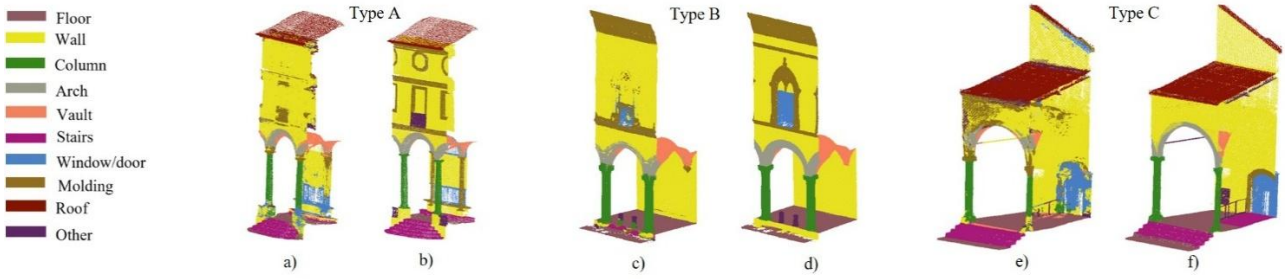


Figure 7. Confusion matrix – "A_SMG_portico" cloud classification by RF_CHC model.



Accuracy	0,70	0,86	0,69
F1-score	0,70	0,86	0,70

Figure 8. Class-wise F1-score of type A - B - C: RF_CHC model (a) (c) (e), ground truth (b) (d) (f).

	mIoU	Accuracy	F1-score
DGCNN	0,376	0,784	0,794
DGCNN Md+3Dfeat	0,599	0,914	0,915
OACNNs	0,237	0,889	0,379
PTv1	0,556	0,963	0,736
PTv2	0,199	0,909	0,275
RF_CHC	0,493	0,612	0,637

Table 3. Comparison between RF_CHC model and DL algorithms.

This makes the model particularly suitable for application scenarios characterized by computational constraints or requirements for model interpretability (Figure 8).

5. Conclusions

The conducted research demonstrates that an RF-based approach can be effectively applied to the automatic semantic classification of CH architectural point clouds, while simultaneously highlighting the crucial importance of training dataset diversification in ensuring adequate generalization capability.

The comparative analysis between the RF4PCC-optimized model - trained on a specific monastic context - and the new RF_CHC model revealed that classification performance is strongly influenced by the typological variety of the architectures used during training. While the former model exhibited reasonable generalization capability for certain recurring architectural elements (columns, vaults), it also showed significant limitations in the classification of morphologically complex categories or those with ambiguous semantic boundaries, such as arches, moldings, and openings.

The development of the RF_CHC model, through the integration of six additional point clouds representative of heterogeneous architectural contexts and the implementation of an automatic feature selection process, made it possible to overcome most of these limitations. The observed improvements in accuracy (up to 57%) and weighted F1-score (up to 63%) compared to the consolidated RF4PCC model confirm the validity of the adopted strategy and demonstrate that training set diversification, along with automatic feature and hyperparameter selection, is fundamental to achieving robust classifiers that are transferable across different architectural typologies.

The comparison with DL algorithms revealed articulated performance dynamics. Although the most advanced DL models (DGCNN Mod+3Dfeat and PTv1) achieved higher accuracy values, the RF_CHC model demonstrated competitive performance and, above all, greater performance consistency compared to several DL algorithms. With an mIoU of 0,49, RF_CHC outperformed three of the five tested algorithms, including DGCNN (0,376), OACNNs (0,237), and PTv2 (0,199), highlighting how some DL approaches - despite apparently high accuracy - may suffer from significant performance drops in key metrics such as weighted F1-score and mIoU.

Despite these advances, challenges remain in distinguishing elements with similar geometric characteristics or under conditions of high morphological complexity, suggesting the need for future developments. These may include further expansion of the dataset with more diverse case studies - in line with current research directions (Matrone et al., 2025) - the integration of contextual semantic information and the exploration of advanced DL approaches. Nevertheless, the RF_CHC model represents a significant step forward toward the development of reliable automatic tools for the documentation and analysis of cultural architectural heritage.

Acknowledgements

This article is the result of the joint research efforts of the authors. The specific written contributions of the authors are attributed as follows: G.A. wrote the paragraphs 1, 3, 3.5, 4.2 and 5; V.C. wrote paragraphs 2, 3.1, 3.2 and 4.3; D.C. wrote paragraphs 3.3, 3.4 and 4.1. The graphical elaborations were jointly carried out by the authors. The research is part of a collaboration agreement between the University of Naples Federico II and the Religious Provinces of the Franciscan Order in Campania. The scientific coordinators of the project are G. Antuono and V. Cera.

References

Antuono, G., Cera, V., Ciarlo, D., 2025. Optimized approach to feature selection for semantic segmentation using Random Forest. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-M-9-2025, 41-48. doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-41-2025.

Bello, S.A., Yu, S., Wang, C., Adam, J.M., Li, J., 2020. Review: Deep learning on 3D point clouds. *Remote Sensing*, 12(11), 1729. doi.org/10.3390/rs12111729.

Liu, W., Sun, J., Li, W., Hu, T., Wang, P., 2019. Deep learning on point clouds and its application: A survey. *Sensors*, 19(19), 4188. doi.org/10.3390/s19194188.

Bergstra, J, Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.*, 13, 281-305.

Betsas, T., Georgopoulos, A., Doulamis, A., Grussenmeyer, P., 2025a. Deep Learning on 3D Semantic Segmentation: A Detailed Review. *Remote Sensing*, 17(2), 298. doi.org/10.3390/rs17020298.

Betsas, T., Tsarpalis, H., and Georgopoulos, A., 2025b. Assessing Generalization Capability of 3D Semantic Segmentation Algorithms using 3D Point Clouds of Cultural Heritage. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-M-9-2025, 111–118. doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-111-2025.

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., Deng, D., Lindauer, M., 2023. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(2), e1484. doi.org/ 10.1002/widm.1484.

Buldo, M., Agustín-Hernández, L., Verdoscia, C., 2024. Semantic Enrichment of Architectural Heritage Point Clouds Using Artificial Intelligence: The Palacio de Sástago in Zaragoza, Spain. *Heritage*, 7(12), 6938-6965. doi.org/10.3390/heritage7120321.

Cao, Y., Teruggi, S., Fassi, F., Scaioni, M., 2022. A comprehensive understanding of machine learning and deep learning methods for 3d architectural cultural heritage point cloud semantic segmentation. *Ital. Conf. Geomatics Geospat. Tech.*, Springer, 329-341.

Cera, V., Antuono, G., Campi, M., D'Agostino, P., 2025. Data Quality, Semantics, and Classification Features: Assessment and Optimization of Supervised ML-AI Classification Approaches for Historical Heritage. *Heritage*, 8(7), 265. https://doi.org/10.3390/heritage8070265.

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J. T., Blum, M., Hutter, F., 2015. Efficient and robust automated machine learning. In *Proc. of 29th International Conference on Neural Information Processing Systems*, 2. Cambridge: MIT Press, 2755-2763.

Gil, A., Arayici, Y., 2025. Point Cloud Segmentation Based on the Uniclass Classification System with Random Forest Algorithm for Cultural Heritage Buildings in the UK. *Heritage*, *8*(5), 147. doi.org/10.3390/heritage8050147.

Gîrbacia, F., 2024. An Analysis of Research Trends for Using Artificial Intelligence in Cultural Heritage. *Electronics*, 13(18), 3738. doi.org/10.3390/electronics13183738.

Grilli, E., Ozdemir, E., Remondino, F., 2019. Application of machine and deep learning strategies for the classification of heritage point clouds. *Int. arch. photogramm. remote sens. spat. inf. sci.*, XLII-4/W18, 447-454. doi.org/10.5194/isprs-archives-XLII-4-W18-447-2019.

Grilli, E., Remondino, F., 2020. Machine Learning Generalisation across Different 3D Architectural Heritage. *ISPRS Int. J. Geo-Inf.*, 9(6), 379. doi.org/ 10.3390/IJGI9060379.

Huang, Q., Wang, W., Neumann, U., 2018. Recurrent slice net works for 3d segmentation of point clouds. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2626-2635. https://doi.org/10.48550/arXiv.1802.04402.

Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R., Remondino, F., 2020a. Comparing machine and deep learning methods for large 3D heritage semantic segmentation. *ISPRS Int. J. Geo-Inf.*, 9(9), 535. doi.org/10.3390/ijgi9090535.

Matrone, F., Lingua, A., Pierdicca, R., Malinverni, E. S., Paolanti, M., Grilli, E., Remondino, F., Murtiyoso, A., Landes, T., 2020b. A Benchmark for Large-Scale Heritage Point Cloud Semantic Segmentation. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B2-2020, 1419–1426. doi.org/10.5194/isprs-archives-XLIII-B2-2020-1419-2020.

Matrone, F., Lingua, A. M., Murtiyoso, A., Sridhar, M., Sanchez Aparicio, L. J., Santamaria Maestro, R., Sanz-Honrado, P., Pande, H., Seth Tiwari, P., Agrawal, S., 2025. Cross-domain urban and heritage point cloud benchmarks for semantic segmentation: insights from the 3M (Mapping Mappano Municipality) and ArCHX datasets. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-1/W6-2025, 131-137. doi.org/10.5194/isprs-archives-XLVIII-1-W6-2025-131-2025.

Moyano, J., Musicco, A., Nieto-Julian, J. E., Dominguez-Morales, J. P., 2024. Geometric characterization and segmentation of historic buildings using classification algorithms and convolutional networks in HBIM. *Autom. Constr.*, 167, 105728. doi.org/10.1016/j.autcon.2024.105728.

Morbidoni, C., Pierdicca, R., Paolanti, M., Quattrini, R., Mammoli, R., 2020. Learning from Synthetic Point Cloud Data for Historical Buildings Semantic Segmentation. *J. Comput. Cult. Herit.* 13(4), 34. doi.org/10.1145/3409262.

Lin, H.-I., Nguyen, M. C., 2020. Boosting Minority Class Prediction on Imbalanced Point Cloud Data. *Applied Sciences*, 10(3), 973. doi.org/10.3390/app10030973.

Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proc. of the IEEE conference on computer vision and pattern recognition*, 652-660.

Peng, B., Wu, X., Jiang, L., Chen, Y., Zhao, H., Tian, Z., Jia, J., 2024. Oa-cnns: Omni-adaptive sparse cnns for 3d semantic segmentation. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21305-21315.

Terruggi, S., Grilli, E., Russo, M., Fassi, F., Remondino, F., 2020. A Hierarchical Machine Learning Approach for Multi-Level and Multi-Resolution 3D Point Cloud Classification. *Remote Sensing*, 12(16), 2598. doi.org/10.3390/ RS12162598.

Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., Guibas, L. J., 2019. Kpconv: Flexible and de formable convolution for point clouds. *Proc. of the IEEE/CVF international conference on computer vision*, 6411-6420.

Wang, F., Zhuang, Y., Gu, H., Hu, H., 2019. Automatic generation of synthetic LiDAR point clouds for 3-D data analysis. *IEEE Trans. Instrum. Measure.* 68(7), 2671-2673. doi.org/10.1109/TIM.2019.2906416.

Wu, X., Lao, Y., Jiang, L., Liu, X., Zhao, H., 2022. Point trans former v2: Grouped vector attention and partition-based pooling. *Adv. Neural Inf. Process. Syst.*, 35, 33330-33342.

Yang, S., Hou, M., Li, S., 2023. Three-Dimensional Point Cloud Semantic Segmentation for Cultural Heritage: A Comprehensive Review. *Remote Sensing*, 15, 548. doi.org/10.3390/rs15030548.

Zhao, H., Jiang, L., Jia, J., Torr, P. H., Koltun, V., 2021. Point transformer. *Proc.s of the IEEE/CVF international con ference on computer vision*, 16259–16268.