

Uncertainty quantification of laserscanning point clouds for road asset classification

Shirin Malihi^{1*}, Ioannis Brilakis¹, Nima Arij², Abbas Kiani³, Stefano Cavazzi⁴, Qiuchen Lu⁵, Graham Starkey⁶,
Matt Peck⁷, George Economides⁸

¹ Civil Engineering Department, University of Cambridge, UK (sm2852, ib340)@cam.ac.uk

² Independent Researcher, Amol, Iran - nima.nit96@gmail.ir

³ Babol Noshirvani University of Technology, Iran - a.kiani@nit.ac.ir

⁴ Innovation and Research Department, Ordnance Survey, UK - stefano.cavazzi@os.uk

⁵ Bartlett School of Sustainable Management, University College London (UCL), UK - ql284@cam.ac.uk

⁶ Head of BIM Department, Costain, UK - graham.starkey@costain.com

⁷ Director of Innovation and Fellow, AtkinsRéalis, Visiting Professor, University of Birmingham, UK - matt.peck@atkinsrealis.com

⁸ Digital Twins Department, UK Government's Department for Transport, UK

Keywords: Point cloud, Machine learning, Entropy, Sensitivity analysis, RGB Image, Uncertainty

Abstract

Accurate road extraction from LiDAR data remains challenging, as the absence of spectral cues or spatial heterogeneity increases model uncertainty. This study introduces a comparative, entropy-driven framework for evaluating the performance and reliability of road asset detection using three supervised machine learning algorithms—XGBoost, Random Forest (RF), and Support Vector Machine (SVM). Using a high-density aerial point cloud, a reproducible computational pipeline was developed to select reliable machine learning methods for road asset mapping. Beyond traditional accuracy metrics (Overall Accuracy, F1-score, and Kappa coefficient), uncertainty-based evaluation of the outputs has been conducted using KPIs of entropy and sensitivity to training sets to quantify model reliability and spatial instability. Results reveal that the inclusion of RGB significantly reduces entropy across all models. XGBoost achieved the lowest mean entropy (0.084–0.143) and the most consistent probabilistic behaviour, reflecting confident and well-calibrated model. SVM, while statistically the most accurate (OA and Kappa > 0.97), exhibited higher local entropy (≈ 0.23 – 0.26), implying precise yet less certain classification. RF demonstrated the highest entropy (≈ 0.65 – 0.70) and the greatest variability, underscoring its sensitivity to feature noise. Under the WOR configuration, mean entropy rose markedly—most for RF WOR (≈ 0.93) and moderately for SVM WOR (≈ 0.39)—while XGBoost retained low uncertainty. Spatial entropy maps further highlighted that uncertainty concentrates along road edges with RGB data but expands diffusely under WOR conditions, emphasizing the role of spectral-spatial synergy in constraining ambiguity. Entropy-based evaluation provided insights beyond conventional accuracy metrics, revealing paradoxes between correctness and confidence.

1. Introduction

1.1 Background and problem statement

Roads have always been the vital arteries of human societies and the most probable means of transportation (Khan and Singh, 2023), which are fundamental for economic growth, social connectivity, and public safety (Fiorentini and Losa, 2025). Roads play essential roles in various domains such as navigation (Wen et al., 2019) urban planning (Schreiber, 2025), and crisis response (Aziz et al., 2023). Estimates indicate that by 2050, approximately 25 million kilometers of new roads will be constructed worldwide (Laurance et al., 2017). It is also predicted that the global population will approach 10 billion by 2050 (Olimid and Olimid, 2019). This rapid growth, accompanied by accelerated urbanization, is driving the expansion of road infrastructure to facilitate population mobility and meet the increasing demand for travel and transportation (Laurance et al., 2014). Consequently, precise identification of road infrastructure has become essential for urban planning, traffic management, disaster response, and environmental monitoring (Liu et al., 2024).

To achieve accurate recognition of road infrastructure, precise detection and identification of roads are required (Li et al., 2024). However, automatic road detection remains a challenging research topic. Typically, a road is a linear object resembling other linear features such as road markings, fractures, cracks, dirt paths, side roads, and so on, each with distinct characteristics (Wang et al., 2021a; Wang et al., 2021b). Specifically, semantic labeling and modeling of roads are critical for comprehending the overall structure of a city, as roads form a continuous surface across the urban area, divide

the city into blocks, and provide contextual cues for detecting smaller objects (Boyko and Funkhouser, 2011).

To date, researchers have proposed various road-mapping methods that can be divided into traditional approaches (Song and Civco, 2004) and artificial intelligence-based methods (Bugday, 2018). Traditional road detection techniques rely on manual feature design (Valero et al., 2010). In such methods, researchers have employed shape indices and density features (Song and Civco, 2004), morphological operators (Valero et al., 2010), or orientation histograms (Li et al., 2016) for road extraction. Although these approaches have proven effective in certain cases, their lack of automated learning leads to limitations such as low automation, operational complexity, and time-consuming processes (Lu et al., 2021). Conversely, modern road detection techniques based on artificial intelligence encompass fuzzy logic, artificial neural networks (ANN), adaptive network fuzzy inference systems (ANFIS), machine learning, and deep learning (DL) (Mohammadzadeh and Zoj, 2010). While approaches such as SOFS, ANN, and ANFIS have demonstrated effectiveness in previous studies, machine learning—with its hierarchical feature extraction and self-learning capabilities—has emerged as the dominant method in road detection due to the rapid growth of multisource data (Zhang et al., 2016).

The performance of road detection methods is highly dependent on the quality and diversity of input data (Xu et al., 2018). In this regard, various types of data have been utilized in related research. Optical satellite imagery such as Landsat and Sentinel-2, despite being freely available and providing long-term records, fail to deliver accurate information under adverse weather conditions (Arij et al., 2025). Despite the high spatial resolution and accuracy of RGB imagery in feature detection, its limited spatial coverage and costly, time-intensive acquisition

process reduce its practicality on large scales (Xu et al., 2018). In recent years, LiDAR data have gained significant attention as one of the most precise and efficient resources for road extraction and monitoring (Gargoum and El-Basyouny, 2017). The primary advantage of LiDAR over other data sources lies in its very high spatial accuracy, independence from lighting conditions, and the ability to extract embedded three-dimensional features, which make it superior to other datasets (Gargoum and El-Basyouny, 2017).

Recent studies have demonstrated that the evolution of road extraction methods from remote sensing data has progressed from two-dimensional high-resolution imagery to three-dimensional LiDAR data, in parallel with the advancement of learning-based algorithms (Quackenbush et al., 2013). Following this line of research, (Suleymanoglu et al., 2024) proposed an unsupervised approach for extracting 3D road boundaries from LiDAR point cloud data and imagery acquired from mobile laser scanning (MLS) and image-based mapping systems (MMS). In this method, ground and non-ground points were separated and the initial boundary was detected based on elevation and slope thresholds, and subsequently refined using the DBSCAN algorithm and RANSAC and B-spline algorithms. Finally, (Matikainen et al., 2020) conducted a comprehensive assessment of the application of single-photon LiDAR (SPL) data for land cover classification and road surface extraction. Classification was performed using an object-based approach and the RF algorithm, identifying two road surface classes: "asphalt" and "gravel." The single-channel SPL data exhibited limited accuracy, whereas integrating it with multispectral intensities increased extraction accuracy to 81% and 77%, and overall accuracy to 86%. This study introduced multispectral SPL as a new pathway for automated and efficient large-scale road mapping.

Our systematic review of the existing studies (using SCOPUS®) revealed that most of the current research in this field has focused on improving the accuracy of deep learning models in two-dimensional imagery, developing multi-source data fusion methods (imagery, elevation, and spectral data), and employing three-dimensional LiDAR data to extract road geometry and boundaries. However, the simultaneous application and comparison of multiple machine learning approaches can provide a more comprehensive insight into model performance and enable the selection of optimal methods (Khanam and Foo, 2021). Nevertheless, this aspect has not been considered in most previous studies, and this deficiency is also observed in examining the influence of effects of training datasets on model performance and certainty of the outputted classes. This may lead to reduced generalizability and trustability of models under real-world conditions.

1.2 Theoretical framework of machine learning methods

The fields of LiDAR data analysis and road assets detection have witnessed remarkable progress in recent years under the title of digital twinning. Within this context, machine learning (ML) methods have emerged as powerful tools for large-scale road assets extraction and the analysis of complex behaviours (Döllner, 2020).

In the ML paradigm, models are trained on labeled datasets to discern underlying patterns, which are then used to predict new outcomes. This category encompasses a diverse range of algorithms, such as linear and nonlinear regression, RF, SVM, and Artificial Neural Networks (ANN). Each of these algorithms demonstrates varying levels of performance based on factors like the type of data, the size of the training sets, and the structural complexity of the variables. This research employs several supervised learning algorithms for the purpose

of modeling and comparatively evaluating their performance on aerial road 3D data. Alongside the comparative assessment of model efficacy, uncertainty and sensitivity analysis are performed to ascertain the impact of input variables on the model's output.

The XGBoost, RF, and SVM algorithms were used in this study to improve prediction accuracy and evaluate the effectiveness of machine learning techniques, because these algorithms can handle the geometric and tabular features of LiDAR data efficiently. The XGBoost algorithm efficiently lowers prediction errors and mitigates overfitting since it is based on a boosting technique and sequential decision trees (Chen and Guestrin, 2016). The Random Forest technique allows for the evaluation of feature relevance while generating a stable and noise-resistant model by integrating several decision trees via random sampling (Pal, 2005). The SVM algorithm was also used because of its good performance in high-dimensional spaces and in lack of huge labelled data (Suthaharan, 2016).

Uncertainty is often decomposed into two principal types: aleatoric uncertainty and epistemic uncertainty. Aleatoric uncertainty or data *uncertainty* arises from inherent randomness or noise in the observations (e.g. sensor noise, measurement error, or intrinsic variation in the phenomena). Epistemic uncertainty or *model uncertainty* stems from lack of knowledge: limited or biased training data, model misspecification, or regions of input space not seen during training. Because epistemic uncertainty derives from ignorance, it *can* be reduced by acquiring additional data or better models (Kendall and Gal, 2017).

Sensitivity analysis (SA) has become an essential component in evaluating the uncertainty of machine learning (ML) models for image and spatial segmentation. static metrics such as accuracy, Cohen's kappa, and F1-score are reported to assess how model performance fluctuates under variations in inputs and data quality. In segmentation tasks, such as road extraction, these metrics can exhibit high variance due to class imbalance, label noise, and domain shifts (Ma et al., 2019). accuracy is sensitive to dominant classes and may underestimate uncertainty in minority regions, while Cohen's kappa incorporates chance agreement, making it more informative under imbalanced conditions (Farhadpour et al., 2024). The F1-score, balancing precision and recall, captures uncertainty linked to false positives and negatives—critical in segmentation of rare or noisy classes (Diallo et al., 2024).

This theoretical foundation constitutes the basis of the integrated analytical approach of this research in employing supervised machine learning methods to detect road structures in LiDAR data with known uncertainty and quality metrics. This analytical framework provides the basis for evaluating quality metrics computed based on confusion matrix, uncertainty metrics and the sensitivity of the models, representing an effective step toward improving the efficiency and trustability of road asset detection.

2. Research contributions and innovations

After a thorough review of previous studies of existing road detection methods, we identified several fundamental research gaps in the current approaches. Most previous research, relying on two-dimensional imagery or three-dimensional LiDAR data, has merely aimed at enhancing accuracy indices without examining the capacity of models in terms of *trustability of road structures*.

In this study, a comprehensive experimental framework is proposed to adaptively compare machine learning algorithms—Random Forest, SVM, and XGBoost—under two data configurations (LiDAR + RGB and LiDAR-only) for assessing road asset detection uncertainty.

In all experiments, the sizes of the training and testing datasets were kept constant, and only the type of input data was changed to allow a fair and direct comparison of how each data source affects model performance. Because the initial dataset was relatively large, it was organized to minimize the chance of selecting the same samples during random resampling. Consequently, the ratio and number of training and testing samples remained fixed across all experiments, although the specific data composition in each iteration was randomly drawn from the same overall population. Subsequently, a multivariate sensitivity analysis and an uncertainty analysis were conducted to evaluate the stability and reliability of the models and to determine whether changes in the type of training data lead to significant variations in model accuracy and robustness. Overall, the innovation of this study lies in the design of a comparative and data-driven framework that simultaneously:

1. Sensitivity analysis of using radiometric modality on uncertainty (LiDAR + RGB and LiDAR-only)
2. Sensitivity analysis of changing training data sets on uncertainty of road asset classification
3. Examines entropy metric as an indicator for uncertainty of road asset classification

This framework analyses road assets detection in a novel point of view and quantifies a new quality dimension of this well known problem.

3. Materials and methods

3.1 Materials

The A30 Chiverton to Carland Cross project involved aerial photogrammetry conducted on the 18th of May 2024, utilizing both manned aircraft and unmanned aerial systems. A total of 12 Ground Control Points (GCPs) were strategically deployed across the survey area to ensure accurate project control and verification. The resulting dataset comprises a high-density, coloured 3D point cloud generated exclusively through aerial photogrammetry techniques. All deliverables are georeferenced to the standard OSGB 1936 coordinate system (OSTN15) with a Ground Sample Distance of 0.035 metres. The captured data provides detailed representation of the terrain and surface features, including roads, tracks, and footpaths, as well as street furniture, wooded areas, trees, recreation zones, and vehicles present within the survey extent. In figures 1, whole of the area and selected sections and GCPs are displayed. Figures illustrate the spatial distribution and location details of the temporary GCPs used in this study.

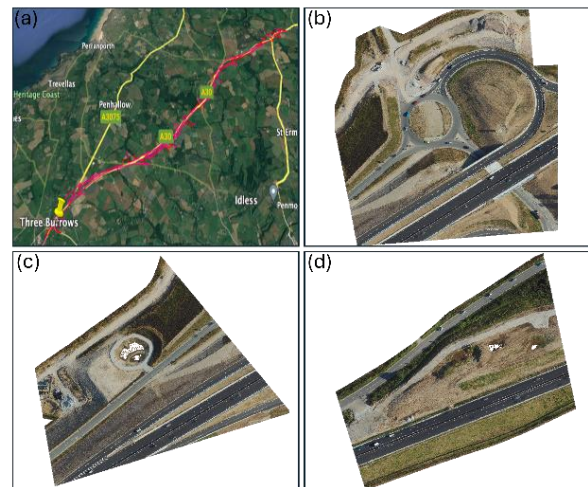


Figure 1: A general view of the study area. (a) shows a map of the A30 route near Cornwall, UK, where the project area is marked and the yellow pin indicates the location of the central reference point. Images (b) to (d) show the first to third study areas, containing 4,700- 3,400 and 3,000 thousand points.

3.2 Methodology

This study employed a structured multi-stage methodology to classify road surfaces from LiDAR point cloud data and to assess model uncertainty. The methodological framework was designed to ensure data consistency, feature representativeness, and reliable probabilistic evaluation. Each stage was implemented in a reproducible computational pipeline, allowing for systematic evaluation of model robustness and sensitivity to training data variations.

3.2.1 Workflow for lidar-based road detection analysis

In this study, a multi-stage computational framework was designed and implemented to enable accurate classification and uncertainty analysis of LiDAR point cloud data. The overall workflow of the proposed framework is illustrated in figure 2. In the first stage, raw LiDAR data acquired from three study areas were imported as LAS files and subjected to a comprehensive preprocessing procedure. This step involved coordinate normalization, noise removal, and initial labeling using manually annotated reference samples corresponding to the classes of main road, soil road, secondary road, sidewalk and marking lines, vegetation, soil, and vehicle. These processes yielded a unified and consistent dataset for the subsequent analytical phases. The three-dimensional points were then converted into numerical matrix structures to allow systematic computation of features for each individual point.

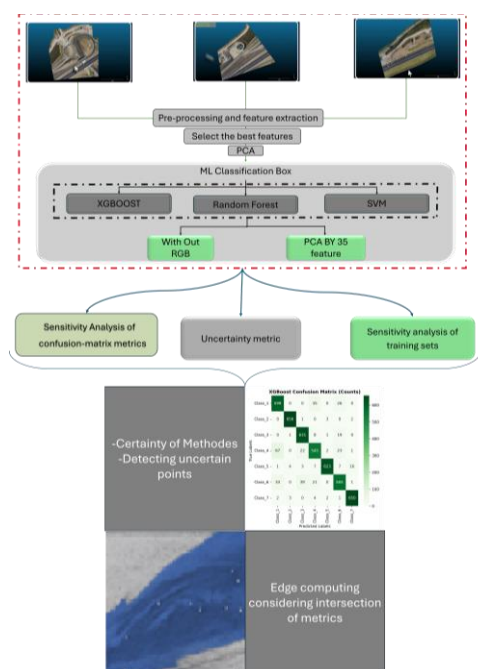


Figure 2: overall workflow

In the second stage, feature engineering was performed to quantitatively describe the geometric, statistical, and topographic patterns of the point cloud (Weinmann et al., 2015). Accordingly, a set of basic features—including normalized spatial coordinates and return intensity—was computed (Banerjee et al., 2018), alongside advanced features derived from local neighborhood analysis using the k-nearest neighbours approach (Hackel et al., 2016). The advanced features comprised height-based indices (mean, standard deviation, skewness, and kurtosis), eigenvalue-based geometric descriptors such as linearity, planarity, sphericity, and anisotropy, as well as features related to surface normal vectors and local point density (Li and Cheng, 2018; Duan et al., 2021). After feature computation, data normalization was conducted using the *StandardScaler*, and optimal features were selected through the statistical *SelectKBest* method based on the ANOVA F-value test (Son et al., 2023). To reduce dimensionality and eliminate collinearity among variables, *Principal Component Analysis* (PCA) was applied, retaining over 95% of the total data variance (Duan et al., 2021). In the final stage, the processed dataset was used to train and evaluate multiple supervised machine learning algorithms, including *XGBoost*, *Random Forest*, and *Support Vector Machine* (*SVM*) classifiers. The dataset was partitioned into training and testing subsets using stratified sampling to ensure balanced class representation. Hyperparameter optimization was performed using the *RandomizedSearchCV* approach (Alamsyah et al., 2024) with 100 iterations and evaluation metrics including *Overall Accuracy* (*OA*), *Balanced Accuracy* (*BA*), *Kappa Coefficient*, *F-Score*, and *Weighted F-Score* and Entropy. The trained models were employed not only to predict the class label of each point but also to estimate class membership probabilities. Prediction uncertainty was quantified using the entropy index from information theory. Finally, the outputs including predicted labels, class probabilities, and entropy-based uncertainty maps were exported in *CSV*, *Parquet*, and *LAS* formats, enabling both quantitative assessment and spatial interpretation of model performance.

To assess the robustness and generalization capability of the classification models, the training and prediction processes for Section 3 were iteratively repeated in five independent runs using different subsets of the training data. This approach

enabled a systematic evaluation of the models' sensitivity to variations in the training samples and provided insights into the variability of the classification outcomes. Subsequently, to quantify model uncertainty, the entropy index and other probability-based measures were employed. These evaluations served as a basis for identifying low-confidence regions within the classification maps, certainty of machine learning methods and detecting paradoxical areas of metrics intersection.

3.2.2 Uncertainty quantification

To ensure accurate, stable, and interpretable classification of LiDAR point cloud data which inherently contain complex nonlinear structures and often imbalanced distributions a strategic selection was made among supervised machine learning algorithms. This selection was based on forming a complementary triad consisting of *XGBoost*, *RF*, and *SVM*. The objective was not merely to achieve the highest accuracy but to provide a multifaceted evaluation and sensitivity analysis, wherein each model, due to its unique computational characteristics, challenges a different aspect of model robustness. Entropy serves as a statistical measure of prediction uncertainty, where higher values denote lower model confidence and greater variability, while lower values indicate more decisive classifications and higher reliability.

4. Results

In this paper, uncertainty-related analyses are primarily employed to examine trends in the data. However, these analyses are not necessarily capable of capturing higher granularities of spatial data structures, where fine-scale variations and spatial dependencies play a critical role in accurate interpretation.

4.1 Performance Analysis of Machine Learning Algorithms

In this section, the performance of the machine learning algorithms employed in the classification process was evaluated using confusion matrix for models under different data scenarios (LiDAR and LiDAR+RGB). This evaluation was conducted using statistical indicators such as Overall Accuracy (*OA*), *F-Score*, *Kappa Coefficient*, Entropy, and other methods to provide a comprehensive picture of each algorithm's efficiency in accurately and reliably extracting road features. The results obtained after optimizing the input parameters using the random search method as shown in Figure 3 show a 7% reduction in the kappa coefficient and other metrics for the *XGBoost* and *RF* methods and a 9% reduction for the *SVM* method in the no-RGB condition.

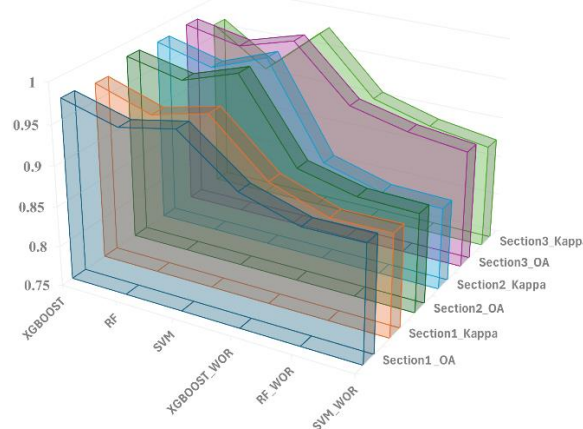


Figure 3: Comparative performance of the machine learning algorithms (XGBoost, Random Forest, and SVM) under two data configurations (LiDAR+RGB and LiDAR-only) across three study sections. "WOR" denotes the results obtained in the absence of RGB data.

Computational efficiency was systematically evaluated (Table 1). XGBoost consistently exhibited the shortest run times, whereas Random Forest demonstrated moderate computational demand. In contrast, SVM required substantially longer execution times. Exclusion of RGB data had minimal impact on XGBoost and RF but notably reduced SVM's processing time, which nevertheless remained higher than that of the other algorithms. All experiments were performed using Google Colab's Python 3 environment on a Google Compute Engine backend (TPU), equipped with 172.9 GB of system RAM and 225.3 GB of available disk space. These high-performance computing resources ensured consistent runtime comparisons among models and eliminated hardware-induced variability during benchmarking.

Table 1: Computational run times of the machine learning algorithms across three study sections under two data configurations. Times are expressed in minutes and seconds for clarity. "WOR" denotes results obtained without RGB data.

methods	Section 1	Section 2	Section 3
XGBOOST	1.36	1.12	0.55
RF	2.1	2	1.15
SVM	63	10.36	4.2
XGBOOST_WOR	1	0.58	0.56
RF_WOR	3	2.26	1.15
SVM_WOR	56	23	17.5

4.2 Uncertainty Analysis

Understanding a model's behavior and reliability requires more than evaluating its overall accuracy. In this context, Activation Neural Analysis provides a quantitative perspective on the model's uncertainty and informational content in its predictions. A fundamental statistic within this approach is entropy, which assesses the degree of unpredictability in the model's output distributions. The research results suggest that the XGBoost model exhibits consistently low and stable mean entropy values across all sections (range from 0.167 to 0.661), reflecting a high level of confidence in its predictions (Figure 4).

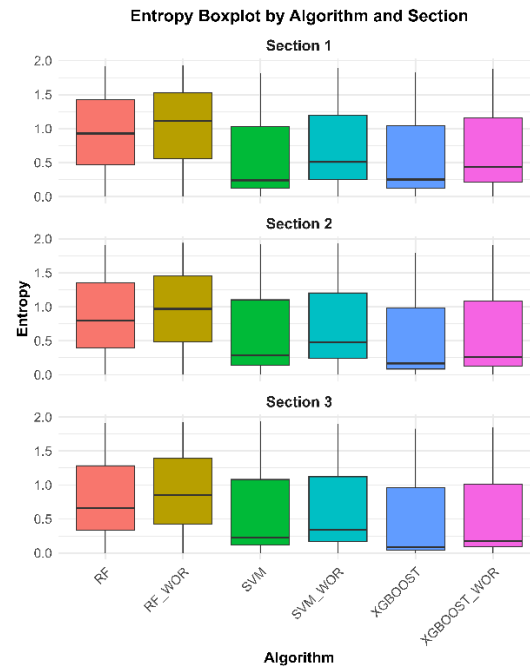


Figure 4: Comparison of the performance of the applied models across the three study sections. For each section (Sections 1–3), the boxplots represent the distribution of error or performance metric values for the evaluated models.

The maximum entropy stays around 1.9 in all scenarios, while the mean entropy somewhat rises (from 0.176 to 0.435) in the RGB-excluded version (XGBOOST_WOR). For the Random Forest model, mean entropy varies more widely across sections (0.230 to 0.929). However, in the absence of RGB data (RF_WOR), this value increases substantially (up to 1.133), indicating a strong dependence of the model on color-related features and a notable decline in prediction confidence when such information is removed. Similarly, the SVM model shows balanced mean entropy in its standard configuration (0.230 to 0.283), but the RGB-excluded version (SVM_WOR) exhibits higher entropy values (up to 0.509), signifying a reduction in predictive confidence and robustness in the absence of color information (Figure 3). The spatial distribution of SVM model uncertainty was further examined through entropy mapping (Figure 5). These maps illustrate how prediction confidence varies across different regions, with higher entropy values marking areas of increased uncertainty (Figure A1).

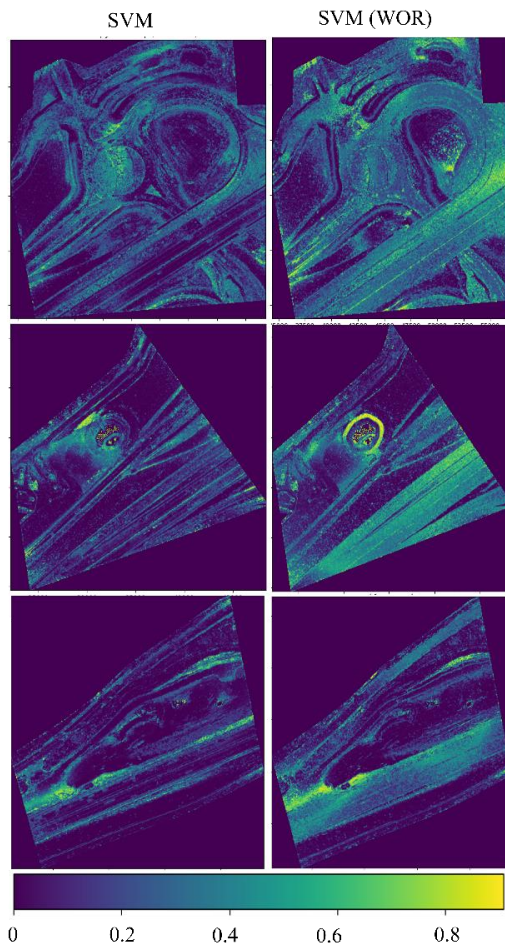


Figure 5: Visualization of advanced LiDAR processing results. Entropy heat maps reveal complex spatial patterns using the SVM model in LiDAR (WOR) and LiDAR–RGB modes, where brighter areas indicate higher variability or uncertainty. In addition, for sensitivity analysis, after selecting section 3, five random resampling iterations of the training data were performed. Classification indices were calculated for each run, with SVM showing the highest classification accuracy in each. Comparison of the mean $\pm$ standard deviation values for each method (Figure 6) confirmed the superior and consistent performance of the SVM classifier across all evaluation metrics, including BA, OA, Kappa, and F1 scores. RF and XGBoost also achieved high mean scores but their results showed slightly higher variability, especially when trained without RGB. This suggests that SVM not only provides the highest classification accuracy, but is also the most robust method in terms of sensitivity to training set composition (Figure 6).

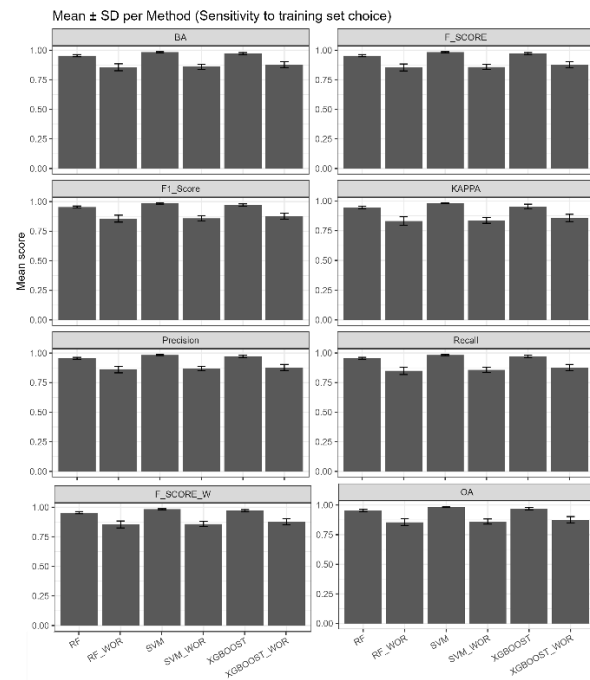


Figure 6: Mean performance ($\pm$ SD) of different Machine Learning methods (Sensitivity to training set choice). Figure 7 illustrates that the standard RF, SVM, and XGBoost models (with RGB features) consistently outperformed their WOR counterparts. Among them, SVM achieved the highest stability and accuracy, showing minimal variance. RF and XGBoost also performed well but exhibited greater dispersion. In contrast, the WOR versions showed reduced accuracy and higher variability. Moreover, outliers reflecting sensitivity to training data were more common in RF and XGBoost, while SVM showed almost none, highlighting its robustness.

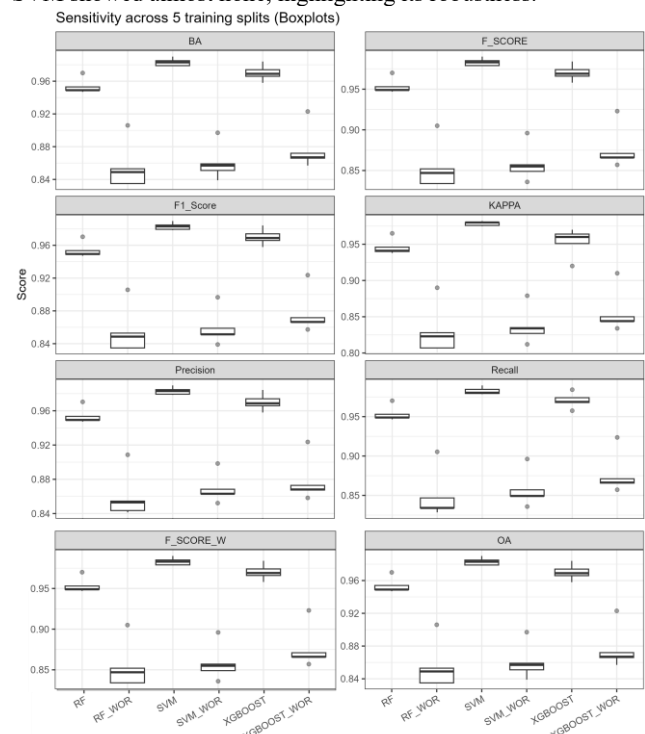


Figure 7: Comparison of the sensitivity of different models across five training splits based on various performance metrics.

The entropy statistics show the confidence of the algorithms' predictions across the five iterations (Table 2). Overall,

XGBoost had the lowest mean entropy, indicating more stable and reliable predictions. Although SVM showed the highest classification accuracy, its mean entropy was slightly higher than XGBoost, indicating local uncertainty in the predictions. RF algorithm had intermediate entropy values, and its predictions were reliable but slightly more variable than XGBoost. Omitting RGB data increased the mean entropy in all algorithms, indicating a decrease in the confidence of the predictions. In this case, SVM_WOR and RF_WOR had the highest average entropy, which emphasizes the sensitivity of the algorithms to feature combinations.

Table 2: Comparison of the performance of machine learning algorithms (XGBoost, RF, SVM) based on the mean entropy across five different sections.

Methods	Section				
	1	2	3	4	5
XGBOOST	0.084	0.143	0.131	0.121	0.119
RF	0.661	0.688	0.628	0.678	0.708
SVM	0.23	0.254	0.266	0.232	0.245
XGBOOST (WOR)	0.176	0.222	0.249	0.245	0.242
RF (WOR)	0.854	0.936	0.918	0.893	0.888
SVM (WOR)	0.339	0.391	0.398	0.398	0.39

5. Discussion

5.1 Influence of Multisource Data Fusion on Classification Accuracy, computation cost and Model Performance

The findings of this study clearly demonstrate that the integration of LiDAR and RGB data significantly enhances classification accuracy, model stability, and spatial coherence (Figure 3). This observation aligns well with (Zhuang et al., 2021), who emphasized the value of multisource data fusion in improving land-surface discrimination. In the present work, all three classifiers exhibited strong performance under the LiDAR + RGB configuration, with SVM achieving the highest OA and Kappa values. This highlights the superior capability of SVM in modeling complex nonlinear decision boundaries, particularly when high-density and spectrally diverse inputs are available.

Nevertheless, when compared to prior studies, some distinctions emerge. For instance, (Tang et al., 2023) reported that Random Forest (RF) and its improved variant outperformed other algorithms when feature redundancy reduction was applied. Conversely, our results indicate that SVM performs exceptionally well in road-surface environments characterized by high point density and well-separated object boundaries. This finding is further supported by (Aljumaily et al., 2023), who showed that geometric features extracted from LiDAR data significantly enhance classification accuracy. Similarly, we observed that factors such as point density, class diversity, adequate object spacing, and high-quality segmentation collectively contributed to improved performance—a trend consistent with (Zhang et al., 2013), who achieved comparable results using SVM-based classification.

Despite its superior accuracy, SVM exhibited the longest computational time among all models (Table 1). This result

corroborates (Tang et al., 2023), who also attributed SVM's high time cost to the quadratic complexity of convex optimization and the pairwise computation of kernel functions, particularly under RBF kernels with high-dimensional inputs. In contrast, RF and XGBoost demonstrated better scalability due to their tree-based and parallel optimization structures, allowing faster training and more practical application to large datasets; while SVM achieved the highest accuracy, its computational inefficiency may limit its suitability for large-scale or near-real-time road-surface mapping tasks. Overall, XGBoost offers near real-time processing, making it suitable for operational mapping. RF remains a balanced choice—efficient, interpretable, and stable—useful for agencies performing routine LiDAR classifications, SVM, while potentially accurate, becomes computationally impractical for large-scale LiDAR datasets.

Under the LiDAR-only (WOR) configuration, the removal of spectral information resulted in a noticeable decline in accuracy—approximately 5–10% across all metrics. This degradation underscores the crucial role of RGB data in improving spectral-spatial separability and reducing class collinearity. Without spectral cues, models struggled to distinguish between materials with similar geometric characteristics, especially in spectrally homogeneous regions such as asphalt and paved surfaces. Interestingly, XGBoost outperformed other models in the WOR scenario, which agrees with (Ramdani and Furqon, 2022), who found that XGBoost can perform robustly even in low-feature environments. This is attributed to its boosting mechanism that sequentially corrects errors and constructs stable hierarchical patterns, whereas SVM, deprived of spectral information, forms weaker and less stable decision boundaries.

5.2 Model Stability, Uncertainty Analysis, and the Role of Spectral Features

Model stability and sensitivity analysis across five independent training splits (Figures 6 and 7) further confirmed the importance of RGB data. The inclusion of spectral features led to narrower boxplots and reduced variance, indicating higher robustness and lower sensitivity to training data variation. Although SVM maintained the highest median OA and F1 scores, it also exhibited slightly higher variance in Precision and Recall, reflecting its dependence on training composition. XGBoost, in contrast, achieved comparable accuracy with lower dispersion, underscoring its strong generalization and consistent convergence. RF showed the largest variability under WOR conditions, revealing greater susceptibility to random sampling and reduced feature diversity. These patterns align with (Reisi Gahrouei et al., 2024), who suggested that SVM is well-suited for high-dimensional data, while ensemble methods such as RF and XGBoost effectively reduce variance in complex environments. The absence of outliers in the LiDAR + RGB configuration (Figure 7) further confirms the stability and reliability of these two models.

A central aspect of this study was the use of entropy as an indicator of model uncertainty. Unlike (Westphal et al., 2025),

who investigated deep learning models with dropout-based stochasticity, the current research employed classical machine learning algorithms (SVM, RF, XGBoost) whose outputs are probabilistically stable. Consequently, entropy was computed directly from the final probability distributions, providing a more faithful representation of model confidence. The results revealed a strong inverse relationship between spectral information and entropy: the inclusion of RGB data consistently reduced entropy values across all models, indicating increased classification confidence. Among the three algorithms, XGBoost exhibited the lowest entropy (0.084–0.143), reflecting its precise decision boundaries and reliable predictions. This observation aligns with (Chen et al., 2015) and (Liu et al., 2025), who demonstrated that the use of Information Gain and log-loss (cross-entropy) optimization in XGBoost promotes more cautious decision-making, thereby lowering uncertainty. In contrast, RF produced the highest entropy due to its inherent stochastic nature, while SVM, despite high accuracy, yielded slightly less confident predictions than XGBoost.

The spatial distribution of uncertainty, visualized through entropy heatmaps (Figure 5), further substantiates these findings. In the LiDAR + RGB mode, most road surfaces exhibited low entropy (blue–green tones), indicating coherent and confident classifications. Conversely, the WOR configuration produced broader regions of high uncertainty (yellow–green hues), particularly along road boundaries and central lanes. This spatial pattern suggests that, in the absence of spectral data, models rely primarily on geometric attributes—such as intensity and elevation—which are less effective in capturing subtle material variations. Sections 1 and 2 displayed the most pronounced improvements under RGB integration, while residual uncertainty in Section 3 was likely caused by LiDAR-specific effects such as shadowing and occlusion.

In summary, the integration of LiDAR and RGB data not only enhances accuracy but also strengthens model stability and confidence. Among the evaluated classifiers, SVM achieved the highest accuracy yet suffered from excessive computational costs, whereas XGBoost offered the most balanced trade-off between speed, accuracy, and confidence. RF, although robust, showed higher sensitivity and uncertainty when spectral information was omitted. Finally, entropy proved to be an effective and interpretable metric for quantifying model confidence, demonstrating a clear correlation with the presence or absence of spectral features.

6. Conclusion

The integrated evaluation of XGBoost, Random Forest (RF), and Support Vector Machine (SVM) across multiple LiDAR-derived datasets—supported by statistical, computational, and entropy-based analyses—provides a holistic understanding of each algorithm's strengths, limitations, and internal consistency. The comparison across *LiDAR + RGB* and *LiDAR-only (WOR)* configurations demonstrates clear and quantifiable trade-offs between accuracy, uncertainty, and efficiency.

Across all sections, the inclusion of RGB information consistently enhanced performance metrics such as Overall Accuracy (OA), F1-score, and Kappa coefficient, while simultaneously reducing mean entropy—a proxy for model uncertainty. Among the models, SVM achieved the highest

statistical accuracy (OA and Kappa exceeding 0.97) and produced coherent, well-defined classification boundaries, confirming its strength in modeling nonlinear feature spaces. XGBoost, while marginally lower in absolute accuracy, yielded the lowest entropy values (≈ 0.1), indicating the most confident and consistent probabilistic predictions. In contrast, RF showed the highest entropy (≈ 0.65 – 0.70) and the widest performance variability, reflecting its sensitivity to feature noise and random training variability.

Under WOR conditions, performance declined across all models, with entropy values increasing substantially—most sharply for RF_WOR (up to 0.93) and moderately for SVM_WOR (≈ 0.39). Interestingly, although SVM retained the highest statistical accuracy in the absence of RGB data, it also exhibited slightly higher entropy than XGBoost, suggesting a paradox between accuracy and confidence: SVM correctly classified more samples overall but did so with greater uncertainty in probability space. Conversely, XGBoost displayed lower entropy and more stable decision boundaries even when its accuracy metrics were marginally lower—revealing that it produced more confident, though not always more correct, classifications.

A second paradox emerges when computational performance is considered. XGBoost, despite being the most confident model (lowest entropy), is also the most computationally efficient, completing classifications in under 1.5 minutes per section. In contrast, SVM, the statistically most accurate algorithm, required up to 63 minutes per section, highlighting a trade-off between accuracy and computational scalability. RF occupied the middle ground, balancing runtime and interpretability but suffering from higher uncertainty and lower consistency.

Entropy and sensitivity analyses further confirm that models with low entropy also exhibit reduced variability across training splits, validating entropy as an effective measure of predictive stability. In visual terms, entropy heat maps showed concentrated uncertainty at road boundaries when RGB data were included but diffuse, spatially widespread uncertainty under WOR conditions—especially for RF. These spatial patterns reinforce that spectral features play a decisive role in constraining uncertainty propagation in LiDAR-based classification.

In conclusion, while SVM remains the most statistically accurate, XGBoost demonstrates the best balance of accuracy, confidence, and efficiency, making it the most reliable algorithm for operational large-scale road classification, especially in edge-computing or near-real-time applications. The observed paradoxes—high accuracy but moderate confidence (SVM), and moderate accuracy but high confidence and speed (XGBoost)—highlight the importance of combining both statistical and entropy-based metrics for comprehensive model evaluation. These findings collectively affirm that entropy complements traditional performance indicators, enabling deeper insight into model reliability, uncertainty localization, and the true robustness of machine learning approaches in spatial data analysis.

Based on the findings, we recommend that practitioners prioritize XGBoost for large-scale road asset mapping tasks where both reliability and computational efficiency are critical, while considering SVM when highest possible classification accuracy is required and computational resources are constrained.

References

Alamsyah, N., Budiman, B., Yoga, T. P., and Alamsyah, R. Y. R.: XGBoost hyperparameter optimization using

- RandomizedSearchCV for accurate forest fire drought condition prediction, *Jurnal Pilar Nusa Mandiri*, 20, 103-110, 2024.
- Aljumaily, H., Laefer, D. F., Cuadra, D., and Velasco, M.: Point cloud voxel classification of aerial urban LiDAR using voxel attributes and random forest approach, *International Journal of Applied Earth Observation and Geoinformation*, 118, 103208, 2023.
- Arij, N., Latifi, H., Fakhri, A., and Esmaili, R.: Four decades of spatio-temporal trends in Miankaleh Wetland's water body and vegetation as revealed by remote sensing time series, *Ecological Informatics*, 103374, 2025.
- Aziz, Z., Alzaabi, E., and Syazli Fathi, M.: Prioritisation of resilience criteria and performance indicators for road emergencies crisis response: An analytic hierarchy process (AHP) approach, *Journal of Financial Management of Property and Construction*, 28, 178-205, 2023.
- Banerjee, P., Bhunia, A. K., Bhattacharyya, A., Roy, P. P., and Murala, S.: Local neighborhood intensity pattern—a new texture feature descriptor for image retrieval, *Expert Systems with Applications*, 113, 100-115, 2018.
- Boyko, A. and Funkhouser, T.: Extracting roads from dense point clouds in large scale urban environment, *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, S2-S12, 2011.
- Buğday, E.: Capabilities of using UAVs in forest road construction activities, *European Journal of Forest Engineering*, 4, 56-62, 2018.
- Chen, T. and Guestrin, C.: Xgboost: A scalable tree boosting system, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794, 2015.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., and Zhou, T.: Xgboost: extreme gradient boosting, *R package version 0.4-2*, 1, 1-4, 2015.
- Diallo, R., Edalo, C., and Awe, O. O.: Machine learning evaluation of imbalanced health data: a comparative analysis of balanced accuracy, MCC, and F1 score, in: *Practical Statistical Learning and Data Science Methods: Case Studies from LISA 2020 Global Network*, USA, Springer, 283-312, 2024.
- Döllner, J.: Geospatial artificial intelligence: potentials of machine learning for 3D point clouds and geospatial digital twins, *PFG—Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88, 15-24, 2020.
- Duan, Y., Yang, C., Chen, H., Yan, W., and Li, H.: Low-complexity point cloud denoising for LiDAR by PCA-based dimension reduction, *Optics Communications*, 482, 126567, 2021.
- Farhadpour, S., Warner, T. A., and Maxwell, A. E.: Selecting and interpreting multiclass loss and accuracy assessment metrics for classifications with class imbalance: Guidance and best practices, *Remote Sensing*, 16, 533, 2024.
- Fiorentini, N. and Losa, M.: Road Detection, Monitoring, and Maintenance Using Remotely Sensed Data, 2025.
- Gargoum, S. and El-Basyouny, K.: Automated extraction of road features using LiDAR data: A review of LiDAR applications in transportation, 2017 4th International Conference on Transportation Information and Safety (ICTIS), 563-574, 2017.
- Hackel, T., Wegner, J. D., and Schindler, K.: Fast semantic segmentation of 3D point clouds with strongly varying density, *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, 3, 177-184, 2016.
- Kendall, A. and Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision?, *Advances in neural information processing systems*, 30, 2017.
- Khan, M. J. and Singh, P. P.: Advanced road extraction using CNN-based U-Net model and satellite imagery, *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, 5, 100244, 2023.
- Khanam, J. J. and Foo, S. Y.: A comparison of machine learning algorithms for diabetes prediction, *Ict Express*, 7, 432-439, 2021.
- Laurance, W. F., Campbell, M. J., Alamgir, M., and Mahmoud, M. I.: Road expansion and the fate of Africa's tropical forests, *Frontiers in Ecology and Evolution*, 5, 75, 2017.
- Laurance, W. F., Clements, G. R., Sloan, S., O'connell, C. S., Mueller, N. D., Goosem, M., Venter, O., Edwards, D. P., Phalan, B., and Balmford, A.: A global strategy for road building, *Nature*, 513, 229-232, 2014.
- Li, F., Zhang, J., Huang, J., Huang, D., Yang, J., Song, Y., and Zeng, G.: Heavy metals in road dust from Xiandao District, Changsha City, China: characteristics, health risk assessment, and integrated source identification, *Environmental Science and Pollution Research*, 23, 13100-13113, 2016.
- Li, L., Wang, W., Khatoon, A., and Wang, M.: Road detection by using gradient magnitude image and adaptive thresholding based watershed, *Egyptian Informatics Journal*, 25, 100446, 2024.
- Li, Q. and Cheng, X.: Comparison of different feature sets for TLS point cloud classification, *Sensors*, 18, 4206, 2018.
- Liu, R., Wu, J., Lu, W., Miao, Q., Zhang, H., Liu, X., Lu, Z., and Li, L.: A review of deep Learning-Based methods for road extraction from High-Resolution remote sensing images, *Remote Sensing*, 16, 2056, 2024.
- Liu, X., Mou, X., Zhang, Y., Chen, H., Wang, X., Sun, X., Wang, C., Shong, R., and Li, Y.: EDHD-FG-XGBoost: Entropy-driven high-dimensional feature gain XGBoost model for short-term wind power prediction and application, *AIP Advances*, 15, 2025.
- Lu, X., Zhong, Y., Zheng, Z., and Zhang, L.: GAMSNet: Globally aware road detection network with multi-scale residual learning, *ISPRS Journal of Photogrammetry and Remote Sensing*, 175, 340-352, 2021.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A.: Deep learning in remote sensing applications: A meta-analysis and review, *ISPRS journal of photogrammetry and remote sensing*, 152, 166-177, 2019.
- Matikainen, L., Karila, K., Litkey, P., Ahokas, E., and Hyypä, J.: Combining single photon and multispectral airborne laser

scanning for land cover classification, *ISPRS Journal of Photogrammetry and Remote Sensing*, 164, 200-216, 2020.

Mohammadzadeh, A. and Zoj, M. J. V.: A self-organizing fuzzy segmentation (SOFS) method for road detection from high resolution satellite images, *Photogrammetric Engineering & Remote Sensing*, 76, 27-35, 2010.

Olimid, A. P. and Olimid, D. A.: Societal challenges, population trends and human security: evidence from the public governance within the United Nations publications (2015-2019), *Revista de Stiinte Politice*, 53-64, 2019.

Pal, M.: Random forest classifier for remote sensing classification, *International journal of remote sensing*, 26, 217-222, 2005.

Quackenbush, L. J., Im, I., and Zuo, Y.: Road extraction: A review of LiDAR-focused studies, *Remote sensing of natural resources*, 155-169, 2013.

Ramdani, F. and Furqon, M. T.: The simplicity of XGBoost algorithm versus the complexity of Random Forest, Support Vector Machine, and Neural Networks algorithms in urban forest classification, *F1000Research*, 11, 1069, 2022.

Reisi Gahrouei, O., Côté, J.-F., Bournival, P., Giguère, P., and Béland, M.: Comparison of deep and machine learning approaches for quebec tree species classification using a combination of multispectral and LiDAR data, *Canadian Journal of Remote Sensing*, 50, 2359433, 2024.

Schreiber, F.: The long road from urban experimentation to the transformation of urban planning practice: The case of tactical urbanism in the city of Barcelona, *Journal of Urban Mobility*, 7, 100100, 2025.

Son, S., Lee, W., Jung, H., Lee, J., Kim, C., Lee, H., Cho, S., Jang, J., Lee, M., and Ryu, H.-C.: Experimental analysis of various blockage performance for LiDAR sensor cleaning evaluation, *Sensors*, 23, 2752, 2023.

Song, M. and Civco, D.: Road extraction using SVM and image segmentation, *Photogrammetric Engineering & Remote Sensing*, 70, 1365-1371, 2004.

Suleymanoglu, B., Soykan, M., and Toth, C.: 3d road boundary extraction based on machine learning strategy using lidar and image-derived mms point clouds, *Sensors*, 24, 503, 2024.

Suthaharan, S.: Support vector machine, in: *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*, Springer, 207-235, 2016.

Tang, Q., Zhang, L., Lan, G., Shi, X., Duanmu, X., and Chen, K.: A classification method of point clouds of transmission line corridor based on improved random forest and multi-scale features, *Sensors*, 23, 1320, 2023.

Valero, S., Chanussot, J., Benediktsson, J. A., Talbot, H., and Waske, B.: Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images, *Pattern Recognition Letters*, 31, 1120-1127, 2010.

Wang, W., Li, L., and Han, Y.: Crack detection in shadowed images on gray level deviations in a moving window and

distance deviations between connected components, *Construction and Building Materials*, 271, 121885, 2021a.

Wang, W., Berholm, F., Hu, K., Zhao, L., Feng, S., Tu, A., and Fan, E.: Lane line extraction in raining weather images by ridge edge detection with improved MSR and Hessian matrix, *Information Technology and Control*, 50, 722-735, 2021b.

Weinmann, M., Jutzi, B., Hinz, S., and Mallet, C.: Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers, *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 286-304, 2015.

Wen, C., Sun, X., Li, J., Wang, C., Guo, Y., and Habib, A.: A deep learning framework for road marking extraction, classification and completion from mobile laser scanning point clouds, *ISPRS journal of photogrammetry and remote sensing*, 147, 178-192, 2019.

Westphal, F., Lidberg, W., Busarello, M. D. S. T., and Ågren, A. M.: Uncertainty quantification for LiDAR-based maps of ditches and natural streams, *Environmental Modelling & Software*, 106488, 2025.

Xu, Y., Xie, Z., Feng, Y., and Chen, Z.: Road extraction from high-resolution remote sensing imagery using deep learning, *Remote Sensing*, 10, 1461, 2018.

Zhang, J., Lin, X., and Ning, X.: SVM-based classification of segmented airborne LiDAR point clouds in urban areas, *Remote sensing*, 5, 3749-3775, 2013.

Zhang, L., Zhang, L., and Du, B.: Deep learning for remote sensing data: A technical tutorial on the state of the art, *IEEE Geoscience and remote sensing magazine*, 4, 22-40, 2016.

Zhuang, Z., Li, R., Jia, K., Wang, Q., Li, Y., and Tan, M.: Perception-aware multi-sensor fusion for 3d lidar semantic segmentation, *Proceedings of the IEEE/CVF international conference on computer vision*, 16280-16290,

Appendix A

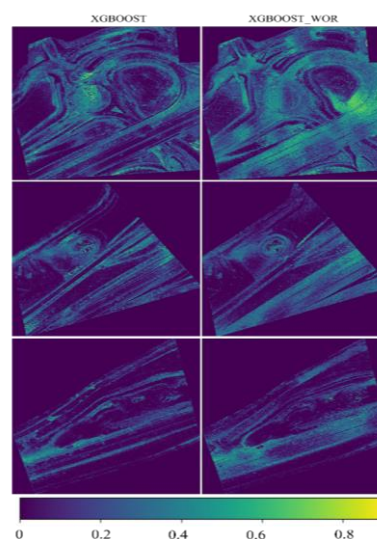


Figure A1: Results of the 7-class classification entropy using the XGBOOST method under two scenarios: with RGB and without RGB (XGBOOST (WOR)). The shapes are in columns corresponding to sections 1 to 3 respectively