

Road surface material classification from unmanned aerial vehicle red, green, blue and multispectral imagery using spectral features and ensemble machine learning model

Marcin Bilski¹ 

¹ Institute of Civil Engineering, Faculty of Civil and Transport Engineering, Poznan University of Technology, ul. Piotrowo 5, 61-138 Poznan, Poland
E-mail: marcin.bilski@put.poznan.pl

ABSTRACT

The aim of this study was to evaluate the impact of incorporating multispectral data acquired from an unmanned aerial vehicle (UAV) as an extension of standard RGB imagery on the accuracy of road surface material classification using machine learning algorithms. The research was conducted on six road sections differing in pavement type and technical condition. The performance of three ensemble learning algorithms: Random Forest, XGBoost, and LightGBM was investigated. Three input data configurations were considered: RGB imagery, a seven-band dataset comprising RGB channels and four multispectral bands (MS7), and an extended dataset (MS9) additionally including the NDVI and NDRE vegetation indices. For the best-performing model, the influence of texture features calculated within a moving 5×5 pixel window and edge information derived using the Sobel operator was also evaluated. The results demonstrated that the inclusion of multispectral data significantly improved classification performance compared with RGB imagery alone. The highest accuracy was achieved by the LightGBM model using the MS9 dataset, reaching an overall accuracy (OA) of 81.31%. The incorporation of texture features further increased the classification accuracy to 93.30% OA. The findings confirm the positive contribution of UAV-acquired multispectral data to the effectiveness of road surface material classification.

Keywords: unmanned aerial vehicle (UAV), multispectral imaging, road pavements, machine learning, texture features, Random Forest, XGBoost, LightGBM.

INTRODUCTION

In recent years, the rapid development of unmanned aerial vehicle (UAV) technology, image processing techniques, and machine learning methods has created new opportunities for the acquisition and analysis of spatial data in civil engineering. High-resolution aerial imagery and photogrammetric products, such as orthomosaics and 3D models, can now be automatically analysed using artificial intelligence algorithms, enabling efficient identification of infrastructure objects, assessment of their technical condition, and monitoring of changes occurring in their surroundings. Increasingly, the integration of remote sensing data with machine learning techniques

is recognized as one of the key tools supporting modern transportation infrastructure management systems [1–4].

The identification of road surface materials and pavement conditions using remote sensing data represents an important research direction at the intersection of photogrammetry, remote sensing, and transportation infrastructure management systems. Traditional pavement assessment methods rely primarily on field inspections or specialized survey vehicles, which are associated with high costs, considerable time requirements, and limited data update frequency. The development of unmanned aerial vehicles (UAVs) and automated image analysis methods has significantly enhanced the possibilities for acquiring information on road

infrastructure [5, 6]. Very high-resolution imagery, supported by advanced machine learning (ML) algorithms, including deep learning (DL) approaches, now enables not only the extraction of road networks but also the assessment of pavement conditions and the detection of surface defects [7, 8]. Most existing studies [5–9], however, focus on road detection, crack identification, rutting and pothole recognition, and pavement degradation monitoring. Considerably less attention has been devoted to the identification of the materials from which road surfaces are constructed. Previous studies [10–12] have demonstrated the potential of multispectral and hyperspectral data for pavement material identification. Analyses based on airborne and satellite imagery have shown that asphalt, concrete, aggregates, and unpaved surfaces exhibit distinct spectral signatures that enable their automatic discrimination. Particular attention has been given to the near-infrared and short-wave infrared spectral regions, which often provide greater material separability than conventional RGB bands. At the same time, increasing interest has been observed in the use of UAVs as platforms for acquiring data for surface material classification [13, 14]. Compared with satellite imagery, UAV systems provide substantially higher spatial resolution, which is particularly important in urban environments and transportation infrastructure areas where individual pavement elements are often relatively small [15]. Despite these advantages, the number of studies directly addressing the classification of road surface materials using UAV-based multispectral imagery remains limited. Information on road construction materials is important not only for transportation infrastructure management but also for environmental analyses related to surface sealing, urban heat balance assessment, and storm water runoff modelling [16, 17]. The authors of [12] demonstrated the possibility of distinguishing asphalt pavements with different levels of degradation using hyperspectral data. More recent studies have focused on integrating remote sensing data with machine learning methods. For example, in [18], an automated framework combining remote sensing data, GIS, and machine learning techniques was developed for extracting information on road surfaces. Using high-resolution multispectral imagery, the authors achieved 89% classification accuracy for asphalt and concrete pavements, highlighting the considerable potential of airborne data for road infrastructure inventory applications. However, the practical use of hyperspectral data is

limited by their high acquisition costs, processing complexity, and lower availability compared with RGB and multispectral sensors commonly mounted on UAV platforms. Despite their high classification potential, hyperspectral datasets are still relatively rarely employed in operational applications. Multispectral imaging represents a compromise between conventional RGB imagery and advanced hyperspectral data. Multispectral sensors capture additional portions of the electromagnetic spectrum, allowing a more comprehensive characterization of surface spectral properties than standard RGB cameras. Studies on land-cover classification have shown that combining RGB and multispectral imagery can improve classification performance, particularly in heterogeneous environments [19, 20]. The application of multispectral imaging to road surface material identification is based on the phenomenon of differential electromagnetic radiation reflectance exhibited by various materials. Asphalt, cement concrete, stone paving blocks, and unpaved surfaces are characterized by distinct spectral response curves resulting from differences in their mineral composition, physical properties, and degree of weathering. These differences are often most pronounced outside the visible spectrum, particularly within the red-edge and near-infrared wavelength regions [21]. Spectral indices such as NDVI, NDRE, and other band combinations are most commonly used for vegetation condition assessment. However, they can also serve as additional features describing relationships among spectral bands [22]. In remote sensing image classification, such indices are frequently employed as auxiliary variables that enhance class separability [23]. In the context of road material classification, their role is not necessarily related to the direct analysis of vegetation conditions but rather to enriching the model feature space and improving the discrimination of road surfaces from adjacent areas, bare soil, and vegetation.

In recent years, ensemble learning algorithms have become among the most widely used methods for remote sensing data classification [24]. Their popularity stems from their high effectiveness in analysing multidimensional image datasets, their ability to model complex nonlinear relationships, and their strong generalization capabilities [25]. A particularly important algorithm in remote sensing applications is Random Forest, originally proposed by Breiman [26]. According to the review conducted by Belgiu and Drăguț [27], Random Forest is among the most frequently applied

classifiers for satellite and airborne imagery. The authors emphasize that the method achieves high classification accuracy while reducing the risk of model overfitting and simultaneously providing measures of the relative importance of input variables. In recent years, increasing attention has also been devoted to gradient boosting algorithms. XGBoost, developed by Chen and Guestrin [28], was designed for efficient processing of large datasets and optimization of the learning process. Through the use of regularization techniques and advanced decision-tree construction mechanisms, the algorithm achieves high performance in both classification and regression tasks. LightGBM, in turn, employs histogram-based tree growth and a leaf-wise growth strategy, significantly reducing training time while maintaining high classification accuracy [29]. Several studies have demonstrated the high effectiveness of both ensemble learning methods and modern machine learning and deep learning techniques in remote sensing classification tasks [30, 31]. Beyond classification performance, ensemble learning algorithms also provide the ability to identify the most informative spectral features. Feature importance analysis has become one of the key directions in the development of machine learning methods for remote sensing applications [27, 32–34]. In addition to evaluating classification accuracy, it enables the determination of which spectral bands and derived indices contribute most strongly to the model decision-making process. Such information can be used both to optimize future data acquisition systems and to improve understanding of the physical relationships between material properties and their spectral responses. For road surface materials, this issue remains relatively underexplored, particularly with respect to multispectral data acquired from UAV platforms. This aspect is especially important for UAV-based multispectral imagery, where the contribution of individual spectral bands and derived indices remains insufficiently investigated in the context of road surface material classification [13, 34].

RESEARCH PROBLEM

Despite the growing number of studies employing UAV imagery and machine learning methods for transportation infrastructure monitoring, existing research has focused primarily on pavement condition assessment and the automated

detection of defects such as cracks, potholes, and surface deformations. Considerably less attention has been devoted to the classification of road surface materials using remote sensing data, particularly data acquired from UAV platforms. Although multispectral sensors provide information beyond that captured by conventional RGB cameras, their actual contribution to the classification accuracy of road surface materials remains insufficiently understood. Previous studies have demonstrated the potential of multispectral and hyperspectral data for distinguishing different pavement types. However, relatively few investigations have examined the extent to which supplementing standard RGB imagery with additional spectral information improves classification performance. Consequently, it remains unclear which spectral regions provide the most useful information for discriminating among asphalt, concrete, stone, and unpaved road surfaces. Another limitation of existing research is the relatively small number of studies comparing the performance of modern ensemble learning algorithms in remote sensing applications. Although algorithms such as Random Forest, XGBoost, and LightGBM are increasingly applied in land-cover classification, infrastructure object detection, and road surface analysis [23, 35–37], studies enabling their direct comparison for road surface material classification based on UAV-derived multispectral data are still scarce. In particular, there is a lack of research comparing these methods using the same dataset and an identical set of input features. At the initial stage of the study, semantic segmentation architectures, particularly U-Net, were considered because of their widespread use in UAV image analysis. However, preliminary experiments indicated that the available dataset, comprising only six independent study sites, was insufficient to achieve satisfactory model generalization. Since the primary objective of this study was to evaluate the contribution of multispectral information rather than to compare different artificial intelligence architectures, the experimental framework was based on ensemble learning algorithms (Random Forest, XGBoost, and LightGBM). This approach enabled the influence of successive groups of input features, including multispectral bands, spectral indices, and texture descriptors, to be assessed while keeping the classifier family unchanged and minimizing the influence of model architecture on the obtained results. Therefore, the research problem addressed in this study

concerns the evaluation of the impact of extending standard RGB data with additional multispectral bands on the accuracy of road surface material classification. Particular emphasis is placed on comparing the results obtained from different input data configurations and assessing the capabilities of selected ensemble learning algorithms for road surface material identification.

RESEARCH METHODOLOGY

Study area and UAV data acquisition

The study was conducted on six road sites located in the Wielkopolska Voivodeship, Poland (Figure 1). The selected areas represented the most common pavement types found within the transportation network, including cement concrete pavements, precast concrete block pavements, stone block pavements, and asphalt pavements exhibiting varying levels of deterioration. In addition, the investigated sites contained accompanying features such as unpaved shoulders, roadside ditches, road markings, stormwater inlets and metal manhole covers, as well as surrounding vegetation. The characteristics of the individual study sites are presented in Table 1. The study sites were intentionally selected to represent the most common pavement materials occurring in the Polish road network, while simultaneously ensuring variation in pavement age, technical condition, surrounding land cover, and the presence

of accompanying infrastructure. This allowed the evaluation of the proposed methodology under different surface characteristics while maintaining comparable acquisition conditions.

Data were acquired between July and September 2025 using a DJI Mavic 3 Multispectral unmanned aerial vehicle (UAV). The platform was equipped with a 20 MP RGB camera (5280×3956 pixels) and four 5 MP multispectral cameras (2592×1944 pixels each), capturing reflected radiation in the Green (560 ± 16 nm), Red (650 ± 16 nm), Red Edge (730 ± 16 nm), and Near Infrared (860 ± 26 nm) spectral bands. In addition, the system incorporated a sunlight sensor that enabled compensation for variations in illumination conditions during flight operations. The DJI Mavic 3 Multispectral was also equipped with a GNSS RTK (real-time kinematic) receiver, allowing the coordinates of image projection centres to be recorded with centimetre-level accuracy. To ensure high georeferencing accuracy of the orthomosaics and derived photogrammetric products, between five and nine ground control points (GCPs) were established at each study site. The coordinates of the GCPs were determined using the GNSS RTK surveying method in the ETRF2000-PL/CS2000 horizontal reference system (Zone 5), while elevation data were referenced to the PL-EVRF2007-NH vertical datum. The control points were distributed evenly across each survey area, covering both the central portions of the scene and its boundaries, in accordance with recommendations for low-altitude

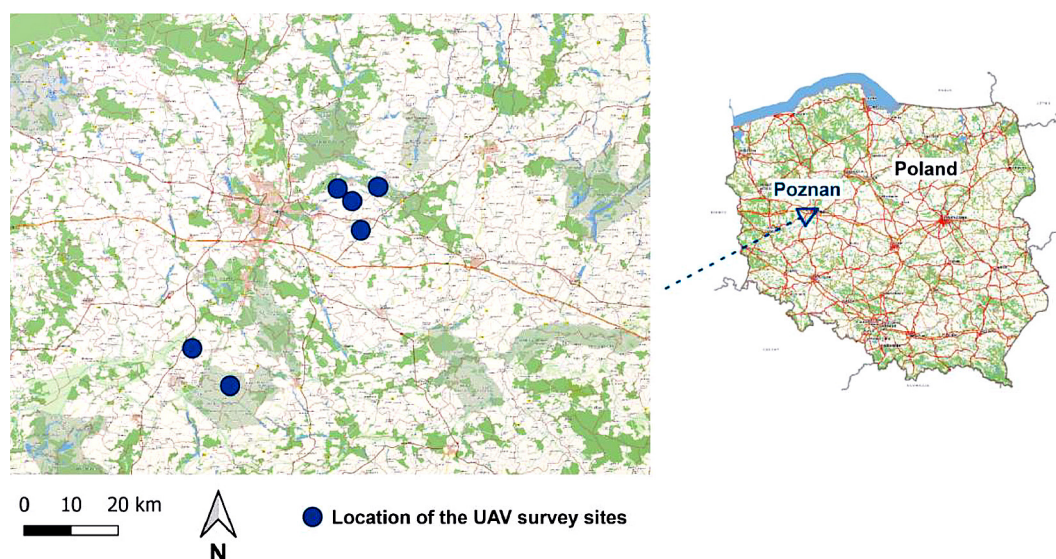


Figure 1. Location of the UAV survey sites used for orthomosaic acquisition and subsequent machine-learning-based classification analysis

Table 1. Characteristics of the UAV survey sites, including pavement type, site features, and geometric parameters

Site ID	Predominant pavement material	Site characteristics	Length [m]	Area [m ²]
A1	Asphalt pavement (good condition)	Asphalt road section with an unpaved road segment and a roadside ditch	239	4183
A2	Asphalt pavement (poor condition)	Asphalt road in poor condition located in an agricultural area, with a railway crossing	147	4946
A3	Asphalt pavement with drainage ditch (poor condition)	Asphalt road section in poor technical condition, with a roadside ditch and an asphalt cycle path	253	10434
C1	Concrete pavement	Concrete road with an asphalt cycle path, pedestrian crossing, and concrete block driveways	310	4235
CB1	Concrete block pavement	Concrete block parking area with sidewalks and linear drainage inlets	-	3196
SB1	Stone block pavement	Rural cobblestone road partially covered by tree canopy	362	7844

UAV photogrammetry [38, 39]. A subset of the points was used as control points during photogrammetric processing, whereas the remaining points served as independent check points for the assessment of georeferencing accuracy.

All UAV surveys were conducted under comparable weather conditions, namely on sunny days with limited cloud cover, between 13:00 and 17:00 local time. During data acquisition, local shadowing caused by tree canopies and road infrastructure elements was present; however, illumination conditions were generally consistent across all investigated sites. All flight missions were performed at a nominal altitude of 30 m above ground level (AGL), with a planned forward overlap of 70% and a side overlap of 80%. These parameters are consistent with recommendations for acquiring high-resolution UAV photogrammetric data intended for transportation infrastructure analysis and land-cover classification applications [40, 41]. For each study site, both RGB and multispectral imagery were collected. The acquired datasets were subsequently processed in Agisoft Metashape Professional version 2.2.0. The workflow included image alignment, dense point cloud generation, digital elevation model (DEM) creation, and orthomosaic production. Depending on the size of the study area and the number of acquired images, the photogrammetric reconstruction yielded approximately $6 \cdot 10^4$ to $1.6 \cdot 10^6$ tie points and from approximately $1.1 \cdot 10^7$ to $9.6 \cdot 10^8$ dense point cloud points. Project georeferencing was based on GCPs distributed uniformly throughout the study areas. Reconstruction accuracy was evaluated using both control points and independent check points, following commonly adopted validation procedures for UAV

photogrammetric products [38, 42]. The resulting orthomosaics were characterized by very high spatial resolution. For the RGB datasets, the ground sampling distance (GSD) ranged from 0.78 to 0.98 cm/pixel, whereas for the multispectral datasets it ranged from 1.33 to 1.57 cm/pixel (Table 2). Georeferencing accuracy was assessed using the root mean square error (RMSE) calculated for both control points and check points. The obtained RMSE values ranged from 2.2 cm to 4.1 cm for RGB control points and from 6.1 cm to 8.1 cm for multispectral datasets. For the independent check points, the errors did not exceed 12.8 cm (Table 3). The achieved georeferencing accuracy was several times higher than the dimensions of the analysed pavement features and ensured the correct spatial positioning of the reference data used during the training of the classification models. Given that the analyses focused on the classification of pavement materials within relatively homogeneous road surface areas, the obtained accuracy level was considered sufficient for further data processing. The achieved RMSE values are comparable to, or lower than, those reported in previous studies employing UAV photogrammetry supported by GNSS RTK-measured ground control points [38].

Preliminary data processing and reference dataset preparation

Following the completion of photogrammetric processing, the RGB and multispectral orthomosaics were exported separately from Agisoft Metashape Professional. Due to the physical separation of the RGB and multispectral sensors mounted on the UAV platform, slight spatial offsets

Table 2. UAV data acquisition parameters for the surveyed sites

Site ID	Number of images [-]	RGB GSD [cm/pix]	MS GSD [cm/pix]	Number of GCPs [-]
A1	212	0.83	1.43	5
A2	67	0.93	1.57	5
A3	178	0.98	1.43	7
C1	122	0.78	1.33	7
CB1	74	0.84	1.44	9
SB1	601	0.80	1.36	5

were observed between the resulting products, particularly along the boundaries of objects characterized by high spectral contrast. In addition, the two datasets exhibited different ground sampling distances (GSDs). To ensure pixel-level alignment across all image channels, an image co-registration procedure was performed. This process involved aligning the RGB orthomosaics to the multispectral data grid and resampling the RGB imagery using cubic interpolation. Subsequently, the RGB and multispectral layers were merged into a unified seven-band orthomosaic comprising three RGB channels (Red, Green, and Blue) and four multispectral channels (Green, Red, Red Edge, and Near Infrared, NIR). The data were further clipped to a common area of interest in order to remove regions lacking complete information and to ensure full spatial consistency among all channels. As a result, seven-band orthomosaics with a spatial resolution corresponding to that of the multispectral imagery were obtained and used as the basis for subsequent classification analyses.

Reference samples were prepared in QGIS version 3.40 through the manual digitization of polygons representing road surface materials and elements of their immediate surroundings. In total, 1,442 reference polygons were digitized, representing 12 classes of road surface materials and associated surrounding features:

- concrete pavement,
- asphalt pavement,
- grey concrete block pavement,

- red concrete block pavement,
- dark-grey concrete block pavement,
- curbs and edging elements,
- stone block pavement,
- white road markings,
- red road markings,
- metal elements (storm water inlets and rails),
- aggregate shoulders and unpaved surfaces,
- vegetation.

The twelve classes were deliberately selected to represent not only the investigated pavement materials but also the most common objects occurring in their immediate surroundings. This class composition enabled the evaluation of the proposed classification framework under realistic operational conditions. This approach is commonly used for obtaining reliable training data in the analysis of very high spatial resolution remote sensing imagery, where the accuracy of reference labels is critical to the quality of subsequent classification results [43]. An example of the distribution of the reference polygons used for the preparation of the training dataset is presented in Figure 2.

The vector data stored in the GeoPackage format were subsequently converted into raster class masks corresponding to the geometry of the orthomosaics. These masks served as reference data for the training and validation of the classification models. In the next stage, the orthomosaics and their corresponding reference masks were divided into overlapping image patches of

Table 3. Georeferencing accuracy of RGB and multispectral orthomosaics

Site ID	RGB control <i>RMSE</i> [m]	RGB check <i>RMSE</i> [m]	MS control <i>RMSE</i> [m]	MS check <i>RMSE</i> [m]
A1	0.038	0.042	0.080	0.072
A2	0.032	0.032	0.079	0.124
A3	0.041	0.036	0.072	0.128
C1	0.022	0.022	0.061	0.111
CB1	0.038	0.018	0.081	0.088
SB1	0.034	0.048	0.064	0.079

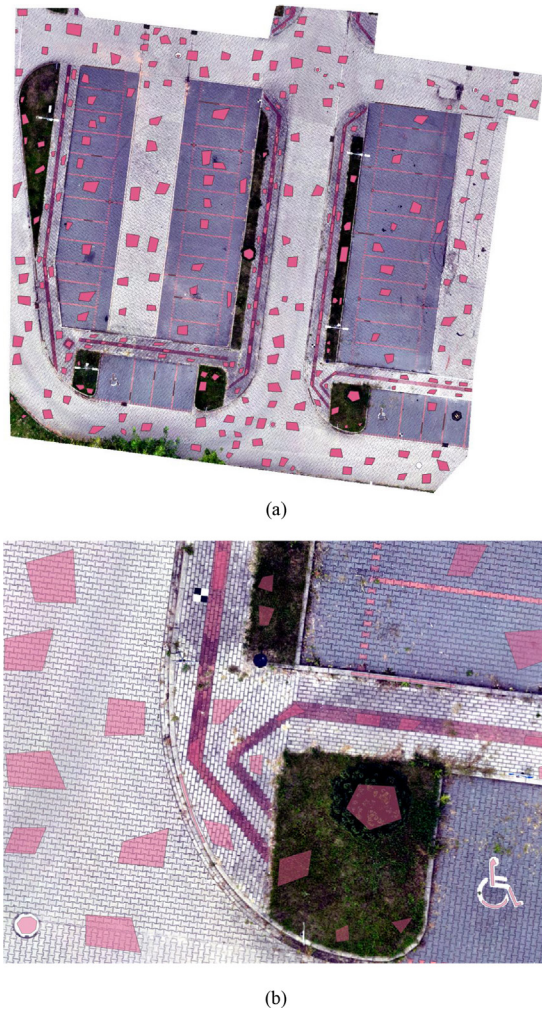


Figure 2. Preparation of reference data used for pavement-material classification: (a) seven-band orthomosaic with reference polygons (b) enlarged view showing manually delineated reference polygons representing pavement materials and surrounding land-cover classes

512 × 512 pixels. Patch extraction was performed using a sliding-window approach with a stride of 128 pixels, enabling an increase in the number of training samples while preserving local spatial context. Only patches containing at least 1,000 reference pixels were retained for further analysis, ensuring adequate representation of the target classes within each sample. For each patch, spectral values were normalized independently for each channel. In addition, the NDVI (normalized difference vegetation index) and NDRE (normalized difference red edge index) were calculated using the Red, Red Edge, and NIR bands. These indices are among the most widely used spectral measures for describing vegetation presence and condition, as well as supporting the

discrimination of land-cover objects [44–47]. As a result, three input data configurations were prepared: a three-band RGB dataset, a seven-band MS7 dataset, and an extended nine-band MS9 dataset additionally containing NDVI and NDRE values calculated according to Equations 1 and 2, respectively [44, 46]:

$$NDVI = (NIR - Red) / (NIR + Red) \quad (1)$$

where: *NIR* – near-infrared reflectance, *Red* – red-band reflectance.

$$NDRE = (NIR - RedEdge) / (NIR + RedEdge) \quad (2)$$

where: *NIR* – near-infrared reflectance, *RedEdge* – reflectance in the red-edge spectral band.

To ensure an independent assessment of model generalization performance, a study site-based data partitioning strategy was adopted. Samples derived from five locations (A1, A2, C1, CB1, and SB1) were used for model training, whereas data from location A3 were reserved exclusively for testing. This approach reduces the risk of spatial autocorrelation between the training and test datasets, thereby enabling a more reliable evaluation of classification performance on previously unseen road sites [48].

Machine learning algorithms and classification workflow

To evaluate the influence of spectral and spatial information on the classification of road surface materials, three ensemble machine learning algorithms were investigated: Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM). The three algorithms were deliberately selected because they belong to the same family of ensemble learning methods while representing different tree construction and boosting strategies. This enabled a consistent comparison of successive input feature configurations while minimizing the influence of fundamentally different classifier architectures on the obtained results. The primary objective of this study was to quantify the contribution of multispectral information and derived spatial features to road surface material classification performance rather than to benchmark fundamentally different artificial intelligence architectures. In the first stage, each model was trained using three different input data configurations:

RGB imagery, a seven-band multispectral dataset (MS7), and an extended nine-band dataset (MS9) enriched with the normalized difference vegetation index (NDVI) and the normalized difference red edge index (NDRE). During the initial stage of the study, semantic segmentation architectures, particularly U-Net, were also considered. However, preliminary experiments conducted using the MS7 dataset indicated poor model generalization for the available dataset, which comprised only six independent study sites. Consequently, the final experimental framework focused on ensemble learning algorithms, which proved more suitable for the available dataset and additionally allow straightforward interpretation of feature importance. Based on the obtained results, the best-performing model was selected for further analysis. In the second stage, additional experiments were conducted using this model by incorporating texture features and edge information derived from the Sobel operator. In total, fifteen classification configurations were evaluated. The selected algorithms were chosen because of their widespread application in remote sensing studies and their proven ability to model complex nonlinear relationships in high-dimensional datasets [26, 27, 29]. Tree-based ensemble methods are particularly well suited for UAV image classification, as they are robust to multicollinearity, effectively handle multidimensional feature spaces, and provide high classification performance even when relatively limited training datasets are available. The baseline dataset consisted of three RGB channels (Red, Green, and Blue). The MS7 configuration included seven spectral bands, comprising the three RGB channels together with four multispectral bands: Green, Red, Red Edge, and Near Infrared (NIR). The third configuration, denoted as MS9, extended the MS7 dataset by incorporating two spectral vegetation indices, NDVI and NDRE, calculated from the Red, Red Edge, and NIR bands according to Equations 1 and 2.

Subsequently, additional experiments were performed using the best-performing model identified in the first stage to investigate the contribution of spatial information to classification

performance. Texture features were extracted for each spectral band using a moving window of 5×5 pixels. Specifically, the local mean, standard deviation, and local contrast, defined as the difference between the maximum and minimum pixel values within the moving window, were calculated. Such texture descriptors are widely employed in remote sensing image classification because they capture local surface heterogeneity and improve the discrimination of objects exhibiting similar spectral characteristics [49, 50]. In the second variant of the extended experiments, the feature set was further enriched with edge information representing local image intensity variations. For each spectral band, gradient magnitude values were computed using the Sobel operator, providing information on edge strength and spatial variability. This approach enabled the simultaneous exploitation of spectral, textural, and geometric information describing the analyzed road surfaces. The number of features included in each experimental configuration is summarized in Table 4.

The Random Forest algorithm was implemented as an ensemble of decision trees constructed using randomly selected training samples and random subsets of input features [26]. To mitigate the effects of class imbalance, class weighting was applied using the balanced subsample strategy. Model training was performed iteratively with up to 1,500 decision trees, and the optimal number of estimators was determined based on the Macro F1-score, calculated according to Equation 4, on the validation dataset.

The XGBoost models were trained using a learning rate of 0.03, a maximum tree depth of 4, a subsampling ratio (subsample) of 0.75, and a feature sampling ratio (colsample bytree) of 0.75 [28]. The maximum number of boosting iterations was set to 10,000. Training was automatically terminated using an early stopping criterion when the validation loss did not improve for 200 consecutive iterations.

For the LightGBM models, a learning rate of 0.03, 31 leaves, observation and feature sampling ratios of 0.75, and an L2 regularization parameter of 5.0 were applied [29].

Table 4. Number of input features for each experimental variant

Dataset variant	RGB	MS7	MS9	RGB + textures	MS7 + textures	MS9 + textures	RGB + textures + Sobel	MS7 + textures + Sobel	MS9 + textures + Sobel
Number of features	3	7	9	12	28	36	15	35	45

The maximum number of trees was also set to 10,000, and training was terminated when no improvement in the validation error was observed for 200 consecutive iterations.

Prior to model training, the available data were divided into training and validation subsets. The models were trained using the training dataset, while a validation dataset comprising 20% of the training samples, selected using stratified sampling to preserve class proportions, was used for hyperparameter tuning and monitoring the training process. The independent test dataset consisted exclusively of samples derived from location A3, which was completely excluded from model training. This site-based train-test split was intentionally adopted to evaluate the generalization capability of the models on a previously unseen road section while reducing the risk of optimistic performance estimates caused by spatial autocorrelation between neighbouring image patches [51]. Classification performance was evaluated using three widely adopted metrics: OA, Macro F1-score (calculated as the arithmetic mean of the class-wise F1-scores), and Mean Intersection over Union (mIoU), which are widely used for assessing the quality of remote sensing image classification results [52]. These metrics were calculated according to Equations 3–5.

$$OA = (TP + TN) / (TP + TN + FP + FN) \quad (3)$$

where: TP (True Positives) denotes the number of correctly classified samples belonging to a given class, TN (True Negatives) represents correctly identified samples not belonging to that class, FP (False Positives) refers to samples incorrectly assigned to the class, and FN (False Negatives) denotes samples belonging to the class that were incorrectly classified as another category.

$$Macro\ F1 = \frac{1}{n} \sum_{i=1}^n F1_i \quad (4)$$

where: $F1_i = 2 \cdot ((Precision_i \cdot Recall_i) / (Precision_i + Recall_i))$, $Precision_i = TP_i / (TP_i + FP_i)$, $Recall_i = TP_i / (TP_i + FN_i)$.

$$mIoU = \frac{1}{n} \sum_{i=1}^n IoU_i \quad (5)$$

where:

$$IoU_i = TP_i / (TP_i + FP_i + FN_i)$$

where: TP_i , FP_i and FN_i denote the numbers of true positives, false positives, and false negatives for class i , respectively.

In addition to the aggregated performance metrics, confusion matrices and class-specific statistics, including precision, recall, and Macro F1-score, were also analysed. This made it possible to identify the classes most frequently confused by the models and to assess the influence of the input data configuration and the applied additional features on the discrimination performance of individual road surface materials.

After the training process was completed, the trained models were used to perform spatial prediction on selected orthomosaic fragments stored in the GeoTIFF format. For each analysed area, input datasets corresponding to the configurations used during model training were prepared. The data were normalized in the same manner as during the preparation of the training samples, using the 2nd and 98th percentile values calculated independently for each channel. Classification was performed on a pixel-by-pixel basis, assigning each pixel to one of the twelve analysed classes. In the configurations incorporating spatial information, additional features were computed for each channel, including local mean, standard deviation, and local contrast calculated within a moving window of 5×5 pixels. For selected experimental variants, the feature set was further extended with edge information derived using the Sobel operator. The prediction results were saved as raster classification maps in GeoTIFF format while preserving georeferencing information, and as PNG visualizations. In addition, classification overlays on RGB imagery were generated to facilitate the qualitative assessment of the spatial distribution of individual classes. To reduce classification noise, a post-processing procedure was applied, consisting of a 3×3 median filter, morphological opening and closing operations, filling of small gaps, and removal of objects smaller than 30 pixels [26, 52]. For each classification result, both the raw classification map and the post-processed map were retained.

The complete workflow adopted in this study is presented in Figure 3.

RESULTS AND ANALYSIS

Comparison of classification models under different input data configurations

In the first stage of the analysis, the performance of three ensemble learning algorithms:

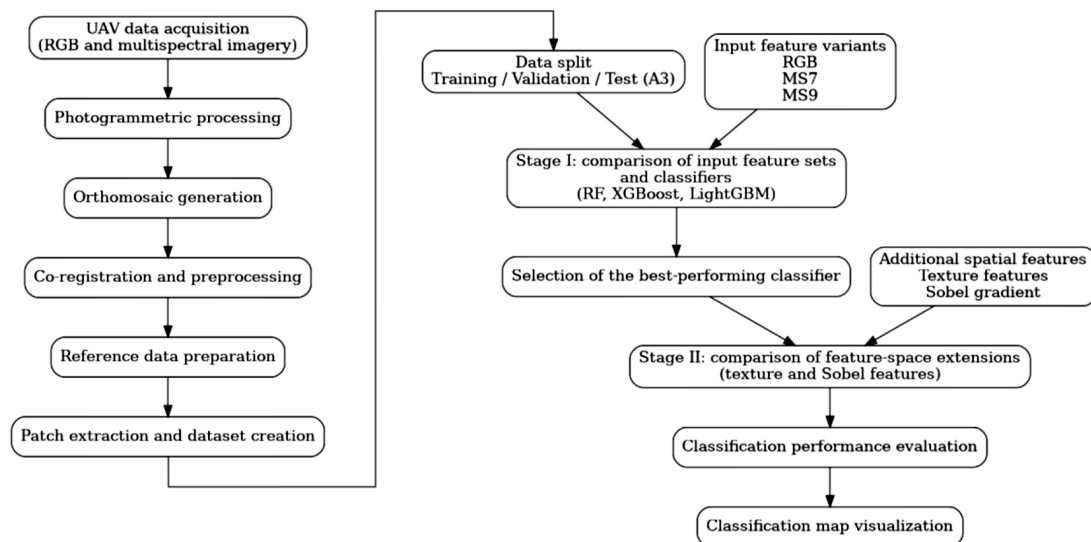


Figure 3. Workflow of the proposed pavement-material classification framework

Random Forest, XGBoost, and LightGBM, was compared using three input data configurations: RGB imagery, the seven-band multispectral dataset (MS7), and the extended MS9 dataset, which additionally included the NDVI and NDRE indices. The results obtained for the individual models are summarized in Table 5. The lowest classification performance was observed for models utilizing only RGB data. OA values ranged from 60.13% to 61.46%, while the Macro F1-score varied between 56.62% and 58.03%. Similarly, the mIoU metric did not exceed 45.11%, indicating a limited ability to discriminate among road surface materials based solely on information contained within the visible spectrum. For this data configuration, no clear superiority of any individual algorithm was observed. The highest OA value was achieved by XGBoost (61.46%), whereas Random Forest obtained the highest Macro F1-score and mIoU values (58.03% and 45.11%, respectively). A substantial improvement in classification performance was observed after incorporating multispectral data. For the MS7 dataset, OA values increased to approximately 74–75%, while the Macro F1-score exceeded 73% for all analysed algorithms. The best results within this group were achieved by LightGBM, which obtained an OA of 75.02%, a Macro F1-score of 75.05%, and an mIoU of 62.97%. A further increase in classification accuracy was achieved after extending the dataset with the NDVI and NDRE indices. For the MS9 configuration, all models achieved OA values exceeding 79%, while mIoU surpassed 67%. Once again, the highest performance was obtained

by LightGBM, for which OA reached 81.31%, the Macro F1-score was 81.49%, and mIoU attained 70.73%. Very similar results were also achieved by XGBoost (OA = 80.93%, Macro F1-score = 81.08%, mIoU = 70.15%), whereas Random Forest exhibited slightly lower accuracy. A comparison of the results obtained for the individual algorithms indicates that the differences between models were relatively small within the same data configuration. In contrast, the type of input features had a substantially greater influence on the final classification performance. The obtained results clearly demonstrate that extending RGB imagery with multispectral information and additional spectral indices significantly improved the ability of the models to discriminate among road surface materials. Considering all analysed data configurations and evaluation metrics, the highest classification performance was achieved by the LightGBM model using the MS9 dataset. These findings indicate that the combination of multispectral information with additional vegetation indices provided the greatest class separability among the analysed road surface materials.

Influence of multispectral and vegetation-index features on classification performance

To assess the contribution of additional spectral information, changes in classification performance between successive input data configurations were analysed. The comparison was based on the results obtained for RGB imagery,

Table 5. Classification performance of Random Forest, XGBoost and LightGBM models using different input data configurations

Model	Dataset	OA [%]	Macro F1 [%]	mIoU [m%]
Random Forest	RGB	61.13	58.03	45.11
XGBoost	RGB	61.46	57.58	44.86
LightGBM	RGB	60.13	56.62	43.95
Random Forest	MS7	73.81	73.42	61.03
XGBoost	MS7	74.94	74.88	62.68
LightGBM	MS7	75.02	75.05	62.97
Random Forest	MS9	79.38	79.41	67.88
XGBoost	MS9	80.93	81.08	70.15
LightGBM	MS9	81.31	81.49	70.73

the seven-band multispectral dataset (MS7), and the extended MS9 dataset, which additionally included the NDVI and NDRE indices. The analysis was performed using the OA and mIoU metrics for all investigated models. The incorporation of multispectral data resulted in a substantial improvement in classification performance regardless of the algorithm used (Table 6). Transitioning from RGB imagery to the MS7 configuration increased OA by 12.68 to 14.89 percentage points. An even greater improvement was observed for mIoU, which increased by 15.92 to 19.02 percentage points. These results indicate that the additional spectral bands provided valuable information for discriminating road surface materials that was not available within the visible spectrum alone. Further enhancement of the dataset through the inclusion of NDVI and NDRE indices led to an additional increase in classification performance. For all models, OA improved by approximately 5.5–6.3 percentage points, while mIoU increased by 6.8–7.8 percentage points. This suggests that spectral indices constitute valuable complementary features to multispectral data, improving the separability of classes characterized by similar spectral properties. The largest overall performance gain was achieved by the LightGBM model. Compared with the RGB configuration, OA increased from 60.13% to 81.31%, while mIoU improved from

43.95% to 70.73%, corresponding to gains of 21.18 and 26.78 percentage points, respectively. These findings confirm that the quality and richness of the input data had a greater impact on classification performance than the choice of a particular ensemble learning algorithm.

Influence of texture features and edge information on classification performance

In the second stage of the study, the impact of extending the feature set with spatial information describing local image structure was investigated. The analysis was conducted using the LightGBM model, which achieved the highest classification performance in the first stage of the experiments. For each image channel, three texture features were calculated within a moving window of 5×5 pixels: local mean, standard deviation, and local contrast. In a subsequent experimental variant, the feature set was further extended with edge information derived using the Sobel operator. The incorporation of texture features resulted in a substantial improvement in classification performance across all analysed data configurations (Table 7). The largest relative improvement was observed for the RGB dataset, where OA increased from 60.13% to 77.09%, while mIoU improved from 43.95% to 65.66%. This indicates

Table 6. Classification accuracy gain after input dataset expansion

Model	ΔOA RGB $\rightarrow$ MS7 [%]	ΔOA MS7 $\rightarrow$ MS9 [%]	$\Delta mIoU$ RGB $\rightarrow$ MS7 [%]	$\Delta mIoU$ MS7 $\rightarrow$ MS9 [%]
Random Forest	12.68	5.57	15.92	6.85
XGBoost	13.48	5.99	17.82	7.47
LightGBM	14.89	6.29	19.02	7.76

that texture features provided valuable information about the local surface structure, partially compensating for the absence of additional spectral bands. A similar pattern was observed for the multispectral datasets. For the MS7 configuration, the inclusion of texture features increased OA from 75.02% to 89.07% and mIoU from 62.97% to 82.02%. Even better results were achieved for the MS9 configuration, where OA increased from 81.31% to 93.30% and mIoU from 70.73% to 88.08%. This corresponds to improvements of approximately 12 and 17 percentage points, respectively, compared with the baseline model. The additional inclusion of edge information derived from the Sobel operator did not result in a meaningful improvement in classification performance. For the RGB and MS7 datasets, only minor changes in the analysed metrics were observed, not exceeding 1 percentage point. For the MS9 configuration, a slight decrease in classification performance was recorded. These findings suggest that most of the useful spatial information had already been captured by the texture features, while the additional use of image gradients provided little further benefit. The highest performance among all evaluated configurations was achieved by the LightGBM model using the MS9 dataset enriched with texture features. This configuration attained an OA of 93.30%, a Macro F1-score of 93.61%, and an mIoU of 88.08%. The obtained results confirm that the combination of multispectral information, vegetation indices, and features describing local image structure provides the most effective data representation for road surface material classification.

The addition of texture features substantially improved the classification performance for all input configurations. In contrast, incorporating edge features derived using the Sobel operator did not

provide any further improvement, indicating that texture descriptors alone were sufficient to capture the spatial characteristics of road surface materials.

Class-wise performance of the LightGBM model using MS9 and texture features

To complement the overall performance metrics presented in the previous sections, a class-wise analysis was conducted for the best-performing classification model, i.e. the LightGBM model trained using the MS9 dataset enriched with texture features. The evaluation included precision, recall, and F1-score calculated for each class, providing a more detailed assessment of the model's ability to discriminate between individual road surface materials and surrounding objects. The results are summarized in Table 8.

In addition to class-wise metrics, weighted average values are reported to account for the unequal distribution of samples among the analysed classes. Unlike the macro average, the weighted average assigns greater importance to classes represented by a larger number of pixels in the test dataset. The results presented in Table 8 indicate consistently high classification performance across all analysed classes. The highest F1 values were obtained for “red road markings”, “red concrete block pavement”, “metal elements”, and “vegetation”, all exceeding 97%. In particular, the “red road markings” class achieved perfect precision, recall, and F1. Although this class was represented by a relatively smaller number of test samples, its distinctive spectral characteristics likely facilitated reliable discrimination from the remaining pavement materials and surrounding objects. The lowest F1 values were observed for “concrete curbs and edging elements” (83.59%) and “grey concrete block pavement” (85.35%),

Table 7. Impact of texture features and edge features extracted using the Sobel operator on the classification performance of the LightGBM model

Dataset variant	Number of features [-]	OA [%]	Macro F1 [%]	mIoU [%]
RGB	3	60.13	56.62	43.95
RGB + textures	12	77.09	76.99	65.66
RGB + textures + Sobel	15	77.74	77.73	66.33
MS7	7	75.02	75.05	62.97
MS7 + textures	28	89.07	89.50	82.02
MS7 + textures + Sobel	35	89.09	89.63	82.20
MS9	9	81.31	81.49	70.73
MS9 + textures	36	93.30	93.61	88.08
MS9 + textures + Sobel	45	93.03	93.23	87.72

Table 8. Class-wise precision, recall, and F1 obtained for the LightGBM model trained using the MS9 dataset enriched with texture features

Class	Precision [%]	Recall [%]	F1 [%]
1: Concrete pavement	92.25	93.60	92.92
2: Asphalt pavement	96.01	94.57	95.28
3: Grey concrete block pavement	89.45	81.60	85.35
4: Red concrete block pavement	98.83	98.26	98.55
5: Dark-grey concrete block pavement	89.03	98.19	93.39
6: Concrete curbs and edging elements	81.21	86.12	83.59
7: Stone block pavement	88.07	95.68	91.72
8: White road markings	97.77	94.67	96.20
9: Red road markings	100.00	100.00	100.00
10: Metal elements	96.30	99.29	97.77
11: Aggregate shoulders and unpaved surfaces	90.39	90.18	90.29
12: Vegetation	97.92	98.73	98.32
Macro average	93.10	94.24	93.61
Weighted average	93.37	93.30	93.29

indicating that these surface types were the most challenging to discriminate. This can be attributed to their greater spectral and textural similarity to neighbouring pavement materials and surrounding surface classes. Nevertheless, all analysed classes achieved F1 values above 83%,

confirming the robust discrimination capability of the proposed classification approach.

The corresponding confusion matrix is shown in Figure 4. Most samples were correctly assigned to their respective classes, as reflected by the dominant diagonal values. The remaining

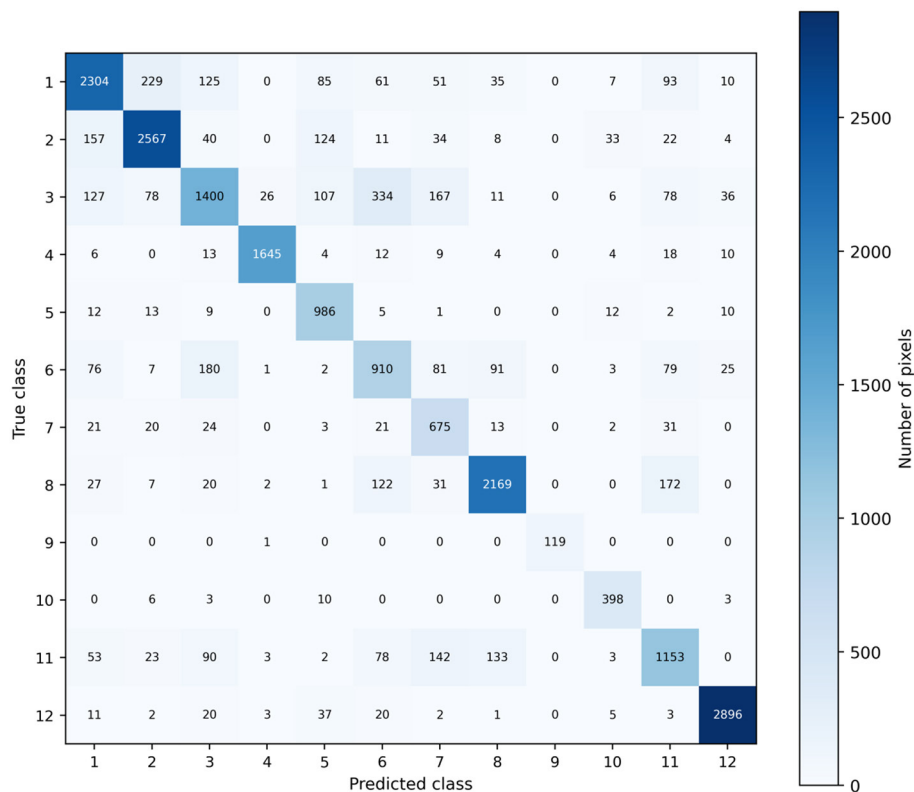


Figure 4. Confusion matrix for the LightGBM model trained using the MS9 dataset enriched with texture features (numerical class labels correspond to the class definitions presented in Table 8)

misclassifications mainly involved spectrally and texturally similar pavement materials and surrounding surface classes, which is consistent with the lower F1 values reported in Table 8. Overall, the confusion patterns indicate that the proposed feature configuration provided reliable discrimination of the analysed road surface materials under varying field conditions.

Feature importance analysis

To further interpret the contribution of individual input variables, feature importance was analysed for the LightGBM model trained using the MS9 dataset enriched with texture features. The ranking of the 15 most important predictors is presented in Figure 5. Feature importance values were obtained from the trained LightGBM model and should be interpreted as an indication of the relative contribution of individual variables to the tree-based decision process, rather than as a direct physical measure of material separability. The feature names used in Figure 5 refer to the original spectral and index-based variables and to texture descriptors calculated within a 5×5 pixel moving window. The prefix raw denotes original pixel values, whereas mean5, std5, and contrast5 denote the local mean, standard deviation, and local contrast, respectively. The MS9 feature set included three RGB channels, four multispectral channels, and two spectral indices. In this notation, B1–B3 correspond to the RGB channels, B4–B7 correspond to the multispectral Green,

Red, Red Edge, and NIR bands, while B8 and B9 correspond to NDVI and NDRE, respectively.

The ranking shows that the most informative predictors were mainly texture-derived variables, particularly local mean and standard deviation values calculated for NDRE, NDVI, and NIR-related features. This indicates that the model did not rely only on the absolute spectral response of individual pixels, but also used information describing the local spatial variability of spectral and index-based values. Such information is particularly useful for road surface material classification, because many pavement materials may have similar colour in RGB imagery but differ in local texture, roughness, ageing pattern, and spectral response outside the visible range. The high importance of features derived from NDVI and NDRE does not imply that these indices were used only for vegetation identification. In this study, the indices also acted as additional band-ratio descriptors that enhanced the separation between road surfaces, unpaved shoulders, vegetation, and adjacent objects. The presence of NIR- and Red Edge-related variables among the most important predictors confirms that spectral information beyond the visible range contributed substantially to the classification process. Overall, the feature importance analysis supports the quantitative results presented in Sections 4.2 and 4.3, showing that the best classification performance was obtained when spectral, index-based, and local texture information were combined.

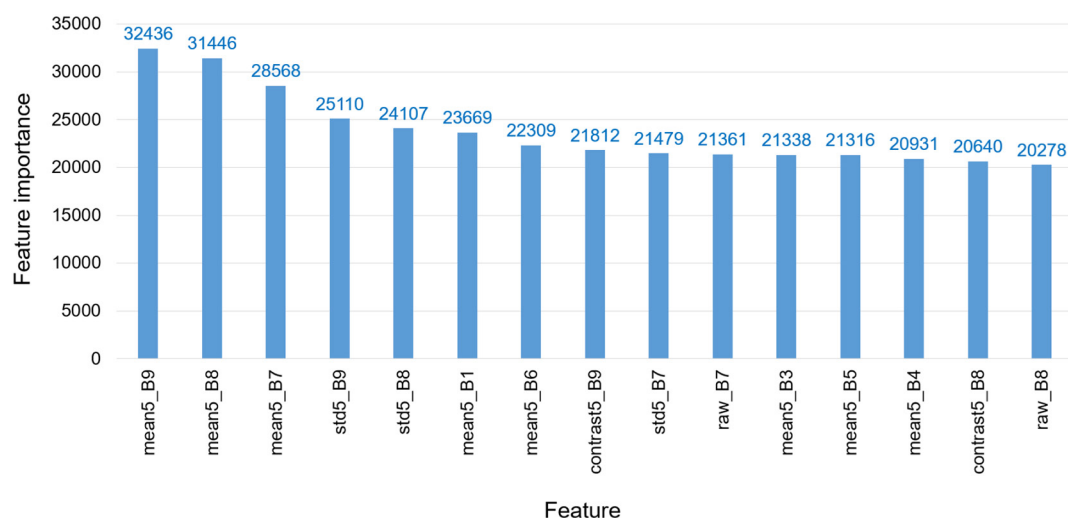


Figure 5. Feature importance ranking for the LightGBM model trained using the MS9 dataset enriched with texture features. B1–B3 denote RGB bands, B4–B7 denote multispectral Green, Red, Red Edge and NIR bands, while B8 and B9 denote NDVI and NDRE, respectively

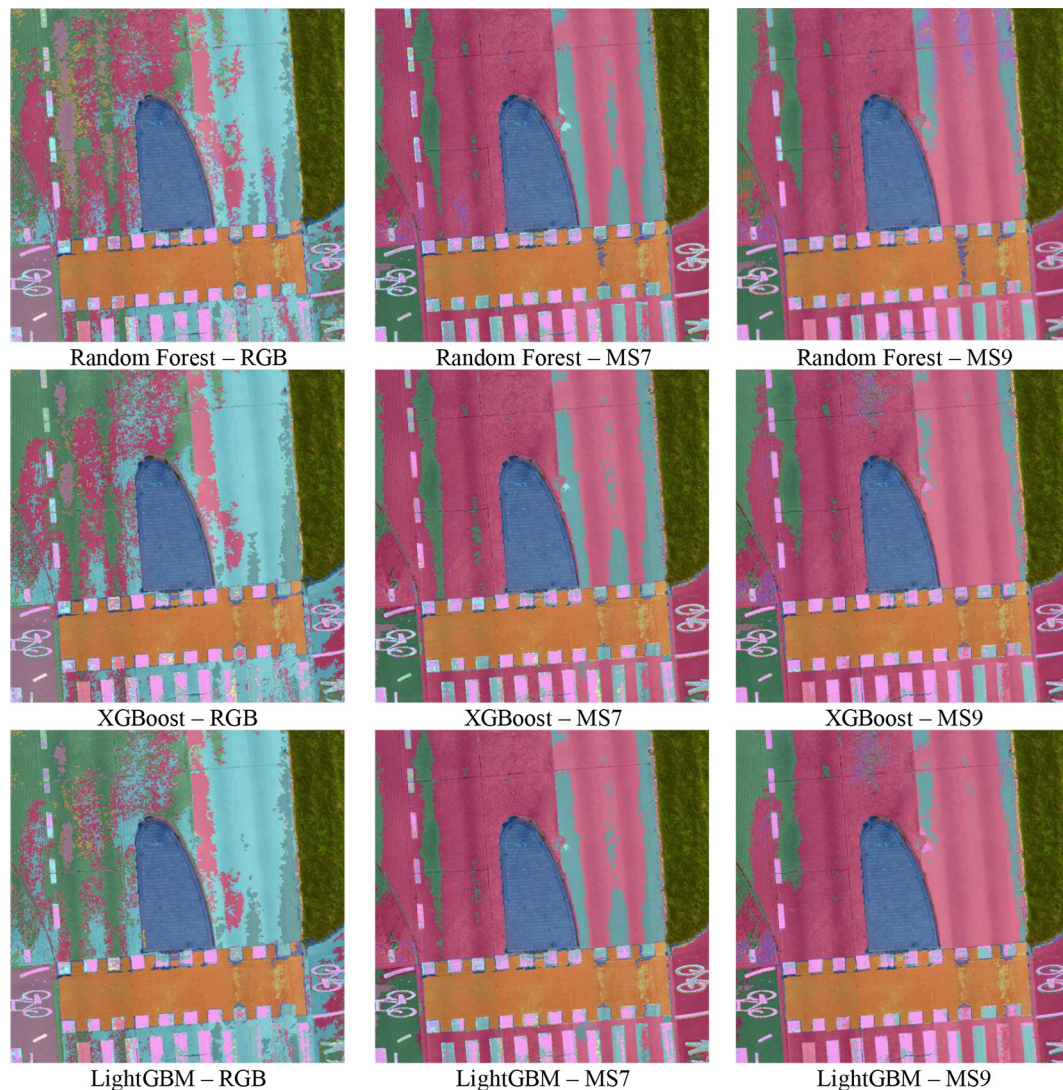


Figure 6. Comparison of pavement-material classification results obtained using Random Forest, XGBoost, and LightGBM models with RGB, MS7, and MS9 input datasets for the C1 study site

Spatial assessment of classification results

In addition to the quantitative evaluation of model performance, a qualitative assessment of the spatial distribution of classification results was also conducted. For this purpose, predictions were generated for selected orthomosaic fragments, and the resulting class maps were produced in raster format. Figure 6 presents the classification results for site C1 obtained using three ensemble learning algorithms (Random Forest, XGBoost, and LightGBM) and three input data configurations (RGB, MS7, and MS9).

The visual analysis confirms the observations derived from the accuracy metrics presented in the previous sections. For all analysed models, the use of RGB imagery alone resulted in a greater number of misclassified pixels and lower spatial

homogeneity of the mapped classes. This effect was particularly evident for asphalt and concrete pavements exhibiting different levels of deterioration and surface wear. The incorporation of multispectral data (MS7) substantially improved the spatial consistency of the classification results and reduced the “salt-and-pepper” effect observed in the RGB-based classifications. Further extension of the dataset with the NDVI and NDRE indices (MS9) led to even better separation of individual classes and a reduction in the number of isolated misclassified pixels. This improvement was observed regardless of the classification algorithm employed, confirming the significant contribution of additional spectral information to the discrimination of road surface materials. The influence of texture features calculated within a moving window of 5×5 pixels is illustrated in Figure 7.

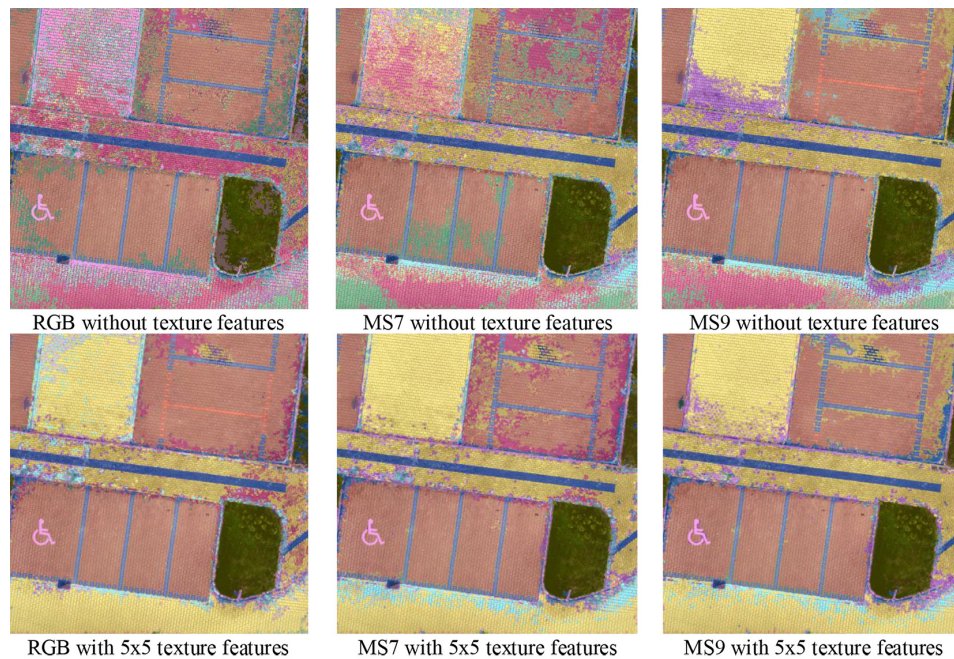


Figure 7. Influence of texture features on pavement-material classification results obtained using the LightGBM model and different input data configurations (RGB, MS7, and MS9) for the CB1 study site

For all analysed input data configurations, the incorporation of additional texture features led to a noticeable reduction in classification noise and an increase in the spatial homogeneity of the mapped classes. This effect was particularly evident over extensive areas of homogeneous pavement materials, where the number of isolated misclassified pixels was substantially reduced. The greatest improvement was observed for the multispectral datasets MS7 and MS9. The inclusion of information describing the local image structure enabled a more accurate delineation of boundaries between neighbouring classes and reduced the occurrence of the “salt-and-pepper” effect commonly associated with pixel-based classification. This improvement is particularly evident in parking areas and transportation corridors, where the addition of texture features resulted in more homogeneous and spatially coherent class regions. These qualitative observations are consistent with the quantitative results presented in Table 7. The inclusion of texture features increased OA from 81.31% to 93.30% for the MS9 configuration and improved mIoU from 70.73% to 88.08%, confirming the high utility of texture information for road surface material classification.

To assess the generalization capability of the proposed approach, classification was also performed for the independent test site A3, which was not used during either model training or validation.

The results obtained using the LightGBM model are presented in Figure 8. The analysed area includes asphalt pavement, unpaved shoulders, and vegetation, enabling the evaluation of classification performance under heterogeneous land-cover conditions. The visual analysis confirms the observations derived from the quantitative assessment. The configuration based solely on RGB imagery exhibited a greater number of misclassified pixels, particularly within transition zones between the road surface, roadside areas, and vegetation. The incorporation of multispectral data resulted in a clear improvement in the spatial consistency of the classification and a reduction of the “salt-and-pepper” effect. The most coherent results were obtained for the MS9 configuration enriched with texture features calculated within a 5×5 pixel moving window. This variant provided the most accurate delineation of class boundaries and the lowest number of isolated classification errors. The obtained results indicate that the combination of multispectral information, NDVI and NDRE indices, and texture features enables the effective transfer of learned relationships to previously unseen study areas. These findings confirm the strong generalization capability of the LightGBM model and demonstrate the suitability of the proposed methodology for road surface material classification based on UAV-acquired multispectral imagery.

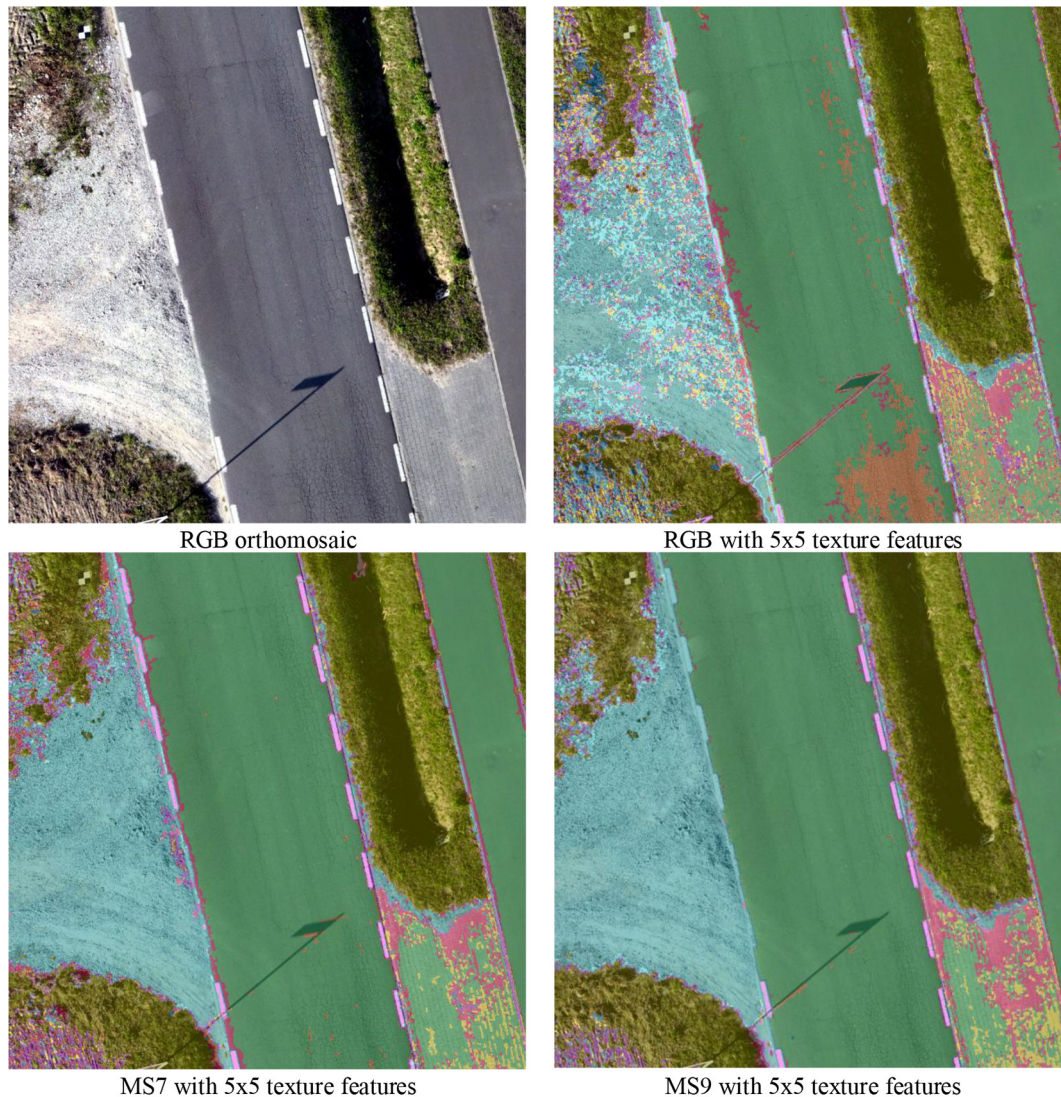


Figure 8. Classification results obtained for the independent test site (A3), which was not used during model training or validation, using the LightGBM model with different feature configurations

DISCUSSION

The results obtained in this study confirm that UAV-acquired multispectral imagery substantially improves road surface material classification compared with conventional RGB imagery. The increase in accuracy after extending the RGB dataset to the MS7 and MS9 configurations indicates that spectral information outside the visible range provides additional discriminatory information for pavement and surrounding surface classes. In particular, the Red Edge and NIR bands, together with NDVI and NDRE, improved the separation of asphalt pavements, unpaved shoulders, vegetation, and adjacent infrastructure elements. This confirms that the role of spectral indices in this study was not limited to vegetation detection, but also involved enriching the feature

space with band-ratio information useful for separating visually similar surfaces.

The obtained results complement recent studies on UAV-based road monitoring and AI-supported analysis of transportation infrastructure [53, 54]. While previous works have demonstrated the usefulness of UAV imagery and camera-based AI methods for road condition assessment and object detection, the present study extends this research direction towards road surface material classification. Its main contribution is the quantitative evaluation of how multispectral bands, vegetation-index features, and texture descriptors affect classification accuracy. Among the analysed ensemble learning algorithms, LightGBM achieved the best overall performance. However, the differences between Random Forest, XGBoost, and LightGBM were

smaller than the improvements obtained by changing the input feature configuration. This indicates that, in the analysed dataset, the information content of the input data had a stronger influence on classification performance than the choice of a particular tree-based ensemble algorithm. The strongest improvement was obtained after adding texture descriptors calculated within a 5×5 pixel moving window. These features increased the spatial consistency of the classification maps and reduced isolated misclassified pixels. The feature importance analysis further confirmed that local texture descriptors derived from NDVI, NDRE, NIR, and Red Edge-related variables were among the most informative predictors. This indicates that the model benefited from the combined use of spectral response and local spatial variability. In contrast, the additional Sobel edge features did not improve the classification results, suggesting that the local mean, standard deviation, and contrast already captured most of the relevant spatial information for the analysed road surface classes. The results also provide partial evidence of robustness to local illumination disturbances. Although all UAV surveys were conducted under generally favourable weather conditions, local shadows caused by trees and road infrastructure were present in the imagery. This is particularly visible in the independent A3 test site, where shadowed areas occur within the road scene. The classification results indicate that the MS9 configuration enriched with texture features maintained coherent asphalt classification even in areas affected by local shadowing. This suggests that the combination of multispectral information and local texture descriptors can reduce the negative effect of local illumination variability. However, this observation should be interpreted cautiously. The dataset was acquired under generally comparable weather conditions, and therefore the proposed approach cannot yet be considered fully validated for strongly variable illumination, wet pavement surfaces, overcast conditions, or different seasons.

A further issue concerns the comparison with deep learning semantic segmentation methods. Architectures such as U-Net and DeepLabV3+ are widely used in UAV image segmentation and can achieve high accuracy when large and diverse annotated datasets are available. In the present study, preliminary experiments with a U-Net architecture were conducted using the MS7 dataset. However, the model showed weak

learning and poor generalization: despite training on 4178 image patches, the best validation mIoU reached only 9.62%. This result was far below the performance obtained using tree-based ensemble models and indicated that the available dataset was insufficient for effective training of a deep semantic segmentation network. The limited number of independent study sites, the relatively homogeneous acquisition conditions, and the restricted variability of annotated scenes most likely prevented the U-Net model from learning robust spatial representations. For this reason, U-Net and DeepLabV3+ were not included in the main comparative experiment. Instead, the study focused on ensemble learning algorithms, which provided stable performance for the available dataset and allowed direct interpretation of feature importance. This does not diminish the potential of deep learning methods for road surface material classification. Rather, the obtained U-Net results indicate that their effective application in this task requires a larger and more diverse annotated UAV dataset. Future research should therefore include multi-region and multi-season UAV acquisitions, more variable illumination and surface moisture conditions, and a broader comparison with deep learning segmentation architectures, including U-Net, DeepLabV3+, and transformer-based models. Such an extended dataset would make it possible to evaluate whether deep learning models can outperform feature-based ensemble methods when sufficient training data are available.

Several limitations of the study should be acknowledged. First, the dataset consisted of six study sites located within one geographical region and acquired during a single vegetation season. Although the use of an independent test site allowed the assessment of model transferability to an unseen road section, further validation on datasets acquired in other regions, seasons, and illumination conditions is needed. Second, all UAV surveys were conducted under generally favourable weather conditions, although local shadows caused by vegetation and road infrastructure were present. Therefore, the robustness of the proposed approach under stronger illumination variability, image noise, wet pavement conditions, and seasonal changes should be further investigated. Third, the reference data were prepared by manual polygon digitization, which ensured high label quality but also limited the scale of the annotated dataset. Future studies

should therefore include larger multi-site datasets, additional road material classes, and a direct comparison between ensemble learning and deep learning segmentation approaches under comparable training conditions.

CONCLUSIONS

This study evaluated the usefulness of UAV-acquired RGB and multispectral imagery for road surface material classification using ensemble machine learning algorithms. The results demonstrated that extending RGB imagery with multispectral bands, spectral indices, and local texture descriptors substantially improves classification performance. The best result was obtained for the LightGBM model trained using the MS9 dataset enriched with texture features, which achieved an Overall Accuracy of 93.30%, a Macro F1 of 93.61%, and an mIoU of 88.08%. The findings confirm that the integration of spectral and spatial information provides an effective feature representation for UAV-based pavement material classification. The main findings of the study can be summarized as follows:

Multispectral imagery improved the classification of road surface materials compared with RGB data alone, confirming the importance of spectral information beyond the visible range.

The use of NDVI and NDRE indices enhanced the separability of pavement surfaces and adjacent land-cover classes, showing that index-based features can provide useful complementary information.

Texture features calculated within a local moving window had the strongest positive effect on classification performance, improving both accuracy and the spatial consistency of the classification maps.

Among the tested ensemble learning algorithms, LightGBM provided the most accurate results, although the improvement related to input feature configuration was more pronounced than the differences between classifiers.

The proposed feature-based approach achieved accuracy above 90% on an independent road section, indicating its potential for practical UAV-based pavement material classification; however, further validation is required for other regions, seasons, illumination conditions, and pavement surface states.

Acknowledgements

This research was funded by the National Science Centre (NCN), Poland, under the MIN-IATURA 9 programme, project no. 2025/09/X/ST8/00014, entitled “The use of multispectral imaging to analyse the properties of road surfaces and adjacent ecosystems”. The author gratefully acknowledges the financial support provided by the National Science Centre, which enabled the research presented in this paper.

REFERENCES

1. Zhang T, Su J, Xu Z, Luo Y, Li J. Sentinel-2 satellite imagery for urban land cover classification by optimized Random Forest classifier. *Appl. Sci.* 2021; 11(2):543. <https://doi.org/10.3390/app11020543>
2. Arya D, Maeda H, Ghosh SK, Toshniwal D, Mraz A, Kashiyama T, Sekimoto Y. Deep learning-based road damage detection and classification for multiple countries. *Autom. Constr.* 2021; 132:103935. <https://doi.org/10.1016/j.autcon.2021.103935>
3. Aela P, Chi HL, Fares A, Zayed T, Kim M. UAV-based studies in railway infrastructure monitoring. *Autom. Constr.* 2024; 167:105714. <https://doi.org/10.1016/j.autcon.2024.105714>
4. Schnebele E, Tanyu BF, Cervone G, Waters N. Review of remote sensing methodologies for pavement management and assessment. *Eur. Transp. Res. Rev.* 2015; 7:7. <https://doi.org/10.1007/s12544-015-0156-6>
5. Roberts R, Inzerillo L, Di Mino G. Using UAV based 3D modelling to provide smart monitoring of road pavement conditions. *Inf.* 2020; 11(12):568. <https://doi.org/10.3390/info11120568>
6. Małek A. Assessment of the use of unmanned aerial vehicles for road pavement condition surveying. *Roads and Bridges - Drogi i Mosty.* 2023; 22(4):331–345. <https://doi.org/10.7409/rabdim.023.017>
7. Wu W, Qurishee MA, Owino J, Fomunung I, Onyango M, Atolagbe B. Coupling deep learning and UAV for infrastructure condition assessment automation. 2018 IEEE International Smart Cities Conference (ISC2). 2018. <https://doi.org/10.1109/ISC2.2018.8656971>
8. Ibrahim HB, Salah M, Zarzoura F, El-Mewafi M. Smart monitoring of road pavement deformations from UAV images by using machine learning. *Innov. Infrastruct. Solut.* 2024; 9:16. <https://doi.org/10.1007/s41062-023-01315-2>
9. Outay F, Mengash HA, Adnan M. Applications of unmanned aerial vehicle (UAV) in road safety, traffic and highway infrastructure management: Recent advances and challenges. *Transp. Res. Part A Policy*

- Pract. 2020; 141:116–129. <https://doi.org/10.1016/j.ttra.2020.09.018>
10. Mohammadi M. Road classification and condition determination using hyperspectral imagery. ISPRS Archive. 2012; 39(B7):141–146. <https://doi.org/10.5194/isprsarchives-XXXIX-B7-141-2012>
11. Cui YP. Methods and Results of Identifying a Road Material Dataset from GF–6 Remote Sensing Data in the Langfang Area. JGCDD. 2022; 6(4):607–618. <https://doi.org/10.3974/geodp.2022.04.12>
12. Andreou C, Karathanassi V, Kolokoussis, P. Investigation of hyperspectral remote sensing for mapping asphalt road conditions. Int. J. Remote Sens. 2011; 32(21):6315–6333. <https://doi.org/10.1080/01431161.2010.508799>
13. Pesek O, Krisztian L, Landa M, Metz M, Neteler M. Convolutional neural networks for road surface classification on aerial imagery. PeerJ Comput. Sci. 2024; 10:e2571. <https://doi.org/10.7717/peerj-cs.2571>
14. Zhou Q, Liu Z, Huang Z. Mapping Road Surface Type of Kenya Using OpenStreetMap and High-resolution Google Satellite Imagery. Sci Data. 2024; 11(1):331. <https://doi.org/10.1038/s41597-024-03158-7>
15. Inzerillo L, Acuto F, Di Mino G, Uddin MZ. Super-resolution images methodology applied to UAV datasets to road pavement monitoring. Drones. 2022; 6(7):171. <https://doi.org/10.3390/drones6070171>
16. Lu L, Weng Q, Xiao D, Guo H, Li Q, Hui W. Spatiotemporal variation of surface urban heat islands in relation to land cover composition and configuration: A multi-scale case study of Xi'an, China. Remote Sens. 2020; 12(17):2713. <https://doi.org/10.3390/rs12172713>
17. Öztürk Ş, Yılmaz K, Dinçer AE, Kalpakçı V. Effect of urbanization on surface runoff and performance of green roofs and permeable pavement for mitigating urban floods. Nat. Hazards. 2024; 120:12375–12399. <https://doi.org/10.1007/s11069-024-06688-w>
18. Aghasi F, Smadi S, Nlenanya I, Gelder BK. Integrating machine learning, remote sensing, and a geographic information system to determine road surface type and construction year from high-resolution and multispectral aerial imagery: A case study in Ames, Iowa. TRR. 2025; 2679(12):230–248. <https://doi.org/10.1177/03611981251352515>
19. Zheng Z, Yuan J, Yao W, Kwan P, Yao H, Liu Q, Guo L. Fusion of UAV-acquired visible images and multispectral data by applying Machine-Learning methods in crop classification. Agronomy. 2024; 14(11):2670. <https://doi.org/10.3390/agronomy14112670>
20. Aydin I., Sefercik UG. Multispectral UAV-based LULC mapping performance improvement by integrating precise NDSM data and machine learning algorithms. Earth Sci. Inform. 2025; 18:326. <https://doi.org/10.1007/s12145-025-01841-w>
21. Zahiri Z, Laefer DF, Gowen A. Characterizing building materials using multispectral imagery and LiDAR intensity data, J. Build. Eng. 2021; 44:102603 <https://doi.org/10.1016/j.jobe.2021.102603>
22. Jinru X, Baofeng S. Significant remote sensing vegetation indices: A review of developments and applications, J. Sens. 2017; 353691. <https://doi.org/10.1155/2017/1353691>
23. Park J, Lee Y, Lee J. Assessment of Machine Learning algorithms for land cover classification using remotely sensed data. Sens. Mater. 2021; 33(11):3885–3902. <https://doi.org/10.18494/SAM.2021.3612>
24. Karpiński P, Rzeczkowski J. Comparison of unsupervised machine learning segmentation algorithms in the analysis of unmanned aerial vehicle – based multispectral crop images. ASTRJ. 2026; 20(2):94–113. <https://doi.org/10.12913/22998624/210743>
25. Maxwell AE, Warner TA, Fang F. Implementation of machine-learning classification in remote sensing: an applied review. Int. J. Remote Sens. 2018; 39(9):2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>
26. Breiman L. Random Forests. Mach. Learn. 2021; 45:5–32. <https://doi.org/10.1023/A:1010933404324>
27. Belgiu M, Drăguț L Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 2016; 114:24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
28. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016; 785–794. <https://doi.org/10.1145/2939672.2939785>
29. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY. LightGBM: A highly efficient gradient boosting decision tree. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017.
30. Ballanti L, Blesius L, Hines E, Kruse B. Tree species classification using hyperspectral imagery: A comparison of two classifiers. Remote Sen. 2016; 8(6):445. <https://doi.org/10.3390/rs8060445>
31. Papoutsis I., Bountos NI, Zavras A, Michail D, Tryfonopoulos C. Benchmarking and scaling of deep learning models for land cover image classification. ISPRS J. Photogramm. Remote Sens. 2023; 195; 250–268. <https://doi.org/10.1016/j.isprsjprs.2022.11.012>
32. Descals A, Verger A, Yin G, Filella I, Peñuelas J. Local interpretation of machine learning models in remote sensing with SHAP: the case of global climate constraints on photosynthesis phenology. IJRS. 2023; 44(10):3160–3173. <https://doi.org/10.1080/01431161.2023.2217982>
33. Höhl A, Obadić I, Fernández-Torres MÁ, Najjar H, Oliveira DAB, Akata Z, Dengel A, Zhu XX.

- Opening the Black Box: A systematic review on explainable artificial intelligence in remote sensing. *IEEE Geosci. Remote Sens. Mag.* 2024; 12(4):261–304. <https://doi.org/10.1109/MGRS.2024.3467001>
34. Yao J, Xu Q, Yu F, Yu S. The remote sensing method for large-scale asphalt pavement aging assessment with automated sample generation and deep learning. *Sci. Rep.* 2025; 15: 45432. <https://doi.org/10.1038/s41598-025-29966-4>
35. Dugalam R, Prakash G. Development of a random forest based algorithm for road health monitoring. *Expert Syst. Appl.* 2024; 251:123940. <https://doi.org/10.1016/j.eswa.2024.123940>
36. Shubham, Banerjee D, Upadhyay D, Rawat RS. Enhanced Road Surface Classification using CNN and Random Forest Models. 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS). 2024; 812–817. <https://doi.org/10.1109/ICSCSS60660.2024.10625525>
37. Zhao Y, Hasan MRM, Zhang K, You L, Jamshidi A, Jiang J. XGBoost-based intelligent framework for asphalt pavement skid resistance assessment under different variables. *Smart Constr.* 2025; 4:0029. <https://doi.org/10.55092/sc20250029>
38. James MR, Robson S, d'Oleire-Oltmanns S, Niethammer U. Optimising UAV topographic surveys processed with structure-from-motion: Ground control quality, quantity and bundle adjustment, *Geomorphology.* 2017; 280:51–66. <https://doi.org/10.1016/j.geomorph.2016.11.021>
39. Nesbit PR, Hugenholtz CH. Enhancing UAV–SfM 3D model accuracy in high-relief landscapes by incorporating oblique images. *Remote Sens.* 2019; 11(3):239. <https://doi.org/10.3390/rs11030239>
40. Aasen H, Honkavaara E, Lucieer A, Zarco-Tejada PJ. Quantitative remote sensing at ultra-high resolution with UAV spectroscopy: A review of sensor technology, measurement procedures, and data correction workflows. *Remote Sens.* 2018; 10(7):1091. <https://doi.org/10.3390/rs10071091>
41. Tmušić G, Manfreda S, Aasen H, James MR, Gonçalves G, Ben-Dor E, Brook A, Polinova M, Arranz JJ, Mészáros J, et al. Current practices in UAS-based environmental monitoring. *Remote Sens.* 2020; 12(6):1001. <https://doi.org/10.3390/rs12061001>
42. Shahbazi M, Théau J, Ménard P. Recent applications of unmanned aerial imagery in natural resource management. *GISci. Remote Sens.* 2014; 51(4):339–365. <https://doi.org/10.1080/15481603.2014.926650>
43. Blaschke T. Object based image analysis for remote sensing. *ISPRS.* 2010; 65(1):2–16. <https://doi.org/10.1016/j.isprsjprs.2009.06.004>
44. Rouse, JW, Haas, RH, Schell, JA, Deering, DW. Monitoring Vegetation Systems in the Great Plains with ERTS (Earth Resources Technology Satellite). In *Proceedings of the 3rd Earth Resources Technology Satellite Symposium*, Greenbelt, MD, USA, 10–14 December 1973; SP-351. pp. 309–317
45. Barnes EM, Clarke T, Richards SE, Colaizzi PD, Haberland J, Kostrzewski M, Waller PM, Choi CY, Riley E, Thompson TL, Lascano RJ, Li H, Moran MS, Robert PC, Rust RH, Larson WE. Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data. 2000.
46. Kang Y, Hu X, Meng Q, Zou Y, Zhang L, Liu M, Zhao M. Land cover and crop classification based on red edge indices features of GF-6 WFV time series data. *Remote Sens.* 2021; 13(22):4522. <https://doi.org/10.3390/rs13224522>
47. Kocur-Bera, K, Małek A. Assessing the feasibility of using remote sensing data and vegetation indices in the estimation of land subject to consolidation. *Sensors* 2024; 24, 7736. <https://doi.org/10.3390/s24237736>
48. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillerá-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography.* 2017; 40:913–929. <https://doi.org/10.1111/ecog.02881>
49. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics.* 1973; SMC-3(6):610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
50. Löfstedt T, Brynolfsson P, Asklund T, Nyholm T, Garpebring A. Gray-level invariant Haralick texture features. *PLoS ONE.* 2019; 14(2): e0212110. <https://doi.org/10.1371/journal.pone.0212110>
51. Pelletier C, Webb GI, Petitjean F. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sens.* 2019; 11(5):523. <https://doi.org/10.3390/rs11050523>
52. Maxwell AE, Warner TA, Guillén LA. Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—Part 1: Literature review. *Remote Sens.* 2021; 13(13):2450. <https://doi.org/10.3390/rs13132450>
53. Liu C, Szirányi T. Road condition detection and emergency rescue recognition using on-board UAV in the wildness. *Remote Sens.* 2022; 14(17):4355. <https://doi.org/10.3390/rs14174355>
54. Nowakowski M, Kurylo J, Dang PH. Camera based AI Models Used with LiDAR Data for Improvement of Detected Object Parameters. In: *Modelling and Simulation for Autonomous Systems. Lecture Notes in Computer Science.* Springer; 2024:287–301. https://doi.org/10.1007/978-3-031-71397-2_18