Explainable Artificial Intelligence for Machine Learning-Based Photogrammetric Point Cloud Classification

Muhammed Enes Atik, Zaide Duran, and Dursun Zafer Seker

***Abstract*—Point clouds are one of the most widely used data sources for spatial modeling. Artificial intelligence approaches have become an important tool for understanding and extracting semantic information of point clouds. In particular, the explainability of machine learning approaches for 3-D data has not been sufficiently investigated. Moreover, existing studies are generally limited to object classification issues. This is a pioneer study that addresses the classification of photogrammetric point clouds in terms of explainable artificial intelligence. In this study, the explainability of black-box machine learning models in the context of the classification of photogrammetric point clouds was investigated. Each point in the point cloud is defined using geometric and spectral features. In addition, the effect of selecting the most important of these features on the classification performance of ML models such as Random Forest, XGBoost, and LightGBM was examined. The explainability of ML models was analyzed with Shapley additive explanation (SHAP), an explainable artificial intelligence approach. SHAP analysis was compared with filter-based information gain (IG) and ReliefF methods for feature selection. Using the features selected with SHAP analysis, overall accuracy (OA) of 85.50% in the Ankeny dataset, 91.70% in the Building dataset, and 83.28% in the Cadastre dataset was achieved with LightGBM. The evaluation with XGBoost shows an OA of 85.22% for Ankeny, 91.21% for Building, and 82.47% for Cadastre. The evaluation with RF shows an OA of 83.70% for Ankeny, 89.08% for Building, and 79.36% for Cadastre.**

***Index Terms*—Classification, explainable artificial intelligence (XAI), feature selection, machine learning, photogrammetry, point cloud.**

I. Introduction

THREE dimensional (3-D) objects can be represented in different ways; together, they form complex point cloud structures [1]. Due to their widespread use and comprehensive 3-D description, point clouds are popular data sources. The main interest of photogrammetry is the 3-D reconstruction of objects from images [2]. Recently, dense point clouds of large areas can be produced using algorithms such as structure from motion (SfM) [3]. Direct processing of point clouds has become one of the main research areas in recent years, as it prevents information loss and provides maximum utilization of multi-dimensional information of points. The geometric information in point clouds is valuable as a basis for many applications. Point cloud classification is one of the basic tasks in point cloud processing. The purpose of point cloud classification is to identify each point with a label to determine the general or local characteristics of the point cloud [4]. The features produced using the 3-D structure of the point cloud are used for point cloud classification using machine learning. Although deep learning extracts features with hidden layers, methods using geometric features as input data have also been developed. Large-scale point cloud segmentation is performed independently for each point or voxel, using handcrafted features derived from its local neighbors. However, due to the complexity of 3-D scenes caused by irregular point sampling, varying point density, and very different object types, there are several difficulties in calculating appropriate local geometric features [5]. Defining the local geometric features correctly is a problem to be solved, since many additional features are used besides the 3-D coordinate information of the point cloud only. Particularly, machine learning algorithms require features other than 3-D coordinates to define a point. To determine the distinctions between the classes well, it is necessary to determine the geometric features appropriately.

Explainable artificial intelligence (XAI) techniques give a favor for solving the challenges of using artificial intelligence on point cloud classification by providing more interpretable AI models. Not only does XAI allow understanding of the complexity of artificial intelligence models due to the black box, but it also assures decision-makers over the model. XAI is being used especially in crucial industries, where it's important to understand the model's functioning and key characteristics [6]. XAI provides a great contribution to decision-makers and researchers when comparing and conducting AI algorithms, especially through the data, likewise point cloud complexity and the needs of expertise perspectives. Because XAI methods aid in increasing the machine learning models' transparency and trustworthiness [7]. The XAI models can produce the important values of spectral bands or features to contribute to the proper models' decisions.

As more features are used to generate more information and increase the distinctiveness of algorithms, investigating the effects of the features on point cloud classification has been considered an additional step in recent studies [8], [9]. Because

Manuscript received 11 September 2023; revised 8 January 2024 and 11 February 2024; accepted 22 February 2024. Date of publication 27 February 2024; date of current version 8 March 2024. *(Corresponding author: Muhammed Enes Atik.)*

The authors are with the Department of Geomatics Engineering, Faculty of Civil Engineering, Istanbul Technical University, Istanbul 34469, Türkiye (e-mail: atikm@itu.edu.tr; duranza@itu.edu.tr; seker@itu.edu.tr).

Digital Object Identifier 10.1109/JSTARS.2024.3370159

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

each feature can have a different effect on point cloud classification, it is necessary to determine the minimum number of attributes accurately representing the data. Therefore, feature selection is an essential step in point cloud classification. Feature selection algorithms are used to improve the accuracy of classifiers, increase computational efficiency in terms of time and memory consumption, and find compact and robust subsets of relevant and informative features. [10]. Although feature selection and XAI approaches are used extensively for 2-D data, their use in 3-D data is still a very new field of research. Studies have been carried out on the use of geometric features for machine learning-based classification of photogrammetric point clouds. However, the effect of these geometric features on the predictions of black-box machine learning models has not been examined by XAI methods.

In this study, research on the use of XAI methods in the classification of photogrammetric point clouds with machine learning is presented. The important values of the 3-D geometric features obtained from the point cloud were calculated by XAI and filter-based feature selection methods. The effect of selection of optimum features on the classification performance of ML models was investigated. In addition, class-based impact analysis of each feature was performed with SHAP, one of the XAI approaches. Experiments were carried out using three aerial photogrammetric point cloud datasets. Ensemble classifiers RF, XGBoost, and LightGBM were used as ML models. Particularly, a unique contribution to the literature on the use of XAI methods for ML-based point cloud classification is presented.

II. Related Works

Many methods have been developed for point cloud classification in the literature. Initially, rule-based methods were used to distinguish between different land cover classes [11], [12]. However, rule-based approaches have limited ability to describe complex relationships between classes. Recently, successful results have been obtained in point cloud classification with machine learning algorithms. The main problem in studies is how to define a point. Geometric features are used for point identification in many studies [8], [13], [14], [15], [16]. Weinmann et al. [8] published extensive research on different features, classifiers, and feature selection methods for machine learning-based point cloud classification. In [14], point cloud classification was performed by calculating 19 geometric features. In addition to the geometric features, there are also studies using the spectral features of the points. Thus, not only the geometric structure of the point cloud is used, but also the color information of the objects. Sevgen and Abdikan [17] presented a study on the classification of mobile LiDAR point clouds with LightGBM using hand-crafted features. Moorthy et al. [18] tested Random Forest (RF), XGBoost, and lightGBM algorithms with individual tree and field data from tropical and deciduous forests, using geometric features calculated at multiple spatial scales. By using multiple spatial scales, it is aimed to eliminate the need for optimal neighborhood size selection. LiDAR point clouds are preferred as data source in many studies in point cloud classification. Public datasets generally contain LiDAR point clouds. The classification of photogrammetric point clouds is a current research topic. Becker et al. [16] presented a study on the controlled classification of photogrammetric point clouds with machine learning. In the study, color information was added to the point feature vector along with 15 geometric features. Sharma and Gark [19] applied RF classification using geometric features to extract buildings from photogrammetric point cloud. Zeybek [20] applied UAV point cloud data classification with RF using spectral and geometric features. In the study published by Carbonell–Rivera et al. [21], Mediterranean shrub species were classified from UAV point clouds with different ML classifiers using geometric, spectral, and neighbor-based features. Weidner et al. [22] presented smartphone and UAV photogrammetry point cloud datasets to classify with RF. The geometric and spectral features of each point were used to classify datasets containing multiple slope morphologies and lighting conditions.

Deep learning has become popular in the field of computer vision because it requires less human processing in processing large and complex data [23], [24]. Recently, deep learning approaches have been used for point cloud classification. Deep learning-based point cloud classification is used in many research areas such as autonomous driving, virtual reality, cultural heritage, and augmented reality [4]. Deep learning approaches also calculate distinctive rules and features within the network. Although most deep learning approaches take 3-D coordinates as input, few approaches take additional features as input. In the study conducted by Atik and Duran [10], a comprehensive research was presented by giving geometric and spectral features as input to RandLA-Net [25] and Superpoint graph [26] in addition to 3-D geometric features. Kurdi et al. [27] designed a DL-based classification network using eleven geometric features to classify LiDAR point clouds. Ozturk et al. [28] proposed a fusion strategy that combines the use of optical images and geometric features obtained from point clouds to improve the road segmentation performance of a deep learning model. Although deep learning approaches come to the fore with successful results, processing Big Data such as point cloud with deep learning requires high hardware requirements.

One of the first attempts to use XAI methods for point cloud classification is the PointHop algorithm [1]. PointHop consists of two steps: local-to-global feature creation through iterative one-stop information exchange and classification. Gupta et al. [29] proposed a model-agnostic approach for point cloud classification. According to the study, edges and vertices are important features in 3-D data, while planar surfaces are less important. Zheng et al. [30] developed a salience map to measure the importance of each point in a point cloud. In this study, they show that the contribution of a point is roughly proportional to the slope of the loss with respect to the point under a scaled spherical coordinate system and can be scored by approximating the point drop with a point shift. In another study published by Matorne et al. [31], a new multimodal fusion framework, BubblEX, was proposed to learn 3-D point features. BubblEX consists of two modules, the "Visualization Module," which enables the visualization of features learned from the network,

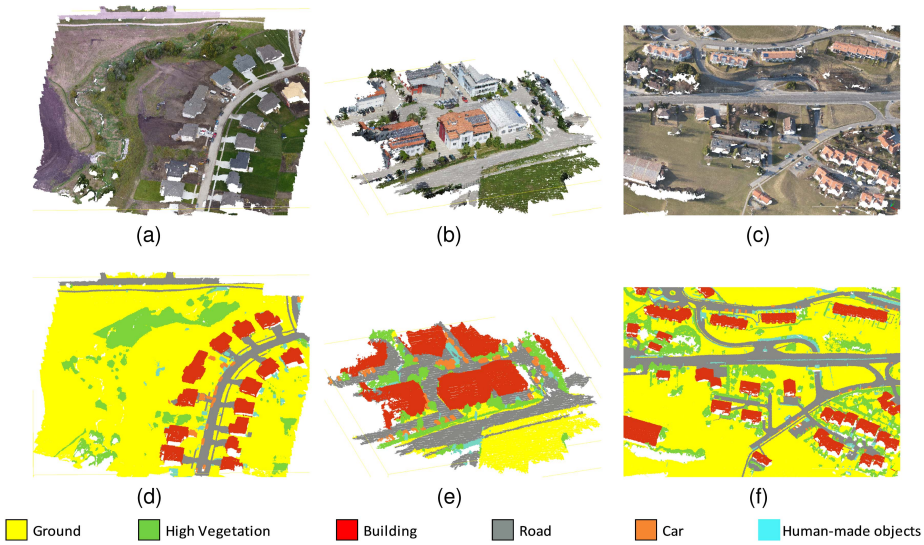

Fig. 1. Point clouds and their ground truths used in this study. (a) Original data (Ankeny). (b) Original data (Building). (c) Original data (Cadastre). (d) Ground truth (Ankeny). (e) Ground truth (Building). (f) Ground truth (Cadastre).

TABLE I
POINT CLOUD DATASETS USED FOR EXPERIMENTS

Dataset	Color	Num. of points	GSD ($\frac{\text{cm}}{\text{pixel}}$)	Area (m^2)	Density ($\frac{\text{points}}{\text{m}^2}$)
Ankeny	RGB	$\approx$ 9.0 M	2.3	55419	161
Building	RGB	$\approx$ 3.4 M	1.8	12767	270
Cadastre	RGB	$\approx$ 5.8 M	5.1	208033	28

and the "Interpretability Module," which explains the contribution of neighboring points to feature extraction. Tan [32] proposed the fractal projection forest (FPF) method, which takes advantage of fractal properties to improve the performance of machine learning-based classifiers. FPF proposed two perturbation-based explanations: Gini Importance and grouped feature ablation.

Although current studies have started to apply explainable AI to point clouds in the last few years, these studies have generally been limited to object classification studies. To the best of our knowledge, this is the first study to apply XAI methods to photogrammetric point cloud classification. There is a gap in the literature on examining the effect of the geometric structure and the spectral information of photogrammetric point clouds on machine learning models. By presenting an XAI-based approach to the use of geometric and spectral features in machine learning approaches, a novel contribution is presented to the point cloud classification literature.

III. MATERIAL AND METHODS

A. Datasets

In this study, point clouds produced by aerial photogrammetry were used for analysis. The datasets used are publicly available [16] and produced with Pix4Dmapper Pro. The datasets consist of three parts: Ankeny, Building, and Cadastre. Ground sampling distances (GSDs) of aerial images vary in each dataset. The entire dataset contains six classes: ground, high vegetation, building, road, car, and human-made objects. The GSD varies significantly between datasets. Ankeny has 2.3 cm/piksel GSD, Building has 1.8 cm/piksel GSD and Cadastre has 5.1 cm/piksel GSD. The area and density are calculated manually. The details about all datasets are given in Table I. The point clouds and ground truth labels are presented in Fig. 1.

B. Geometric Features

Photogrammetric point clouds generally contain only spatial 3-D coordinates and spectral color information. In order to increase the discrimination power of machine learning algorithms, local geometric features of the point cloud can be calculated using 3-D coordinates. Geometric features are produced from the covariance matrix calculated from the local neighborhood area of the central point [13]. Other points within a certain radius around a point in the point cloud are called adjacent points. The area containing neighboring points is also defined as the support area. The geometric features of a point are calculated depending on the relationship of the point with its neighboring points around it. The coordinates of a point in the point cloud can be defined as $P = (x, y, z)$. The covariance matrix is calculated from neighboring points of P point within a certain support radius. Covariance is a metric that measures the amount of deviation from the mean [33]. The covariance matrix of a sphere centered at point $P = (x, y, z)$ is

$$C = \begin{bmatrix} \text{Cov}(x_i, \bar{x}) & \text{Cov}(x_i, \bar{y}) & \text{Cov}(x_i, \bar{z}) \\ \text{Cov}(y_i, \bar{x}) & \text{Cov}(y_i, \bar{y}) & \text{Cov}(y_i, \bar{z}) \\ \text{Cov}(z_i, \bar{x}) & \text{Cov}(z_i, \bar{y}) & \text{Cov}(z_i, \bar{z}) \end{bmatrix} \quad (1)$$

where Cov(*x, y*) is the covariance of *x, y* computed by using the following:

$$\mathrm{Cov}(x,y) = \frac{1}{n-1}\sum_{i=1}^{k}(x_i-\bar{x})(y_i-\bar{y}) \tag{2}$$

where *n* refers to afterward, the eigenvalues of the covariance matrix *C* are calculated. The eigenvalues are ordered from largest to smallest as $\lambda_1 > \lambda_2 > \lambda_3$. After the eigenvalues are obtained, the geometric features are calculated. In this study, nine geometric features, whose equations are given below, were calculated. In addition to these nine features, height of point (Z), roughness, mean curvature, gaussian curvature, normal change rate, number of neighbors, and volume density are added to the feature vector. The eigenvalue-based geometric features are calculated as follows [8]:

$$Sum\ of\ Eigenvalues = \lambda_1+\lambda_2+\lambda_3 \tag{3}$$

$$Linearity = (\lambda_1-\lambda_2)/\lambda_1 \tag{4}$$

$$Planarity = (\lambda_2-\lambda_3)/\lambda_1 \tag{5}$$

$$Sphericity = \lambda_3/\lambda_1 \tag{6}$$

$$Omnivariance = \sqrt[3]{\lambda_1\lambda_2\lambda_3} \tag{7}$$

$$Anisotropy = (\lambda_1-\lambda_3)/\lambda_1 \tag{8}$$

$$Eigenentropy = \sum_{i=1}^{3}\lambda_i \ln \lambda_i \tag{9}$$

$$Surface\ variation = \lambda_3/(\lambda_1+\lambda_2+\lambda_3) \tag{10}$$

$$Verticality = 1-\lambda_3/(\lambda_1+\lambda_2+\lambda_3). \tag{11}$$

C. Color Features

In order to increase the discriminative power of the feature set, color features of each point were added to the feature set in addition to geometric features. There are studies showing that color information improves the photogrammetric point cloud classification performance of the algorithms [34]. Color features are obtained in the RGB color space from the photogrammetric point cloud for each point. Using aerial images, the corresponding RGB information for each point is calculated along with the 3-D coordinates. The datasets used in the study also include color features.

D. SHAP Analysis

Since the gain value showing the total loss reduction in ensemble tree models is inconsistent in describing the importance of a feature, Lundberg and Lee [35] proposed Shapley values to calculate feature importance in order to eliminate this inconsistency. Shapley additive explanations (SHAP) is an XAI approach that explains the relevance and interrelations of input features based on game theory. Shap analysis determines the contribution degrees of selected features to classification according to the final outputs of machine learning models. By assessing the overall relevance of each feature and its local influence, SHAP analysis determines how interpretable the black box ML models.

The Shapley value is calculated generally as follows:

$$\varphi_i = \sum_{S\subseteq N\setminus\{i\}} \frac{|S|!(N-|S|-1)!}{N!}\left[g_x(S\cup\{i\})-g_x(S)\right] \tag{12}$$

$$g_x(S) = E\left[g(x)\mid x_K\right] \tag{13}$$

where N is the set of the all features. ϕ_i represents the contribution of feature i. S is defined as any feature subset not including the feature i. g(x) refers to the expected value for each feature in the subset S. The expected value of each observation is obtained by summing the Shapley values of all the features [36].

There are many different types of SHAP (e.g., DeepSHAP, Kernel SHAP, LinearSHAP, and TreeSHAP) which is implemented by Lundberg and Lee. TreeSHAP, which is used to describe the predictions of machine learning algorithms in this study, uses a linear explanatory model and Shapley values to predict the initial prediction model [36]. The linear explanatory model is calculated as follows:

$$h\left(z'\right) = \emptyset_0 + \sum_{\mathrm{I}=1}^{K}\emptyset_i z_i'. \tag{14}$$

In (14), $h(z')$ is explanation of the model. K is the number of features. $\emptyset$ refers to the feature attribution.

E. Filter-Based Feature Selection

The features calculated to increase the distinctiveness of the algorithms do not have the same effect on the classification. Some features are more suitable for classification, while others are less suitable. Thus, in order to improve the classification performance of the algorithm, increase efficiency in terms of time and memory consumption, and preserve effective features, feature selection methods have been developed to detect important features. Finding the minimum number of features that will adequately describe the data is referred to as feature selection [37]. Feature selection methods can be grouped as filter-based, wrapper-based, and embedded methods. Wrapper-based methods and embedded-based methods may have better performance than filter-based methods because they work based on a classifier. However, they only identify features that are optimized for a specific classifier. Filter-based methods are model-agnostic, simple, and efficient. Popular filter-based feature selection methods such as information gain (IG) and ReliefF were used in this study.

1) Information Gain: IG [38] is an entropy-based feature selection algorithm that measures the amount of information provided by features. The importance of the features in terms of classification and which features are appropriate to use are decided by applying inflammation gain. The feature's importance is calculated as follows:

$$G(D,t) = -\sum_{i=1}^{m}P(C_i)\log P(C_i) + P(t)\sum_{i}^{m}P(C_i|t)\log P(C_i|t)$$

$$+ P(\bar{t}) \sum_{i}^{m} P(C_i|\bar{t}) \log P(C_i|\bar{t}) \tag{15}$$

where C refers to a set of feature, $C_i|t$ is a feature set without feature t. The $G(D,t)$ represents the importance of the feature. The t features with the highest $G(D,t)$ should be selected. m is the number of class [38].

2) ReliefF: ReliefF [39] is a filter-based feature selection algorithm that provides high efficiency in solving multiclassification problems that weight relevant features and eliminates irrelevant features. ReliefF aims to determine the importance of features based on the distinction between randomly selected and close samples from the training set. Randomly selected points have k neighbor points with the same label and k neighbor points with different labels. If a feature has different values for points with the same label, the importance of that feature is reduced. If it has different values for points with different labels, the importance of the feature is increased. R_f score is computed as follows:

$$R_f(x_i) = \frac{1}{N} \sum_{t=1}^{N} \left\{ -\frac{1}{k} \sum_{x_i \in NH(y)} \text{diff}(x_{t,i}, x_{j,i}) + \sum \frac{1}{k} \frac{P(y)}{1 - P(y_i)} \sum_{x_j \in NM(x_i,y)} \text{diff}(x_{t,i}, x_{j,i}) \right\} \tag{16}$$

where $R_f(x_i)$ refers to the score of x_i. y_i is the class label of the sample x_t. $P(y)$ defines probability of a sample being from class y. $x_{t,i}$ defines the values of x_t on feature x_i and diff$(\cdot)$ is the function used to calculate the difference between $x_{t,i}$ and $x_{j,i}$. $NH(x_i, y)$ is neighbors with the same class label. $NM(x_i, y)$ is neighbors with different class label [37]. N is the number of samples in the input data.

F. Machine Learning Classifiers

Traditional machine learning techniques provide promising results for many classification problems [40]. Ensemble learning is a machine learning approach developed with the idea of creating more powerful models by integrating multiple models to solve a problem. Especially by using more than one classifier instead of a single classifier, its variance is reduced and more reliable results are obtained [41]. Gradient bossing machines combine weak classifiers to iteratively create a strong classifier, improving scores compared to the previous iteration. This update process makes gradient boosting machines superior to single-step decision tree building methods such as RF. XGBoost and LightGBM, based on gradient boosting decision tree (GBDT), have become popular algorithms in machine learning in recent years because they provide less training time and higher accuracy [42]. In this study, a bagging method named RF and two boosting methods such as LightGBM and XGBoost were selected as classifiers. RF, XGboost, and LightGBM have also been applied for point cloud classification in recent years [17], [18], [43].

1) Random Forest: There are two basic ensemble learning methods, boosting and bagging. In boosting, models are trained sequentially with a data sample. Each model tries to compensate for the weaknesses of the previous model. In bagging, another ensemble approach, the stability and accuracy of the models are at the forefront. To train a model in bagging, multiple subdatasets are generated from a training set and a subdataset is assigned to each tree in the tree structure. Bagging improves accuracy using random features and can be used to give estimates of the generalized error of the aggregated tree ensemble [44].

RF [45] is the first successful bagging algorithm. RF, as an enhanced version of bagging, is a machine learning method that aggregates a large number of trees and decides based on the predictions of trees. RFs with different tree structures tend to overfitting and outliers. For this reason, prediction voting for classification is an approach to prevent overfitting. For regression problems, the tree estimates are averaged [44].

RF has two basic parameters: number of features used in each node and the number of trees to develop to determine the best split. Although a lower number of features provides faster computation, it can affect classification accuracy by reducing both the correlation between any two trees and the power of each tree in the forest. The number of trees can be as large as possible based on the hardware. To achieve the global optimum, 2/3 of the samples are used to train the bagged trees, and 1/3, which is out of bag (OOB) data, is used for the calculation of the test error of the RF model. The error obtained from this process is called the generalized error

$$PE^* = P_{X,Y}(\text{mg}(X,Y) < 0). \tag{17}$$

Margin function (mg()) measures how much the average number of votes for correct predictions in (X, Y) exceeds the average score for other classes. The margin is higher the more accurately classification can be made [45].

2) Light Gradient-Boosting Machine (LightGBM): The success, precision, and comprehension of the GBDT make it a popular machine learning approach. GBDT, is a decision tree ensemble model achieves cutting-edge performance in a wide range of machine learning tasks [46]. LightGBM [46] is a tree-based method generated by Microsoft under the notion of GDBT for quick and efficient prediction issue solution in huge high-dimensional data. LightGBM is an upgraded version of XGBoost. In LightGBM, gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) are proposed. GOSS is a cutting-edge sampling approach for GBDT that can maintain the accuracy of trained decision trees while lowering the amount of data instances. EFB is an innovative technique for efficiently reducing the amount of features.

Instead of controlling all previous leaves in the decision tree, LightGBM adopts a histogram-based and leaf-based growth strategy with a maximum depth limit. In the histogram-based DT algorithm, a histogram of size s is created with consecutive floating-point eigenvalues to calculate the required statistics. Discrete values in the histogram define segmentation points. The leaf-based growth strategy considers only the leaf with the most IG in the same layer to increase accuracy and prevent overfitting [47].

3) Extreme Gradient Boosting (XGBoost): Extreme Gradient Boosting, often known as XGBoost [48], is a reliable and well-known machine learning technique that is typically used to solve regression and classification problems. It is an enhanced form of the gradient boosting machine learning approach, which combines the predictions of multiple weak models (typically decision trees) to create an efficient predictive model [49]. In (2), traditional loss function and model complexity define the two components of XGBoost's objective function. XGBoost seeks to minimize the regularized objective function in order to learn the function applied to the model. The difference between the prediction y_i and the target $\hat{y}_i$ calculated using the differentiable convex loss function in (18)'s first term. In (19), the second term is utilized to assess model complexity

$$L(\emptyset) = \sum_i l\left(\hat{y}_i, y_i\right) + \sum_k \Omega\left(f_k\right) \tag{18}$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\|\omega\|^2. \tag{19}$$

Tree complexity is adjusted with γ and λ. The extra regularization term smooths the final learning weight and prevents over-fitting. The output of a tree ensemble model is estimated using functions of K additive for the provided dataset D=(x_i, y_i) with n samples and m features. The formula for prediction is as follows:

$$\hat{y}_i = \emptyset\left(x_i\right) = \sum_{k=1}^{K} f_k\left(x_i\right), f_k \in \varphi \tag{20}$$

$$\varphi = \left\{f(x) = \omega_{q(x)}\right\}\left(q : R^m \to T, \omega \in R^T\right. . \tag{21}$$

While x_i is one of the samples, $f_k(x_i)$ is the expected score for the given sample, and y_i is the expected outcome of the model. φ in (5) denotes the collection of regression trees with the independent q tree topology. T is the number of leaves on the tree, and w is their weight.

G. Hyperparameter Tuning

The most suitable hyperparameters for the applied dataset must be determined in order to get the best prediction performance from ML models. The learning process of an ML model can be controlled with hyperparameters. Hyperparameter tuning is defined as the determination of the optimum hyperparameter group that minimizes the loss function. In this study, GridSearch was used for hyperparameter tuning. In this method, a group of possible values for a hyperparameter is created and the aim is to determine the most appropriate values. Each model created is tested on the dataset and the model with the highest classification performance is determined.

The n_estimators, max_features, min_sample_split, and random_state hyperparameters were tuned for RF. In the Building and Cadastre datasets, n_estimators, max_features, min_sample_split and random_state were selected as 300, 10, 10, and 300, respectively. In the Ankeny dataset, n_estimators, max_features, min_sample_split and random_state were selected as 300, 10, 10, and 200, respectively. For XGBoost, n_estimators, max_depth and learning rate hyperparameters were selected as 300, 10, and 0.1, respectively, for all datasets. The hyperparameters n_estimators, num_leaves, max_depth, and min_data_in_leaf have been fine-tuned for LightGBM. In all datasets, n_estimators=300, min_data_in_leaf=20 and num_leaves=100 were selected. max_depth is set to 30 for Building and Cadastre and 50 for Ankeny.

H. Experimental Details

In this study, the effect of determining the optimum features on machine learning-based point cloud classification was investigated. In the study, photogrammetric point clouds were classified using three ensemble machine-learning algorithms: RF, XGBoost, and LightGBM. Scikit-learn [50] is a Python library to implement machine learning algorithms. Photogrammetric point clouds contain color information as well as 3-D coordinates. In order to increase the classification performance of machine learning algorithms, the geometric features of each point in the point cloud are also calculated and included in the feature space of the relevant point. Geometric features describe the local geometry that a point creates together with neighboring points. Experiments were carried out with different support radius values to determine the most suitable geometric features. The specified radius values are 0.5, 1, 1.5, and 2 m. In the second step, in order to determine the effect of using together geometric features and spectral features, the effect of defining a point with only geometric features and only spectral features on classification was also investigated. As a result, the feature space of a point contains a total of 19 features along with geometric and spectral features. The class-based impact of these nineteen features was analyzed using SHAP, one of the XAI approaches. In addition to SHAP analysis, feature importance values were calculated with filter-based IG and ReliefF approaches. By determining the cut-of-point for each method, fewer features were selected and the performances of ML classifiers were examined. Thus, the effect of feature selection on point cloud classification was investigated. The workflow of the study is presented in Fig. 2.

As the training dataset changed, fine-tuned hyperparameters for the ML classifier were determined in each training. Feature selection algorithms were implemented using the WEKA workbench [51]. For the experiments, i7-11800H, 2.30 GHz processor, GTX 3070 graphics card, and 32 GB RAM hardware is used.

Overall accuracy (OA), F1-score, precision, and recall metrics were used to evaluate the study results. OA represents the ratio of correctly predicted samples to all samples. TP is the number of samples that their the estimated label and ground truth are positive. FP is the number of samples that their the estimated label is positive, but the ground truth is negative. FN is the number of samples that their the estimated label is negative, but the ground truth is positive. The evaluation metrics are calculated as follows:

$$\text{Overall accuracy} = \sum_{i=1}^{k} \frac{N_{ii}}{N} \tag{22}$$

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{23}$$

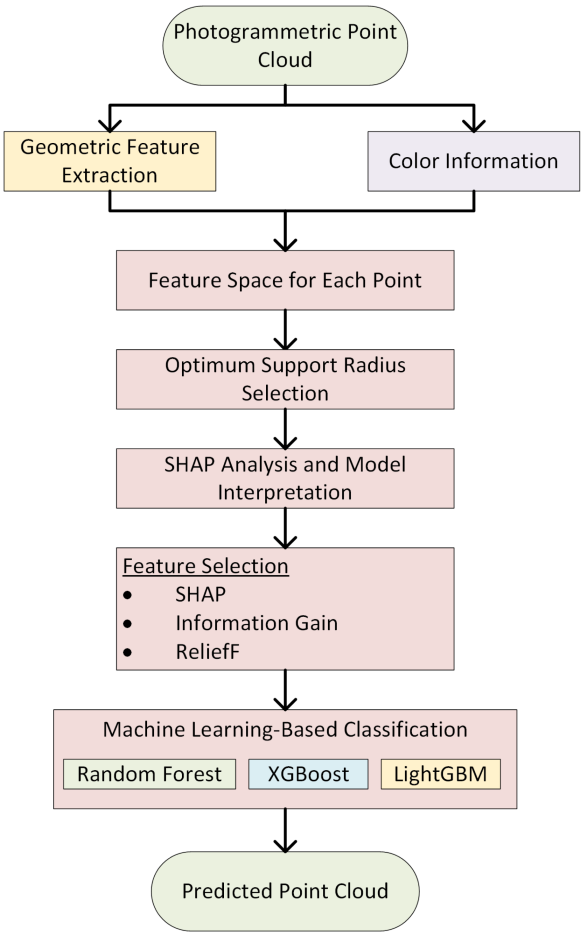

Fig. 2. Workflow of the study.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{24}$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{25}$$

IV. Results

A. Model Interpretation With SHAP Analysis

The SHAP summary graph lists the most relevant features in descending order, explaining the contribution of features to the prediction of the classification model. Since each color represents a class, the effect of each feature on the prediction of the relevant class can be interpreted in the classification process. SHAP analysis was performed for each algorithm depending on the dataset. There is a similar ranking for all three models in the Ankeny dataset. Height is seen as the most important feature. For the RF model, Height is the most effective feature in the Building dataset [see Fig. 3(b)], while blue is the most effective feature in the Cadastre dataset [see Fig. 3(c)]. In the Ankeny dataset, height contributed the most to the car class prediction of the RF model [see Fig. 3(a)]. In the Building dataset, it is effective in the high vegetation class. In the XGBoost model, a similar ranking to RF is seen for all three datasets. Height is the most important feature for Ankeny [see Fig. 4(a)] and Building [see Fig. 4(b)] datasets, and blue for Cadastre [see Fig. 4(c)]. On the other hand, sphericity, surface variation, and volume density do not contribute to XGBoost model estimation in any dataset. In LightGBM, height has the highest importance in the Ankeny dataset (Fig. 5(a)), while blue is the most important feature in Building [see Fig. 5(b)] and Cadastre [see Fig. 5(c)] datasets. Surface variation and volume density did not contribute to the predictions of LightGBM model. In addition, it is seen that the height feature makes a significant contribution to LightGBM's building estimations. In all datasets, the effects of features on the human-made object class are low. Color information is often among the influential features in all datasets. Considering the results of the study showing the effect of color information on point cloud classification, it is understood that significant results were obtained with the SHAP analysis.

TABLE II
OA of the Algorithms on the Datasets Based on the Support Radius

Classifier	Data set	0.5 m	1 m	1.5 m	2 m
RF	Ankeny	81.55	83.22	83.26	**83.35**
	Building	86.47	88.05	88.55	**88.84**
	Cadastre	–	76.93	78.47	**78.93**
XGBoost	Ankeny	82.77	84.36	84.72	**85.09**
	Building	87.88	90.01	90.84	**91.27**
	Cadastre	–	80.33	81.96	**82.04**
LightGBM	Ankeny	83.04	84.65	85.00	**85.21**
	Building	88.30	90.53	91.35	**91.70**
	Cadastre	–	81.07	82.82	**82.88**

The highest OA values are marked as bold.

B. Optimum Support Area Selection

In order to avoid strong assumptions about 3-D neighborhoods, many studies have focused on determining the optimum neighborhood size for each point and, thus increasing the distinctiveness of geometric features [8]. In this study, the support area was changed from 0.5 to 2 m in order to determine the support area to calculate the most suitable geometric features. For all algorithms, the highest OA value in all three datasets was reached when a 2 m support radius was selected. Generally, similar results are obtained with 1.5 and 2 m support areas. With LightGBM, the highest OA values are achieved in all three datasets. When all geometric features are used, LightGBM achieves 85.21% OA in the Ankeny dataset, 91.70% OA in Building dataset and 82.88% OA in the Cadastre dataset. Point cloud density is an important parameter when determining the support area size. When the support radius of 0.5 m is selected in the Cadastre dataset, the geometric features of many points cannot be calculated because there are not enough neighbor points. For this reason, results for 0.5 m could not be presented in the Cadastre dataset. Classification results are presented in Table II. The experiments were carried out using the appropriate support radius determined in this section.

C. Effect of Color and Geometric Features

Each point in the point cloud is defined using two types of features: color information and geometric features. The effects

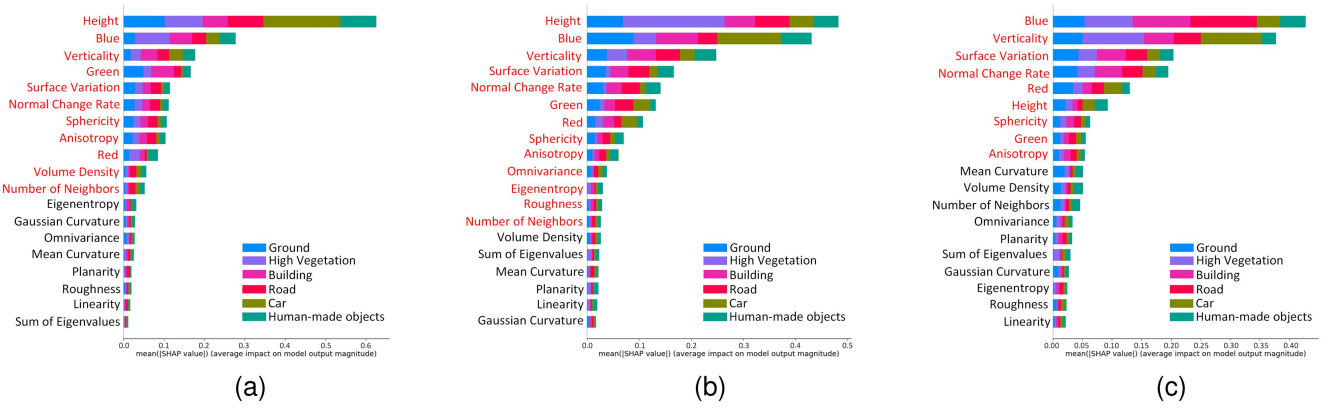

Fig. 3. SHAP values calculated with RF. The selected features are highlighted in red. (a) Ankeny. (b) Building. (c) Cadastre.

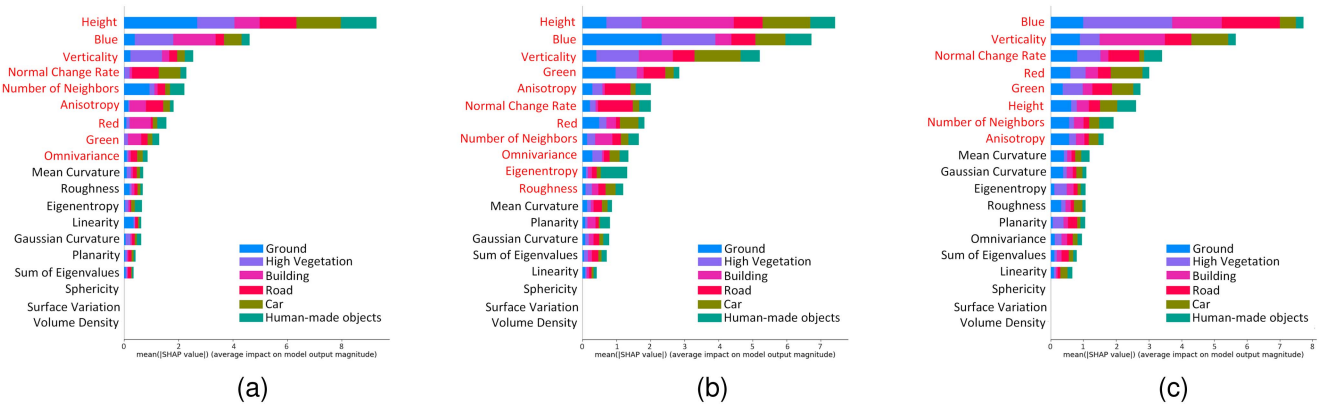

Fig. 4. SHAP values calculated with XGBoost. The selected features are highlighted in red. (a) Ankeny. (b) Building. (c) Cadastre.

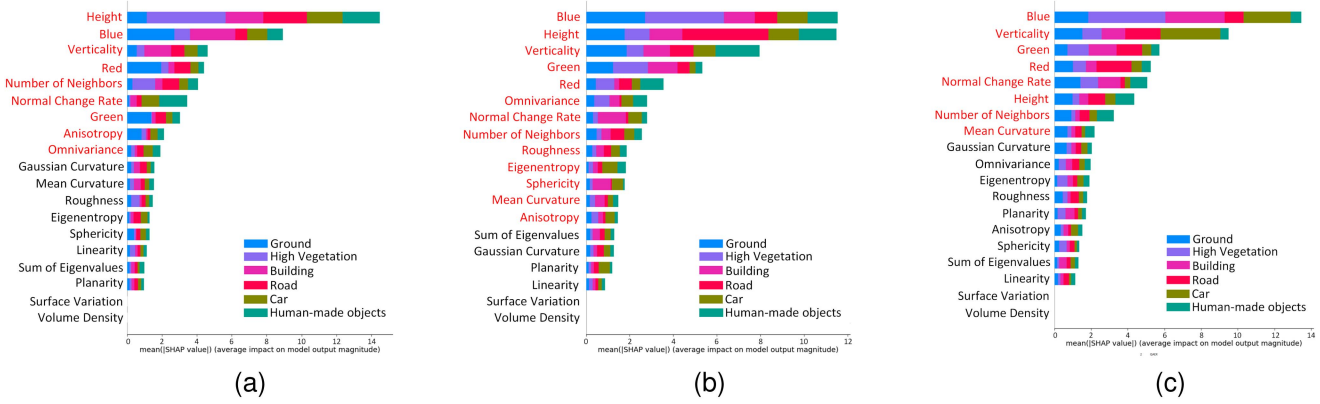

Fig. 5. SHAP values calculated with LightGBM. The selected features are highlighted in red. (a) Ankeny. (b) Building. (c) Cadastre.

of these feature groups separately were also examined within the scope of the study. According to Table III, classification accuracy is significantly reduced when only RGB information is used. In the Ankeny dataset, the RF algorithm has an OA of less than 60%. If only spectral features are used, similarly colored classes are mixed with each other. When only geometric features are used, classification is performed with higher accuracy than classification with RGB features. Only in the Cadastre dataset, the RF and XGBoost algorithms show lower classification performance when only geometric features are used. In the Building and Ankeny dataset, more than 10% improvement is achieved in all algorithms compared to the case where only RGB is used.

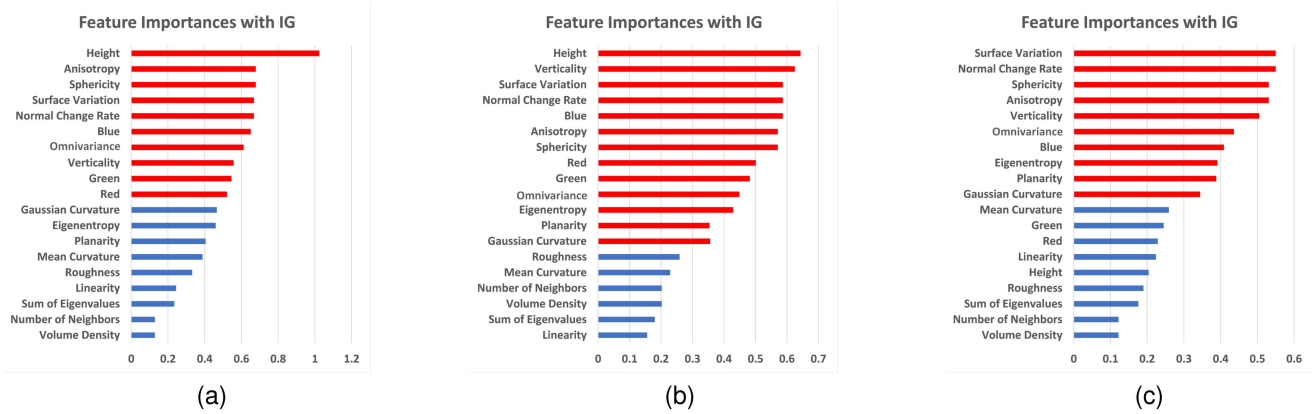

Fig. 6. Feature importance calculated by IG for the three datasets. The selected features are marked as red color. (a) Ankeny. (b) Building. (c) Cadastre.

TABLE III
OA OF THE ALGORITHMS ON THE DATASETS ACCORDING TO COLOR AND GEOMETRIC FEATURES

Classifier	Dataset	RGB	GEO	RGB+GEO
RF	Ankeny	64.47	73.21	**83.35**
	Building	71.96	81.96	**88.84**
	Cadastre	66.43	62.35	**78.93**
XGBoost	Ankeny	61.29	75.10	**85.09**
	Building	68.61	84.42	**91.27**
	Cadastre	64.50	64.15	**82.04**
LightGBM	Ankeny	59.63	74.64	**85.21**
	Building	67.47	84.72	**91.70**
	Cadastre	61.57	64.50	**82.88**

The highest OA values are marked as bold.

Finally, in classifications applied with feature spaces where RGB and geometric features are used together, all algorithms reach the highest classification accuracy in all three datasets. In the Building dataset, 88.84%, 91.27% and 91.70% OA were obtained with RF, XGBoost, and LightGBM models, respectively.

D. Feature Selection for Classification

To determine the contributions of spectral and geometric features classification accuracy, three feature selection strategies were applied to create optimal subsets. The feature importance values are ordered from the largest to the smallest. While determining the relevant features, the feature area was scanned by adding a certain number of features and the feature set with the highest accuracy was experimentally determined. The optimal number of features is different for each algorithm and dataset. As a result of SHAP, 11, 13, and 9 features were selected for the RF model, 9, 11, and 8 features were selected for the XGBoost, and 9, 13, and 8 features were selected for the LightGBM for Ankeny, Building, and Cadastre datasets, respectively. Since the filter-based methods are independent of the classifier, only one importance assessment was made for each dataset. In the three datasets, both IG and ReliefF identify the height attribute as having a high impact on classification. Only in the Cadastre dataset, when feature selection with IG is applied, height has low importance. It is seen that spectral information (RGB) is generally among the selected features. The feature importance and selected features are presented in Figs. 6 and 7.

In the Ankeny dataset, all three algorithms achieve the highest classification performance when the features determined by SHAP are used. With LightGBM, 85.50% OA and 75.43% F1-score were obtained. Algorithms have similar performance when the features determined by ReliefF and SHAP are used in the Building dataset. In the Cadastre dataset, when the methods selected with IG are used, the algorithms reach higher OA. 83.88% OA was achieved with RF, 85.22% OA with XGBoost, and 85.70% OA with LightGBM. The details of the classifications based on the selected features with IG, ReliefF, and SHAP are presented in Table IV. The classified point clouds are illustrated in Figs. 8–10.

V. DISCUSSION

In this section, the results of the research are evaluated in terms of the feature selection in point cloud classification with machine learning. The presented work includes a comprehensive set of experiments that allow to evaluate the classification of photogrammetric point clouds from different perspectives. The results of the study allow general inferences to be made regarding the feature selection with XAI and filter-based algorithms for machine learning-based point cloud classification.

The support area of the point is the basic parameter in determining the geometric features. According to the results of this study, as the support area increases, the distinctiveness of the geometric features increases. The classification performance of algorithms with a radius of 0.5 m support is significantly reduced. Because the number of neighbors of the points is insufficient to define the local geometry (see Table II).

It is understood that the point cloud GSD and classification accuracy are inversely proportional. While ML classifiers achieved the highest OA in the Building dataset with 1.8 cm GSD, the lowest OA was obtained in the Cadastre dataset with 5.1 cm GSD. In aerial images with high GSD, details decrease and objects become difficult to distinguish from each other.

For feature selection, we focused on both geometric and spectral features. As in other studies [16], using geometric and spectral features together significantly improves the performance of ML classifiers. The distinguishability of classes with similar

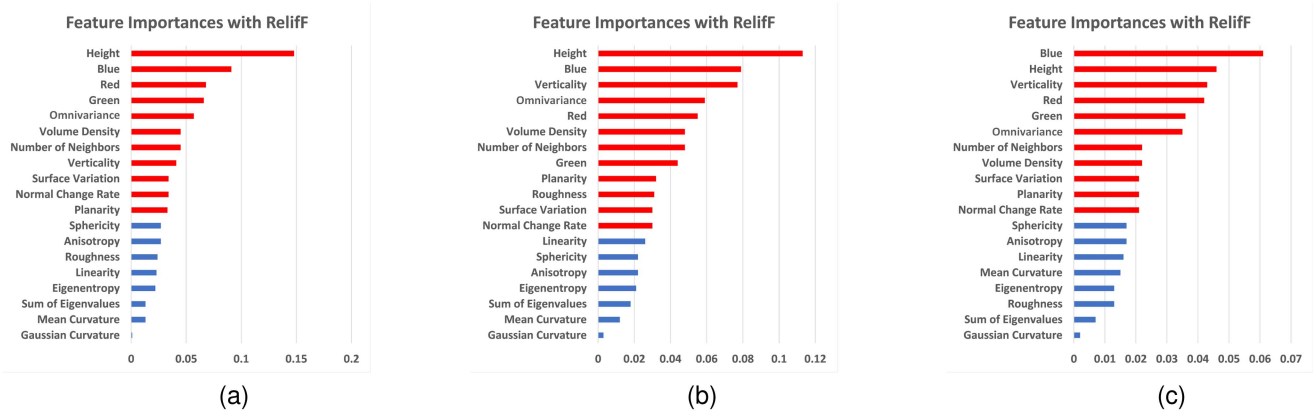

Fig. 7. Feature importance calculated by ReliefF for the three datasets. The selected features are marked as red color. (a) Ankeny. (b) Building. (c) Cadastre.

TABLE IV
CLASSIFICATION RESULTS OF THE ALGORITHMS ACCORDING TO SELECTED FEATURE IG, RELIEFF AND SHAP ANALYSIS

Classifier	Dataset	Features Selected with IG				Features Selected with ReliefF				Features Selected with SHAP			
		Precision	Recall	F1	OA	Precision	Recall	F1	OA	Precision	Recall	F1	OA
RF	Ankeny	66.51	85.03	70.76	82.66	68.04	86.71	72.48	83.51	67.79	86.51	72.31	**83.70**
	Building	73.92	85.05	76.78	88.40	75.01	86.22	78.04	**89.11**	75.03	86.27	78.07	89.08
	Cadastre	70.68	55.47	55.60	**83.88**	59.06	76.65	60.89	79.30	59.01	75.54	60.56	79.36
XGBoost	Ankeny	67.20	87.09	72.21	83.34	69.58	88.96	74.91	85.16	69.47	88.86	74.84	**85.22**
	Building	76.31	87.25	79.43	90.19	78.04	88.71	81.27	**91.22**	78.14	88.78	81.36	91.21
	Cadastre	67.29	58.46	59.87	**85.22**	62.17	79.88	64.30	82.36	62.50	80.01	64.54	82.47
LightGBM	Ankeny	67.31	87.47	72.48	83.33	69.86	89.52	75.41	85.30	69.92	89.42	75.43	**85.50**
	Building	76.86	87.87	80.02	90.55	78.90	89.70	82.22	**91.77**	79.01	89.85	82.31	91.70
	Cadastre	67.10	59.27	60.92	**85.70**	62.94	80.63	65.20	83.31	63.20	80.89	65.57	83.28

The highest OA values are marked as bold.

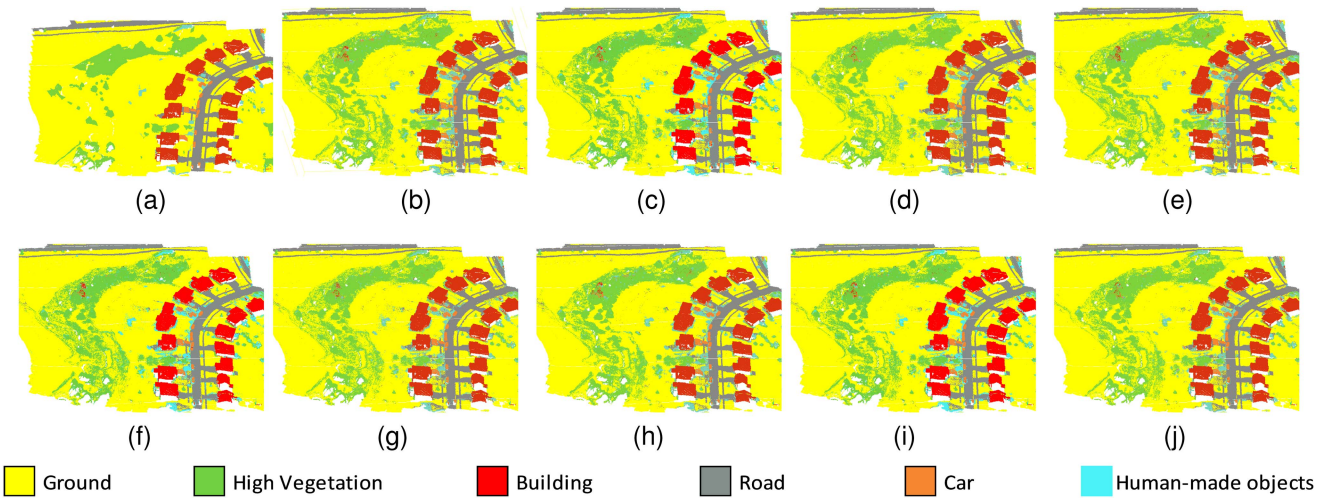

Fig. 8. Classified point clouds for Ankeny dataset. (a) Ground Truth. (b) RF (SHAP). (c) XGBoost (SHAP). (d) LightGBM (SHAP). (e) RF (IG). (f) XGBoost (IG). (g) LightGBM (IG). (h) RF (ReliefF). (i) XGBoost (ReliefF). (j) LightGBM (ReliefF).

geometric features is increased by spectral features. Ground and road classes can be easily distinguished from each other using spectral features. When only spectral features are used, classes such as building and road with similar reflectance values may not be distinguished. Feature selection increases the accuracy of ML classifiers in all three datasets. Thus, it has been shown that it is possible to achieve higher accuracy classification by using fewer features. Features with low importance negatively affect the classification performance of ML algorithms. SHAP and ReliefF produce very similar results. According to F1-score values, it showed slightly higher performance in most of the classification scenarios performed with the features selected by SHAP analysis. Higher accuracy classification is achieved with the features selected with IG in the Cadastre dataset.

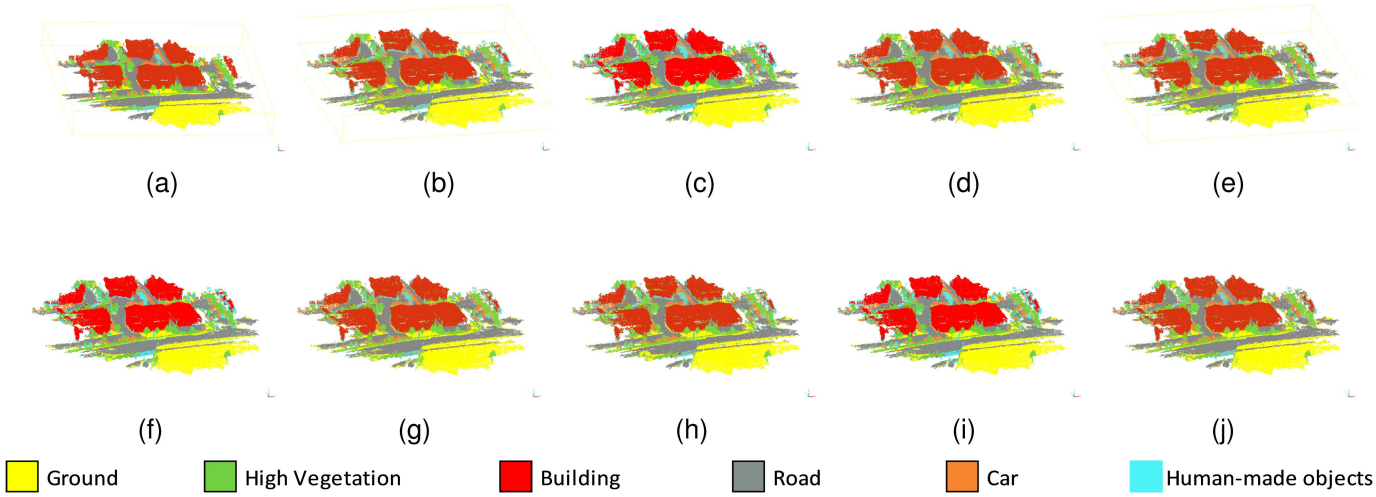

Fig. 9. Classified point clouds for Building dataset. (a) Ground Truth. (b) RF (SHAP). (c) XGBoost (SHAP). (d) LightGBM (SHAP). (e) RF (IG). (f) XGBoost (IG). (g) LightGBM (IG). (h) RF (ReliefF). (i) XGBoost (ReliefF). (j) LightGBM (ReliefF).

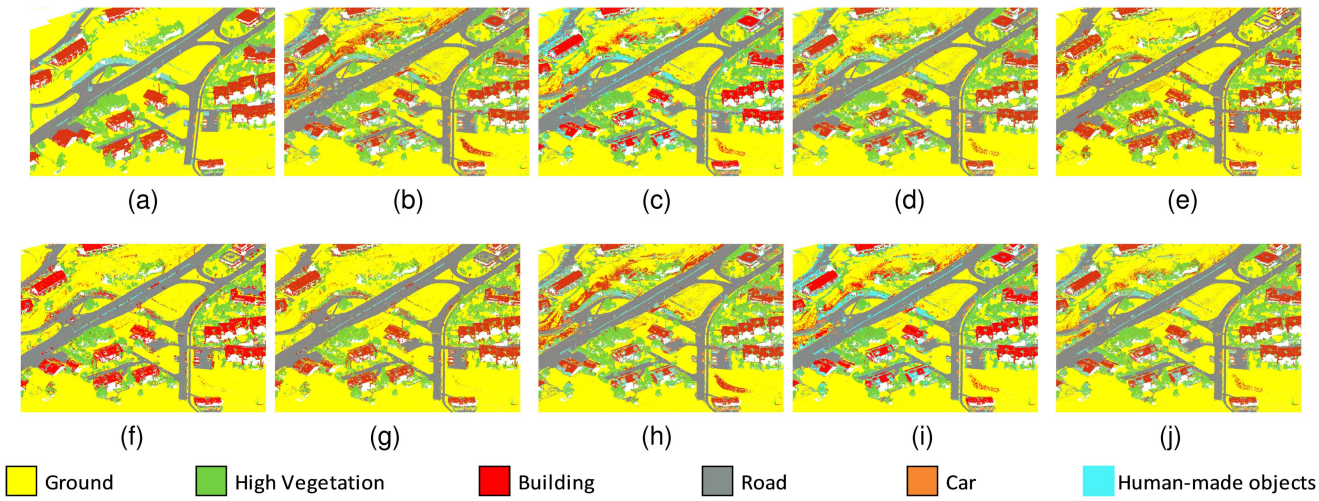

Fig. 10. Classified point clouds for Cadastre dataset. (a) Ground truth. (b) RF (SHAP). (c) XGBoost (SHAP). (d) LightGBM (SHAP). (e) RF (IG). (f) XGBoost (IG). (g) LightGBM (IG). (h) RF (ReliefF). (i) XGBoost (ReliefF). (j) LightGBM (ReliefF).

When the selected features are examined, the geometric features that most affect point cloud classification are height and features describing surface changes. Linearity, sum of eigenvalues, and volume density appear to be less effective features. Cadastre dataset is a dataset that covers a larger area and has more topographic change. The height differences between classes are not as clear as in the other two datasets. For this reason, it is less important than other features in both SHAP analysis and filter-based feature selection algorithms. In addition, it appears that spectral information is generally a feature with high impact. The results in Table IV confirm this consequences. Photogrammetric point clouds with spectral information obtained from high-resolution aerial images have significant advantages for point cloud classification.

Although SHAP analysis and filter-based feature selection algorithms have similar results, filter-based feature selection algorithms do not provide information about the impact of features on class predictions. On the other hand, SHAP analysis helps to explain the working principles of black-box ML models. Particularly, the effects of handcrafted features on the decision-making process of ML algorithms can be examined through SHAP analysis. While ReliefF and IG algorithms generally present the feature effect, SHAP analysis also reveals the class-based effects of geometric and spectral features. In terms of classification results, the features selected by SHAP analysis offer competitive results with filter-based feature selection algorithms, and in addition, it can quantitatively calculate the effects of each feature in the feature set on the classification process. This shows the advantage of SHAP analysis as an XAI approach over classical filter feature selection algorithms for point cloud classification with ML. For this reason, it was concluded that it would be more appropriate to use SHAP to perform more detailed analysis and achieve high accuracy in photogrammetric point cloud classification.

VI. Conclusion

The need for efficient and effective processing of high-dimensional 3-D data encourages the development of consistent models for the correct interpretation and classification of such complex data with high accuracy. Although traditional feature selection algorithms are used for point cloud classification, they are far from providing a detailed and class-based analysis. In this context, XAI approaches that have come to the fore recently have great potential for interpreting and effectively classifying the geometric structure of point clouds. Although geometric features are widely used for point cloud classification in existing studies, the effects of these features on the decisions of classifier algorithms is an important gap in the literature. In this study, the effects of SHAP analysis and filter-based feature selection algorithms on the photogrammetric point cloud classification performances of ensemble ML models were investigated. Spectral features as well as geometric features produced from the 3-D coordinates of the point cloud were used to define a point. With point clouds being one of the main data sources for describing the environment, the features used for information extraction and the explainability of artificial intelligence-based models are becoming an important research area. This study is one of the pioneer works which examine the effects of geometric features on point cloud classification using XAI approaches.

XAI approaches have inspiring results for the explainability of black-box ML models in point cloud classification. In future studies, thanks to explainable deep learning networks and effective feature selection, the computational load required by deep learning networks can be reduced and high-accuracy point cloud classification can be achieved.

References

[1] M. Zhang, H. You, P. Kadam, S. Liu, and C. C. Kuo, "PointHop: An explainable machine learning method for point cloud classification," *IEEE Trans. Multimedia*, vol. 22, no. 7, pp. 1744–1755, Jul. 2020.

[2] O. Akyol and Z. Duran, "Low-cost laser scanning system design," *J. Russian Laser Res.*, vol. 35, pp. 244–251, 2014.

[3] J. Iglhaut, C. Cabo, S. Puliti, L. Piermattei, J. O'Connor, and J. Rosette, "Structure from motion photogrammetry in forestry: A review," *Curr. Forestry Rep.*, vol. 5, pp. 155–168, 2019.

[4] H. Zhang et al., "Deep learning-based 3 D point cloud classification: A systematic survey and outlook," *Displays*, vol. 79, 2023, Art. no. 102456. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0141938223000896

[5] M. E. Atik and Z. Duran, "Classification of aerial photogrammetric point cloud using recurrent neural networks," *Fresenius Environ. Bull.*, vol. 30, pp. 4270–4275, 2021.

[6] V. Belle and I. Papantonis, "Principles and practice of explainable machine learning," *Front. Big Data*, vol. 4, 2021, Art. no. 688969.

[7] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[8] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet, "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers," *ISPRS J. Photogrammetry Remote Sens.*, vol. 105, pp. 286–304, 2015.

[9] M. E. Atik, Z. Duran, and D. Z. Seker, "Machine learning-based supervised classification of point clouds using multiscale geometric features," *ISPRS Int. J. Geo- Inf.*, vol. 10, 2021, Art. no. 187.

[10] M. E. Atik and Z. Duran, "Selection of relevant geometric features using filter-based algorithms for point cloud semantic segmentation," *Electron. (Switzerland)*, vol. 11, 2022, Art. no. 3310.

[11] J.-Y. Rau, J.-P. Jhan, and Y.-C. Hsu, "Analysis of oblique aerial images for land cover and point cloud classification in an urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1304–1319, Mar. 2015.

[12] N. Yastikli and Z. Cetin, "Classification of raw Lidar point cloud using point-based methods with spatial features for 3 D building reconstruction," *Arabian J. Geosciences*, vol. 14, 2021, Art. no. 146.

[13] Z. Duran, K. Ozcan, and M. E. Atik, "Classification of photogrammetric and airborne Lidar point clouds using machine learning algorithms," *Drones*, vol. 5, 2021, Art. no. 104.

[14] G. Vosselman, M. Coenen, and F. Rottensteiner, "Contextual segment-based classification of airborne laser scanner data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 128, pp. 354–371, 2017.

[15] M. E. Atik and Z. Duran, "An efficient ensemble deep learning approach for semantic point cloud segmentation based on 3 D geometric features and range images," *Sensors*, vol. 22, 2022, Art. no. 6210.

[16] C. Becker, E. Rosinskaya, N. Häni, E. d'Angelo, and C. Strecha, "Classification of aerial photogrammetric 3 D point clouds," *Photogrammetric Eng. Remote Sens.*, vol. 84, pp. 3–10, 2018.

[17] E. Sevgen and S. Abdikan, "Classification of large-scale mobile laser scanning data in urban area with LightGBM," *Remote Sens.*, vol. 15, no. 15, 2023, Art. no. 3787. [Online]. Available: https://www.mdpi.com/2072-4292/15/15/3787

[18] S. M. Krishna Moorthy, K. Calders, M. B. Vicari, and H. Verbeeck, "Improved supervised learning-based approach for leaf and wood classification from LiDAR point clouds of forests," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3057–3070, May 2020.

[19] M. Sharma and R. D. Garg, "Building footprint extraction from aerial photogrammetric point cloud data using its geometric features," *J. Building Eng.*, vol. 76, 2023, Art. no. 107387. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S235271022301567X

[20] M. Zeybek, "Classification of uav point clouds by random forest machine learning algorithm," *Turkish J. Eng.*, vol. 5, no. 2, pp. 48–57, 2021.

[21] J. P. Carbonell-Rivera, J. Torralba, J. Estornell, L. A. Ruiz, and P. Crespo-Peremarch, "Classification of mediterranean shrub species from UAV point clouds," *Remote Sens.*, vol. 14, no. 1, 2022, Art. no. 199. [Online]. Available: https://www.mdpi.com/2072-4292/14/1/199

[22] L. Weidner, G. Walton, and A. Krajnovich, "Classifying rock slope materials in photogrammetric point clouds using robust color and geometric features," *ISPRS J. Photogrammetry Remote Sens.*, vol. 176, pp. 15–29, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271621000964

[23] A. Akagic, S. Krivić, H. Dizdar, and J. Velagić, "Computer vision with 3 D point cloud data: Methods, datasets and challenges," in *Proc. 28th Int. Conf. Inf., Commun. Autom. Technol.*, 2022, pp. 1–8.

[24] S. O. Atik and C. Ipbuker, "Integrating convolutional neural network and multiresolution segmentation for land cover and land use mapping using satellite imagery," *Appl. Sci.*, vol. 11, no. 12, 2021, Art. no. 5551. [Online]. Available: https://www.mdpi.com/2076-3417/11/12/5551

[25] Q. Hu et al., "RandLA-Net: Efficient semantic segmentation of large-scale point clouds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11108–11117.

[26] L. Landrieu and M. Simonovsky, "Large-scale point cloud semantic segmentation with superpoint graphs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4558–4567.

[27] F. T. Kurdi, W. Amakhchan, Z. Gharineiat, H. Boulaassal, and O. El Kharki, "Contribution of geometric feature analysis for deep learning classification algorithms of urban Lidar data," *Sensors*, vol. 23, no. 17, 2023, Art. no. 7360. [Online]. Available: https://www.mdpi.com/1424-8220/23/17/7360

[28] O. Ozturk, M. S. Isik, M. Kada, and D. Z. Seker, "Improving road segmentation by combining satellite images and LiDAR data with a feature-wise fusion strategy," *Appl. Sci.*, vol. 13, no. 10, 2023, Art. no. 6161. [Online]. Available: https://www.mdpi.com/2076-3417/13/10/6161

[29] A. Gupta, S. Watson, and H. Yin, "3D point cloud feature explanations using gradient-based methods," in *Proc. Int. Joint Conf. Neural Netw.*, 2020, pp. 1–8.

[30] T. Zheng, C. Chen, J. Yuan, B. Li, and K. Ren, "Pointcloud saliency maps," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1598–1606.

[31] F. Matrone, M. Paolanti, A. Felicetti, M. Martini, and R. Pierdicca, "BubblEX: An explainable deep learning framework for point-cloud classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6571–6587, 2022.

[32] H. Tan, "Fractual projection forest: Fast and explainable point cloud classifier," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2023, pp. 4229–4238.

[33] K. F. West, B. N. Webb, J. R. Lersch, S. Pothier, J. M. Triscari, and A. E. Iverson, "Context-driven automated target detection in 3 D data," *Proc. SPIE*, vol. 5426, 2004, pp. 133–143. [Online]. Available: https://api.semanticscholar.org/CorpusID:129084999
[34] N. Yastikli and Z. Cetin, "Contribution of colour information to the classification of 3 d points from aerial images," *Arabian J. Sci. Eng.*, vol. 48, pp. 1–12, 2023.
[35] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, I. Guyon et al., Eds. Red Hook, NY, USA: Curran Associates, 2017, pp. 4765–4774.
[36] I. U. Ekanayake, D. P. Meddage, and U. Rathnayake, "A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP)," *Case Stud. Construction Mater.*, vol. 16, 2022, Art. no. e01059.
[37] B. Wu, C. Chen, T. M. Kechadi, and L. Sun, "A comparative evaluation of filter-based feature selection methods for hyper-spectral band selection," *Int. J. Remote Sens.*, vol. 34, pp. 7974–7990, 2013.
[38] S. Lei, "A feature selection method based on information gain and genetic algorithm," in *Proc. Int. Conf. Comput. Sci. Electron. Eng.*, 2012, pp. 355–358.
[39] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, pp. 23–69, 2003.
[40] A. A. Khan, A. Jamil, D. Hussain, I. Ali, and A. A. Hameed, "Deep learning-based framework for monitoring of debris-covered glacier from remotely sensed images," *Adv. Space Res.*, vol. 71, no. 7, pp. 2978–2989, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0273117722004410
[41] Y. Li and W. Chen, "A comparative performance assessment of ensemble learning for credit scoring," *Mathematics*, vol. 8, no. 10, 2020, Art. no. 1756. [Online]. Available: https://www.mdpi.com/2227-7390/8/10/1756
[42] D. Zhang and Y. Gong, "The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure," *IEEE Access*, vol. 8, pp. 220990–221003, 2020.
[43] X. Shi, Y. Cheng, and D. Xue, "Classification algorithm of urban point cloud data based on lightgbm," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 631, no. 5, Oct. 2019, Art. no. 052041, doi: 10.1088/1757-899X/631/5/052041.
[44] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.
[45] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
[46] G. Ke et al., "'LightGBM: A highly efficient gradient boosting decision tree," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 3146–3154.
[47] D. Zhang and Y. Gong, "The comparison of lightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure," *IEEE Access*, vol. 8, pp. 220990–221003, 2020.
[48] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794.
[49] K. Budholiya, S. K. Shrivastava, and V. Sharma, "An optimized XGBoost based diagnostic system for effective prediction of heart disease," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, pp. 4514–4523, 2022.
[50] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[51] G. Holmes, A. Donkin, and I. H. Witten, "WEKA: A machine learning workbench," in *Proc. Australian New Zealnd Intell. Inf. Syst. Conf.*, 1994, pp. 357–361.

Muhammed Enes Atik received the Ph.D. degree in geomatics engineering from Istanbul Technical University, Istanbul, Turkey, in 2022.

He is currently working as an Assistant Professor with the Department of Geomatics Engineering, Istanbul Technical University. His research interests include photogrammetry, deep learning, machine learning, point cloud segmentation, LiDAR, UAVs, and remote sensing.

Zaide Duran received the Ph.D. degree in geomatics engineering from Istanbul Technical University, Istanbul, Turkey, in 2003.

Since 2022, she has been a Full Professor with the Department of Geomatics Engineering, Istanbul Technical University. Her expertise is on photogrammetry, laser scanning, artificial intelligence, cultural heritage documentation, and geographic information systems. She has many national and international publications in these fields.

Dursun Zafer Seker received the Ph.D. degree in geomatics from Istanbul Technical University, Istanbul, Turkey, in 1993.

Since 2004, he has been a Full Professor with the Department of Geomatics Engineering, Istanbul Technical University. His expertise is on photogrammetry, remote sensing, coastal zone management, watershed management and spatial data modeling and analysis from both the theoretical and empirical viewpoint. In these fields, he has been involved with several research projects both national and international, where these projects were interdisciplinary. He has authored more than 120 SCI international papers and more than 300 conference proceedings.