A NEW APPROACH FOR THE CLASSIFICATION OF GROUND POINTS FROM AIRBORNE LIDAR DATA IN FORESTED AREAS

NOVI PRISTUP KLASIFIKACIJI TERENSKIH TOČAKA ŠUMSKOGA PODRUČJA TEMELJEM PODATAKA PRIKUPLJENIH ZRAČNIM LASERSKIM SUSTAVOM

Zümrüt KURTULGU[1]*, Atınç PIRTI[2]

SUMMARY

LiDAR systems are being used increasingly widely and effectively in the field of forestry. They play an important role in many applications such as creating detailed maps of forests, obtaining high-precision elevation data, obtaining information about the height, density and distribution of trees, mapping the topography under the forest in detail, landslide and erosion control, road planning and water management. Underforest topography maps can be created with high precision thanks to LiDAR ground points. In our study, we present a new approach by using LiDAR data to create the physical characteristics of forest land. In this approach, firstly, airborne LiDAR beams were divided into datasets according returns. Three datasets were created: second return, last return and first/last return. Secondly, each dataset was positionally placed in the grid structure. Filtering was done according to mean height values of points in the cell. Thirdly, DBSCAN clustering algorithm, one of the machine learning methods, was used. The epsilon value, one of the parameters used in the DBSCAN algorithm, was determined according to the silhouette index, and LiDAR ground points were classified. The classified LiDAR ground points were compared with the existing ground control points. As a result, the combination of the second return and the last return dataset showed successful results with a kappa value of 82.27% and an F1 score value of 0.71. Also, digital terrain models were created and compared. To demonstrate the effectiveness of the proposed approach, data was compared with the CSF algorithm, which is one of the traditional filtering methods. After the accuracy evaluations, we were able to classify more LiDAR ground points with our proposed approach. Thus, we think that LiDAR ground data can create a detailed and accurate topography map, define forest features and contribute to the decision-making process for forestry activities.

KEY WORDS: second/last return, grid-based filter, DBSCAN clustering algorithm

INTRODUCTION
UVOD

Light detection and ranging (LiDAR) technology has been widely used in forestry and natural resource management, particularly forest inventory and management, investigating forest ecology, and road construction (Reutebuch et al. 2005; Singh et al. 2015; Zhao et al.2016; Matinnia et al. 2017;

Lou et al. 2023). LiDAR technology enables three-dimensional (3D) point cloud data of objects and the Earth's surface. Also, it can penetrate underneath a canopy of trees. Hence, LiDAR data allows the generation of accurate and precise digital elevation models (DEMs). Precise and accurate DEM generation is crucial in forestry applications, such as extracting canopy height models (CHM) or DEM-deri-

[1] Zümrüt Kurtulgu, Muğla Sıtkı Koçman University, Land Registry and Cadastre Programme,Muğla, Türkiye
[1] Atınç Pirti, Yildiz Technical University, Department of Surveying Engineering Davutpaşa, İstanbul, Türkiye
*Corresponding author: Zümrüt Kurtulgu email: zumrutkurtulgu@mu.edu.tr

ved topographic features (i.e. slope and aspects) (Zhao et al. 2016; Fu et al. 2022; Zhang et al. 2023). Therefore, point filtering to distinguish ground points from non-ground points and generate DEM is the primary and the most critical step in processing raw LiDAR data because it affects the accuracy of generated DEM. LiDAR uses laser pulses. The laser allows us to collect data about objects thanks to the rays reflected from objects on the Earth and their returning. Luo et al. (2023) created a digital elevation model for forest areas using full waveform LiDAR data and hyperspectral image data. Points are classified into different groups depending on surface laser hits. The main group is the first and last return (Dong and Chen 2017). The filtering process was performed assuming that the last return represented the ground (Pingel et al. 2013). Matinnia et al. (2017) produced an accurate and high-resolution digital elevation model from the first and last laser pulses. Chen et al. (2017) created LiDAR by recording the last return points and dividing points into different elevation layers, and obtained digital terrain model by filtering each layer.

With advancing technology, many filtering algorithms have been developed to classify ground points and create DEMs using LiDAR data. While the majority of these filtering algorithms work quite well on flat terrain and sparse vegetation, it is still difficult to filter them in areas with very rugged terrain, steep slopes, dense vegetation, and rugged terrain (Zhao et al. 2016; Cai et al. 2023; Luo et al. 2023). The most important challenge is determining appropriate parameter values for filtering. Yu et al. (2022) applied a ground filter in their proposed algorithm, thanks to the five parameters required by the user. The accuracy of ground filtering algorithms was evaluated and digital terrain models were compared by using the most appropriate parameters in lands with vegetation density and slope (Klápště et al. 2020). The IPTD algorithm was proposed to filter ground points in the forested areas in both topographically and environmentally complex regions. This algorithm was successful in finding steep slopes and peaks thanks to four parameter data (Zhao et al. 2016). Grid resolution, time step, hardness, slope adaptation factor, iteration and classification threshold value parameters were defined for CSF filtering algorithm (Zhang et al. 2016). In this case, finding the right parameters for each terrain type can be time-consuming.

The creation, accuracy, or quality of the DEM impacts usage scenarios (Kraus and Pfeifer 2001; Chen et al. 2017). If a high-resolution DEM is created, the resulting topographic surface and soil subsidence can be easily identified (Matinnia et al. 2017). By using the geometric or radiometric features of LiDAR, the computer's ability to learn from data was also utilized. An improved DBSCAN clustering algorithm is proposed to detect signal photons from ICESat-2 LiDAR data to determine ground elevations in urban areas. By adjusting minPts and eps values in each cluster, ground

points were able to be obtained (Alzaghoul et al. 2021). Machine learning method was used to create landslide hazard map thanks to geomorphological factors derived from DEM (Chang et al. 2019). Deng et al. (2022) proposed an obstacle detection algorithm. With the developed DBSCAN clustering algorithm, an adaptive neighborhood value that changes with distance was determined for each point cloud by the linear interpolation method. Another method based on clustering algorithms was proposed by Feng et al. (2017). With this method, the accuracy when processing boundary points is increased, which revealed its minimum acceptable distance, thus achieving good performance both in terms of accuracy of processing boundary points and time complexity (Feng et al. 2017).

We proposed a new approach for LiDAR ground point extraction in a forested area with dense vegetation. In this approach, firstly, point cloud is divided into appropriate datasets. Three datasets consisting of the last return, the second return and the combination of the first return and the last return were obtained from the airborne LiDAR point cloud data. A grid-based filter was applied to each dataset. Then, machine learning was used in the study. The silhouette value used to determine the number of clusters in the clustering process was this time automatically used to determine the epsilon parameter, and LiDAR ground points were classified. A digital terrain model was obtained with determined ground points and compared with the reference model. When compared to the previous studies, four issues were taken into consideration with the proposed approach. (1) Instead of using the lowest elevation points in the grid-based filtering process, the average elevation value was taken as the threshold value to deal with slope changes. (2) Fast analysis will be possible thanks to a single parameter value entered by the user. (3) For each dataset, the parameter value can be calculated automatically according to the characteristics of the data. (4) It will increase the accuracy of the resulting digital terrain model as more LiDAR ground points are classified after filtering. Additionally, our proposed approach was compared with the CSF algorithm, which is one of the traditional filtering methods. As a result, a highly accurate forest understory topography map can be created with the proposed approach. However, we think it will help decision-making in forestry practices. Additionally, this approach will be tested on other land types and the results will be shared.

MATERIALS AND METHODS
MATERIJALI I METODE

Study area – *Područje istraživanja*

The area where the study was carried out is the Kirazlı Weir Nature Park within Belgrad Forest in the Eyüpsultan district of Istanbul. The region took its name from the Kirazlı

dam built on the Kirazlı stream in 1818. The area is suitable for outdoor activities such as picnics, walking and cycling. In the forest area of the nature park, there is a deciduous forest area including the species such as sessile oak (*Quercus petraea*), larch (*Pinus nigra*), shrub (*Erica arborea*), hornbeam (*Carpinus betulus*) and complex terrain types. The study area is shown in Figure 1. In addition, measurements were made with total station (TS) to evaluate the accuracy of the model in this region, and 941 ground control points were obtained.

Airborne Laser Systems (ALS) – *Zračni laserski sustavi (ALS)*

Airborne LiDAR point cloud was created in August 2013 with a Riegl Q680i scanner and a platform mounted on a helicopter, at a frequency of 16 points per square meter. Approximately 1352010 point data were obtained in the scanning. The surveys were performed in the study region, and the platform height was 600 m. The study area of the nature park amounts to approximately 20 decares. The area where the dam and its surroundings are measured is shown in Figure 2. While ALS is a reliable method for collecting terrain data, creating a high-quality and efficient digital terrain model can be difficult. Therefore, correct classification of points is important for creating a terrain model. For this, less important elements should be removed by data reduction (Liu 2008). Therefore, in our proposed method, data, which can include ground points, are selected from

multiple rotations through the data reduction process. Additionally, discrete return LiDAR shows better accuracy in height measurements at individual tree and stand levels (Chen 2007).

Grid-based filter algorithm – *Algoritam filtriranja temeljen na mreži*

Grid-based filters, one of the surface-based filters, create a reference surface with discriminant functions, after which the point category is determined according to the distance between each point and the corresponding reference surface. Active shape models are used to determine the reference ground surface (Qin et al. 2023). Whether this parameter, defined as cell size or window size, is large or small is important in terms of filtering or reducing the data. Martinez et al. (2019) wanted to identify bridges and water bodies by using LiDAR features such as radiometric and geometric variables. A 2 m x 2 m DEM grid was created and the Z value was determined with the different components created. If the height difference is less than 1 m, these points are considered bridges. Deng L. et al. (2022) used a predefined height threshold value to filter ground points after obtaining a ground surface elevation model. Thus, the size of the data in the dataset was reduced. Du and Wu (2022) calculated the data density by dividing the data field into a limited number of cells with a standard grid structure.
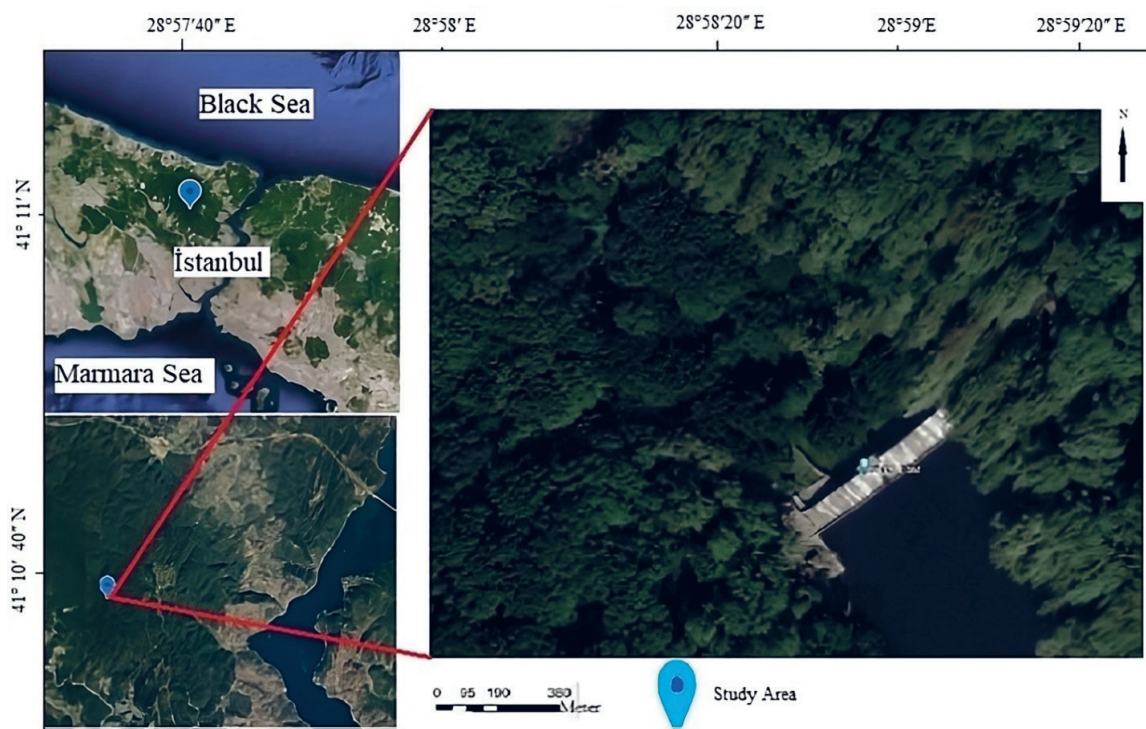

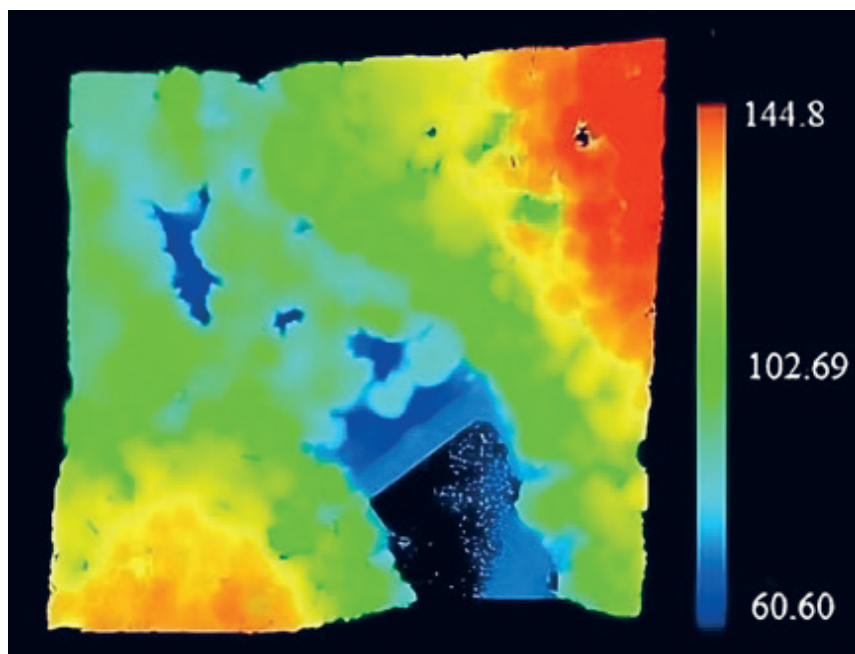**Figure 1.** Study area
Slika 1. Područje istraživanja

Figure 2. LiDAR data of the study area
Slika 2. LiDAR podaci područja istraživanja

DBSCAN clustering algorithm – *DBSCAN algoritam klasteriranja*

This algorithm is one of the density-based clustering algorithms, which is one of the spatial data clustering techniques that produces useful models from complex databases, searching for clusters of any size or shapes and identifying outliers. A cluster may consist of a core dataset that is close to each other and a non-core dataset that is close to this dataset (Ester et al. 1996; Schubert et al. 2017). DBSCAN uses only two parameters that determine whether a region is considered dense enough to belong to a cluster (Gunawan and De Berg 2013). These include: minPts – the minimum number of points clustered together for a region to be considered dense; Eps – the radius of the virtual circle defined separately around each point.

The algorithm randomly checks all points starting from one point in the dataset. If the controlled point has already been included in a cluster, it moves to the next point without taking any action. If the point is not clustered, it finds the neighboring points in the epsilon neighborhood of the point by performing a region query. If the number of neighboring points is more than minPts, it transfers this point and its neighboring points to a new cluster. If there are at least minPts points within the epsilon radius to the point, then it considers all these points to be part of the same set. After that, clusters are expanded by repeating neighborhood calculation for each neighboring point. If it has a distance less than epsilon, the point remains inside the cluster. When points are further than epsilon from each other, the point is transferred to a new cluster (Ester et al.

1996; Schubert et al. 2017). Many researchers have tried to find parameter values. Wu et al. (2021) analyzed minPts value with the numbers 2, 4, …, 20 when using clustering algorithm. It has been confirmed that it can separate data of unequal density into clusters. As a result, it was observed that minPts value had little effect on the clustering results when choosing different values. Zhang et al. (2023) used DBSCAN algorithm to detect trees. Using values eps = 2.15 and minPts = 480, they clustered understory vegetation and low branch leaves well. In addition, when eps = 0.6 and minPts = 3, the lower vegetation was clustered. Thus, although high accuracy analysis was achieved with appropriate parameters, it took time to find these parameters. In our proposed approach, the cluster radius will be determined for the value with the highest eps value and silhouette value within cluster, but minPts value will be determined manually.

METHODS
METODE

We present a new approach that classifies ground points to determine topographic features of the terrain using aerial LiDAR data in a densely vegetated forest area. Three separate datasets were obtained from an ALS point cloud: second return, last return and first/last return. Then, Z score algorithm was applied to these three datasets to remove outliers or noise points. Each dataset was then positionally placed in the grid structure. In order not to ignore height differences in the field, a classification was made with average height values of points in a cell. Points below average

Table 1. The number of return points from ALS datasets
Tablica 1. Broj povratnih točaka iz ALS skupova podataka

	Second return *Drugi povrat*	Last return *Posljednji povrat*	First + last return *Prvi + posljednji povrat*
Points – *Točke*	374469	13809	518396
Min Z	60.6	61.57	61.4
Max Z	129.05	124.54	144.78

height value of the points in a cell were classified as possible ground points; otherwise they were classified as non-ground points. Then, clustering algorithm was used. Clustering serves as a partitioning task, where objects in same cluster are similar to each other, but other clusters are different from each other. Epsilon value, one of the parameters used in the DBSCAN algorithm, which is one of clustering algorithms, was calculated according to the dataset. Here it was calculated by taking into account height differences of points. The silhouette value, which has verification and interpretation features, was used for this calculation. When the silhouette value was high, epsilon value was automatically determined. With this epsilon value obtained, datasets were analyzed with minPts value entered by the user. Thus, they were classified as ground and non-ground points. This proposed approach was performed in Python software. The determined LiDAR ground points were compared with ground control points we obtained. As a result of the analysis, more LiDAR ground points were extracted in a dataset obtained by combining the second return and the last return datasets. Thus, changes in the land surface can be monitored. Additionally, a comparison was made based on the accuracy assessment made with a CSF algorithm, which is one of ground filtering algorithms. In the comparison of LiDAR ground points obtained with the proposed approach and ground control points, mean error, RMSE, kappa and F1 score values, which are vertical accuracy statistics, were calculated. As a result, with the approach we proposed, statistical values obtained by combining the first return, the second return and the last return datasets from LiDAR point cloud showed close accuracies, and similar terrain models were obtained in comparison with CSF algorithm.

As the first step in the proposed approach, discrete points were obtained in ALS point cloud. For our approach, they were classified into three datasets: last return, second return, and first/last return. The obtained point count data are shown in Table 1. After all discrete points were determined, noise and outlier points were determined. The number of noise and outliers was 2213. The algorithm was applied for individual datasets. The flow diagram of the proposed approach is shown in Figure 3.

As the second step, Z-score test was applied to each dataset. In this way potential outliers with non-compliant low va-
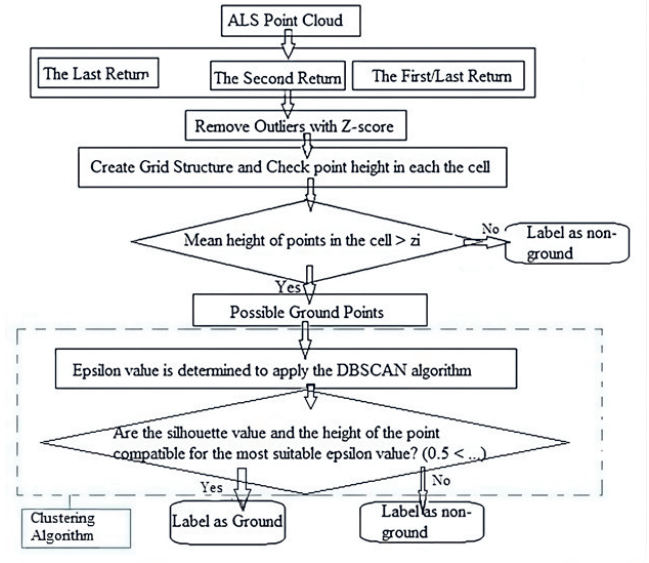
Figure 3. Flow diagram of the proposed approach
Slika 3. Dijagram toka predloženog pristupa

lues in datasets should be removed. Thus, the points that produce both intense noise caused by reflective objects (leaf, water etc.) and less accurate predictions for forest models will be filtered (Gao et al. 2021). The Z-score test is a type of standard deviation that allows us to find the probability of a value occurring in a normal distribution or to compare two samples from different populations (Curtis et al. 2016). Using the Z-score test, we can find outliers in the desired range and remove them from our dataset. For this, the following equation is used:

$$Z\text{-score value} = (X — \mu) / \sigma \qquad (1)$$

Where X value is our dataset from which we want to calculate the Z-score; μ is the dataset mean; and σ is the standard deviation of the dataset (Curtis et al. 2016).

As the third step, a grid-based filter is applied to LiDAR datasets. Each dataset is transferred to a grid plane according to its spatial values, with a cell size of 3 m x 3 m. Choosing a smaller grid size will increase the number of points in the cell. Additionally, experiments were made by gradually increasing and decreasing cell size, and the most ideal size was selected for each dataset. After placing points on a grid structure, height difference relationship of points in each cell was taken into account. The average height values of points in each cell were calculated as the threshold value. If height value of each point within the cell is higher than the average height value of that cell, the point is classified as a non-ground point in the dataset, otherwise it is classified as a possible ground point. However, slope change on the ground is also taken into account when calculating average height. For example, if there are seven dots in a cell, and the height values of these points are 66, 68, 72, 75, 78, 82, 90,

the average height value in this cell is 75.85 when calculated according to Formula 2. Then, points with higher average height values will be removed from this cell and classified as non-ground points. That is, points with height values of 78, 82 and 90 are not ground points. There is a height difference of approximately nine meters at the possible ground points for this cell. In this case, the slope of the ground is also included in the process. If Z is the average height within cell, then the average height value is found according to the procedure in Equation 2.

$$Z = \frac{1}{n}(\sum_{n=1}^{n} zi) \tag{2}$$

Some of points in cell are called boundary points or neighbor points. Since we wanted to obtain ground points in our datasets, if there were points corresponding to the border point or corner point, these points were classified as possible seed points if they were lower than the average height value of points in the cell. Additionally, these points were classified by transferring them to the next neighboring cell. If height of any point of cluster is higher than its neighboring points, any point of cluster should be classified as a non-ground point, otherwise it should be classified as a ground point. In classifying points with clustering and segmentation, the dataset was divided into regular segments and the resulting segments were compared with the neighboring segments in terms of topological and geometric aspects. As long as a ground point is adjacent to some boundary points, the entire ground point can be found by clustering (Deng W. et al. 2022).

As the fourth step, DBSCAN algorithm, one of clustering algorithms, was used. Clusters that would form neighbors to each other in terms of location and geometry were evaluated. To find the epsilon parameter used by the DBSCAN algorithm, a separate value was found for each dataset by taking into account height differences of points. The silhouette index was used to interpret and verify consistency within datasets to be created. Therefore, the highest silhouette value calculated for each dataset obtained was determined as the epsilon value. Silhouette values ranged from -1 to +1, where a high value indicates that object matches well with its own cluster and poorly with the neighboring clusters. A clustering with an average silhouette width above 0.7 is considered "strong", a value above 0.5 is considered "reasonable", and a value above 0.25 is considered "weak", although this increases with increasing the dimensionality of data. Therefore it may become difficult to reach high values (URL 1). The silhouette value may not perform well if datasets have irregular shapes or are of different sizes (Monshizadeh et al. 2022). In this case, it is important to carefully select the dataset used for analytical methods that are affected by point cloud density, such as vegetation (Petras et al. 2023). Silhouette value can be calculated with any distance measurement, such as Euclidean distance or Manhattan distance. However, our approach is based on height differences. Table 2 shows the values and calculated parameters obtained as a result of applying the proposed approach. After applying the DBSCAN algorithm, outliers were found. These points were classified as non-ground points. The second parameter, minPts, was determined in a smaller number, and it was investigated whether each point was a ground point.

Silhouette value is defined for each sample and consists of two values: a: The average distance between a sample and all other points in the same class; b: The average distance between a sample and all other points in the next closest cluster (Rousseeuw 1987).

Silhouette coefficient for a single sample cluster =

$$\frac{b-a}{\max(a,b)} \tag{3}$$

Wu et al. (2021) proposed a method for automatic extraction of rock masses in a 3D point cloud. They used silhouette index to evaluate effectiveness of clustering when applying the K-means algorithm. When a low minPts value was used to cluster datasets with varying densities,

Table 2. Results obtained after the proposed approach
Tablica 2. Rezultati dobiveni nakon primjene predloženog pristupa

Process steps _Koraci obrade_		Second return _Drugi povrat_	Last return _Posljednji povrat_	First + last return _Prvi + posljednji povrat_
	Points – _Točke_	374469	13799	518394
Contour-based filter _Filtar temeljen na obrisima_	Ground – _Terenske_	142625	6674	203931
	Non-ground – _Ostale_	231844	7125	314463
	minPts	5	5	15
DBSCAN	Eps (m)	0.4	0.6	0.1
	Silhouette score	0.6	0.6	0.5
	Outliers points – _Točke ekstrema_	1	8	3
Comparison with TS _Usporedba s TS_	Ground points – _Terenske točke_	445	98	177

HDBSCAN algorithm created smaller clusters in regions with dense data points (Malzer and Baum 2020). Fu et al. (2022) automatically determined trees with the parameter: Eps = 0.72 and minPts = 9. Additionally, if eps value is set too small, the dataset will not cluster in a meaningful way. If eps value is high, it will combine meaningless and meaningful data and place them in the same cluster.

RESULTS AND DISCUSSION
REZULTATI I RASPRAVA

Statistical comparison and evaluation metrics –
Statistička usporedba i metrika procjene

The point cloud obtained in a forested area with dense vegetation was divided into datasets. The number of ground points obtained in the last return dataset after the application of the proposed approach was the lowest and it is shown in Table 2. It can be seen that the last return will not be considered as ground point. In the clustering stage, silhouette values were calculated as 0.6 for the second return dataset, 0.6 for the last return dataset and 0.5 for the first/last return dataset. These values are within a reasonable or acceptable range. For the second return dataset, silhouette value was calculated as 0.6, eps value was calculated as 0.4, and minPts value was calculated as 5. In the last return dataset, while silhouette value was 0.6, eps value was calculated as 0.6 and minPts value was calculated as 5. In the first/last return dataset, while silhouette value was 0.5, eps value was calculated as 0.1 and minPts value was 15. Since point density was scattered during the clustering stage in this dataset, the most appropriate value was determined as a result of trials. El Yabroudi et al. (2022) stated that minPts value may be small for a dense dataset, but if the dataset is sparse, minPts should be larger. As eps value increases, minPts value will also increase as a larger area is

scanned. However, in the first/last return dataset, minPts value (minPts=15) we used for DBSCAN algorithm was chosen higher than in other datasets, and ground points were classified by creating smaller clusters with eps value (eps=0.1). DBSCAN algorithm also has the feature of filtering outliers. After clustering, outliers in datasets were removed from the dataset. This is shown in Table 2. After the three datasets created were filtered with the proposed approach, ground points were determined. The classified LiDAR ground points were compared with 941 ground control points. As a result of the comparison, 445 ground points from the second return dataset, 98 ground points from the last return dataset and 177 ground control points from the first/last return dataset were matched. Statistical values obtained as a result of the proposed approach were calculated. This is shown in Table 3. In addition, as the comparison points from the datasets are different as a result of statistical calculations, statistical results calculated with the combination of datasets obtained are shown in Table 3.

More ground control points were obtained in the second/last return and the first/second/last return datasets. Statistically, average error and RMSE values of the second return and the second/last return datasets are close to each other. Total error, kappa and F1 score values are better than other datasets. As a result of filtering using only the last return dataset, although the average error value and RMSE values seem successful compared to other datasets, total error value, kappa value and F1 score values are unsuccessful. When making measurements in a forest area, it is important not to use the last return dataset alone because the number of laser pulses penetrating vegetation to reach the ground will be inconsistent due to changing density of the canopy level, which will cause irregular ground points and negatively affect the terrain model (Cai et al. 2023). In the first/last return dataset, it is the least successful value according to the average error value and RMSE values. Moreo-

Table 3. Statistical and accuracy values of the data after the proposed approach
Tablica 3. Statističke vrijednosti i točnost podataka nakon primjene predloženog pristupa

	Second return *Drugi povrat*	Last return *Posljednji povrat*	First + last return *Prvi + posljednji povrat*	Last + second return *Posljednji + drugi povrat*	First + second + last return *Prvi + drugi + posljednji povrat*	CSF
min (m)	zatijelo	–0.076	–0.869	–0.890	–0.890	–0.497
max (m)	0.860	0.481	0.748	0.860	0.860	0.493
Mean error (m) *Srednja pogreška*	–0.090	–0.077	–0.189	–0.091	–0.125	–0.019
RMSE (m)	0.303	0.250	–0.349	0.300	0.326	0.213
Ground points *Terenske točke*	445	98	177	567	622	476
Total error *Ukupna pogreška*	20.13%	43.71%	33.67%	17.73%	8.98%	16.88%
Kappa	79.87%	56.29%	66.33%	82.27%	91.02%	83.12%
F-1 Score	0.64	0.19	0.31	0.71	0.49	0.45

Table 4. Filtered result comparison matrix
Tablica 4. Matrica usporedbe filtriranih rezultata

		Filtered results *Filtrirani rezultati*	
		Ground points *Terenske točke*	Non-ground points *Ostale točke*
Reference *Referenca*	Ground points – *Terenske točke*	a	b
	Non-ground points – *Ostale točke*	c	d

ver, although its kappa value showed an acceptable success with 66.33%, it failed with an F1 score value of 0.31. As a result of the statistical result obtained by combining all datasets, ground control point was obtained more times than others. While total error and kappa value were successful, F1 score value was reasonably accurate with 0.49. To evaluate analysis results of data, total error, and kappa and F1 score values were calculated according to Table 4.

$$\text{Toplam Error} = (b + c) / (a + b + c + d) \qquad (3)$$

$$\text{Kappa} = 2 \times (a \times d - b \times c) / (a + c) \times (c \times d) + (a + b) \times (b \times d) \quad (4)$$

$$\text{F1 Score} = 2 \times a / (2 \times a + c + b) \qquad (5)$$

Kappa value indicates the ratio of correct predictions made in system to all predictions. In this case, it determines the overall performance of the model at a reasonable level. The F1 score value does not make an incorrect model selection in unevenly distributed datasets. It also evaluates performance of the model in terms of both accuracy and sensiti-

vity. When evaluating accuracy, it is more logical to do it based on datasets (URL 2).

Finally, ALS point cloud was classified with the CSF ground algorithm using parameters of 0.3 grid resolution, 500 iterations and 0.5 classification threshold. As a result of classification, an average error value of 19 cm and RMSE of 0.213 were calculated. It was successful compared to RMSE and average error values of other datasets. However, while kappa value provided a success rate of 83.12%, F1 score value showed reasonable success with 0.45. The terrain model created from values obtained with total station (TS) and CSF algorithm is shown in Figure 3. A terrain model of these obtained values was created, which is shown in Figure 2. Here, high data volume and method selection were important.

Additionally, there was a ground point that our proposed approach misclassified based on accuracy evaluation. To understand this situation, we filtered each of the datasets classified as non-ground points again in Table 2. The results
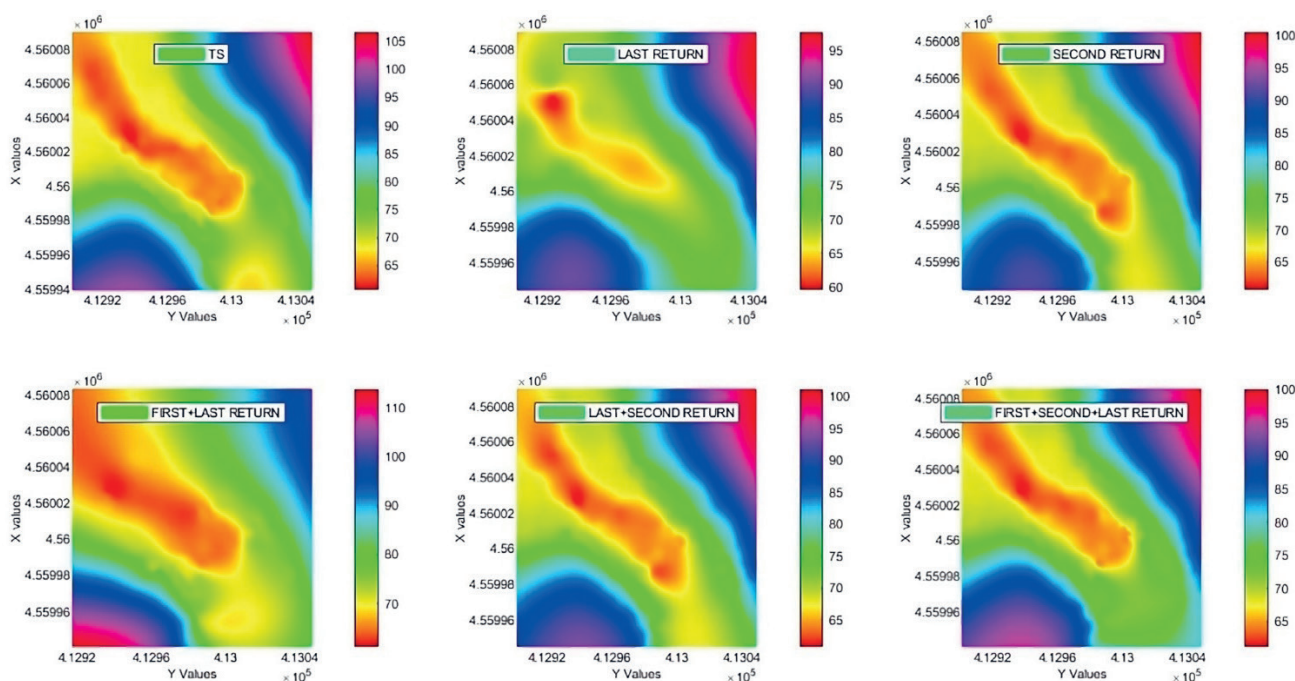
Figure 4. Terrain models created with TS with all datasets
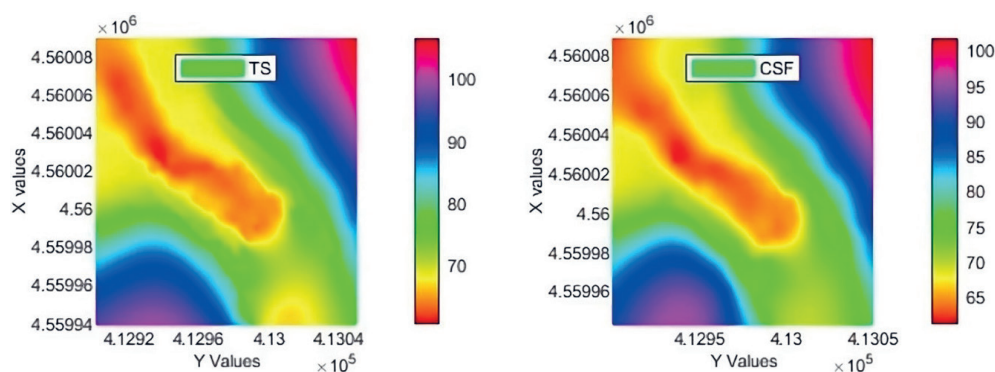Slika 4. Modeli terena stvoreni pomoću TS-a sa svim skupovima podataka

Figure 5. TS and CSF terrain models
Slika 5. TS i CSF modeli terena

obtained are shown in Table 2. The number of points found out of the existing 941 ground control points included: 15 ground control points in the second return dataset, 8 in the last return dataset, and 27 ground control points in the first/last return dataset. Since there were misclassified ground points, more LiDAR ground points could be obtained thanks to the proposed approach by increasing the experimentally determined cell size. In addition, 64 ground points were obtained from non-ground points made with the CSF algorithm. The model made with a total station was chosen as the reference model and was obtained with Matlab software.

Even though they were successful according to RMSE values with the last return dataset compared to the reference land model, both accuracy values were not similar to the model and they were unsuccessful. Differences were obvious when comparing the first/last return dataset with the reference model. The similarity with the reference model was high in the second return dataset and the second/last return dataset. There were differences in 68-75 m altitude range for the first/second/last return dataset. CSF terrain model, on the other hand, was more similar to the terrain model obtained from the first/second/last return dataset.

CONCLUSIONS
ZAKLJUČCI

It was aimed to extract ground points using aerial LiDAR data in a deciduous forest area with a complex terrain type. Both surface-based filter and machine learning methods were used. Position and height values of the points were used with grid-based filtering. Machine learning, on the other hand, helped us to either classify data according to developed models or make predictions for future outcomes based on the models. Although learning-based methods perform well in identifying key points, they can cause incomprehensible misclassifications in some cases (Qin et al. 2023). Correct datasets must be selected to prevent miscla-

ssification and loss of time. The method we applied for the last return, the second return and the first/last return datasets was the result obtained by combining the second/last turn analysis of ideal data in a forest area with a complex structure for place points. Moreover, this approach can be developed further. Thanks to the proposed approach, extracting more LiDAR ground points in densely vegetated forest areas can help the following activities: 1) it can help assess fire risks in forest areas through analysis of factors such as tree density, height and slope; 2) it can create water flow patterns and assess erosion risks; 3) it can help with road and infrastructure planning in forest areas; 4) it can create a forest inventory thanks to correctly classified above-ground points; 5) if the study area is a deciduous forest, it can be important in mapping different habitat types and biodiversity by obtaining information in terms of both tree species, plant species and animal species.

REFERENCES
LITERATURA

- Alzaghoul, E., Al-Zoubi, M. B., Obiedat, R.,Alzaghoul, F. 2021: Applying machine learning to DEM raster images. Technologies, 9(4). doi.org/https://doi.org/10.3390/technologies9040087

- Cai, S., Liang, X.,Yu, S. 2023: A Progressive Plane Detection Filtering Method for Airborne LiDAR Data in Forested Landscapes. Forests, 14(3). doi.org/https://doi.org/10.3390/f14030498

- Chang, K. T., Merghadi, A., Yunus, A. P., Pham , B. T., Dou, J.2019: Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques.Scien-

tific reports, 9(1). https://doi.org/https://doi.org/10.1038/s41598-019-48773-2

- Chen, Q. 2007: Airborne lidar data processing and information extraction. Photogrammetric engineering and remote sensing, 73(2).

- Chen, Z., Gao, B., Devereux, B. 2017: State-of-the-art: DTM generation using airborne LIDAR data. Sensors, 17(1), 150. https://doi.org/https://doi.org/10.3390/s17010150.

- Curtis, A. E., Smith , T. A., Bulat, A. Z.,Elefteriades, J. A. 2016: The Mystery of the Z-Score. Aorta, 4(04), 124-130. https://doi.org/DOI: 10.12945/j.aorta.2016.16.014.

- Deng, L., Guo, T., Wang, H., Chi, Z., Wu, Z.,Yuan, R. 2022:Obstacle Detection of Unmanned Surface Vehicle Based on Lidar Point Cloud Data. In OCEANS 2022, Hampton Roads. Hampton Roads, VA, USA: IEEE. https://doi.org/https://doi.org/10.1109/OCEANS47191.2022.9977340.

- Deng, W., Tang, G.,Wang, Q. 2022: A novel fast classification filtering algorithm for LiDAR point clouds based on small grid density clustering. Geodesy and Geodynamics, 13(1), 38-49. https://doi.org/https://doi.org/10.1016/j.geog.2021.10.002

- Dong, P., Chen, Q. 2017: Lidar Remote Sensing and Applications. CRC Press.

- Du , M.,Wu, F. 2022:Grid-Based Clustering Using Boundary Detection. Entropy. https://doi.org/https://doi.org/10.3390/e24111606

- El Yabroudi, M., Awedat, K., Chabaan, R. C., Abudayyeh, O., Abdel-Qader, I. 2022: Adaptive DBSCAN LiDAR Point Cloud Clustering For Autonomous Driving Applications. In 2022 IEEE International Conference on Electro Information Technolog. https://doi.org/https://doi.org/10.1109/eIT53891.2022.9814025

- Ester, M., Kriegel, H. P., Sander, J., Xu, X. 1996: A density-based algorithm for discovering clusters in large spatial databases with noise. In kdd, 96(34), 226-231.

- Feng, L., Liu, K., Tang,, F., Meng, Q. 2017: GO-DBSCAN: Improvements of DBSCAN Algorithm. International Journal of Computer Theory and Engineering, 9(3). https://doi.org/DOI: 10.7763/IJCTE.2017.V9.1129.

- Fu, H., Li, H., Dong, Y., Xu, F., Chen, F.2022:Segmenting individual tree from TLS point clouds using improved DBSCAN. Forests, 13(4). https://doi.org/https://doi.org/10.3390/f13040566.

- Gao, R., Park, J., Hu, X., Yang, S.,Cho, K.2021:Reflective Noise Filtering of Large-Scale Point Cloud Using Multi-Position LiDAR Sensing Data. Remote Sensing, 13. https://doi.org/https://doi.org/10.3390/rs13163058.

- Gunawan, A.,De Berg, M. 2013:A faster algorithm for DBSCAN.

- Klápště, P., Fogl, M., Barták, V., Gdulová, K., Urban, R., Moudrý, V.2020:Sensitivity analysis of parameters and contrasting performance of ground filtering algorithms with UAV photogrammetry-based and LiDAR point clouds. International Journal of Digital Earth, 13(12), 1672-1694. https://doi.org/https://doi.org/10.1080/17538947.2020.1791267

- Kraus, K., Pfeifer, N.2001:Advanced DTM Generation From Lidar Data. International Archives of Photogrammetry and Remote Sensing And Spatial Information Sciences, 34(3/W4), 23-30.

- Liu, X. 2008:Airborne LiDAR for DEM generation: some critical issues. Progress in physical geography, 32(1), 31-49. https://doi.org/https://doi.org/10.1177/0309133308089496.

- Luo, W., Ma, H., Yuan, J., Zhang, L., Ma, H., Cai, Z.,Zhou, W.2023:Luo, W., Ma, H., Yuan, J., Zhang, L., Ma, H., Cai, Z., Zhou, W. 2023:High-Accuracy Filtering of Forest Scenes Based on Full-Waveform LiDAR Data and Hyperspectral Images. Remote Sensing, 15(14). https://doi.org/https://doi.org/10.3390/rs15143499.

- Malzer, C.,Baum, M.2020:A hybrid approach to hierarchical density-based cluster selection. In 2020 IEEE international conference on multisensor fusion and integration for intelligent systems. Germany. https://doi.org/https://doi.org/10.1109/MFI49285.2020.9235263.

- Martinez, L. S., Jabato, M. J., Sáenz de Tejada, G. J., Rodríguez-Cuenca, B.2019:Automatic classification of bridges and continental water bodies from 3D point clouds (aerial lidar). The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, 1047-1051.

- Matinnia, B., Parsakhoo, A., Mohammadi, J., Jouibari, S. S. 2017:Monitoring geometric properties of an existing forest road using airborne Lidar data. Journal of Forest Science, 63(11), 490-495. https://doi.org/https://doi.org/10.17221/95/2017-JFS

- Monshizadeh, M., Khatri, V., Kantola, R.,Yan , Z. 2022:A deep density based and self-determining clustering approach to label unknown traffic. Journal of Network and Computer Applications, 20. https://doi.org/https://doi.org/10.1016/j.jnca.2022.103513

- Petras, V., Petrasova, A., McCarter, J. B., Mitasova, H.,Meentemeyer, R. K. 2023:Point Density Variations in Airborne Lidar Point Clouds. Sensors, 23(3). https://doi.org/https://doi.org/10.3390/s23031593

- Pingel, T., Clarke , K. C.,McBride, W. A. 2013:An improved simple morphological filter for the terrain classification of airborne LIDAR data. ISPRS journal of photogrammetry and remote sensing, 77, 21-30. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2012.12.002

- Qin, N., Tan, W., Guan, H., Wang, L., Ma, L., Tao, P., . . . Li, J.2023:Towards intelligent ground filtering of large-scale topographic point clouds: A comprehensive survey. International Journal of Applied Earth Observation and Geoinformation, 125. https://doi.org/https://doi.org/10.1016/j.jag.2023.103566.

- Rousseeuw, P. J.1987:Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65. https://doi.org/https://doi.org/10.1016/0377-0427(87)90125-7.

- Schubert, E., Sander, J., Ester, M.,Kriegel, H. 2017:DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. ACM Transactions on Database Systems, 42(3), 1-21. https://doi.org/https://doi.org/10.1145/3068335.

- Singh, K., Chen, G., McCarter, J.,Meentemeyer, R. 2015: Effects of LiDAR point density and landscape context on estimates of urban forest biomass. ISPRS Journal of Photogrammetry and Remote Sensing, 101, 310-322. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2014.12.021.

- Wang, C., Ji, M., Wang, J., Wen, W., Li , T.,Sun, Y. 2019:An Improved DBSCAN Method for LiDAR Data Segmentation with Automatic Eps Estimation. Sensors, 19. https://doi.org/http://dx.doi.org/10.3390/s19010172.

- Wu, X., Wang, F., Wang, M., Zhang, X., Wang, Q., Zhang, S.2021:A new method for automatic extraction and analysis of discontinuities based on TIN on rock mass surfaces. Remote Sensing, 13(15). https://doi.org/https://doi.org/10.3390/rs13152894.

- Yu, D., He, L., Ye, F., Jiang, L., Zhang, C., Fang, Z., . . . Liang, Z. 2022:Unsupervised ground filtering of airborne based 3D meshes using a robust cloth simulation. International Journal of Applied Earth Observation and Geoinformation, 111. https://doi.org/https://doi.org/10.1016/j.jag.2022.102830.

- Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X.,Yan, G. 2016:An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation. Remote Sensing, 8(6). https://doi.org/https://doi.org/10.3390/rs8060501.

- Zhang, Y., Tan, Y., Onda , Y., Hashimoto, A., Gomi, T., Chiu, C.,Inokoshi, S.2023:A tree detection method based on trunk point cloud section in dense plantation forest using drone LiDAR data. Forest Ecosystems, 10. https://doi.org/https://doi.org/10.1016/j.fecs.2023.100088.
- Zhao, X., Guo, Q., Su, Y., Xue, B.2016:Improved progressive TIN densification filtering algorithm for airborne LiDAR data in forested areas. ISPRS Journal of Photogrammetry and Remote Sensing, 117, 79-91. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2016.03.016.
- URL 1: https://scikit-learn.org/stable/modules/clustering.html, 2023.
- URL 2: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html ,2023.

SAŽETAK

LiDAR sustavi sve se učestalije i učinkovitije koriste u svim aspektima šumarstva. Imaju važnu ulogu u mnogim primjenama kao što je primjerice izrada detaljnih šumskih karata, dobivanje visoko preciznih podataka o visini, gustoći i rasporedu stabala, a koriste se i za detaljno mapiranje topografije šumskoga tla, kontrolu klizišta i erozijskih procesa, pri planiranju šumskih cesta te upravljanju vodenim površinama. Topografske karte šuma i šumskih zemljišta mogu se izraditi s velikom preciznošću zahvaljujući LiDAR tehnologiji. U našem istraživanju predstavljamo novi pristup korištenja LiDAR podataka pri kreiranju fizičkih karakteristika šumskog zemljišta. U ovom pristupu, LiDAR podaci prikupljeni iz zraka podijeljeni su u skupove podataka prema povratima. Kreirana su tri skupa podataka: drugi povrat, posljednji povrat i prvi/posljednji povrat te je svaki skup podataka pozicioniran u mrežnu strukturu, a daljnje filtriranje izvršeno je prema srednjim vrijednostima visine točaka u ćeliji. Tijekom analize i obrade podataka korišten je DBSCAN algoritam klasteriranja, što predstavlja jednu od metoda strojnog učenja. Epsilon vrijednost, jedan od parametara korištenih u DBSCAN algoritmu, određena je prema indeksu obrisa terena, a LiDAR terenske točke zatim su klasificirane. Klasificirane LiDAR terenske točke uspoređene su s postojećim terenskim kontrolnim točkama. Kao rezultat dobivena je kombinacija skupa podataka drugog povrata i posljednjeg povrata koja pokazuje uspješne rezultate s kappa vrijednošću od 82,27 % i vrijednošću F1 rezultata od 0,71. Isto tako stvoreni su i uspoređeni digitalni modeli terena. Kako bi se pokazala učinkovitost predloženog pristupa, podaci su uspoređeni s CSF algoritmom, koji predstavlja jednu od tradicionalnih metoda filtriranja. Nakon procjene točnosti, bilo je moguće klasificirati više LiDAR terenskih točaka pomoću predloženog pristupa. Stoga smatramo da LiDAR terenski podaci mogu stvoriti detaljnu i točnu topografsku kartu, definirati značajke šume i šumskoga zemljišta te doprinijeti procesu donošenja odluka za različite aspekte šumarskih djelatnosti.

Ključne riječi: drugi/posljednji povrat, filtriranje temeljeno na mreži, DBSCAN algoritam klasteriranja