

ESTIMATION OF POST-TORNADO LOSSES IN THE TRZEBCINY FOREST DISTRICT BASED ON TREE CANOPY SEGMENTATION AND CLASSIFICATION METHODS

Daria Dobrogowska, Wiktoria Doroz, Julia Domachowska, Paulina Zachar 

Department of Photogrammetry, Remote Sensing and Spatial Information Systems,
Faculty of Geodesy and Cartography, Warsaw University of Technology, Poland

KEYWORDS: orthophotomap, segmentation, classification, SVM, Random Forest, tree canopy, loss estimation

ABSTRACT: Tree canopy extraction is a crucial process for damage evaluation and changes caused by natural disasters. This study uses unsupervised machine-learning techniques to assess post-tornado tree canopy loss in the Trzebciny Forest District. Orthophotomaps from 2010 and 2013 were analyzed with Mean Shift Segmentation in ArcGIS Pro, followed by classification using Support Vector Machine (SVM) and Random Forest (RF) algorithms. Accuracy assessment revealed that SVM outperformed RF, with a higher Kappa coefficient and classification precision. Results indicate a significant reduction in forested areas due to tornado damage. The study demonstrates the effectiveness of segmentation and classification methods in rapid post-disaster forest assessment.

1. INTRODUCTION

Wind storms, floods, earthquakes, fires, or tornadoes often disturb forest ecosystems. Then the assessment of post-tornado damage is crucial for effective forest management, aiding in reforestation planning, ecological restoration, and economic evaluations of timber loss. Remote sensing and image processing techniques, particularly those leveraging aerial imagery, offer a powerful tool for assessing forest damage with high spatial and temporal resolution ([Honkavaara et al., 2013](#); [Nasimi and Wood, 2024](#)).

Among the various approaches, tree canopy segmentation and classification methods can play a crucial role in analyzing post-disturbance landscapes. By accurately delineating tree canopy cover before and after a natural disaster, the affected areas can be identified, and the loss can be quantified.

The tree canopy segmentation and classification process has been widely used in forest studies for biomass estimation or change detection ([Tariq et al., 2023](#); [Mosin et al., 2019](#); [Mehmood et al., 2024](#)). Various segmentation approaches, such as pixel-based and object-based methods, have been explored for extracting meaningful information from high-resolution imagery ([Omer et al., 2015](#)). Numerous studies have demonstrated the

effectiveness of segmentation-based classification for forest canopy mapping and post-disturbance assessment.

For instance, the paper ([Ciesielski et al., 2016](#)) estimated changes in post-hurricane forest cover in Poland based on satellite imagery and NDVI indicators. Another research ([Carani and Pingel, 2023](#)) has shown that machine learning approaches, such as convolutional neural networks (CNNs), enhance classification accuracy, offering a more reliable method for detecting tornado damage. Among different machine learning algorithms, Random Forest (RF) and Support Vector Machines (SVM) have drawn attention to image classification ([Sheykhmousa et al., 2020](#), [Heikkinen et al., 2011](#); [Habib et al., 2009](#); [Giacco et al., 2010](#); [Adam et al., 2014](#)). SVM classification has proven effective in identifying windthrow areas, demonstrating its potential for improving forest management after natural disturbances, as shown in the thesis ([Furukawa, 2022](#)).

This study aims to assess post-tornado tree canopy loss in the Trzebczyn Forest District using orthophotomaps, Mean Shift Segmentation, and two classification approaches: Support Vector Machine (SVM) and Random Trees (RT). Integration of those techniques provides a reliable methodology for rapid post-tornado damage assessment in forested landscapes.

2. METHODOLOGY

Tree canopy extraction is an essential process in environmental analysis, aiding in forest management, ecological monitoring, and urban planning. In this study, the segmentation-based classification in ArcGIS Pro was utilised to extract tree canopy areas from orthophotomaps¹. The methodology involves segmenting the raster image, training a supervised classifier, and classifying the image using the Support Vector Machine (SVM) algorithm and Random Forest (Figure 1).

Segmentation was performed using the Segment Mean Shift raster function, which groups similar pixels based on spectral characteristics. A supervised classification approach was employed, requiring the creation of training samples for both tree canopy and non-canopy areas. The Support Vector Machine (SVM) classifier has been chosen due to its ability to handle complex classification tasks with limited training samples. The Random Forest classifier was applied to compare classification accuracy with SVM. The post-processing and refinement steps involved majority filtering to reduce noise and smooth classified regions. The evaluation of classification results was done using reference data. Analysis and interpretation as a last step consist of tree canopy area calculations and change detection for comparison results from different years to assess forest loss estimation.

¹ <https://medium.com/@suryatoy37/tree-canopy-extraction-using-segmentation-process-in-arcgis-pro-81412e32fd4f>

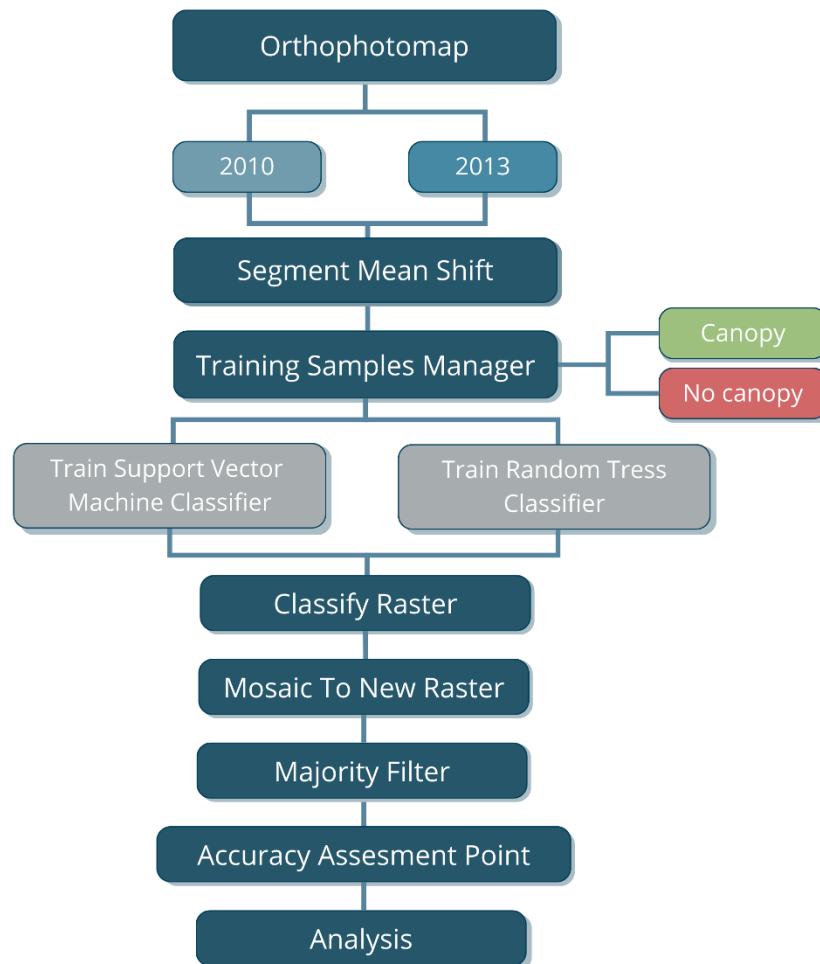


Fig. 1. Flow chart describes the step-by-step methodology applied in this work.

2.1. Data used in the project

The present study utilised orthophotomaps obtained from the geoportal.gov.pl platform, with a ground sample distance (GSD) of 0.5m across two distinct temporal intervals. The first set, consisting of 14 tiles, was taken in June 2010, before the tornado occurred in the study area. This facilitated the documentation of the natural state of forest cover in the region under investigation. The second set of orthophotomap, taken in May 2013, depicts the same area one year subsequent to the tornado event, thereby enabling an assessment of the disaster's impacts and subsequent alterations in land cover. The entire project was done using ArcGIS Pro software.

2.2. Area of interest

The purpose of this research is to evaluate the extent of forest damage in the Kujawsko-Pomorskie voivodeship within Bory Tucholskie Forests in the Trzebciny Forest Area, following the tornado event that occurred on July 14, 2012. The analysed area (Figure 2) is located in the region of Zdroje, Łązek-Szarłata and Stara Rzeką, northwest of Tleń village, within the Świecie and Tuchola districts.



Fig. 2 The area of interest analysed in this work.

2.3. Method used for canopy segmentation

In this project, we employed the built-in "Segment Mean Shift" function in ArcGIS Pro to group adjacent pixels with similar spectral characteristics into segments. The function settings permit the adjustment of several parameters: Spectral Detail (indicates the level of importance given to the spectral differences of features in the imagery), Spatial Detail (indicates the level of importance given to the proximity between features in the imagery) and Minimum Segment Size in Pixels (representing the smallest allowable segment size). Each parameter can be assigned a value from 1 to 20. A higher value is appropriate when there are features to classify separately that have similar spectral characteristics.

Initially, we configured all parameters to a value of 15. This resulted in the segmentation of the raster into excessively large fragments, which led to a significant blending of objects in the image and could cause unreliable classification results.

Consequently, we adjusted all parameters to the maximum value of 20. This modification facilitated the segmentation of the image into significantly smaller chunks, thereby enhancing the accuracy of the classification process (Figure 3).

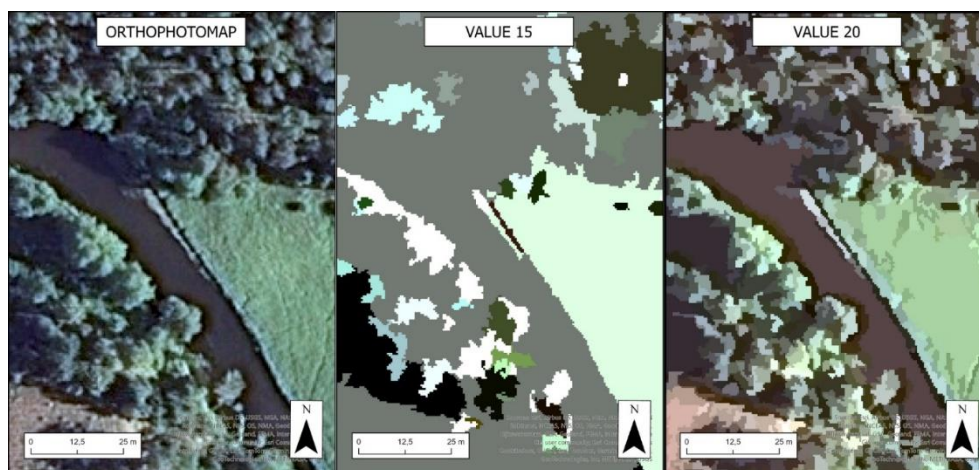


Fig. 3 The result of the orthophotomap segmentation with mean shift method with different parameters.

2.4. Gathering training samples

As part of the supervised image classification, a Training Sample Manager was used to create the training dataset and then train the classifier, allowing pixels to be assigned to specific classes based on the provided samples. A classification schema was developed with two categories: "canopy" and "non-canopy." This schema was designed to distinguish between pixels representing the trees and those that do not, enabling precise categorisation of the image data. The following Table 1 is a summary of the training dataset.

Table 1 Table with percentage coverage of samples used for classification.

Sheet	Surface [ha]	No canopy		Canopy	
		[ha]	%	[ha]	%
2013_d_c_4_4	487.06	0.0075	0.75	0.0027	0.27
2010_d_d_4_4	487.06	0.0275	2.75	0.0032	0.32
2013_a_b_3_2	49031.01	0.0016	0.16	0.0001	0.01
2010_a_b_3_2	49031.01	0.0011	0.11	0.0000	0.00

2.5. Classification with Support Vector Machine (SVM)

The Support Vector Machine (SVM) model is one of the supervised learning algorithms used for classification tasks, and it was used in this paper to train the samples of tree crowns. It works by identifying an optimal hyperplane that maximally separates different classes in a dataset. The algorithm relies on support vectors – critical data points that influence the hyperplane's position and orientation-allowing for robust and efficient classification (Figure 4). SVM excels in high-dimensional spaces because its effectiveness does not degrade as the number of features increases, which is often a challenge for other machine-learning algorithms. The algorithm is also memory efficient, which means that it only uses a subset of the training data, specifically the support vectors, to determine the decision boundary. SVM is robust to data outliers and is also little affected by data distribution and density.

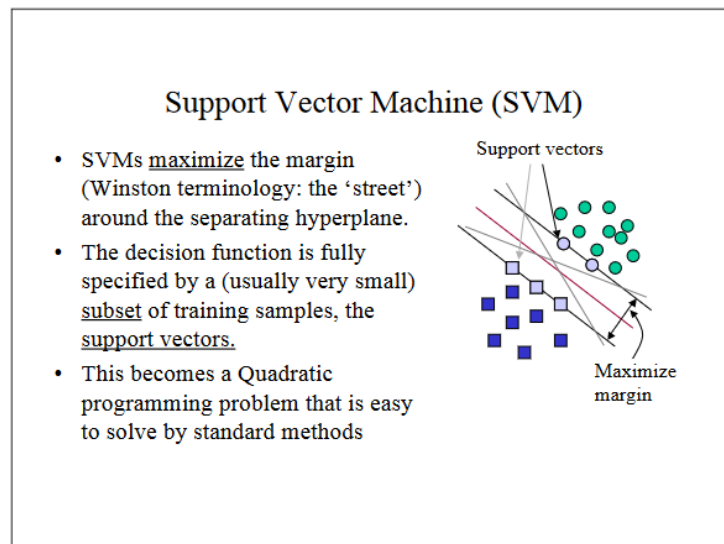


Fig. 4 Overview of the SVM algorithm (source: <https://web.mit.edu/6.034/wwwbob/svm.pdf>)

The image below (Figure 5) shows the results of the classification performed using the SVM algorithm. The 2013 visualisation clearly shows the deforestation caused by the tornado. The green colour shows the forested areas, while the beige colour shows the remaining land cover.

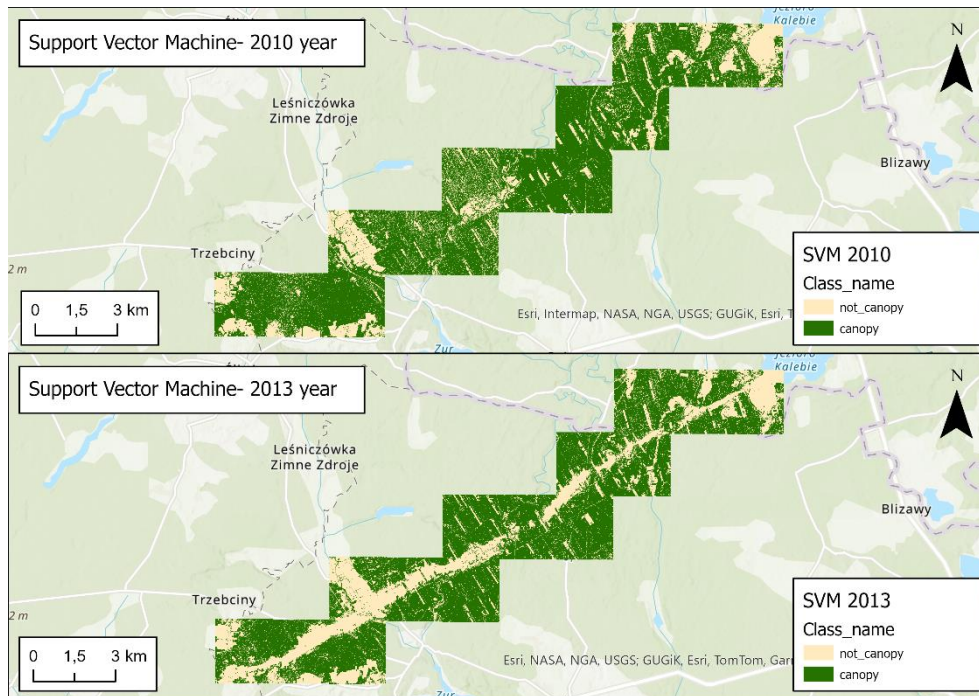


Fig. 5 The result of the SVM classification for two terms.

2.6. Classification with Random Forest

The Random Trees classifier was employed to compare machine learning algorithms within the project. Random Forest (Figure 6) is an algorithm that uses a collection of multiple decision trees to improve prediction accuracy and reduce the risk of overfitting. Each tree in the random forest is trained on a random subset of the data and features, increasing the diversity of the models and improving their ability to generalise the results. Thanks to the bootstrap aggregating technique, the algorithm combines the predictions of individual trees, resulting in more stable and precise results.

Random Forest is based on two core principles: Bagging (Bootstrap) and Random Subspace selection.

- The first one involves creating multiple subsets of training data through random sampling with replacement, with each subset used to train an individual decision tree. The predictions from all trees are combined, resulting in a model that is more accurate and less prone to overfitting.
- The second one randomly selects a subset of features for each decision tree, allowing each tree to focus on different attributes and uncover diverse patterns in the data. This approach enhances prediction accuracy and stability, particularly for high-dimensional datasets or those with many irrelevant features.

In ArcGIS Pro, the following parameters were configured:

- Max_number_of_trees: 50 – A higher number of trees increases the model's resistance to overfitting,
- Max_tree_depth: 30 - depth is another way of saying the number of rules each tree is allowed to create to come to a decision,
- Max_number_of_samples: 1000 – the maximum number of samples that will be used to define each class.

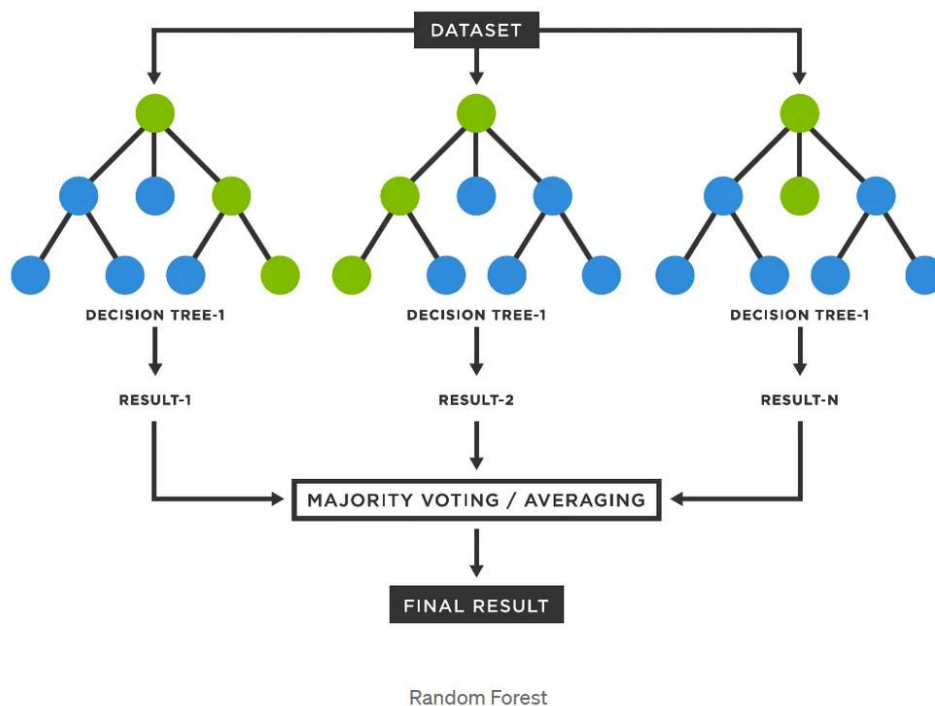


Fig. 6 Overview of the Random Forest method

(source: <https://www.spotfire.com/glossary/what-is-a-random-forest>)

In the figure 7 below, the outcomes of the classification conducted with the random trees algorithm are presented. The 2013 visualisation distinctly highlights the deforestation resulting from the tornado event. Forested regions are depicted in green, whereas the remaining land cover is represented in beige. Notably, the classification from 2013 erroneously identified Lake Kałębie as a forested area.

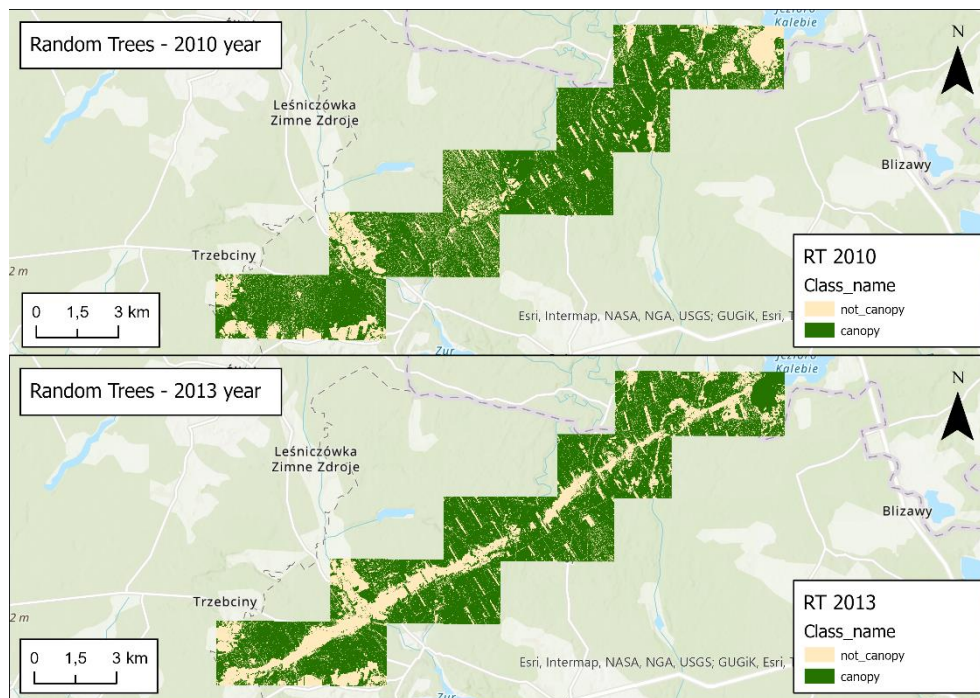


Fig. 7 The result of the Random Forest classification for two terms.

3. RESULTS

In addition to the visual assessment (qualitative analysis), a quantitative analysis was performed. The first step of the analysis was to merge the 14 rasters obtained in the classification process into a single aggregate raster using the “Mosaic to New Raster” function so that an accuracy assessment could be carried out. Due to the presence of individual misclassified pixels, a majority filter was applied to reduce local mismatches. The “Number of Neighbors” parameter was set to 8 due to the relatively high noise level in the data. The use of filtering helped increase the image's readability and make it more suitable for further analysis (Figure 8).

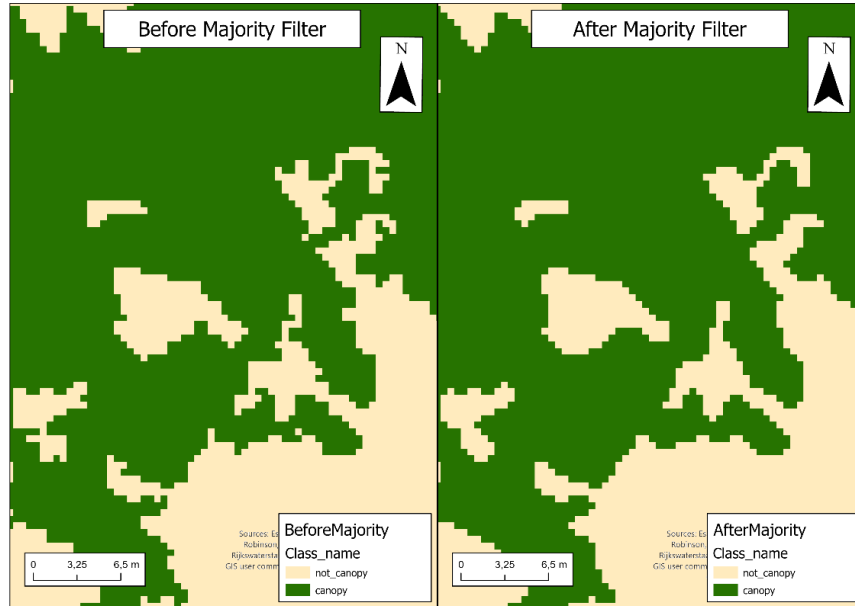


Fig. 8 Result of applying the majority filter.

Five hundred randomly distributed points were generated to assess accuracy, for which a reference value based on field data was assigned (tree canopy = 1, other objects = 0). The obtained points served as the basis for calculating a classification error matrix to assess the quality of the results.

An analysis of classification results for 2010 and 2013 shows that the Support Vector Machine (SVM) algorithm consistently performs better than Random Trees (RT). The Kappa coefficient, which measures the consistency of classification with reality, for SVM was 74% in 2010 and 85% in 2013, while the values for RT were lower: 68% in 2010 and 69% in 2013. This means the SVM classification was more accurate and stable in both cases.

Table 2 SVM classification error matrix for 2010.

2010 - SVM After Majority		no canopy	canopy	Total	User's accuracy [%]	The error of excess [%]
		px				
no canopy	px	188	55	243	77	23
canopy		9	248	257	96	4
Total		197	303	500		
Producer's accuracy [%]		95	82			
The error of omission [%]		5	18			
Omega [%]		87				
Kappa [%]		74				

Table 3 Random Trees classification error matrix for 2010.

2010 – RT After Majority		no canopy	canopy	Total	User’s accuracy [%]	The error of excess [%]
		px				
no canopy	px	158	38	196	81	19
canopy		39	265	304	87	13
Total		197	303	500		
Producer’s accuracy [%]		80	87			
The error of omission [%]		20	13			
Omega [%]		85				
Kappa [%]		68				

Table 4 SVM classification error matrix for 2013.

2013 – SVM After Majority		no canopy	canopy	Total	User's accuracy [%]	The error of excess [%]
		px				
no canopy	px	208	17	225	92	8
canopy		20	255	275	93	7
Total		228	272	500		
Producer's accuracy [%]		91	94			
The error of omission [%]		9	6			
Omega [%]		93				
Kappa [%]		85				

Table 5 Random Trees classification error matrix for 2013.

2013 - RT After Majority		no canopy	canopy	Total	User's accuracy [%]	The error of excess [%]
		px				
no canopy	px	178	27	205	87	13
canopy		50	245	295	83	17
Total		228	272	500		
Producer's accuracy [%]		78	90			
The error of omission [%]		22	10			
Omega [%]		85				
Kappa [%]		68				

For comparison, an error matrix was also generated for the unfiltered data for 2013:

- for the Support Vector Machine, the accuracy was 84%,
- for Random Trees, the accuracy was 68%.

The lower accuracies for unfiltered data confirm the validity of using filtering in the project, which slightly improved the quality of the classification results.

Table 6 SVM classification error matrix for 2013 before applying the majority filter.

2013 - SVM Before Majority		no canopy	canopy	Total	User's accuracy [%]	The error of excess [%]
		px				
no canopy	px	208	19	227	92	8
canopy		20	253	273	93	7
Total		228	272	500		
Producer's accuracy [%]		91	93			
The error of omission [%]		9	7			
Omega [%]		92				
Kappa [%]		84				

Table 7 Random Trees classification error matrix for 2013 before applying the majority filter.

2013 - RT Before Majority		no canopy	canopy	Total	User's accuracy [%]	The error of excess [%]
		px				
no canopy	px	178	29	207	86	14
canopy		50	243	293	83	17
Total		228	272	500		
Producer's accuracy [%]		78	89			
The error of omission [%]		22	11			
Omega [%]		84				
Kappa [%]		68				

Estimating the amount of damage to the tree stand:

Based on the results, it can be concluded that there was a loss of forest area between 2010 and 2013, regardless of the classification algorithm used.

- In the case of Support Vector Machine, the forest area decreased by 432.21 hectares.
- In the case of Random Trees, the forest area decreased by 300.37 hectares.

The differences in the results between the algorithms may be due to the different ways of classifying the pixels. SVM can be better than Random Forest in classifying tree crowns in an orthophotomap due to its ability to handle high-dimensional data and create clear class boundaries, especially when there is a spectral similarity between tree species and surrounding vegetation. SVM effectively deals with small sample sizes and avoids overfitting by focusing on the most relevant support vectors rather than considering an ensemble of models. Additionally, SVM can capture complex, non-linear patterns in images, which is particularly useful for distinguishing tree crowns from other land cover types. Still, both approaches show a noticeable decrease in forest area over the analysed period (Table 8).

Table 8 The summary of forest areas determined from the classification carried out, including calculated differences in forest areas between 2013 and 2010.

	2010 [ha]	2013 [ha]	Difference [ha]
SVM	4946.30	4514.09	432.21
RT	5000.88	4700.50	300.37

Due to the significantly higher classification accuracy obtained with the SVM algorithm, only the results obtained with this method were used for further analyses.

Below in figure 9 there are approximations of characteristic areas where changes detected by the algorithm can be observed. The images are presented in the following order: orthophotomap from 2010, orthophotomap from 2013, obtained differential raster. The green colour represents forest growth in the area, and the red colour is deforestation.

The results of the differential raster for Random Trees (Figure 10) and SVM (Figure 11) are presented below. This raster was generated by subtracting the results of the 2010 classification from the classification performed for 2013.

4. CONCLUSIONS

Analysis of the classification results for 2010 and 2013 made it possible to evaluate the effectiveness of two algorithms: Support Vector Machine (SVM) and Random Trees (RT). The study showed that SVM consistently performed better, with a higher Kappa coefficient and classification accuracy in both years analysed.

One of the significant factors influencing the results was the misclassification of a section of Lake Kałęż as a forest in 2013. This resulted in an overestimation of the area of forest land and affected the differences in forest loss calculations between 2010 and 2013.

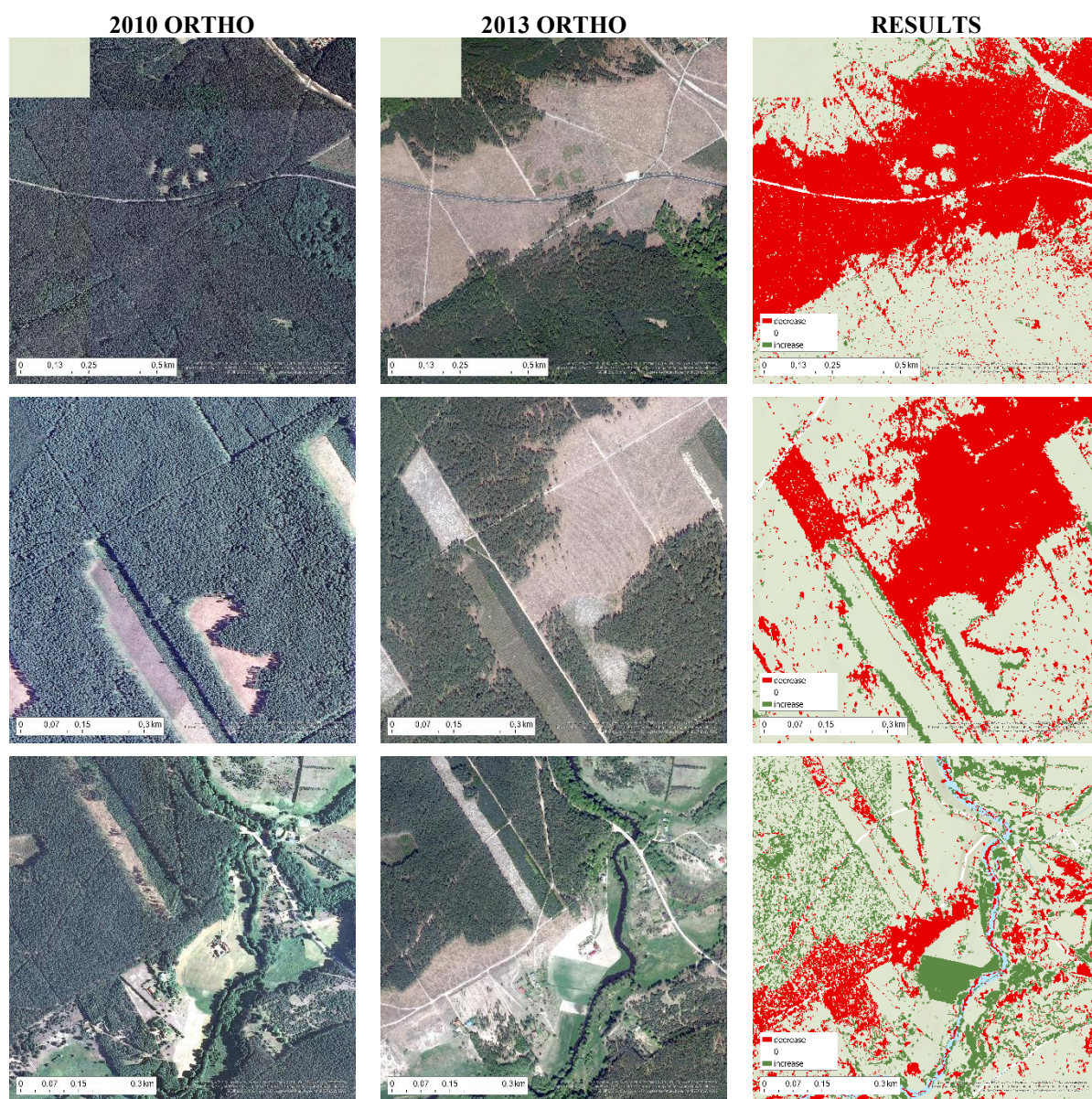


Fig. 9. Examples of the classification obtained using the SVM algorithm in combination with orthophotos.

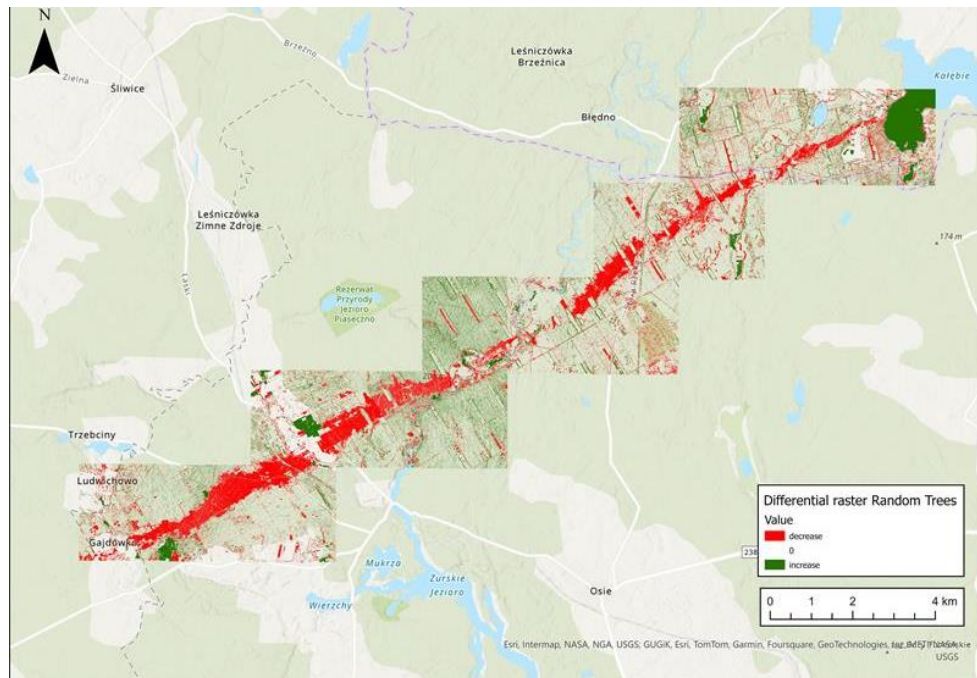


Fig. 10. Differential image obtained for the Random Trees algorithm by subtracting 2013 from 2010.

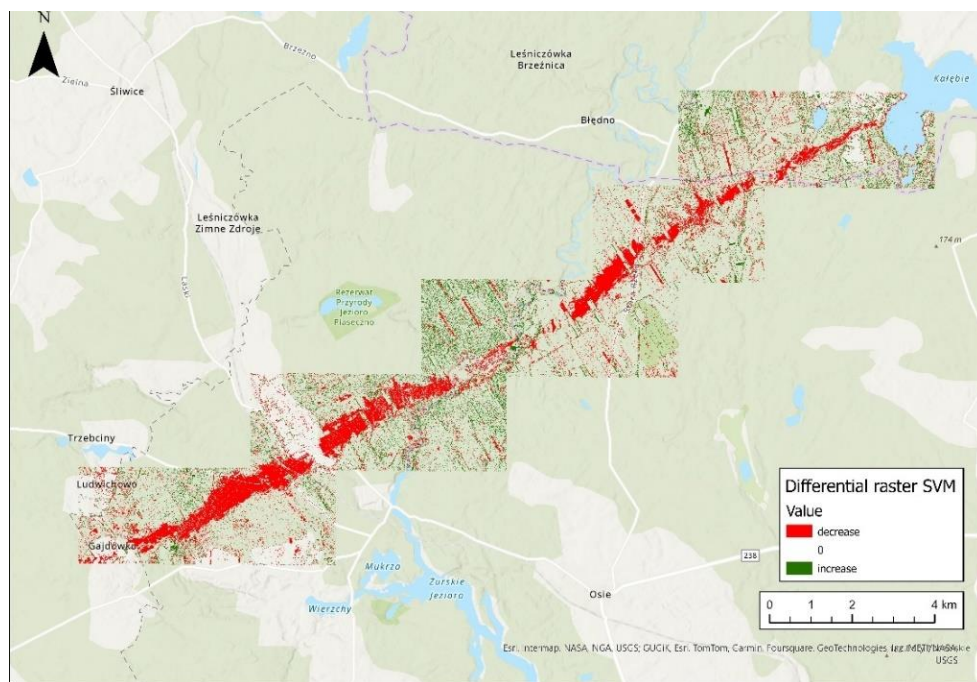


Fig. 11. Differential image obtained for the SVM algorithm by subtracting 2013 from 2010.

In addition, the use of filtering helped improve the classification results, reducing the number of erroneous pixels assigned to the wrong categories. The data after filtering was more in line with reality, confirming the validity of its use in the project.

Image segmentation was also an essential part of the analysis. The initial value of 15 used caused over-generalisation, leading to the merging of different objects into a single class and reducing classification precision. The optimal segmentation value was 20, which allowed more accurate separation of spatial units and improved the quality of the results.

In conclusion, using the SVM algorithm, filtering and appropriately selecting segmentation parameters made it possible to obtain more precise results for the classification of forest area changes in 2010-2013.

According to data from the State Forests, due to the tornado passing through Bory Tucholskie in 2012, losses were estimated at around 550 ha². However, it should be noted that the areas included in the analysis were not the same, which makes a direct comparison of the results obtained inappropriate from a methodological point of view.

REFERENCES

- Adam, E., Mutanga, O., Odindi, J., & Abdel-Rahman, E. M. (2014). Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, 35(10), 3440-3458.
- Carani, S., & Pingel, T. J. (2023). Detection of Tornado damage in forested regions via convolutional neural networks and uncrewed aerial system photogrammetry. *Natural Hazards*, 119(1), 143-166.
- Honkavaara, E., Litkey, P., & Nurminen, K. (2013). Automatic storm damage detection in forests using high-altitude photogrammetric imagery. *Remote Sensing*, 5(3), 1405-1424.
- Ciesielski, M., Bałazy, R., Hycza, T., Dmyterko, E., & Bruchwald, A. (2016). Szacowanie szkód spowodowanych przez wiatr w drzewostanach przy wykorzystaniu obrazowań

² <https://www.drewno.pl/artykuly/8534,bory-tucholskie-traba-powietrzna-spustoszy-la-lasy.html>

satelitarnych i danych Systemu Informatycznego Lasów Państwowych. *sylvan*, 160(05), 371-377.

Furukawa, F. (2022). Application of remote sensing for characterization of windthrow and landslides at multiple scales in forest landscape (thesis).

Giacco, F., Thiel, C., Pugliese, L., Scarpetta, S., & Marinaro, M. (2010). Uncertainty analysis for the classification of multispectral satellite images using SVMs and SOMs. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10), 3769-3779.

Habib, T., Inglada, J., Mercier, G., & Chanussot, J. (2009). Support vector reduction in SVM algorithm for abrupt change detection in remote sensing. *IEEE Geoscience and Remote Sensing Letters*, 6(3), 606-610.

Heikkinen, V., Korpela, I., Tokola, T., Honkavaara, E., & Parkkinen, J. (2011). An SVM classification of tree species radiometric signatures based on the Leica ADS40 sensor. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11), 4539-4551.

Mehmood, K., Anees, S. A., Rehman, A., Tariq, A., Liu, Q., Muhammad, S., ... & Hatamleh, W. A. (2024). Assessing forest cover changes and fragmentation in the Himalayan temperate region: implications for forest conservation and management. *Journal of Forestry Research*, 35(1), 82.

Mosin, V., Aguilar, R., Platonov, A., Vasiliev, A., Kedrov, A., & Ivanov, A. (2019, October). Remote sensing and machine learning for tree detection and classification in forestry applications. In *Image and Signal Processing for Remote Sensing XXV* (Vol. 11155, pp. 130-141). SPIE.

Nasimi, M., & Wood, R. L. (2024). Using Deep Learning and Advanced Image Processing for the Automated Estimation of Tornado-Induced Treefall. *Remote Sensing*, 16(7), 1130.

Omer, G., Mutanga, O., Abdel-Rahman, E. M., & Adam, E. (2015). Performance of support vector machines and artificial neural network for mapping endangered tree species using WorldView-2 data in Dukuduku forest, South Africa. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10), 4825-4840.

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 6308-6325.

Tariq, A., Jiango, Y., Li, Q., Gao, J., Lu, L., Soufan, W., ... & Habib-ur-Rahman, M. (2023). Modelling, mapping and monitoring of forest cover changes, using support vector machine, kernel logistic regression and naive bayes tree models with optical remote sensing data. *Heliyon*, 9(2).

SZACOWANIE STRAT PO TORNADACH W NADLEŚNICTWIE TRZEBCINY NA PODSTAWIE METOD SEGMENTACJI I KLASYFIKACJI KORON DRZEW

Streszczenie

Usuwanie koron drzew jest kluczowym procesem oceny szkód i zmian spowodowanych klęskami żywiołowymi. Niniejsze badanie wykorzystuje nienadzorowane techniki uczenia maszynowego do oceny strat drzewostanów po tornadach w Nadleśnictwie Trzebciny. Ortofotomapy z lat 2010 i 2013 zostały przeanalizowane za pomocą segmentacji metodą przesunięcia średniej w ArcGIS Pro, a następnie klasyfikacji za pomocą algorytmów Support Vector Machine (SVM) oraz Random Forest (RF). Ocena dokładności wykazała, że SVM przewyższa RF wyższym współczynnikiem Kappa i precyzją klasyfikacji. Wyniki wskazują na istotny spadek powierzchni leśnych z powodu szkód spowodowanych tornadem. Badanie pokazuje skuteczność metod segmentacji i klasyfikacji w szybkiej ocenie stanu lasów po katastrofie.

SŁOWA KLUCZOWE: ortofotomapa, segmentacja, klasyfikacja, SVM, las losowy, korony drzew, estymacja strat

Details of authors:

Mgr inż. Daria Dobrogowska
e-mail: daria.dobrogowska@gmail.com

Mgr inż. Wiktoria Doroz
e-mail: wiktoria.doroz.stud@pw.edu.pl

Mgr inż. Julia Domachowska
e-mail: julia.domachowska.stud@pw.edu.pl

Dr inż. Paulina Zachar
e-mail: paulina.zachar@pw.edu.pl

Submitted
Accepted

1.12.2024
31.12.2024

