Integrating Remote Sensing and AI for precision Monitoring of Soil and Vegetation Contamination

Marcin Spiralski [1], Artur Miszczak [2], Grzegorz Siebielec [3], Żaneta Piasecka[1], Ryszard Trojnacki[1], Paweł Kwaśnik[1], Jan Kotlarz[1], Katarzyna Anna Kubiak-Siwinska[1], Mariusz Kacprzak[1], Sylwia Marciniak[1], Karol Adolf Rotchimmel[1], Karol Piotr Beben[1], Jakub Szymanski[1]

[1] Łukasiewicz-Institute of Aviation, Poland, Al. Krakowska 110/114, 02-256 Warszawa, Poland – marcin.spiralski@ilot.lukasiewicz.gov.pl
[2] The National Institute of Research in Horticulture, Skierniewice, Poland – artur.miszczak@inhort.pl
[3] The Institute of Soil Science and Plant Cultivation (IUNG) - State Research Institute, Poland – gs@iung.pulawy.pl

Keywords : multispectral imagery; pollution monitoring; UAV; GEOBiA; machine learning

Abstract:

The paper presents recent advancements in monitoring pesticides and heavy metals using remote sensing techniques. The study is based on the pilot project “Support for Ecological Agricultural Production in Mazovia” (WEPR), which provided a scientific basis for the more comprehensive, larger-scale project “Support for Monitoring the Distribution of Pollutants in Agriculture” (WORZ). Addressing the needs of end-users, the research focuses on the development of reliable methods to assess environmental contaminants that impact soil quality. A combination of high-resolution UAV imagery, multispectral satellite data, and a suite of vegetation indices was utilized to correlate spatial variations in soil and crop conditions with pollutant concentrations. Machine learning algorithms, including Random Forest, were applied to classify contamination levels, while laboratory analyses validated the spectral findings. Despite challenges such as limited sample sizes and class imbalances, the integration of multi-source remote sensing data demonstrated promising results. The future of the study will be focused on the multi-temporal analyses, improving prediction accuracy and dataset and environmental risk mapping. Ultimately, the outcomes of this research show the potential for scalable monitoring systems that align with European sustainability and agricultural precision practices.

1. Introduction

Recently, the development of remote sensing techniques has led to significant advances in monitoring environmental pollutants, including heavy metals and pesticides. Remote sensing technologies, particularly those involving multispectral satellite data and hyperspectral imagery acquired from UAVs and satellites, have been widely utilized in research on the detection of heavy metal pollution. These techniques facilitate the development of spatial distribution models for contaminants, providing valuable insights into the extent and variation of pollution across landscapes. Machine learning algorithms, such as Random Forest, have proven effective in accurately estimating the concentrations of metals, such as Pb, Cu, and Ni, in soils (Omondi and Boitt, 2020; Tan et al., 2020). These algorithms have demonstrated high precision in predicting metal concentrations, further enhancing the application of remote sensing for environmental monitoring. Studies that integrate Sentinel-2 data with UAV imagery have shown that combining various data sources—such as vegetation indices and ancillary geological or land-use information—can significantly improve prediction accuracy. This integration aids in identifying areas at risk of contamination, thus supporting more effective environmental management (Khan and Kumar, 2021; Lassalle et al., 2021).

Studies emphasize the importance of high spectral resolution data for monitoring heavy metals and metalloids in soils and vegetation. Lovynska et al. (2024) reviewed methods for such monitoring, noting that high spectral resolution data can capture subtle chemical signals. However, they emphasized the need for advanced data processing techniques to handle mixed pixels and account for variations in soil properties. Guo et al. (2024) focused on using Sentinel-2 multispectral images to estimate soil copper (Cu) content. They developed two- and three-dimensional spectral indices and applied a decision tree model based on soil pH. Their method demonstrated moderate accuracy, which improved when data were grouped by optimal pH ranges. In Poland, Sobura et al. (2022) applied remote sensing techniques to map the distribution of heavy metals, including nickel, zinc, chromium, and cobalt, around the Barania Góra nature reserve. By utilizing spectral indices such as NDVI and SAVI derived from Sentinel-2 data, they demonstrated that combining satellite images with ground-based data could significantly support regional environmental monitoring. Similarly, in northeastern Kazakhstan, Felegari et al. (2023) modelled heavy metal concentrations in an industrial area using Sentinel-2 data and ensemble learning techniques. Their study further confirms the potential of remote sensing in assessing industrial pollution across large, complex landscapes. In agriculture, Zhu et al. (2023) introduced a spatial-spectral algorithm using hyperspectral leaf scanning (LeafSpec) to detect herbicide injuries in soybean crops at an early stage. Their method successfully identified subtle changes in leaf reflectance shortly after herbicide application, demonstrating its potential as a valuable tool for precision agriculture. This approach highlights the ability of hyperspectral imaging to monitor crop health and enhance agricultural management practices.

While research on pesticide monitoring is less extensive, recent studies have focused on detecting specific chemicals like glyphosate. Analysis of vegetation indices, such as the NDVI derived from Sentinel-2 data, has shown promise in quickly identifying plant responses. This technique provides an effective means of monitoring both the intensity of agrochemi-cal applications and their impact on crop health (Pause, 2019), offering valuable insights for sustainable agricultural practices.

Based on the literature demonstrating the feasibility of the study and the demand from orchard managers, at the Łukasiewicz – Institute of Aviation, we launched the pilot pro-ject "Support for Ecological Agricultural Production in Mazovia" (WEPR) in 2021. The project aimed to develop methods for neutralizing factors that negatively impact crop quality and reduce the ecological and aesthetic value of plant cover in the Mazovia region. The results from WEPR have provided a solid foundation for the subsequent development and enhancement of

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-1373-2025

the WORZ project, "Support for Monitoring the Distribution of Pollutants in Agriculture" in 2025. The ongoing project focuses on creating an innovative, universal system for monitoring pollutant residues in plant biomass and soil after the application of chemical plant protection agents.

The presented study is directly inspired by the needs of orchard growers, particularly in Poland, one of the world's leading exporters of apples, who require reliable methods for assessing pesticide residues and heavy metals—issues that have become significant concerns in maintaining soil quality across Europe (Smith et al., 2015). In Poland, where orchard production plays a key role in agriculture, these challenges are especially pressing. By addressing the presence of harmful substances in the environment, the project aims to improve soil quality and agricultural products, ultimately benefiting public health and enhancing the overall quality of life. The aim of the study was to present the results of the WEPR pilot project which demonstrated integration of the laboratory sample database with remote sensing data to extrapolate the sampling results through spatial analysis. The primary goal was to assess pesticide and heavy metal content in soils, with a focus on ana-lysing their correlation with vegetation indices. The hypothesis guiding this research was that the intensity of land use influences pesticide residue levels, with more intensively managed orchards showing higher levels of chemical application and residues compared to unused areas, where such substances are absent.

2. Methods

The primary objective of the study was to integrate the laboratory sample database with remote sensing data and to try to extrapolate the sampling results using spatial analysis. For the purpose of imagery acquisition, drones were selected due to their high-resolution imagery. Given that the first phase of the research was conducted in orchards, it was essential to obtain imagery with high GSD - several centimetres – which is sufficient, for example, to distinguish grass from individual plants. Additionally, the initial stage focused on the analysis of residues of pesticides and heavy metals in the soil, opting for an intermediate correlation with vegetation indices.

2.1. WEPR project

WEPR was an initiative designed to promote environmentally sustainable farming practices in the Mazovia region. Its primary objective was to assist farmers in transitioning to maintaining ecological methods of production by providing both financial support and technical guidance. This support included measures to reduce the reliance on chemical inputs, promote the use of organic fertilizers, and improve overall resource management through modern digital tools and innovative agricultural practices.

During the WEPR project, soil measurements were conducted in two research areas within the Municipality of Błędów (specifically in Ignaców and Oleśnik, Mazowieckie Voivodeship, Poland), with 40 soil samples collected in spring and autumn from each area, with a total of 160 samples. Research has been conducted by a consortium of 3 research institutes - Łukasiewicz – Institute of Aviation (ILOT), the Institute of Cultivation, Fertilization and Soil Science in Puławy (IUNG), and the Institute of Horticulture in Skierniewice (InHort), with declared support from the Municipality of Błędów and orchard managers cluster "Polska Natura".

2.1.1. Pilot areas

Pilot studies during the WEPR project were conducted in two areas characterized by intensive apple production, located in the "Largest Orchard in Europe" near the town of Błędów (Ignaców and Oleśnik,). This region was selected because of its high orchard density and the associated risks of intensive agricultural practices. These research areas allowed for testing the system's effectiveness in detecting and classifying pollutants under real-world conditions, ensuring its applicability to intensive fruit production environments across Europe. Specific plots were chosen through several variables as digital maps for the municipality of Błędów were selected as potential correlates of soil contamination by residues of pesticides and fertilizers. Due to the absence of high-resolution maps of soil organic matter (humus)—which plays a significant role in the sorption capacity of soils—soil texture maps (sourced from the digital Soil Map at a 1:25,000 scale) were employed as a proxy. Since soil organic matter is generally higher in denser soils with a greater content of fine particles and in more humid conditions, elevation above sea level was chosen as an indirect indicator of soil moisture, given its local correlation with the depth of the groundwater table. Digital layers representing the variability of the selected variables were generated using a circular shape, setting the window radius to 500 meters to correspond with the operational range of a drone.

This circular window, covering approximately 79 hectares, is considered sufficiently extensive to delineate a planned test area of 13.5 hectares. The variability was quantified using the standard deviation normalized to a range between 0 and 1, using following equation:

$$\sigma^* = \frac{\sigma - \sigma_{\min}}{\sigma_{\max} - \sigma_{\min}} \quad (1)$$

where :

σ^* = the normalized value representing the agricultural land use share, value between 0 and 1,

σ = the original percentage of agricultural land use in a specific area,

$\sigma_{\min}$ = the minimum value of agricultural land use share (σ) observed across all analysed areas,

$\sigma_{\max}$ = the maximum value of agricultural land use share (σ) observed across all analysed areas.

Figure 1. Plot areas, source: Łukasiewicz – ILOT

Owing to the limited preliminary data on soil contamination, the variability layers for the different variables were treated equally by computing their mean.Furthermore, to avoid selecting areas with a low proportion of orchards, the resulting mean value was multiplied by the percentage share of agricultural land within the same 500-meter radius. Finally, areas showing the highest variability of the selected variables within chosen zones with soil contamination issues were identified. Three areas - Katarzynów, Ignaców, and Oleśnik have been designated as critical areas. From them, two were chosen as pilot areas.

2.1.2 Remote sensing data

The study combines multispectral imaging techniques with laboratory technologies to assess environmental conditions. Remote sensing was used to spatially extrapolate the contamination results from sampling through the proxy of imagery of the study area, enabling the evaluation of vegetation and soil parameters through different spectral bands.

Band Name/ Function	MicaSense RedEdge-MX Dual (nm)	Band width (nm)	Sentinel-2 (nm)	Band width (nm)
Coastal Blue	444	28	443	20
Blue	475	32	490	65
Green	531	14	-	-
Green	560	27	560	35
Red	650	16	-	-
Red	668	15	665	30
Red Edge 1	705	10	705	15
Red Edge 2	717	10	740	15
Red Edge 3	740	10	783	20
Near Infrared (NIR)	842	57	842	115
Near Infrared (NIR)	-	-	865	20
Water vapour	-	-	945	20
SWIR	-	-	1375	30
SWIR	-	-	1610	90
SWIR	-	-	2190	180

Table 1. Comparison of bands of Micasense RedEdge-MX Dual and Sentinel-2 imagery, source: (Micasense 2025, SentiWiki, 2025)

This integration of field-based imaging and lab-based methods offers a comprehensive approach for validating and enhancing the overall environmental evaluation. The study utilized imaging from UAVs (Unmanned Aerial Vehicles) equipped with 10-band multispectral camera (RedEdge-P, MicaSense). UAVs covered about 300 hectares in total both in 2021 and 2022, capturing high-resolution RGB and multispectral orthophotomaps. Also, Sentinel-2 imagery was acquired for the comparison of multispectral images from both sources. UAV data collection adhered to legal and technical standards, ensuring optimal sunlight conditions. The collected imagery was processed using photogrammetric workstations and software to generate georeferenced orthophoto maps, providing insights into vegetation conditions and enabling spatial analyses.

2.1.3. Field sampling

In addition to the remote sensing data, laboratory analyses provide chemical and physical assessments of the samples.Field data contained information about pesticide residues and heavy metal content in soil. Samples were collected twice from 80 georeferenced points (40 per plot). Ground data included physical and chemical soil properties, pesticide residues (up to 478 detectable substances), and heavy metal concentrations. Laboratory methods, such as chromatography, ICP-MS (inductively coupled plasma mass spectrometry), and other analytical techniques, were employed, alongside UV-VIS (using ASD FieldSpec 4) and XRF (X-ray fluorescence) spectrometry.

2.1.4 Data processing

Prior to the image data processing, spectral signatures of soil samples were acquired. The correlations of spectral signatures with soil sampling results were calculated. Furthermore, using multispectral imagery, 17 remote sensing indices to evaluate the physical and chemical parameters of soil and vegetation, each contributing uniquely to the overall environmental assessment.. The Anthocyanin Reflectance Index (ARI) and Carotenoid Reflectance Index (CRI) were employed to estimate pigment levels, which serve as indicators of plant stress and health. The Green-Red Vegetation Index (GRVI) compared reflectance between the green and red spectral bands to differentiate between healthy and stressed vegetation, while the Modified Soil Adjusted Vegetation Index 2 (MSAVI2) effectively mitigated the influence of the soil background in sparsely vegetated areas. Additionally, the Modified Simple Ratio (MSR) provided insights into vegetation density, and the MERIS Terrestrial Chlorophyll Index (MTCI) estimated chlorophyll concentration, thereby reflecting the photosynthetic activity of the vegetation. Further indices such as the Normalized Difference Red Edge Index (NDRE) and the widely used Normalized Difference Vegetation Index (NDVI) were instrumental in assessing variations in chlorophyll content and overall biomass, respectively. The Normalized Pigment Chlorophyll Index (NPCI) and Photochemical Reflectance Index (PRI) offered valuable information on pigment composition and photosynthetic light use efficiency, while the Ratio Analysis of Reflectance Spectra Algorithm for Chlorophyll A (RARSA) was utilized to estimate chlorophyll-a content directly from spectral data. Complementary to these, the Relative Reflectance Index (RRI) and the Stress Index (SI) provided a broad evaluation of vegetation conditions and the detection of physiological stress. Lastly, the Structural Intensity Pigment Index (SIPI) and Triangular Vegetation Index (TVI) assessed the structural attributes and biomass of the vegetation.

These indices provided comprehensive information about vegetation condition and health (e.g., physiological stress, water stress, photosynthetic activity), land cover and soil contamination levels

Figure 1. Map of NDVI index. Pilot plot „Oleśnik”, 2021.

By correlating spectral reflectance with vegetation conditions, the project identified changes in the studied areas between 2021 and 2022. The data revealed seasonal variations, highlighting the importance of precise monitoring and planning in orchard management.

2.1.5. ML techniques

The classification of the tested areas was carried out using Geographic Object-Based Analysis (GEOBiA) with the use of eCognition 10. Firstly, the images were divided into homogeneous objects through a image segmentation — essentially grouping areas with similar characteristics and creating vector shapes. This step was crucial in breaking down the complex data into smaller, more manageable pieces. The system automatically considered a number of features, such as color, texture, and shape, to help define these objects more accurately. Once the segmentation was complete, different classification methods such as Random Forest (RF), Support Vector Machines (SVM), and K-Nearest Neighbours (KNN) were compared to see which one worked best for identifying contamination levels. The goal was to categorize the areas into five different levels based on pesticide contamination, helping us distinguish between the affected and non-affected zones.

These levels were tied to specific pesticide groups, making it clear where contamination was present. We have determined five pesticide contamination levels, allowing differentiation between contaminated and uncontaminated areas. In the future, monitoring will be expanded to include a broader range of pesticides and heavy metals.

2.2 WORZ system

This system will be using remote sensing tools for the needs of precision agriculture and environmental monitoring. The system aims to identify areas with a high likelihood of contamination, enabling targeted interventions while minimizing environmental impacts. By addressing the presence of harmful substances in the environment, the project focuses on enhancement of soil quality and agricultural products, ultimately contributing to public health and overall quality of life.

As part of the project, a comprehensive database will be created, integrating all the information from the analysed research sites. This will include results from physical and chemical analyses of soil and plants, as well as photogrammetric and remote sensing data. Additionally, advanced algorithms will be developed to classify areas based on the presence and type of pollutants and to identify potential sources of contamination. The database and analytical tools will be integrated into a scalable, cloud-based system, ensuring adaptability and usability at the national level.
and usability at the national level.

2.2.1 Remote Sensing Data

The research will cover 10 test areas with a combined total of at least 2,000 hectares. In these areas, remote sensing data will be collected from various platforms to provide a comprehensive picture of environmental conditions. The primary sources will include high-resolution imagery from Unmanned Aerial Vehicles (UAVs) equipped with 10-band multispectral cameras, as well as satellite imagery from Sentinel-2. In addition, the research will consider the use of Very High Resolution (VHR) satellite data, such as that PlanetScope and hyperspectral imagery from platforms like EnMap and PRISMA.. These different data sources will be processed separately to account for differences in spatial and spectral resolution, and then integrated to produce consistent maps and indices. Drone imagery will play a critical role in the selection of areas that are especially sensitive to contamination. Detailed analyses of the spectral data will allow the research team to detect subtle variations in vegetation health and soil conditions, which are key to identifying regions at risk of pesticide or heavy metal contamination. Optionally, advanced remote sensing data—including LiDAR imagery for generating digital elevation models and three-dimensional terrain maps, high-resolution Hi-res RGB images for pinpointing problematic areas, thermal imagery for monitoring soil thermal properties, and radar data from Sentinel-1 and ICEYE for assessing soil structure and moisture—will be incorporated. PlanetScope may also be utilized for detailed monitoring of short-term changes. This remote sensing approach will serve as the basis for both the initial environmental assessment and for guiding further field investigations.

2.2.2 Field Sampling

Field sampling will be conducted in parallel with the remote sensing campaigns to provide ground truth and additional data for model calibration. In each vegetation season, the research will collect a total of 800 samples across the 10 test areas. Of these 800 samples, 500 will be soil samples and 300 will be vegetation samples. The sampling locations will be carefully selected to represent a range of environmental conditions, including areas flagged as potentially contaminated by the remote sensing analysis. Each sample point will be precisely georeferenced using high-accuracy RTK/RTN systems, ensuring that every sample is accurately linked to its corresponding remote sensing data.

The soil samples will be measured in partners’ certified facilities to measure physical and chemical properties, pesticide residues, and heavy metal concentrations. Vegetation samples will also be examined for similar contaminants. These analyses will be performed using same techniques as in the pilot phase. These are including chromatography, inductively coupled plasma mass spectrometry (ICP-MS), and X-ray fluorescence (XRF) spectrometry. Also in-situ spectral measurements will be conducted using ASD FieldSpec4 spectrometers. The resulting data will form the basis of a comprehensive database that is essential for model development and validation. This database will later be used for machine learning purposes, to identify patterns and predict areas at risk of contamination. The sampling campaign is planned to cover two vegetation seasons, which will provide a seasonal comparison of the environmental conditions.

2.2.3 Data processing

The research will implement an data processing workflow to integrate remote sensing data with the field sampling results. Data preparation will include checking the completeness of attribute tables, extracting values from raster images, and linking these values to the corresponding georeferenced field samples. Statistical analyses will be conducted to explore the relationships between various environmental parameters.

The research will assess correlations between spectral indices and measured contamination levels, and perform variable selection using Principal Component Analysis (PCA) to identify the most significant factors affecting soil and vegetation conditions. Following these analyses, the dataset will be split into training and validation sets for the development of predictive models.

The image classification process will be carried out using object-based image analysis (OBIA) in eCognition Developer 10. Multi-resolution segmentation will be performed to divide the imagery into homogeneous objects based on optimized parameters such as scale, shape, and compactness. Several classification algorithms, including Bayes, Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Decision Trees, Random Trees, and potentially logistic or polynomial regression as well as neural networks, will be tested to classify the objects according to contamination levels. The classification will assign each object to one of several risk categories based on the detected levels of pesticides and heavy metals. Accuracy metrics such as accuracy, precision, recall, and F1-score will be used to evaluate the performance of these models.

During post-processing, spectral data obtained from RedEdge cameras will be resampled to match the resolution of Sentinel-2 data, thereby facilitating the integration of multi-source datasets. Sentinel-2 pixel grid with a 10 m resolution will be created to precisely assign field data and facilitate regionalization based on local soil characteristics and pollution sources. Field transects, guided by existing soil maps, will capture spatial heterogeneity and anomalies in the imagery. Field teams (3–5 members) equipped with FieldSpec spectrometers, XRF devices, and RTK GPS will conduct in-situ measurements, with additional ground spectrometry densifying the sampling grid in critical areas. Spectral data from RedEdge cameras will be resampled to match Sentinel-2 resolution, and all imagery will be orthorectified. Data processing—including the derivation of spectral indices and the development of classification models—will be executed in a cluster computing environment. The integrated dataset will test the hypothesis that pollutants induce stress in vegetation and alter soil properties, leading to detectable changes in spectral signatures, while accounting for sub-pixel variability such as tree density.

In addition, Sentinel-2 imagery will be orthorectified prior to analysis to ensure high geometric precision. All computational processes, including the analysis of remote sensing data, the derivation of spectral indices, and the development of classification models, will be executed in a cluster computing environment. This approach guarantees efficient processing of large datasets with scalability for national-level applications.

Based on the classification results, drone imagery will be used to pinpoint areas that show signs of heightened sensitivity to contamination. In the following vegetation season, a dedicated field campaign will be organized to validate the remote sensing findings in these selected areas. The feedback from this campaign will help to refine the predictive models and improve overall reliability.

Finally, in the third and final year of the project, the research will culminate in the development of an automatic system for environmental monitoring. This system will integrate the established machine learning models, automated data processing routines, and visualization tools into a single platform. The automated system will be designed to continuously process incoming remote sensing and field data, generate contamination risk maps, and provide alerts. This platform will likely be integrated with existing data-sharing and analysis infrastructures, such as CREODIAS, to support broader environmental management efforts.

3. Results

Results from the WEPR project areas are available at the website wepr.pl. It can be accessed both through the databases and map portal, which provides detailed visualizations of spatial data and pollutant distribution. These resources allow users to explore the project's findings through an interactive map and database, facilitating a deeper understanding of pollutant levels across various regions.

The study's most critical finding is the characterization of soil pH. The analysis revealed that a significant portion of soils is slightly acidic (43%) or acidic (34%), with 11% being very acidic. Since soil pH fundamentally affects nutrient availability, microbial activity, and soil structure, maintaining an optimal level is essential. Even slightly acidic conditions, while generally supportive of crop production, require monitoring and occasional liming to prevent further acidification from mineral fertilizers. Acidic soils can lead to deteriorated structure and increased solubility of harmful elements like aluminium, which impairs root growth. Moreover, pH influences the decomposition of organic matter and nutrient cycling, making its proper management a cornerstone of sustainable agriculture and long-term soil conservation.

Residues of plant protection products ranging from one to sixteen different types were detected in the measured samples, indicating varying intensity of treatments applied in the studied area. It was also determined that there were no traces of phosphonic acid derivatives, which could have persisted in the soil and, as a systemic substance, potentially moved from the soil to the plant, resulting in residues in fruits/vegetables for several years after the last application of this fungicide, which was widely used at one time. The issue of detecting phosphonic acid residues is a concern for producers of organic food and food intended for children.

Spatial extrapolation was performed using machine learning techniques that integrated remote sensing data with ground truth measurements. The trained ML models predicted pollutant levels across unsampled areas, resulting in high-resolution contamination maps that enhance our understanding of spatial pollutant distribution. These maps are invaluable tools for identifying areas with elevated pollution levels, which can inform future environmental interventions and agricultural practices. The probability percentage (classification confidence) varied depending on the class (Fig. 2).. For some classes, it is close to 100%, while others have been misclassified.

Overall Accuracy (OA) for Ignaców in 2021 was 0.33, while in 2022, it was 0.49. For Oleśnik, OA in 2021 was 66%, while in 2022 it was 48%. The highest classification accuracy was achieved for the Oleśnik area in 2021, reaching 66%, while the lowest accuracy was observed for the Ignaców area in 2021, at 31% . The variation in accuracy across different regions reflects the challenges inherent in modelling pollutant distribution with remote sensing data, particularly in areas with complex land use or insufficient data points. Despite the significant potential of remote sensing data, the classification results are not entirely satisfactory. The primary reason for this was the low number of measurement points with an uneven distribution of cases across pollution classes. This limitation underscores the importance of increasing the density and spatial coverage of ground truth data to improve the robustness and reliability of the ML models.

It should also be noted that this accuracy was calculated for the training points on which the algorithm was trained. The accuracy scores thus reflect the model’s performance based on the available data and may not fully represent its ability to generalize to new, unseen areas. This highlights the need for further testing and validation, particularly in regions with limited ground truth data.

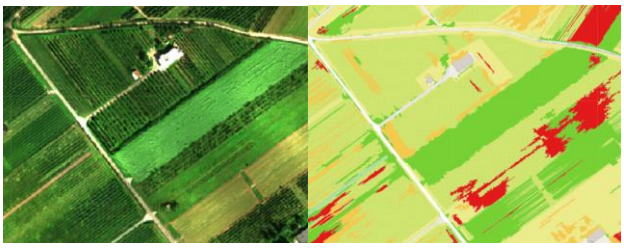

Figure 2. Example of classification, green areas – no contamination of pesticides, yellow areas – medium contamination, red areas – high contamination of pesticides source: Łukasiewicz - ILOT

4. Discussion

To determine a reliable accuracy, soil samples would need to be collected and analysed from other locations within the study areas and compared with the classification results. The small number of samples in this class caused difficulties in achieving correct classification. Also, it seems that a multitemporal analysis (covering multiple time periods), and consequently, increasing the training dataset, should address the issue of misclassifications. Class imbalance generally leads to challenges in the training phase and reduces predictive capability. This should be taken into account, as it is a common phenomenon in environmental data modelling. The contamination map validation and the commentary on the results were also obtained from orchardists operating within the study area. In the Ignaców region, no contamination was recorded on the plot that had remained unused for several years and where no plant protection products were applied. This finding therefore confirms some of the classes on this area. Moreover, elevated contamination levels were observed in the vicinity of an terrain depression, as indicated by the local topography. This, in turn, suggests that leaching and the accumulation of plant protection product residues from the land surface are contributing factors.

The results of similar published studies indicate that the synergistic use of satellite data and UAV imaging, combined with analytical methods (indices analysis, ML, spectral unmixing), represents a promising strategy for the effective monitoring of both soil heavy metals and pesticide applications. This integrated approach facilitates the creation of environmental risk maps that can support remediation efforts and inform decision-making in natural resource management.

Therefore, the new project WORZ recommends scaling multispectral UAV analyses along with spectral sampling and ground truth acquisition for broader applications.

That is why, continuation of the project was proposed, for the scale of whole Poland to achieve much bigger dataset for the purposes of pollutants estimation using ML. WORZ is in line with global trends such as the European Green Deal, the Soil Mission, and other initiatives to reduce pesticide usage. In the research by Sobura et al (2022), among the seven analysed heavy metals, a clear relationship was observed between nickel (Ni), cobalt (Co), chromium (Cr) and zinc (Zn), and the values of spectral indices included in this study.

These trends collectively underscore the importance of sustainable farming practices and environmental protection. By promoting the precise and reduced application of chemical treatments, the project contributes to a decrease in the overall environmental footprint of agricultural practices. The reduced dependency on broad-spectrum chemical applications supports biodiversity, enhances soil health, and aligns with the broader goals of sustainable and resilient agricultural systems across the EU. Also, it may be a significant step to the competitiveness of European agriculture in the face of threats of increasing costs (e.g. fertilization) or new competition from food products made in Mercosur countries.

Having explored the market demand for tools for precise estimation of pollution in agriculture, the project has a clear commercialization pathway. The outcomes are set to be transferred to the market through the establishment of a dedicated spin-off company. This move is designed to ensure that the innovative solutions developed during the project remain accessible and continue to evolve beyond the research phase. The spin-off not only guarantees a sustainable business model for the technology but also plays a crucial role in disseminating the benefits of the project to farmers, agronomists, and other stakeholders in the agricultural sector.

5. Conclusions

1. The integration of UAV and satellite imagery with laboratory and field data proves their potential to be an effective approach for monitoring pesticide residues and heavy metals in agricultural soils.
2. The WEPR pilot project has successfully demonstrated the feasibility of using advanced remote sensing techniques and machine learning for environmental contaminant mapping in intensive orchard settings.
3. While classification accuracy was influenced by limited sampling and data imbalance, the findings suggest that expanding the dataset and incorporating multi-temporal analyses can significantly improve predictive performance.
4. The refined methodologies from WEPR provide a robust foundation for the WORZ project, aiming for national-scale implementation and enhanced environmental monitoring.
5. The research outcomes align with European initiatives such as the European Green Deal and the Soil Mission, promoting sustainable agricultural practices and improved public health.
6. Future efforts should focus on further refining analytical techniques and integrating additional environmental variables to improve the monitoring system’s accuracy and reliability.

Acknowledgments

WEPR research project was a collaborative initiative involving leading scientific and research institutions, such as the Łukasiewicz Research Network – Institute of Aviation (ILOT), the National Institute of Research in Horticulture (InHort), and the Institute of Soil Science and Plant Cultivation - State Research Institute (IUNG), in partnership with local government bodies and orchard producers. The project was financed from Mazovia Voivodeship funds. WORZ project is financed from the funds of The National Centre for Research and Development (NCBR) under agreement number INFOSTRATEG6/0010/2023.

References

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry & Remote Sensin*g, 58, pp. 239–258.

Zeaiter, M., Rutledge, D., 2009. Pre-processing Methods. In: S.D. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, Elsevier, Oxford, pp. 121-231. ISBN 978-0-44-452701-1. doi: 10.1016/B978-044452701-1.00074-0.

Kaźmierowski, C., Piekarczyk, J., Cierniewski, J., Kuśnierek, K., Królewicz, S., Terelak, H., Stuczyński, T., Maliszewska-Kordybach, B., 2012. Diffuse reflectance spectroscopy for monitoring arable soil chemistry on the national scale. 4th International Congress EUROSOIL 2012, Soil Science for the Benefit of Mankind and Environment, 2-6 July 2012, Bari, Italy.

Wężyk, P., et al., 2012. Klasyfikacja pokrycia terenu metodą OBIA z wykorzystaniem zobrazowań satelitarnych Rapideye.

Simmons, K., 2017. Field X-Ray Fluorescence Measurement: Operating Procedure. U.S. Environmental Protection Agency.

Chen, Y., Hou, C., Tang, Y., Zhuang, J., Lin, J., He, Y., Guo, Q., Zhong, Z., Lei, H., Luo, S., 2019. Citrus Tree Segmentation from UAV Images Based on Monocular Machine Vision in a Natural Orchard Environment. *Sensors*, 19(24), 5558. https://doi.org/10.3390/s19245558.

Zhang, C., Valente, J., Kooistra, L., Guo, L., Wang, W., 2021. Orchard management with small unmanned aerial vehicles: A survey of sensing and analysis approaches. Precision *Agriculture*, 22(6), pp. 2007–2052. https://doi.org/10.1007/s11119-021-09813-y.

Delavarpour, N., Koparan, C., Nowatzki, J., Bajwa, S., Sun, X., 2021. A Technical Study on UAV Characteristics for Precision Agriculture Applications and Associated Practical Challenges. *Remote Sensing*, 13(6), 1204. https://doi.org/10.3390/rs13061204.

Khan, A., Kumar, R., 2021: Modeling the distribution of heavy metals in lands irrigated by wastewater using satellite images of Sentinel-2. *Eur. J. Remote Sens.,* 54, 78-91. doi.org/10.1016/j.ejrs.2021.03.002.

Omondi, E., Boitt, M., 2020: Modeling the Spatial Distribution of Soil Heavy Metals Using Random Forest Model —A Case Study of Nairobi and Thirirka Rivers’ Confluence. *J. Geogr. Inf. Syst.*, 12, 597-619. doi.org/10.4236/jgis.2020.126035.

Pause, M., 2019: Monitoring Glyphosate-Based Herbicide Treatment Using Sentinel-2 Time Series—A Proof-of-Principle. *Remote Sens*., 11(21), 2541. doi.org/10.3390/rs11212541.

Lassalle, G., et al., 2021: Mapping leaf metal content over industrial brownfields using airborne hyperspectral imaging and optimized vegetation indices. *Sci. Rep.*, 11, 1-8. doi.org/10.1038/s41598-020-79439-z

Tan, K., Wang, H., Chen, L., Du, Q., Du, P., Pan, C., 2020: Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest. *Remote Sens. Environ.,* 240, 111712. doi.org/10.1016/j.rse.2019.111712.

Smith, P., House, J.I., Bustamante, M., Sobocká, J., Harper, R., Pan, G., West, P.C., Clark, J.M., Adhya, T., Rumpel, C., Paustian, K., Kuikman, P., Cotrufo, M.F., Elliott, J.A., McDowell, R., Griffiths, R.I., Asakawa, S., Bondeau, A., Jain, A.K., Meersmans, J., Pugh, T.A.M., 2016. Global change pressures on soils from land use and management. *Glob Chang Biol* 22, 1008–1028. doi:10.1111/gcb.13068

SentiWiki, 2025. S2 Mission. Available at: https://sentiwiki.copernicus.eu/web/s2-mission (accessed February 8, 2025).

MicaSense 2025. What is the center wavelength and bandwidth of each filter for MicaSense sensors. Available at: https://support.micasense.com/hc/en-us/articles/214878778-What-is-the-center-wavelength-and-bandwidth-of-each-filter-for-MicaSense-sensors (accessed February 8, 2025).

Lovynska, V.; Bayat, B.; Bol, R.; Moradi, S.; Rahmati, M.; Raj, R.; Sytnyk, S.; Wiche, O.; Wu, B.; Montzka, C., 2024: Monitoring Heavy Metals and Metalloids in Soils and Vegetation by Remote Sensing: A Review. *Remote Sens*., 16, 3221. doi.org/10.3390/rs16173221.

Guo, H.; Wu, F.; Yang, K.; Yang, Z.; Chen, Z.; Chen, D.; Xiao, R., 2024. Sentinel-2 Multispectral Satellite Remote Sensing Retrieval of Soil Cu Content Changes at Different pH Levels. *Agronomy*, 14(10), 2182. doi.org/10.3390/agronomy14102182.

Sobura, S.; Widłak, M.; Hejmanowska, B.; Muszyńska, J., 2022: The Application of Remote Sensing Techniques and Spectral Analyzes to Assess the Content of Heavy Metals in Soil – A Case Study of Barania Góra Reserve, Poland. *Geomatics and Environmental Engineerin*g, 16(4), 187–213. doi.org/10.7494/geom.2022.16.4.187.

Felegari, S.; Sharifi, A.; Khosravi, M.; Ma, C.; Wang, Y.; Meng, F.; Karuppannan, S., 2023: Modeling the Distribution of Heavy Metals in Industrial Areas Using Sentinel-2 Data: A Case Study from Ust-Kamenogorsk, Northeastern Kazakhstan. *Eur. J. Remote Sens.*, doi.org/10.1016/j.ejrs.2021.03.002.

Zhu, X.; Huang, L.; Xu, H.; Liu, Y., 2023: Early Detection of Dicamba and 2,4-D Herbicide Drifting Injuries on Soybean with a New Spatial–Spectral Algorithm Based on LeafSpec, an Accurate Touch-Based Hyperspectral Leaf Scanner. *Remote Sens.,* 15(24), 5771. doi.org/10.3390/rs15245771.