Article Type: Research
J Name: Modern Phytomorphology
Short name: MP
ISSN: ISSN 2226-3063/eISSN 2227-9555
Year: 2024
Volume: 18
Page numbers: 183 - 187
DOI: 10.5281/zenodo.200121
(10.5281/zenodo.2024-18-PDFNo.)
Short Title: Supervised machine learning in crop recognition through remote sensing: A case study for Ukrainian croplands

RESEARCH ARTICLE

Supervised machine learning in crop recognition through remote sensing: A case study for Ukrainian croplands

Pavlo Lykhovyd[1]*, Raisa Vozhehova[1], Iryna Bidnyna[2], Oleksandr Shablia[1], Oleksandr Averchev[3], Nataliia Avercheva[3], Valerii Kozyriev[1], Tetiana Marchenko[1], Liubov Leliavska[4], Oleksandr Haydash[5], Maiia Hetman[2], Olena Piliarska[1]

[1] Institute of Climate-Smart Agriculture of NAAS, Odesa, Ukraine
[2] National Academy of Agrarian Sciences of Ukraine, Kyiv, Ukraine
[3] Kherson State Agrarian and Economic University, Department of Agriculture, Kropyvnytskyi, Ukraine
[4] Institute of Water Problems and Land Reclamation of NAAS, Kyiv, Ukraine
[5] State Institution the Institute of Grain Crops of NAAS, Dnipro, Ukraine

***Corresponding author:** Pavlo Lykhovyd, Institute of Climate-Smart Agriculture of NAAS, Odesa, Ukraine
E-mail: pavel.likhovid@gmail.com

Received: 02.10.2024, Manuscript No: mp-24-149385 **| Editor Assigned:** 05.10.2024, Pre-QC No. mp-24-149385(PQ) **| Reviewed:** 21.10.2024, QC No. mp-24-149385(Q) **| Revised:** 27.10.2024, Manuscript No. mp-24-149385(R) **| Accepted:** 02.11.2024 **| Published:** 08.11.2024

Abstract

Automated crop recognition is an important branch of modern agriculture. It provides wide opportunities for cropland mapping, crop rotations analysis, cropland structure and agricultural land use monitoring, etc. Remote sensing is a prospective and powerful technique for crop recognition through the implementation of various vegetation indices, e.g., normalised difference vegetation index, in combination with technologies of machine learning and computer vision. Current study is devoted to real-world testing of the accuracy of recent development in supervised machine learning for crop recognition in Ukraine, namely, software application Agroland Classifier, which has been built based on the results of scientific research at the Institute of Climate Smart Agriculture of NAAS. The application utilizes several supervised machine learning approaches, namely, multiple canonical discriminant analysis and logistic regression, to distinguish between such crops as winter wheat, winter barley, winter rapeseed, grain maize, soybeans, and sunflower. The testing was carried out using randomly chosen labelled fields with known cultivated crops, 100 fields per each crop. Testing was carried out throughout all the territory of Ukraine. The input values of monthly normalised difference vegetation index were retrieved from Agromonitoring Crop Map platform. It was established that the highest precision of crop recognition was associated with wheat (overall accuracy of 82.0%, F1 score 0.90), while the worst results were recorded for soybeans (50.0% of true guesses, F1 score 0.67). It was also observed that the recognition accuracy is highly dependent on soil-climate conditions of the crops cultivation. Further detailed testing and algorithms improvement are required and will be held on.

Keywords: Accuracy, Artificial intelligence, Discriminant analysis, F1 score, Logistic function, Normalised difference vegetation index, Sigmoid function

Introduction

In the context of the current global food crisis, which is especially threatening with the background of climate change and increasing military activities, modern agricultural science must provide relevant and sufficiently substantiated solutions for solving the problem of malnutrition, low food availability and food quality deterioration. It is important to provide rational use of available natural resources so that no harm is done. Land cover and soil are crucial for food production, as agricultural activity is almost impossible without them. Thus, the development of responsible, climate smart and environmentally friendly approaches to land use is essential for sustainable development and food security. Current agro technologies and land use practices are often threatening ecological sustainability, and lead to irrational nature use, and at the same time provide no befits for food crisis solution (Yang et al., 2024). Rational land use is impossible without croplands account and conditional assessment, in which remote sensing technologies are quite promising, helping to obtain necessary information without soil, land and plant disturbance, timely, and with high precision. Cropland monitoring and mapping are important for rational land use policy formation, optimization of land and soil cover use, improvement of crop rotation structure, general enhancement of ecologically responsible crop cultivation, tracing the directions of promising developments in crops cultivation and marking current trends and faults in crop production. For example, irrational and hazardous strategies for the replacement of natural forests with agricultural lands in the recent decades in Africa and South America were discovered by the means of crop mapping technologies and analysis (Potapov et al., 2022).

One of the recent fields of remote sensing application in agricultural and environmental sciences is cropland monitoring in general and crop recognition and mapping in particular. As far as on land direct surveys are still carried out, they require too much manpower, costs, and time to map a comparatively small parcel of land area. With remote sensing, this operation could be performed in hours, if not in minutes, for areas of almost any scale. Recent studies show that in recent years the amount of scientific research in the field of application of remote sensing techniques in combination with geoinformation technologies and mathematical approaches to data analysis is continuously growing from just 1 to 2 studies per year published since the 2000s to 28 to 77 studies per year in 2018-2022 (Alami Machichi et al., 2023). This is an indirect approval of the importance of remote sensing application for crop recognition and mapping. Most of the studies are carried out in China and the United States, with slightly less share in developed western European countries. Spatial imagery from the Sentinel and Landsat satellites is predominant as input for the study conduction; as for the mathematical computation procedures, the artificial intelligence-based approaches are the most prominent. It is not surprising, especially considering the recent success of artificial intelligence in other branches of agriculture (Lykhovyd et al., 2024). As for crop classification and recognition, machine learning approaches are utilized, namely, parametric learning, non-parametric learning, supervised and non-supervised deep learning models within the frameworks of convolutional neural networks of different architecture and structure, support vector machine, long short term memory networks, as well as more conventional models like random forest. Also, different approaches to pre-processing of the spatial images are applied, influencing the overall accuracy rates of recognition. In general, crop mapping and crop guessing precision fluctuate from 75% to 90%-95% (Alami Machichi et al., 2023). Using different approaches to image processing and using different input vegetation indices or their combination have a strong influence on crop recognition accuracy. The most widely used vegetation index is the normalized difference vegetation index, which in our opinion, is the most optimal one for the purpose due to its high availability on different platforms and great flexibility of use for crop canopy cover identification and characterization (Tenreiro et al., 2021).

The aim of this study was to evaluate the overall accuracy of Agroland Classifier an HTML based cloud application with supervised machine learning approaches for major crops recognition based on the remotely sensed normalised difference vegetation index values during the growing season. The application was developed by the scientific team from the Institute of Climate Smart Agriculture of NAAS, Ukraine, mainly for the needs of the stakeholders and farmers, which provide their activity in the Steppe zone of the country. However, in this study, the assessment was performed throughout all the territory of Ukraine, using the labelled fields and ready to use spatial imagery.

Materials and Methods

The testing of the Agroland Classifier application, which uses several supervised machine learning algorithms (multiple canonical discriminant functions and logistic regression sigmoid functions) to recognize agricultural plants was carried out for six major crops, cultivated in Ukraine, namely winter wheat, winter barley, winter rapeseed, sunflower, soybeans, and maize (Lykhovyd, 2024). Each crop was represented by 100 labelled fields, located in different regions of Ukraine (600 investigated fields in total, the initial data set could be provided by the authors on reasonable request). The marginal coordinates in Google Maps format for each crop studied are presented in tab. 1. The values of the normalised difference vegetation index were retrieved from the online platform Agromonitoring Crop Map. The platform provides the smoothed monthly time series of the index, based on the common assessment procedures of cloudless satellite images from Sentinel 2 and Landsat 8 sensors with a resolution of 250 m. Index values were entered into the corresponding cells of the application, as guided by the built-in program instructions. The classification results achieved were fixed and further generalized for each crop to calculate the general percentage of correct guessing and the magnitude of the absolute error. Interpretation of the accuracy testing was carried out as guided by (Vivas et al. 2020) as follows: error <10% is evaluated as a highly accurate prediction; 10%-20% good prediction; 20%-50% reasonable prediction (more useful for theoretical rather than practical implementation); >50% inaccurate prediction (the model is unreliable, should be revised). At the end, the F1 score, representing the harmonic mean of precision and recall, was calculated (Taha and Hanbury, 2015). The closer to 1.0 the F1 score is, the higher the precision of the predictive model. In addition, the highest incidence of mixing up crops was estimated for each recognition group. General methodological workflow of the study is depicted in fig. 1.

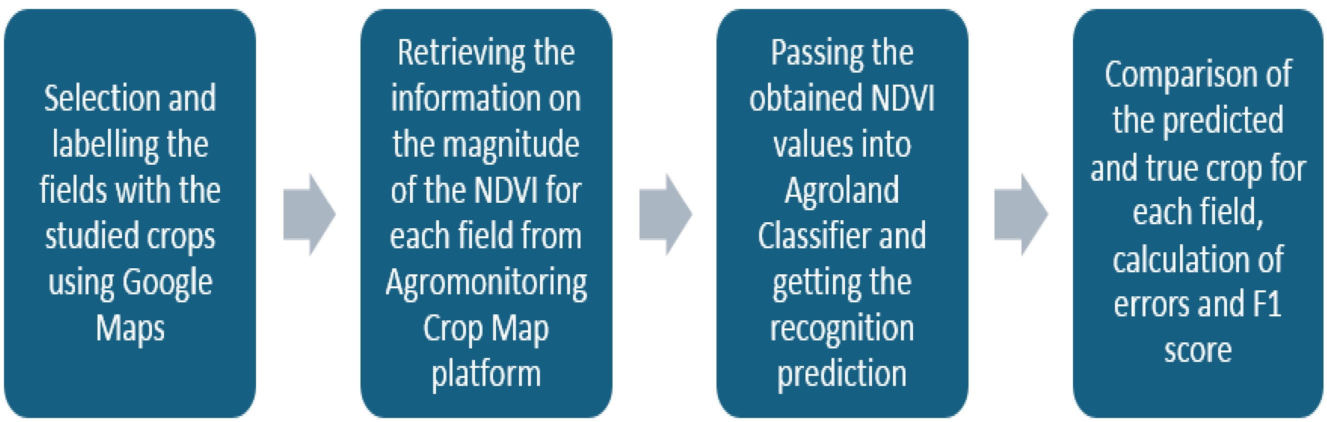

Figure 1. Methodological workflow of testing the Agroland Classifier accuracy in crop recognition.

Results and Discussion

As a result of the software testing, it was established that the accuracy of the studied crops recognition was strongly dependent on the type of agricultural plants. The highest accuracy was recorded for sunflower and winter wheat crops, while the worst percentage of true guessing was attributed to soybeans (Tab. 1). There are several additional factors that must be considered while analysing the study results. First, it should be stressed that the study was carried out throughout Ukraine, while the machine learning models, used in the Agroland Classifier, have initially been developed for the conditions of the Steppe zone. Therefore, in the course of the research it was established that crop guessing worsened with the movement away from the zone of the Steppe of Ukraine. As soil and climate conditions are extremely different in the south central and west northwest parts of Ukraine, it is not surprising that the models failed to distinguish the crops in unfamiliar conditions, because Agroland Classifier uses supervised machine learning, and in the changing conditions the models are highly likely to need additional calibration and adjustment. Second, it should be stressed that the tests were carried out without subdivision on the irrigated and rainfed conditions of the cultivation of agricultural plants. As long as irrigation significantly shifts the patterns of the normalised difference vegetation index distribution throughout the growing season, it could become an additional source of error and incorrect guessing of the crops. This hypothesis can be supported by the fact that the highest percentage of error was recorded for soybeans and maize, which are crops that are cultivated massively under irrigated conditions. While sunflower and winter wheat, which are less represented on the irrigated lands, were recognized the best. High errors for rapeseed crops could also be put upon the features of the crop blooming, when its reflectance characteristics because of bright yellow colour of the crop's inflorescences are easily distorted. As for winter barley, it was not surprising to see that this crop was frequently mixed up with winter wheat, as their growing patterns are quite similar.

Table 1. The results of Agroland Classifier Testing.

Crops	Google Coordinates (max-min amplitude)	Error Percentage/F1	Highest Coincidence	Prediction Probability
Maize	46.212-50.087, 26.089-35.384	48.0/0.68	Sunflower	Reasonable
Soybeans	45.451-48.913, 29.209-35.543	50.0/0.67	Sunflower	Reasonable
Sunflower	46.236-48.547, 30.374-35.289	19.0/0.90	Maize/soybeans	Good
Wheat	46.051-49.563, 26.524-34.965	18.0/0.90	Rapeseed	Good
Barley	46.344-50.776, 25.046-34.421	47.0/0.69	Wheat	Reasonable
Rapeseed	46.309-51.490, 24.881-35.211	41.0/0.74	Wheat/barley	Reasonable
Generalized	45.451-51.490, 24.881-35.543	37.2/0.76	Reasonable	

In terms of F1 score, the tendency is the same. The best F1 score was recorded for winter wheat and sunflower, while the worst for soybeans and maize. However, the F1 score testifies good prediction accuracy of the models for wheat, sunflower, and rapeseed, as well as good general accuracy of 0.76. Compared to other scientific approaches and machine learning algorithms, it should be pointed out that supervised machine learning is usually inferior to non-supervised deep learning. This is mainly because of some limitations, connected with supervised learning, which uses labelled datasets to build a model. Supervised algorithms are unable to 'learn' in the full sense of this word. Therefore, as was mentioned before, they are highly vulnerable to changes in conditions and require continuous calibration. General accuracy of the Agroland Classifier was estimated as 62.8%. This figure varies depending on the crop. The generalized F1 score reached 0.76, which is good enough to justify the implementation of the software in scientific work and in some cases even in practice.

Compared to other similar studies, the algorithms embedded in the Agroland Classifier framework provide competitive results, which are just slightly inferior to other supervised learning techniques. For example, the study by (Latif 2019) testifies that the robust decision tree based approach to crop recognition, implemented in the conditions of Pakistan, using UAV time series of the normalised difference vegetation index provided an overall accuracy of just about 76%. Another study, which describes the results of machine learning based crop recognition using remote sensing imagery, claims that object based crop classification resulted in an overall accuracy of 75.4% (Tian et al., 2021). Much better results were reported by (Zheng et al. 2015), where a more robust machine learning technique of a support vector machine was used to distinguish between irrigated crops. The overall accuracy reached 86%, and it was also stated that the labelled learning approach not only can reduce the training data set but also provides more consistent and accurate predictions in crop recognition. In the study by (Shao et al. 2010), Kappa coefficients for the crop recognition model, which was built on the basis of MODIS NDVI time series, fluctuated within 0.69-0.74 depending on ecological conditions of the studied regions? Using an artificial antibody network to recognize crops depending on the NDVI patterns resulted in an output precision of 83.5%-87.1% in the study by (Hao et al. 2016). There is a lack of scientific evidence on the use of simpler machine learning methods, e.g. logistic and discriminant functions, for crop recognition and classification. Furthermore, some researchers provided scientific evidence for the use of novel, specifically developed, vegetation indices to perform crop recognition, for example, the Normalised Difference Yellow Vegetation Index (NDYVI), which was shown to outperform NDVI in the classification between wheat and rapeseed (0.85 vs. 0.83), maize and soybeans (0.91 vs. 0.84). However, it should be noted that pairwise recognition patterns are much easier held by models, comparing to multi crop recognition systems, thus, additional evidence in multi-crop studies in different agro ecological conditions are required to substantiate the replacement of conventional spatial vegetation indices with NDYVI (Wei et al., 2024).

Conclusions

The machine learning algorithms, realized within the framework of Agroland Classifier, provides good recognition accuracy for winter wheat and sunflower. As for the other crops, the accuracy falls within the range of reasonable prediction, with the highest error fixed for soybeans, and the smallest for winter rapeseed. It was also established that maize and soybeans are most frequently mixed up with sunflower, while winter barley is mixed up with wheat, and winter wheat – with rapeseed. The machine learning algorithms, laid in the basis of Agroland Classifier are far from perfect, but their general performance (generalized error of 37.17%) is reasonable. Another important fact is that the application was originally developed for the Steppe zone of Ukraine, and the test was performed throughout all the territory of Ukraine. Therefore, soil and climatic differences could be to some extent to blame in lowering the program performance in crop recognition.

References

Alami Machichi M, Mansouri LE, Imani Y, Bourja O, Lahlou O, Zennayi Y. (2023). Crop mapping using supervised machine learning and deep learning: a systematic literature review. *Int J Remote Sens*. **44:**2717-2753.

Hao P, Wang L, Zhan Y, Niu Z. (2016). Using moderate-resolution temporal NDVI profiles for high-resolution crop mapping in years of absent ground reference data: a case study of Bole and Manas counties in Xinjiang, China. *ISPRS Int J Geo-Inf*. **5:**67.

Latif MA. (2019). Multi-crop recognition using UAV-based high-resolution NDVI time-series. *J Unmanned Veh Syst*. **7:**207-218.

Lykhovyd P, Hranovska L, Bidnyna I, Marchenko T, Averchev O, Leliavska L. (2024). A review on the use of artificial intelligence and deep learning algorithms in crops phytosanitary monitoring. *Mod Phytomorphol*. **18:**64-69.

Lykhovyd PV. (2024). Normalized difference vegetation index as a marker of winter crops identification in the systems of automated crops mapping. *Agrarian Innov*. **23:**98-104.

Potapov P, Turubanova S, Hansen MC, Tyukavina A, Zalles V, Khan A. (2010). Mapping cropland and major crop types across the Great Lakes Basin using MODIS-NDVI data. *Photogramm Eng Remote Sens*. **76:**73-84.

Shen Q, Cortez J. (2022). Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nat Food*. **3:**19-28.

Taha AA, Hanbury A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. **15:**1-28.

Tenreiro TR, García-Vila M, Gómez JA, Jiménez-Berni JA, Fereres E. (2021). Using NDVI for the assessment of canopy cover in agricultural crops within modelling research. *Comput Electron Agric*. **182:**106038.

Tian Y, Yang C, Huang W, Tang J, Li X, Zhang Q. (2021). Machine learning-based crop recognition from aerial remote sensing imagery. *Front Earth Sci*. **15:**54-69.

Vivas E, Allende-Cid H, Salas R. (2020). A systematic review of statistical and machine learning methods for electrical power forecasting with reported MAPE score. *Entropy*. **22:**1412.

Wei Y, Lu M, Yu Q, Li W, Wang C, Tang H. (2024). The normalized difference yellow vegetation index (NDYVI): a new index for crop identification by using GaoFen-6 WFV data. *Comput Electron Agric*. **226:**109417.

Yang Y, Tilman D, Jin Z, Smith P, Barrett CB, Zhu YG. (2024). Climate change exacerbates the environmental impacts of agriculture. *Science*. **385.**

Zheng B, Myint SW, Thenkabail PS, Aggarwal RM. (2015). A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *Int J Appl Earth Obs Geoinf*. **34:**103-112.