вести щоденник переживань. Але важливо, щоб ці системи розуміли наш контекст: досвід війни, втрат та постійної невизначеності. Це потребує спеціальної адаптації та перевірки. У підсумку, якщо впроваджувати ШІ з розумом і контролем з боку людини, він може стати потужним помічником, не втрачаючи при цьому головного — безпеки та потреб самої людини.

Список використаних джерел

1. Fitzpatrick K. K., Darcy A., Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Mental Health. 2017. Vol. 4, no. 2. Article e19. DOI: https://doi.org/10.2196/mental.7785.

2. Randomized Trial of a Generative AI Chatbot for Mental Health Treatment / M. V. Heinz et al. NEJM AI. 2025. Vol. 2, no. 4. DOI: https://doi.org/10.1056/AIoa2400802.

3. Типтюк Ю. А., Т. В. Іванова. Використання систем штучного інтелекту для психологічної підтримки в умовах воєнного стану в Україні. Журнал соціальної та практичної психології. 2025. С. 171-177.

4. Палько Т.В. Парадокс антропоморфізму: психологічні межі емоційної прив'язаності до неігрових персонажів та чат-ботів з підтримкою ШІ. The III International scientific and practical conference «Technology development: shaping modern thinking and scientific approaches», January 19-21, 2026, Krakow, Poland, 265 p.

УДК 004.85:631.559

MACHINE LEARNING APPROACHES TO CROP YIELD PREDICTION USING MULTISOURCE AGRICULTURAL DATA

Mykhailo Negria

Institute of Special Communications and Information Protection,
National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine

Abstract. This paper looks at how machine learning can be used to forecast crop production by combining information from different sources. It looks at the key things that affect crop production, the kinds of information used in today's prediction tools, how the most widely used machine learning methods work, and how these methods are specifically applied to data from farming. The benefits of using machine learning compared to traditional statistical methods are discussed, along with the difficulties that come with putting these technologies into practice. Special care is taken to mix weather, soil, and satellite information, check how good the weather prediction tools are, and help make better farming choices.

Keywords: machine learning, crop yield prediction, multisource data, remote sensing, vegetation indices,

Research Relevance: Agriculture is facing more challenges because of changing weather patterns, higher costs to produce food, and the need to keep food supplies steady and reliable. Under these conditions, accurate yield predictions are becoming an important tool for planning fertilization, irrigation, storage, logistics, and sales. At the same time, the amount of data available is increasing quickly: weather stations, soil sensors, farm equipment, and Earth observation satellites produce a lot of information each day. This research is important because there's a need for better forecasting tools that can handle different types of data, help reduce uncertainty in planning for farming, and make using resources more efficient.

Research Objective: This study aims to look into how machine learning can be used to predict crop yields using data from different sources.It also seeks to find out the benefits, challenges, and future possibilities of using smart models in farming.

Crop yield comes from many factors working together, like the weather, the type of soil, the kind of crop planted, and the farming methods used. Traditional methods like linear regression or process-based crop simulation models need strong assumptions about how these relationships work or require careful calibration

for each individual field. In real situations, the connections between things are not always straight forward, and how one thing affects another can change based on what other factors are involved. For example, the same amount of rain can be good when plants are growing quickly but bad when it's time to collect the harvest. That's why it's especially important to use methods that can learn these relationships directly from the data.

Several groups of data are typically used to predict crop yield. Meteorological data include air temperature, precipitation, humidity, and solar radiation. Soil data include information about how much humus is present, how acidic or basic the soil is, its texture, and how much moisture it can hold. Remote sensing data gives information about plants, like the Normalized Difference Vegetation Index (NDVI), which shows how well plants are doing during the growing season. Agronomic data include when the seeds were planted, the fertilizers used, the crops grown in previous years, and the type of crop variety planted. Each of these sources has its own limits when used by itself. Weather data don't show what's really happening to the crops.Satellite images can be blocked by clouds and don't show the real condition of the soil. By putting them together, you get a better understanding of the situations in which a crop grows.

Modern open data services allow small agricultural businesses to easily access forecasts from multiple sources. Satellite missions like Sentinel-2, which are part of the European Copernicus program, capture images in different color ranges with a detail level as fine as 10 meters and return images every five days, allowing farmers to check the health of each field throughout the entire growing season. Meteorological reanalysis datasets, like ERA5, provide ongoing records of temperature, rainfall, and sunlight even in places that don't have close weather stations. Cloud platforms like Google Earth Engine let researchers work with big data collections without needing powerful computers at their own location. So now, the biggest challenge isn't getting the data, but figuring out which variables are important and making sure they match up in time and place.

Before modeling, the data must be prepared. This step involves dealing with missing data, getting rid of extreme values, aligning data from various sources to the same time and location standards, and scaling numerical data to a common range. Feature engineering is especially important because raw daily data is often grouped together to create indicators for different growth stages, like the total active temperatures, rainfall during flowering, or the highest NDVI value. How well you prepare usually has a bigger impact on the outcome than which algorithm you choose.

You can divide algorithms into several main groups. Linear regression models are easy to understand and explain, which is why they are commonly used as a starting point for comparing other models. Decision tree ensembles, like random forests, create many trees using different groups of data and combine their predictions, which helps them handle noise and avoid overfitting. Gradient boosting techniques, like XGBoost, create trees one after another, and each new tree helps fix the mistakes made by the earlier ones, which usually leads to very good results when working with table-like data. Support vector regression can work well even when there are not many data points. Deep learning methods have a special role: convolutional neural networks can work directly with satellite images, and recurrent networks can study weather and vegetation data over time during the season. Their weakness is that they require big sets of training data and are harder to understand.

The method you choose depends on how much data you have, how accurate the results need to be, the resources you can use for calculations, and whether you need to explain the outcome to an expert. For small farms with not much data, simpler models might work well.But bigger systems that predict for whole regions or countries might mix different algorithms to get better results. The quality of models is checked using things like the root mean square error (RMSE), the mean absolute error (MAE), and the coefficient of determination (R^2). To get a good estimate, cross-validation is used, and it's important that the data used for validation comes from different areas or different years than the data used for training. Otherwise, the model might look correct but could not work well in new situations.

An smart system for predicting crop yields has different steps that work together.First, it gathers information from weather stations, sensors, and satellites.Then, it cleans and prepares the data, and builds important features from it.After that, it uses this data to train and check the model.Once the model is ready, it creates a forecast.Finally, the results are shown to the users in an easy way. Weather predictions can change

throughout the season as more information comes in, which helps experts make timely changes to their plans for adding nutrients to the soil or giving plants water. Feature importance analysis is useful because it helps to see which factors have the biggest impact on crop yield in a certain area and during different stages of the plant's growth.

Yield prediction is very useful in areas where the moisture levels are not stable, like the southern part of Ukraine.In this region, dry periods greatly affect the outcome of the farming season. Early predictions made several weeks before the harvest allow companies to guess how much they will produce, plan how to use their storage and transportation resources, and decide on selling strategies and future agreements. In precision agriculture, forecasts can be made for different areas of a field, which helps apply fertilizers and water more carefully and lowers the cost of these materials. So, machine learning models are part of a larger system that helps make decisions, not just a tool used on its own.

Even though there are many benefits, using machine learning in farming still has several difficulties. Agricultural datasets are usually small, spread out unevenly in different areas, and often have missing data entries. A model that learns from data in one area might not work well in another place because the weather and ground conditions are different there. Extreme weather events happen very seldom, which is why they are not well shown in the data used to train models. Also, complicated models can be hard to understand, which makes agronomists and farm managers less confident in them. So, the outcomes from machine learning models should be seen as a tool to help make decisions, not as a substitute for the expertise of professionals.

So, machine learning provides a real way to bring together different kinds of farming data into one smooth system for making predictions. Right now, ensemble methods are considered the best choice for handling tabular data from multiple sources, and deep learning models show great potential for working with satellite images and time series data. More studies should aim to gather more data from different areas, make sure models work well across different regions, and create clear and easy-to-understand tools for predicting outcomes that farmers and agricultural businesses can use in their work.

References

1. Breiman L. Random Forests. Machine Learning. 2001. Vol. 45, No. 1. P. 5–32. DOI: https://doi.org/10.1023/A:1010933404324.

2. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. P. 785–794. DOI: https://doi.org/10.1145/2939672.2939785.

3. van Klompenburg T., Kassahun A., Catal C. Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture. 2020. Vol. 177. Article 105709. DOI: https://doi.org/10.1016/j.compag.2020.105709.

4. Khaki S., Wang L., Archontoulis S. V. A CNN-RNN Framework for Crop Yield Prediction. Frontiers in Plant Science. 2020. Vol. 10. Article 1750. DOI: https://doi.org/10.3389/fpls.2019.01750.

5. Gorelick N., Hancher M., Dixon M., Ilyushchenko S., Thau D., Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment. 2017. Vol. 202. P. 18–27. DOI: https://doi.org/10.1016/j.rse.2017.06.031.

6. Pedregosa F. et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011. Vol. 12. P. 2825–2830.