

Site-Specific Forecasting of Agricultural Crop Yield as a Technology and Service

Vladyslav Hnatiienko¹, Vitaliy Snytyuk¹

¹ Taras Shevchenko National University of Kyiv, Volodymyrs'ka str. 64/13, Kyiv, 01601, Ukraine

Abstract

The aim of the research is to improve the accuracy of crop yield forecasting by developing and applying data processing methods and neural network models. A yield forecasting technology is proposed that will include pattern recognition models for analyzing satellite images, data processing methods, and deep neural networks in combination with other artificial intelligence models. This technology is used to analyze the effectiveness and feasibility of agrotechnical measures, thereby supporting rational decision-making in agricultural production.

Keywords

Crops, yields, site-specific forecasting, forecasting models and methods.

1. Introduction

Digital agronomy is at the stage of active development, and farm owners are increasingly incorporating digital farm management into their strategies, which allows them to remotely monitor and control field work. Experts apply artificial intelligence and conduct research to deepen knowledge and develop effective digital agronomy technologies. In modern studies of the effectiveness of yield forecasting methods, a root mean square error (RMSE) of 10-15% of the average yield is achieved. Most models are used to predict the total yield of a field, without the ability to perform site-specific forecasting. The ones that allow to build detailed maps of predicted yields are generally tested on small samples, which makes it impossible to reliably assess their effectiveness.

The current challenge is to develop a forecasting technology that provides predictions for individual field areas. To solve this problem, it is necessary to analyze the data features, relationships, and degrees of influence of various agronomic indicators on the yield. The scientific value of the results lies in simplifying and optimizing future research by providing insights into which agronomic data to use and why. Furthermore, detailing forecasts to individual plots will open up new avenues for future research. From a practical point of view, forecasting will enable budget planning, risk analysis, and appropriate agronomic measures.

The expected results of the study will offer open-access innovative solutions, contributing to the development of digital agronomy.

2. Analysis of preliminary results

Numerous studies have been conducted to enhance the accuracy of crop yield forecasting. These efforts leverage a diverse array of information sources, such as plant genetic data, environmental data, and satellite imagery. To process and analyze this data, researchers employ a variety of models, ranging from traditional statistical approaches to advanced deep neural networks.

8th International Scientific and Practical Conference Applied Information Systems and Technologies in the Digital Society AISTDS'2024, October 01, 2024, Kyiv, Ukraine

* Corresponding author.

† These authors contributed equally.

✉ hnatiienko.vladyslav@knu.ua (V. Hnatiienko); snytyuk@knu.ua (V. Snytyuk)

ORCID 0009-0000-2678-5158 (V. Hnatiienko); 0000-0002-9954-8767 (V. Snytyuk)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

In particular, in [1], data on plant genotype, weather conditions, and soil indicators were used to predict the yield. A deep neural network was used for forecasting, achieving an RMSE of 11-12% of the average yield. However, the study only considers forecasting the total yield of an entire field without detailing its individual plots, which is a limitation for possible applications of this approach. A similar drawback applies to the study [2], which used the Random forest model. The RMSE values were 11.9% for wheat, 16.7% for corn across the United States, 13.9% for potatoes, and 5.8% for corn in the Northeast coastal region of the United States. The authors note that the chosen model is often overfitted, which can lead to difficulties in generalization. Another problem is low reliability: the model is effective on average, which allows analyzing the general features of big data, but there is a high probability of significant errors in individual forecasts.

Methods that use satellite images for forecasting are effective. For example, the authors of [3] predict potato yields based on identifying the relationship between the values of vegetation indices and yields with deviations of 5-9% from the average, but the elements of the training and test sets belonged to the same field, which makes it difficult to assess the generalizability of this approach. In [4], 5 fields form the training set, and the other 5 fields are used to test the model. The results indicate that the most effective model is Random forest, operating with RMSE values from 0.284 to 0.473 t/ha, or about 9-14% of the average sunflower yield. However, due to the small size of the test sample, it is difficult to assert the reliability of these results. According to the authors, the study had a good period for collecting information: 16 satellite images were collected on sunny days during the ripening period, providing favorable conditions for model training. Usually, due to the constant presence of clouds, only 3 to 7 images can be collected, which greatly complicates forecasting using such models.

Thus, we argue that the primary disadvantages of modern forecasting technologies are their insufficient accuracy and the significant dependence of results on weather conditions. Moreover, most studies focus on predicting total yield, while those that attempt site-specific forecasting often rely on highly limited samples during experimental validation, making it impossible to reliably evaluate the effectiveness of the proposed methods [5]. Additional challenges stem from the uncertainty regarding the feasibility of using different data sources: it remains unclear which factors exert the greatest influence on plant yields. Consequently, studies incorporate a wide variety of data, including plant genotypes, weather conditions, terrain variations, nitrogen and phosphorus soil content, and satellite imagery—data that originate from diverse sources and vary greatly in complexity and acquisition cost.

Agronomic experts and farm owners frequently encounter the challenge of insufficient accuracy and reliability in modern forecasting services. Forecast deviations from actual values range from 3–5% to as high as 30–40%, highlighting the inefficiency of current methods [6]. As a result, many abandon forecasting altogether, which hinders the advancement of agricultural production. This abandonment deprives developers of data analysis and decision-making technologies of adequate funding for further development and prevents farm owners from realizing the potential profits these technologies could have delivered.

3. Analysis of data sources and processing methods

3.1. Satellite images

Most of the data on plants and their maturation conditions are a set of constant values that are known at the beginning of the maturation period: genotype, sowing density, sowing date, field coordinates, etc. Such data are static, reflecting only the initial conditions, and forecasting based on them often leads to significant deviations from actual values.

For refined site-specific forecasting, satellite images are the most important source of data, as they are accumulated throughout the entire ripening period and allow tracking the dynamics of plant development, recording any deviations in time. When applying machine learning in yield prediction tasks, the parameters whose values are obtained from satellite images are assigned the highest

weighting coefficients [7, 8]. Table 1 presents the list of parameters and their feature importance scores for the LightGBM model. The importance scores are calculated during training: the parameters are used to build decision trees, and those that contribute to a greater reduction in error receive a higher importance score.

Table 1

Estimates of the importance of parameters for yield forecasting

Number	Feature	Importance score
1	GLI_mean	286
2	GLI_max	259
3	CLr_mean	254
4	NDVI_max	223
5	NDVI_mean	217
6	GLI_min	211
7	CLr_max	195
8	NDWI_max	189
9	CLg_max	174
10	CLr_min	156
11	NDVI_min	154
12	CLg_mean	147
13	NDWI_mean	146
14	NDWI_min	127
15	CLg_min	122
16	mid	61
17	Density	29
18	CAPE_180-0_mb_above_gnd_max	17
19	fungicide_58	12
20	mid_early	7
21	Temperature_2_m_elevation_corrected_max	5
...

Table 1 indicates:

- GLI, CLr, NDVI, CLg, NDWI are vegetation indices, the values of which are calculated for each field area based on satellite data;
- mid, mid_early are parts of the categorical variable hybrid, which indicates the plant hybrid of a given plot; the value 1 is set for the corresponding hybrid, and 0 for all others;
- Density is the density of planting on the field;
- CAPE_180-0_mb_above_gnd_max, Temperature_2_m_elevation_corrected_max, fungicide_58 - other parameters from the sets of meteorological data and data on plant characteristics in the field.

Monitoring services often provide data in the form of maps with vegetation index values, but the primary source of information is the intensity of reflected solar radiation in different spectral ranges, which is recorded for each field once a day. Most satellites have sensors that measure the reflected radiation for ten standard wavelengths belonging to the visible, near-infrared, and mid-infrared spectrums. These values are presented in Table 2.

A snapshot can be labeled $X_b^d = \{x_{1b}^d, x_{2b}^d, \dots, x_{nb}^d\}$, where

- n - the number of field areas, each of which is represented by a separate pixel in the image;

- $b \in \{B02, B03, \dots, B12\}$ - is the wavelength for which the intensity value is recorded;
- $d \in \{1, 2, \dots, T\}$ - the day on which the picture was taken, where T - is the number of days, which can vary depending on the conditions (usually $T = 100$).
- X_b^d - image for the wavelength b and day of observation d ;
- x_{ib}^d - intensity value for the area i for the wavelength b and day d .

Table 2
Designation of standard wavelengths

Number	Symbolic designation	Wavelength, nanometers
1	B02	492.4
2	B03	559.8
3	B04	665.2
4	B05	704.1
5	B06	740.5
6	B07	782.8
7	B08	832.9
8	B8A	864.7
9	B11	1613.7
10	B12	2202.4

To select a set of images suitable for further processing, we denote the set of distorted elements of the set X_b^d as X_{dist}^d (the correctness of these elements of each image is determined by an expert [5, 6]) and check the following criterion:

if the number of elements of the set of distorted values X_{dist}^d exceeds 10% of the total number of areas in the image: $|X_{dist}^d| > 0.1 \cdot n$, then the image X_b^d is considered unsuitable for further use.

Since during most observations a certain part of the field is covered by clouds, most images are classified as unsuitable for analysis, even if a significant part of the data in these images is correct. To solve this problem, cloud recognition can be applied with subsequent recovery of lost information by interpolation methods. Cloud recognition can be performed using object recognition technology based on multiprojection analysis [9].

Solving the problem of information loss will enable data representation in the form of time series and the application of deep learning methods for forecasting. At the current stage of the study, due to insufficient data, only minimum, average, and maximum values are utilized:

$$\min_{d \in D} \{x_{ib}^d\}, \forall i, b, \quad (1)$$

$$mean(\{x_{ib}^d \mid d \in D\}) = \frac{1}{|D|} \sum_{d \in D} x_{ib}^d, \forall i, b, \quad (2)$$

$$\max_{d \in D} \{x_{ib}^d\}, \forall i, b, \quad (3)$$

where D is the set of days for which the images are considered suitable.

This creates uncertainty that reduces the information content of the dataset. First, most of the information about the sequence of values is lost x by reducing it to only three general characteristics. Second, often the values (1) and (3) values often do not correspond to the actual values of the minimum and maximum intensity of reflected radiation, which is due to the removal of images for the days on which these key indicators could be recorded. In addition, the average value of (2) can differ significantly for a complete sequence. Thus, recovering complete time series X_b^d ,

$b \in \{B02, B03, \dots, B12\}$, $d \in \{1, 2, \dots, T\}$, will allow us to track the dynamics of changes in parameter values during the ripening period, which will significantly increase the amount of information about the state of plants.

3.2. Features of data display and sample balancing

During training, the weights of the neural network are adjusted to identify the most informative features and calculate the output values based on them. These features form vectors that can be used to analyze and process data. For example, when analyzing such vector representations in natural language processing tasks, it was found that for synonyms, antonyms, word pairs in singular and plural forms, and other semantically related words, the cosine distance of the corresponding vector representations is significantly smaller than for unrelated words [10]. A similar approach is used in image classification: vector representations of images are calculated to select key features, and then images are divided into classes. Often, this can be done even with a linear classifier due to the fact that the vector representations of images of one class are at a small distance and far enough removed from the images of other classes.

This approach can be applied to analyze data collected during the plant maturation period. For instance, it can be used to detect anomalies: vector representations of field areas containing distorted information (e.g., regions covered by clouds or parts of the field occupied by equipment instead of plants) are expected to deviate significantly from the majority of the data. Clustering techniques can identify typical patterns, and the centroids of these clusters can serve as reference points for detecting outliers.

Another potential application is the detection of fields with atypical data. For instance, if a dataset includes a field with plants of an uncommon hybrid or other distinct characteristics, its data points should be significantly distant from the majority of the training set. This concept can serve as the basis for developing a method to construct balanced and representative samples. To ensure representativeness, the training dataset should include a wide variety of plant species grown under diverse conditions, thereby reducing the likelihood that a field will appear atypical compared to the training data when the technology is implemented.

In trained models, vector representations typically capture the key features of the input data. For example, in image recognition, objects can still be identified even if parts are missing or deformed—the recognition process relies on the most significant features, while disregarding distortions in less critical ones [11, 12]. Therefore, it can be inferred that analyzing the sensitivity of vector representations to variations in input data can help identify the most important parameters: higher sensitivity indicates greater importance.

When constructing training sets with a sufficient amount of data, the sample is typically balanced to ensure an even distribution of data. This often involves maintaining an equal number of representatives from each original class, which, in the context of forecasting, translates to an equal number of observations corresponding to low, medium, and high yields (divided into an arbitrary number of ranges).

Additionally, balancing the input data is also crucial. For instance, if the training set is dominated by plants of a single hybrid, this can lead to overfitting and reduced prediction accuracy for plants of other hybrids. Such balancing can be achieved through the analysis of vector data representations, enabling the identification and adjustment of class imbalances [13].

3.3. Use of additional sources of information

The current forecasting method utilizes input data that includes satellite images, meteorological indicators, and supplementary information about the plants in the field, such as hybrid type, seeding density, and dates of chemical application. Additionally, the potential effectiveness of incorporating other parameters—such as the elevation of field sections above sea level, section coordinates, and

sowing dates—should be thoroughly analyzed to assess their contribution to improving forecasting accuracy.

Studies on yield forecasting also use data on plant yields in previous years [14] and data on patterns of climate change in previous years [15]. Incorporating information about predecessor crops, along with meteorological data and satellite imagery from previous years, can enhance forecasting accuracy.

4. Approaches to yield forecasting

4.1. Uniformity of ripening

An essential parameter for yield prediction is the uniformity of plant maturation within a field. Harvesting combines are calibrated to collect plants at a specific maturity stage, typically targeting the stage that represents the majority of plants. However, yield losses occur when plants that are either over-mature or under-mature are not harvested under optimal conditions. Since combines are generally adjusted according to predefined standards, incorporating the distribution of plant maturity across the field into the input data could improve prediction accuracy. Agronomic experts often rely on the NDVI vegetation index to evaluate plant maturity [16]. Consequently, the formation of this parameter can be achieved using the following algorithm:

1. Dividing NDVI values into ranges with the help of expert opinion;
2. Determining the distribution of the area of the field parts in these ranges;
3. A categorical variable is created based on the distribution, assigning a value of 1 to the category with the largest area and 0 to all others.

Instead of using a categorical variable with possible values {0, 1}, a set of variables can be employed to represent the full distribution—specifically, the percentage of the field area falling within each range of NDVI values.

While NDVI is primarily used during the final stages of ripening, it can also be utilized for early-stage forecasting by predicting future NDVI values based on the dynamics of its changes over time. This approach enables more accurate predictions of plant maturity and yield at earlier stages.

4.2. Using deep learning models for time series forecasting

When data is presented as time series, the most effective artificial intelligence methods for forecasting are recurrent neural networks (RNNs) and transformers.

Recurrent neural networks, including architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), are widely used for efficiently processing sequential data while preserving the context of previous observations. These models are particularly effective for forecasting based on sequences of observations collected over multiple years [17, 18].

Transformers, a more recent development in time-series forecasting, are gaining traction due to their flexibility and ability to process sequences in parallel. While their application in yield forecasting remains an emerging field, transformers have been increasingly utilized in recent research [19, 20].

4.3. Using separate models for different sources of information

In previous studies, high accuracy was achieved by combining models, specifically the computer vision model U-Net and the ensemble model LightGBM. U-Net was employed to forecast yield based on satellite images by segmenting fields into nine performance categories, ranging from the lowest to the highest. LightGBM was then used to refine these predictions by incorporating additional data, such as meteorological indicators and field-specific plant characteristics. This approach effectively distributed tasks, leveraging the strengths of each model for their most suitable functions.

The effectiveness of this approach warrants further investigation with alternative models and different methods of dataset construction. For segmentation tasks, modern models like YOLO [21] and SAM [22] could potentially outperform U-Net, offering improved accuracy and efficiency.

Currently, each forecasting iteration processes approximately 1 hectare of a field. The U-Net model is designed to perform predictions separately for each segment of the data, after which the results are aggregated. While this approach allows for dataset augmentation and facilitates model training even with limited data, it restricts insights into the overall field condition.

Given that weather conditions and plant maturation data are already incorporated in the LightGBM forecasting stage, a new model is required to analyze satellite images of the entire field and extract key features. This could involve developing a dedicated artificial intelligence model to either transform the data into usable formats or independently derive general characteristics deemed important by agronomic experts, such as the average NDVI value across the entire image, the range (difference between maximum and minimum values) of specific vegetation indices, and similar metrics.

One limitation of this approach is that the models are trained separately. After training they are simply combined: using the first model, M_1 (U-Net), the yield is predicted Y_1 based on satellite images. X_1 :

$$Y_1 = M_1(X_1) \quad (4)$$

and when forecasting with the second model M_2 (LightGBM), the output values of the first model are used together with additional data X_2 to generate the final yield forecast:

$$Y = M_2(Y_1, X_2) \quad (5)$$

Each model is trained independently. The process is shown schematically in Fig. 1.

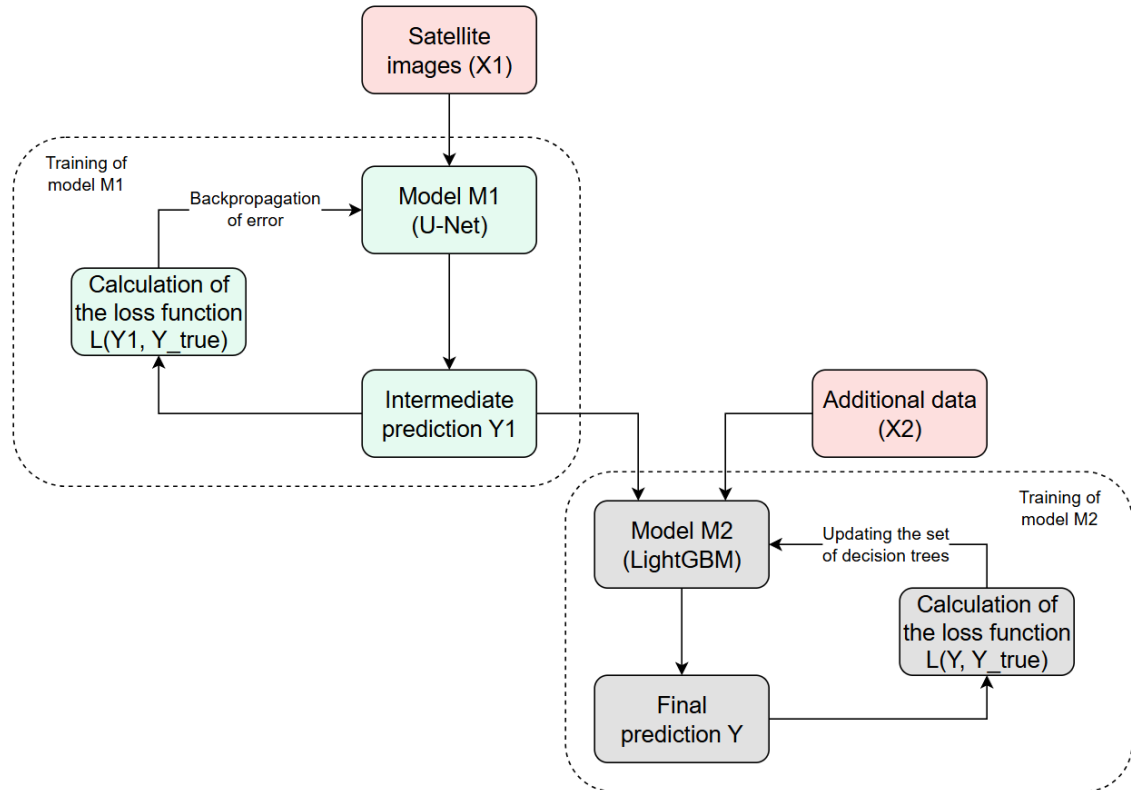


Figure 1: Schematic of the independent model learning process

Since LightGBM does not allow specifying a differentiated loss function, an alternative model, such as a multilayer perceptron (MLP), can be used instead. This substitution enables the use of a customized loss function tailored to the specific requirements of the task.

The learning process can also be adapted to allow simultaneous error propagation for both models. This means that the training pipeline can be designed to integrate the outputs of both models, ensuring that updates to the parameters of one model account for the influence of the other, thereby improving overall synergy and performance M_1 and M_2 and facilitate their simultaneous learning and effective interaction. For this purpose, a common loss function is calculated:

$$Loss = L(Y, Y_{true}) = L(M_2(M_1(X_1), X_2), Y_{true}), \quad (6)$$

where L - is a loss function (e.g., root mean square error), Y - is the final forecast, and Y_{true} - is the set of actual values.

For example, learning by gradient descent will work like this:

1. For model M_2 the gradient with respect to its weights θ_2 :

$$\nabla_{\theta_2} Loss = \frac{\partial L}{\partial Y} \cdot \frac{\partial Y}{\partial \theta_2} \quad (7)$$

2. For model M_1 the gradient with respect to its weights θ_1 :

$$\nabla_{\theta_1} Loss = \frac{\partial L}{\partial Y} \cdot \frac{\partial Y}{\partial Y_1} \cdot \frac{\partial Y_1}{\partial \theta_1} \quad (8)$$

3. Updating the weighting coefficients M_1 and M_2 :

$$\begin{aligned} \theta_1 &\leftarrow \theta_1 - \eta \nabla_{\theta_1} Loss \\ \theta_2 &\leftarrow \theta_2 - \eta \nabla_{\theta_2} Loss \end{aligned} \quad (9)$$

where η - is the learning rate.

Thus, model M_1 is trained to generate intermediate outputs Y_1 that maximize the accuracy of model M_2 's predictions. In turn, model M_2 is trained to optimally utilize the intermediate outputs Y to minimize the deviation of the final forecast Y from the actual values Y_{true} . The backward propagation of errors affects the weights of both models, enabling joint learning where each model is optimized to improve the final forecast Y . This process is illustrated schematically in Fig. 2.

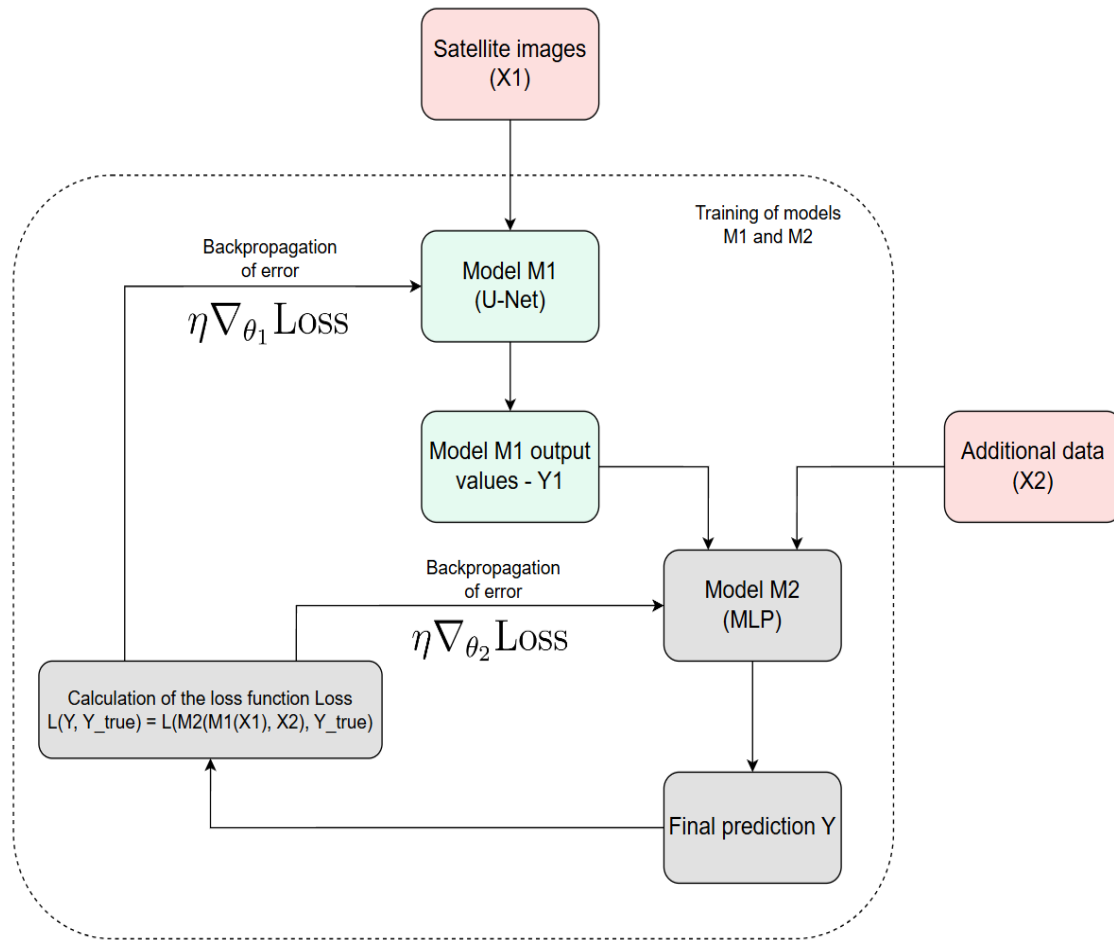


Figure 2: Diagram of the simultaneous model training process

Alternatively, a unified model capable of analyzing all types of information simultaneously might prove even more effective. For instance, study [23] introduced a transformer-based model designed to integrate various information sources, including both static and dynamic indicators. This approach enabled the model to outperform commonly used methods, such as Random Forest and XGBoost, in terms of forecasting accuracy, making it a promising direction for further research and application in yield prediction.

5. Application prospects

5.1. Differentiated application of chemicals

One of the primary methods for preparing plants for harvest is desiccation, an artificial drying process that equalizes moisture levels in the field and accelerates ripening. This practice addresses the problem of uneven ripening, which can otherwise lead to significant harvest losses. However, desiccation is not always economically justified, as the costs of the substances and their application may exceed the value of the yield saved. Forecasting technology can play a key role in evaluating the feasibility of desiccation [24].

Assume the existence of a highly accurate yield prediction model. This model can be retrained by incorporating a binary variable, *des*, into the training set, where 1 indicates that desiccation was performed, and 0 indicates it was not.

The impact of desiccation varies depending on the conditions: in some cases, it results in a substantial yield increase, while in others, the improvement is negligible. With a sufficiently large dataset, these effects will be reflected in the data. Once trained, the model can be used to predict the potential benefits of desiccation, enabling informed decision-making for its application:

1. Yield forecast without desiccation: $Y_{no_des} = M(X, des = 0)$
2. Yield forecast with desiccation: $Y_{with_des} = M(X, des = 1)$
3. Increase in yields: $\Delta Y = Y_{with_des} - Y_{no_des}$

The value of the crop and the cost of purchasing desiccants and spraying may vary, but agronomic experts can get accurate information about them. The only uncertainty is the potential yield of the field. Thus, if forecasting accuracy is high, the potential benefit of desiccation can be calculated with high accuracy:

$$Profit = (\Delta Y \cdot cost_{yield}) - cost_{des}, \quad (10)$$

where $cost_{yield}$ - is the cost per unit of harvest, $cost_{des}$ - cost of desiccants and their application.

To further enhance the training dataset, the model can be improved by incorporating detailed information about desiccation application methods. This involves adding data on two key aspects:

1. Selective Spraying: Desiccation can be applied only to specific areas where it is necessary.
2. Variable Intensity: Different chemical application intensities can be used for different areas, tailored to the needs of the plants in each plot.

With a sufficiently large dataset, a model can be trained to predict yield improvements based on the method of application. To achieve this, the dataset should include an additional variable, $des_{intensity} \in [0, 2]$, representing the intensity of substance application for each plot (e.g., 0 liters for no application, and 2 liters for the standard maximum intensity).

Once the model is trained, the most effective desiccation strategy can be determined using the following algorithm:

1. generation of application variants: since each field consists of numerous plots, and an arbitrary amount of substance can be applied to each within the standard range, it is advisable to limit the generated variants using heuristics based on accepted desiccation practices;
2. limitation of the generated options: even with the previous limitation, the set of options may be too large, so some rules need to be applied, such as discarding similar options (which options are considered similar should be determined separately) and search algorithms to minimize computation;
3. forecasting yields for each of the application methods. Formally, for each variant V_i containing information about the intensity of desiccation $des_{intensity}$ on each field plot;
4. selection of the best option: the option is selected V^* for which the predicted yield is maximized: $V^* = arg \max_{V_i} Y_i$

This approach enables optimized desiccation by incorporating selective spraying and variable application rates, thereby ensuring more efficient resource use.

This approach can be extended to apply to other agronomic procedures, such as the use of pesticides and other chemicals.

5.2. Prediction in the early stages of maturation

If the forecasting technology is successfully developed, it can be extended to smaller, limited datasets. By considering the input data as a time series X^d , $d \in \{1, 2, \dots, T\}$, the sample size can be reduced by decreasing the value of T , thereby training the model on shorter time periods. Although this may reduce accuracy, it enables yield predictions at earlier stages of plant development.

One key advantage of early-stage forecasting is the ability to promptly detect problems. For example, a low predicted yield in a specific area may indicate the presence of diseases or pests, allowing agronomic experts to address potential issues proactively and prevent significant yield losses. This application can be further enhanced with remote monitoring and plant health analysis techniques, such as automated lesion detection [25].

Overall, early problem detection and early planning capabilities can significantly improve decision-making and optimize agricultural production processes.

6. Conclusions

The site-specific yield forecasting technology developed through these methods and approaches has the potential to significantly enhance agricultural efficiency. The proposed data processing techniques are designed to effectively leverage available data, even under challenging conditions such as heavy cloud cover in satellite images, enabling accurate forecasts across diverse scenarios.

The method emphasizes the formation of balanced and representative samples, ensuring that forecasting models can generalize effectively across varying conditions and plant types. By integrating advanced deep learning models and their combination methods, the accuracy of forecasts is expected to improve significantly.

Implementing this technology will enable better resource management by adapting field care to specific conditions, thereby reducing the risks of yield losses due to uneven maturation, pests, or diseases. Its successful application could facilitate innovations such as optimized, differentiated fertilization and precise early-stage yield forecasting. Practical implementation is anticipated to validate the effectiveness of these technologies, paving the way for broader applications across different regions and crop types.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] Khaki, S. & Wang, L. Crop yield prediction using deep neural networks (2019). *Front. Plant Sci.* 10, 621.
- [2] Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, Butler EE, Timlin DJ, Shim KM, Gerber JS, Reddy VR, Kim SH. Random Forests for Global and Regional Crop Yield Predictions. *PLoS One.* 2016 Jun 3;11(6).
- [3] Al-Gaadi KA, Hassaballa AA, Tola E, Kayad AG, Madugundu R, Alblewi B, Assiri F. Prediction of Potato Crop Yield Using Precision Agriculture Techniques. *PLoS One.* 2016 Sep.
- [4] Amankulova, K., Farmonov, N., Mukhtorov, U. & Mucsi, L. Sunflower crop yield prediction by advanced statistical modeling using satellite-derived vegetation indices and crop phenology. *Geocarto Int.* 38, 1.
- [5] Hnatiienko, H., Snytyuk, V., Tmienova, N., Voloshyn, O. Application of expert decision-making technologies for fair evaluation in testing problems // Selected Papers of the XX International Scientific and Practical Conference "Information Technologies and Security" (ITS 2020), Kyiv, Ukraine, December 10, 2020 / CEUR Workshop Proceedings, 2021, 2859, pp. 46-60.
- [6] Hnatiienko H., Tmienova N., Kruglov A. (2021) Methods for Determining the Group Ranking of Alternatives for Incomplete Expert Rankings. In: Shkarlet S., Morozov A., Palagin A. (eds) *Mathematical Modeling and Simulation of Systems (MODS'2020)*. MODS 2020. *Advances in Intelligent Systems and Computing*, vol 1265. Springer, Cham. https://doi.org/10.1007/978-3-030-58124-4_21. Pp. 217-226.
- [7] Voloshin, A.F., Gnatiienko, G.N., Drobot, E.V. A Method of Indirect Determination of Intervals of Weight Coefficients of Parameters for Metricized Relations Between Objects // *Journal of Automation and Information Sciences*, 2003, 35(1-4).
- [8] Hnatiienko H., Snytyuk V. A posteriori determination of expert competence under uncertainty / Selected Papers of the XIX International Scientific and Practical Conference "Information Technologies and Security" (ITS 2019), pp. 82-99 (2019).
- [9] Stepan Bilan, Vladyslav Hnatiienko, Oleh Ilarionov and Hanna Krasovska. The Technology of Selection and Recognition of Information Objects on Images of the Earth's Surface Based on Multi-Projection Analysis / *_CEUR Workshop Proceedings,_ Volume 3538, Pages 23-32, 2023 //*

- Selected Papers of the III International Scientific Symposium "Intelligent Solutions" (IntSol-2023). Symposium Proceedings Kyiv - Uzhhorod, Ukraine, September 27-28, 2023.
- [10] Levy, O., & Goldberg, Y. Linguistic regularities in sparse and explicit word representations. In Proceedings of the eighteenth conference on computational natural language learning. 2014, June. - pp. 171-180.
- [11] Hossain, D., Nilwong, S., Tran, D., Capi, G. Recognition of Partially Occluded Objects: A Faster R-CNN Approach. Journal of Advanced Mechanical Design Systems and Manufacturing. 2018, October.
- [12] Rim, P., Saha, S., & Rim, M. CaltechFN: Distorted and Partially Occluded Digits. ACCV Workshop. 2022.
- [13] Antonevych, M., Tmienova, N., Snytyuk, V. Models and evolutionary methods for objects and systems clustering. CEUR Workshop Proceedings, 2021, 3018, pp. 37-47.
- [14] Khaki, S., & Wang, L. Crop Yield Prediction Using Deep Neural Networks. Frontiers in Plant Science. 2019, May. <https://doi.org/10.3389/fpls.2019.00621>.
- [15] Iizumi, T., Shin, Y., Kim, W., Kim, M., & Choi, J. Global Crop Yield Forecasting Using Seasonal Climate Information from a Multi-Model Ensemble. Climate Services, 11, 13-23. 2018.
- [16] Hnatiienko, H., Domrachev, V., Saiko, V. Monitoring the condition of agricultural crops based on the use of clustering methods // 15th International Conference Monitoring of Geological Processes and Ecological Condition of the Environment, Monitoring 2021, Nov 2021, Volume 2021, Pp.1-5, DOI: <https://doi.org/10.3997/2214-4609.20215K2049>.
- [17] Khaki, S., Wang, L., & Archontoulis, S. V. A CNN-RNN Framework for Crop Yield Prediction. Frontiers in Plant Science. 2020, January. <https://doi.org/10.3389/fpls.2019.01750>.
- [18] Elavarasan, D., & Vincent, P. M. D. Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications. IEEE Access. 2020, May. <https://doi.org/10.1109/ACCESS.2020.2992480>.
- [19] Bi, L., Wally, O., Hu, G., Tenuta, A. U., Kandel, Y. R., & Mueller, D. S. A Transformer-Based Approach for Early Prediction of Soybean Yield Using Time-Series Images. Frontiers in Plant Science. 2023, June. <https://doi.org/10.3389/fpls.2023.1173036>.
- [20] Lin, F., Crawford, S., Guillot, K., Zhang, Y., Chen, Y., Yuan, X., Chen, L., Williams, S., Minvielle, R., Xiao, X., Gholson, D., Ashwell, N., Setiyono, T., Tubana, B., Peng, L., Bayoumi, M., & Tzeng, N.-F. MMST-ViT: Climate Change-aware Crop Yield Prediction via Multi-Modal Spatial-Temporal Vision Transformer. arXiv. 2023, September. <https://doi.org/10.48550/arXiv.2309.09067>.
- [21] Sapkota, R., Du, X., Churuvija, M., et al. Comprehensive Performance Evaluation of YOLO11, YOLOv10, YOLOv9 and YOLOv8 on Detecting and Counting Fruitlet in Complex Orchard Environments. Preprint. 2024, July. <https://doi.org/10.4850/arXiv.2407.12040>.
- [22] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Rolland, C., Gustafson, L., Dollár, P., & Girshick, R. Segment Anything. Meta AI Research, FAIR. arXiv. 2023, April. <https://doi.org/10.48550/arXiv.2304.02643>.
- [23] Liu, Q., Dou, F., Yang, M., Amdework, E., Wang, G., & Bi, J. Customized Positional Encoding to Combine Static and Time-varying Data in Robust Representation Learning for Crop Yield Prediction. Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23), Special Track on AI for Good. 2023.
- [24] Tmienova, N., Snytyuk, V. Method of Deformed Stars for Global Optimization. 2020 IEEE 2nd International Conference on System Analysis and Intelligent Computing, SAIC 2020, 2020, 9239208.
- [25] Bilan, S., Gaina, G., Vlasenko, O., Sutyk, O., & Roiko, Y. Methods for Automatically Determining the Level of Disease Damage to Plant Leaves from Their Raster Image. CEUR Workshop Proceedings, 2023, 3624, pp. 106-115.