Progress in Agricultural Engineering Sciences 21 (2025) 1, 155–167
DOI: 10.1556/446.2025.00202

1828
AKADÉMIAI KIADÓ

Predicting maize growth and biomass: Integrating gradient boosted trees with sentinel images and IoT

ANIKÓ NYÉKI[1], BÁLINT DARÓCZY[2], MIKLÓS NEMÉNYI[1], BÁLINT AMBRUS[1], GERGELY TESCHNER[1] and TAREK ALAHMAD[1*]

[1] Department of Biosystems Engineering and Precision Technology, Albert Kázmér Mosonmagyaróvár Faculty of Agricultural and Food Sciences, Széchenyi István University, H-9200 Mosonmagyaróvár, Hungary

[2] Artificial Intelligence National Laboratory, Institute for Computer Science and Control (SZTAKI), Hungarian Research Network (HUN-REN), Budapest, Hungary

ORIGINAL RESEARCH PAPER

Received: April 7, 2025 • Accepted: May 19, 2025
Published online: June 12, 2025

ABSTRACT

Agricultural big data and high-performance computing have significantly improved crop yield modeling. Maize growth dynamics and yield prediction are crucial for sustainable agriculture. This study introduces an advanced modeling approach utilizing Gradient Boosted Decision Trees (GBDT) combined with a feature selection strategy to predict maize biomass production. A dataset of 200 unique maize plants was observed throughout the vegetation season. Our approach integrates manual measurements, meteorological data, and vegetation indices along with Internet of Things (IoT) field sensors to perform spatio-temporal analysis. Results indicate that maize stalk thickness and height are the most reliable predictors of biomass yield, while environmental variables show minimal impact. The most effective model, period-dependent GBDT, demonstrated superior predictive performance, achieving an average error of 4.39 mm in plant growth predictions. Notably, stalk thickness and height can be estimated six weeks before harvest, while biomass yield two weeks before harvest. This research underscores the potential of machine learning and remote sensing to enhance precision agriculture decision-making.

KEYWORDS

maize growth modeling, Internet of Things (IoT), spatio-temporal data analysis, machine learning, satellite sensing

* Corresponding author. E-mail: alahmad.tarek@sze.hu

1. INTRODUCTION

Maize is one of the world's most significant crops, used for a wide range of purposes including food, fodder, and the production of ethanol (Ranum et al., 2014). The Internet of Things (IoT) has significant implications for agriculture and sustainable crop production, including the ability to anticipate and prepare for potential disturbances that may happen in remote locations, such as insect pest invasions and monitoring soil nutrients, water dynamics, and pest management. Other advantages of utilizing IoT technology in agricultural systems include the ability to predict yield, conserve energy and money, and construct indoor vertical farming systems by utilizing information from an established IoT station (Ali et al., 2023). The Internet of Things (IoT) has the potential to become a key enabler for realizing the vision of smart agriculture (Sinha et al., 2019). Artificial Intelligence (AI) provides the framework and tools to go beyond trivial real-time decision and automation use cases for IoT (Harmani et al., 2022). New sensor networks have transformed parameter measurement from a time-consuming, costly procedure to an efficient "Big Data" gathering process, recording soil type, chemical and physical conditions, and moisture content—all of which are important predictors of crop yield quality. Throughout the growth season, these sensors use low-power gadgets to continuously monitor soil, crops, and environmental data. Crop yield prediction using soil, environmental, and crop parameters is critical for global, regional, and field-level agricultural decision-making. Precision agriculture optimizes yields, quality, and environmental impact by combining sensing, management, and variable rate technology. Cumulative field effects are depicted in simulation models. Using IoT databases, AI, and big data precision technologies allows for accurate yield estimates and in-depth research, which is critical for agricultural developments and environmental evaluation (Alahmad et al., 2023; Nyéki et al., 2020; Zelenák et al., 2022). Sensing has led to the adoption of technology in precision agriculture, and it also enables improved efficiency in agricultural production and practices (Viscarra Rossel and Bouma, 2016; Vuran et al., 2018).

Precision crop yield modeling relies significantly on remote sensing. Satellite remote sensing has been considered one of the most effective technologies for monitoring crop growth conditions and predicting yields in recent decades (Wang et al., 2019). Crop growth monitoring models that integrate various growth-related indicators give significant and precise crop growth and yield predictions (Ines et al., 2013; Xie et al., 2017). Khan et al. (2023), in their study, compared machine learning models for county-level maize production prediction and evaluated the potential of in-season yield estimation in their study. Nine models for maize production prediction were tested using remote sensing data from Google Earth Engine spanning 15 years, with partial least square regression (PLSR) outperforming the others. Support vector regression (SVR) performed best for in-season predictions, with an R^2 of 0.875. The findings indicate that machine learning models based on remote sensing data may accurately estimate both in-season and end-of-season maize yields, providing insights crucial to food security and climate change decision-making.

Sentinel sensors' red and near-infrared spectral bands are frequently used to construct the normalized difference vegetation index (NDVI), which is a reliable tool for crop yield prediction and defining the geographical and temporal aspects of plant vigor and crop biomass production. Crop growth monitoring utilizing sensor-based and IoT technology enables growth-related vegetation time monitoring as well as more accurate crop growth and production predictions (Ren et al., 2008). In the absence of spatial data on the distribution of key model inputs such as crop cultivars and management (e.g., irrigation, planting, fertilization, tillage, weed control), researchers frequently make broad assumptions across large geographic regions that may or may not reflect individual producer decision-making in the context of economic opportunities and policy incentives (Folberth et al., 2019).

Yield prediction in precision, site-specific crop production is nowadays an essential research field, as it has an equally important point of reference for farm management during planning, agrotechnological intervention, and preharvest processes (Nyéki et al., 2013). Farm and in-situ observations, together with existing databases provide the opportunity to not only predict yields using "simpler" statistical methods or decision support systems that are already used as an extension, but they also enable the potential for Machine Learning (ML) to be used. In contrast to process-based methodologies, machine learning models utilize a data-driven approach to approximate the complicated link between input (genotype and environment) and output (crop yield) without depending on incomplete and occasionally inaccurate human knowledge of agricultural science. Several machine learning methods, including multiple linear regression (Ramesh and Vardhan, 2015), partial least squares regression (Foster et al., 2017), random forest regression (Jeong et al., 2016; Sakamoto, 2020), convolutional neural networks (Sun et al., 2019), and deep neural networks (Bhojani and Bhatt, 2020; Khaki and Wang, 2019; Wang et al., 2020), have been successfully implemented to yield remarkable prediction accuracy. ML techniques can provide more accurate predictions with big datasets compared to crop growth simulation models. The Gradient Boosting Decision Tree (GBDT) is a learning method that combines multiple decision trees to generate a prediction model in gradient boosting (Friedman, 2001), but its application in maize growth modeling is less used. Wei et al. (2019) used the GBDT regression that expresses the nonlinear relationship between soil moisture (from in situ measurements) and satellite-based soil moisture indices. The gradient boosting machine (GBM) is a very effective state-of-the-art ML method (Chen and Guestrin, 2016; Friedman, 2001). XGBoost has been used to determine rice yields using weather data, cultivation data, and location information (Maeda et al., 2018).

Understanding how phenotypes are determined by genotype, environment, and their interactions is still one of biology's challenging tasks. Extensive research has been conducted on the relationship between genetics, weather, soil, and management variables and crop yield (Agnolucci et al., 2020; Elahi et al., 2022; Khaki et al., 2020; Lesk et al., 2021; MacCarthy et al., 2017; Verón et al., 2015).

Chang et al. (2023) describes a new crop model that combines process-based and data-driven approaches for predicting maize yield in the US Corn Belt. It obtains 7.16% relative error in estimating 2020 yields using data from 1981 to 2020, providing explainable results. It detects genotype-environment interactions and assists farmers in improving seed selection for increased yields, which is critical for food security in the face of climate change. Tracking daily biomass accumulation during the maize growing season could be a key factor and can be used to estimate the final grain yield.

Abimbola et al. (2022) used in-situ data from several maize farming treatments to calibrate and validate the Hybrid-Maize crop model. Using sensors in North Platte, NE, researchers observed hourly and daily weather and crop statistics from 2019 to 2020. Sensitivity analysis was used to identify key parameters affecting yield, which were then adjusted using a unique approach. Calibration revealed strong relationships between the water received and the parameters, significantly improving yield forecasts. This calibrated model has the potential to improve crop management decision-making tools.

The primary objective of this study is to develop the machine learning framework for maize yield prediction by:

- Validating the performance of GBDT models in maize biomass predicting.
- Integrating multi-source data (field sensors, meteorological data, Sentinel imagery) to enhance predictive accuracy.
- Assessing the relative importance of environmental and morphological factors in maize growth modeling.
- Optimizing early biomass prediction, allowing for informed agronomic interventions before harvest.

Combining IoT sensor data with remote sensing vegetation indices, this study seeks to address key limitations of conventional crop predicting models and contribute to the development of data-driven decision support systems for precision agriculture.

2. MATERIALS AND METHODS

2.1. Study area and maize plant measurements

The study area encompassed Mosonmagyaróvár, Hungary. Field experiments were conducted on a 6-ha experimental field (Fig. 1) belonging to Széchenyi István University, Hungary [N47°54′20.00″; E17°15′10.00″]. The region experiences a continental climate, with an average annual precipitation of 600 mm. The field features flat terrain, fertile alluvial soil, and favorable natural conditions for planting maize (from mid-April to mid-October) and other field crops such as winter wheat, summer barley, soybeans, etc. The main growing stages for maize in

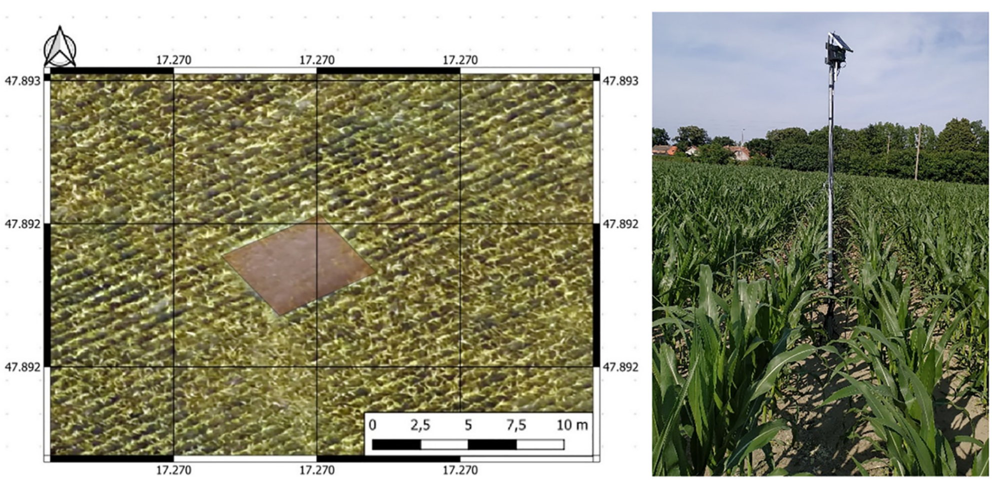

Fig. 1. Location of experimental plot and the maize field with installed sensor station (2019, maize crop)

Hungary are the seeding stage (mid-April to early May), the jointing stage (from mid-May to early June), the heading stage (from late June to late July), the milking stage (from late July to mid-August), and the mature stage (from mid-September to early October). Field data were collected from July to September 2019 every two weeks (July 4 and 18; August 1, 15, and 30; and September 13) from sampling plots measuring 30 × 30 m to spatially approximate the resolution of Sentinel images. In each plot, 20 maize plants were selected along the rows (10 per row) to measure the plant heights and straw. In total, measurements for 200 maize plants of one hybrid were acquired. Stem thickness was measured using a digital caliper (precision ±0.01 mm), and plant height was measured using a measuring tape (precision ±1 cm) from the soil surface to the apex of the plant. Each measurement was performed manually and repeated twice to ensure consistency.

2.2. Satellite observations

To monitor the growth stages of maize, we utilized images from Sentinel satellites, downloaded via Sentinel Hub, using Bottom-of-Atmoshpere (BOA) reflectance values. We selected images with less than 5% cloud cover and those acquired close to the dates of maize plant measurements. The satellite images of acceptable quality used in this study were primarily from June to September.

2.3. Sensor data

Among the 35 meteorological features observed, which remained constant across treatment units, we utilized five key types for each month from April to October, aligning with the maize vegetation season. These included the sum of precipitation, average temperature, average relative humidity, rate of vaporization, and the aridity index. Monthly averages were calculated based on a daily database that recorded these five parameters every month. The meteorological station of the university, located 0.4 km from the observed field, provided the data.

2.4. Machine learning models/statistical methods

Gradient Boosted Trees (GBT), Logistic Regression and Random Forest models were applied in this study. The value of each trait was saved and processed to find out the relationship with respect to maize yield. With machine learning, we used different approaches for yield-based binary classification and to find the relationship between yield and variables: Mean Absolute Error (MAE), ROC-AUC scores to evaluate classification performance (biomass above and below 236 g), and feature importance analysis to determine the most influential parameters, variables.

3. RESULTS

3.1. Manual maize plant measurements

The results showed that the mean yield (cob and grain) is 217.32 g (Fig. 2), with a standard deviation of 42.36 g, indicating a moderate level of variability in yield across the dataset (Table 1). The range from the minimum (60.53 g) to the maximum (326.91 g) yield also highlights

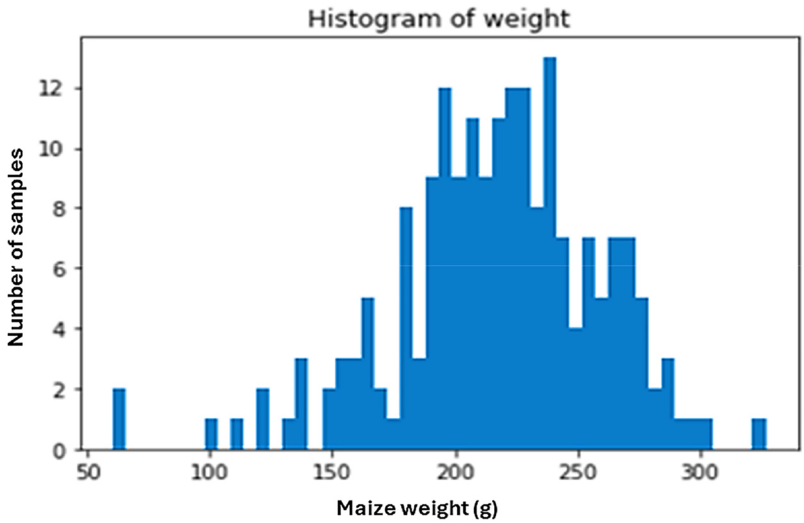

Fig. 2. Histogram of maize weight frequencies (gr)

Table 1. Statistical summary of maize harvest data

	Mean	std	Min	25th percentile	Median	75th percentile	Max
Number	100.50	57.88	1.00	50.75	100.50	150.25	200.00
Yield (g)	217.32	42.36	60.53	196.25	220.22	243.06	326.91
Cob weight (g)	34.81	7.22	15.95	30.30	34.65	39.38	56.00
Grain yield (g)	177.03	47.38	0.00	160.79	184.04	204.46	270.91
Length (cm)	17.46	1.79	11.00	16.3250	17.600	18.80	21.00

significant differences in performance. The average weight of a cob is 34.81 g. The relatively low standard deviation of 7.22 g compared to its mean value suggests that the cob weights are fairly consistent across the dataset. With a mean of 177.03 g and a standard deviation of 47.38 g, the grain yield shows a higher level of variability than cob weight. This suggests that factors influencing grain yield might be more varied or less controlled compared to those affecting cob weight. The length and diameter of the cobs show less variability (standard deviations of 1.79 and 14.94, respectively) compared to yield metrics. The length has a mean of 17.45 with a tight range (11.00–21.00), indicating consistency in cob size. The maximum value in the diameter column (231.90 mm) is notably high compared to the rest of the data. This could be an outlier or a data entry error, as it significantly deviates from the mean diameter of 25.70 mm.

3.2. Modeling results

The GBDT (Gradient Boosted Decision Trees) model demonstrated high performance accuracy. Specifically, the goodness-of-fit for our best model was notably higher when considering free cover. This study focused on examining the accuracy of predicting the final maize biomass weight, using measurements of length and stem thickness. The following sections describe the data set processing in detail.

The biomass yield measurements indicated that neither the moisture data nor the Sentinel satellite data significantly influenced the final maize biomass prediction. The ROC AUC value reached up to 0.779 when changes in thickness and length were combined, provided that the measurements were taken close to the time of final yield determination. The yield's goodness-of-fit was categorized into two groups—above and below 236 g—reflecting the weight distribution. In these instances, variations in soil and satellite measurements did not influence the outcomes, hence showing no predictive effect. The redundancy of the statement on biomass yield measurements and predictive effects of moisture and Sentinel data has been removed for clarity. The Sentinel and soil moisture data exhibited extremely low standard deviation values, suggesting their limited utility, and underscoring the need for additional measurements within the studied period. The most effective model was the period-dependent Gradient Boosted Trees (GBT) or Logistic Regression, which was trained with a quasi-diagonal Hessian. Furthermore, the polynomial kernel-based maximum edge did not significantly underperform compared to the best model.

The relationship between stem thickness and length was analyzed both during the vegetation season and at harvest time. The Gradient Boosted Trees (GBT) models exhibited superior performance, with an average error of 4.39 and 4.4 mm, respectively. These models effectively captured the growth pattern (length-thickness growth) predominantly from the seasonal measurements. While environmental factors were considered, their predictive influence was found to be minimal. Notably, the inclusion of environmental attributes led to a significant decrease in the maximum error observed. This finding prompted further investigation into the impact of in-season environmental variable changes on the models. The analysis of the measurements and database indicated that the variance within individual zones is non-Markovian, and these zonal characteristics are crucial in determining yield and growth. It is feasible to predict maize plant yield, stem thickness, and length early in the growing season. Specifically, the length and thickness of maize plants can be estimated accurately one and a half months before harvest, while the yield (biomass) can be predicted half a month prior to harvest.

3.3. Maize stem diameter and steam length prediction

In our study, we meticulously examined the performance of various machine learning models in predicting maize stem diameter, an essential factor in biomass yield prediction. The dataset (Table 2.) comprises observations on maize stem diameter, evaluated through multiple regression models to ascertain the most accurate predictor within the context of precision agriculture. The dataset encapsulates several regression models, including Linear Regression, Automatic Relevance Determination (ARD), and Huber regression, each assessed on their efficacy in predicting stem diameter based on a set number of measurements. Key performance metrics, such as Mean Absolute Error (MAE), Minimum Absolute Error (minAE), Maximum Absolute Error (maxAE), and the Standard Deviation of Absolute Error (stdAE), are meticulously recorded to gauge the models' predictive accuracy and reliability. For instance, the Linear Regression model, when applied to two measurements of stem diameter, exhibited a mean absolute error of approximately 5.21 mm, with the minimum and maximum absolute errors

Table 2. Performance metrics of regression models for maize stem diameter prediction

	Measurements		MAE	minAE	maxAE	stdAE
Stem diameter (mm)	2	LinearReg	5.206	0.300	22.361	4.270
	2	ARD	5.249	0.028	22.978	4.251
	2	Huber	8.465	0.168	97.796	13.466
	2	PA	8.384	0.008	56.067	8.279
	2	Ada	4.996	0.148	24.364	4.373
	2	RandomForest	4.895	0.060	22.310	4.376
	2	GBT	4.688	0.052	18.235	4.002
	2	LightGBM	5.333	0.171	20.163	4.483
	1	ARD	6.529	0.087	31.412	5.979
	1	RF	7.632	0.144	41.503	7.530
	1	GBT_30_5	6.362	0.102	36.310	6.622
	3	ARD	5.615	0.010	24.188	4.817
	3	RF	5.498	0.040	23.110	4.719
	3	GBT_30_7	5.443	0.050	26.451	5.420
	4	ARD	5.249	0.028	22.978	4.251
	4	RF	4.667	0.010	21.830	4.333
	4	GBT_30_5	4.396	0.298	21.373	4.239
Maize length (cm)	2	LinearReg	6.419	0.282	30.440	5.552
	2	ARD	6.270	0.018	30.084	5.654
	2	Huber	6.004	0.046	28.477	5.407
	2	PA	13.377	0.243	72.349	13.513
	2	Ada	5.998	0.088	30.737	5.697
	2	RandomForest	6.143	0.190	33.870	5.750
	2	GBT	5.781	0.075	27.310	5.369
	2	LightGBM	6.557	0.215	28.483	6.070
	1	ARD	6.529	0.087	31.412	5.979
	1	RF	7.763	0.112	42.003	7.441
	1	GBT_30_5	6.378	0.102	36.310	6.610
	3	ARD	5.615	0.010	24.188	4.817
	3	RF	5.470	0.070	23.920	4.872
	3	GBT_30_7	5.302	0.027	26.455	5.571
	4	ARD	5.249	0.028	22.978	4.251
	4	RF	4.882	0.010	22.220	4.464
	4	GBT_50_5	4.400	0.258	19.892	4.140

*Mae (Mean Absolute Error), minAE (Minimum Absolute Error), maxAE (maximum Absolute Error), stdAE (standard deviation of absolute error), PA: Passive Aggressive Regressor, Ada: AdaBoost, ARD: Automatic Relevance Determination, RF: Random Forest, GBT: Gradient Boosted Trees, LightGBM: Light Gradient Boosting Machine.

being around 0.30 and 22.36 mm, respectively. This level of precision underscores the potential of Linear Regression in modeling maize growth dynamics, although the maximum error margin suggests room for improvement or the necessity for model refinement in specific scenarios. Comparatively, the ARD model, under similar conditions, demonstrated a slightly higher mean absolute error of about 5.25 mm but a notably lower minimum absolute error, suggesting its robustness in certain predictive instances. The Huber model, on the other hand, displayed a significantly higher mean error, indicating a potential limitation in handling outliers or extreme values in stem diameter predictions. These insights are pivotal for advancing our understanding of maize growth dynamics. They underline the criticality of model selection and calibration in predicting maize biomass production, a key determinant of yield in precision agriculture. The integration of such predictive models with IoT sensor data presents a comprehensive approach, enabling real-time monitoring and informed decision-making in crop management.

4. DISCUSSION

In the broader context of our research, these findings complement our advanced modeling approach, which integrates manual measurements, meteorological data, and remote sensing inputs to forecast maize yield. By correlating these results with our primary study, we demonstrate the nuanced interplay between different data sources and modeling techniques in enhancing the accuracy and applicability of yield prediction models in precision agriculture.

Our study employed various regression models, including Linear Regression, ARD, and Huber, to predict maize stem diameter. Compared to other models evaluated in this research, for instance, Linear Regression, ARD, and Huber achieved MAE values of 5.206 mm, 5.249 mm, and 8.465 mm, respectively, the GBT model achieved the lowest MAE (4.396 mm), confirming its robustness in capturing maize stem thickness. This result underscores the relative improvement and precision offered by the GBT approach. The slight discrepancy in error rates underscores the potential advantage of gradient boosting techniques over traditional regression models in capturing the complex nonlinear relationships inherent in agricultural data.

The analysis indicates minimal predictive influence from environmental variables on stem diameter, aligning with our broader research findings where environmental factors had a lesser impact on yield prediction. This observation resonates with the literature, which suggests that while environmental factors are crucial, the direct morphological measurements (like stem thickness and length) tend to offer more immediate and impactful predictive insights for biomass yield in maize.

The integration of IoT and remote sensing data is a cornerstone of our research, aiming to enhance the precision of yield predictions. The literature supports this approach, emphasizing the transformative potential of IoT in agriculture, particularly in facilitating real-time data acquisition and enabling predictive analytics. For instance, studies have highlighted the use of IoT for monitoring soil nutrients, water dynamics, and even pest management, which are integral to precision agriculture and crop yield optimization.

This dataset provides a comparative insight into the performance of various machine learning models. Aligning with literature findings, where machine learning models, especially those utilizing remote sensing data, have shown significant promise in in-season and end-of-season yield predictions, our study further validates the application of these models in a precision agriculture context. For example, Khan et al. (2023) demonstrated that support vector regression performed best for in-season predictions with an R^2 of 0.875, suggesting that meticulous model selection and validation are pivotal for accurate crop yield forecasting.

The convergence of machine learning models with advanced sensor technology represents a significant leap forward in agricultural data analytics. Our findings, in conjunction with the literature, advocate for a multidimensional modeling approach that combines in-situ measurements, remote sensing data, and machine learning for enhanced yield prediction accuracy. Future research could explore the integration of different model types, including gradient boosting and neural networks, to harness the strengths of each in capturing the complexities of crop growth dynamics.

In summary, our study contributes to the evolving field of precision agriculture by providing empirical insights into the efficacy of various regression models for maize stem diameter prediction. By contextualizing our findings within the broader literature, we underscore the importance of model selection, the integration of diverse data sources, and the ongoing exploration of advanced computational methods to refine and enhance yield prediction models in agriculture.

In this work, four periods were defined based on individual maize plant measurements. In each case, the models were trained on the final biomass of yield measurements. In the first period, only the measurements of July were considered, supplemented with Sentinel satellite images and soil parameters for this period. The next periods were the end of July and mid-August, mid-August, and the end of August. In all cases, the sets were randomly assigned to the test and training dataset. In addition, the individual maize plant samples were tested to improve the accuracy of the predictions. When the samples were randomly assigned to the learning (training) or test set, the model performed better, and it was observed that there was no significant improvement based on the neighboring predictions in every case. This result is strongly related to a previous study (Nyéki et al., 2019), in which the location of management zones based on predetermined patterns of soil can provide significant results for maize yield prediction. The considerations support the previous experience that soil variables (and quality within the field) are more determinant than other external factors, so the individual variance (measurements) is not Markovian within the zone (field).

5. CONCLUSIONS

Yield prediction in precision agriculture is an active research topic that needs to understand the mutual relationships within the crops site specifically. These relations govern the overall crop growth and hence the biomass and grain yield. In this study, some of the machine learning approaches were used to understand and model these relationships. Crop, soil and meteorological variables and vegetation status are critical parameters and should be proposed as input traits of an integrated crop model for yield prediction and forecast. In this study, GBDT models outperformed other approaches, achieving an MAE of 4.39 mm for maize stalk thickness predictions. Vegetation indices (NDVI, EVI) showed low variance, indicating limited usefulness for yield forecasting. Soil moisture and meteorological variables had minimal predictive impact, reinforcing the dominance of morphological traits (stalk thickness, plant height) in biomass modeling. Biomass yield could be accurately estimated two weeks before harvest, enabling early intervention strategies for yield optimization.

Unlike previous studies that rely solely on satellite-based predictors (Ren et al., 2008; Wang et al., 2019), this study demonstrates that integrating IoT-driven field sensor data significantly enhances predictive accuracy. Furthermore, our feature importance analysis reveals that environmental conditions, such as soil moisture and temperature, contribute less to yield prediction than morphological attributes. These findings challenge traditional assumptions and highlight the need for a hybrid approach leveraging high-frequency, in-situ plant growth measurements.

ACKNOWLEDGMENTS

The research was supported by the Janos Bolyai Research Scholarship (BO/00578/24) of the Hungarian Academy. The research was carried out by the “Precision Bioengineering Research Group”.

REFERENCES

Abimbola, O.P., Franz, T.E., Rudnick, D., Heeren, D., Yang, H., Wolf, A., Katimbo, A., et al. (2022). Improving crop modeling to better simulate maize yield variability under different irrigation managements. *Agricultural Water Management*, 262: 107429, https://doi.org/10.1016/j.agwat.2021.107429

Agnolucci, P., Rapti, C., Alexander, P., De Lipsis, V., Holland, R.A., Eigenbrod, F., and Ekins, P. (2020). Impacts of rising temperatures and farm management practices on global yields of 18 crops. *Nature Food*, 1(9): 562–571, https://doi.org/10.1038/s43016-020-00148-x

Alahmad, T., Neményi, M., and Nyéki, A. (2023). Applying IoT sensors and big data to improve precision crop production: a review. *Agronomy*, 13(10): 2603, https://doi.org/10.3390/agronomy13102603

Ali, A., Hussain, T., Tantashutikun, N., Hussain, N., and Cocetta, G. (2023). Application of smart techniques, internet of things and data mining for resource use efficient and sustainable crop production. *Agriculture*, 13(2): 397, https://doi.org/10.3390/agriculture13020397

Bhojani, S.H. and Bhatt, N. (2020). Wheat crop yield prediction using new activation functions in neural network. *Neural Computing and Applications*, 32(17): 13941–13951, https://doi.org/10.1007/s00521-020-04797-8

Chang, Y., Latham, J., Licht, M., and Wang, L. (2023). A data-driven crop model for maize yield prediction. *Communications Biology*, 6(1): 439, https://doi.org/10.1038/s42003-023-04833-y

Chen, T. and Guestrin, C. (2016). XGBoost. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, New York, NY, USA, pp. 785–794, https://doi.org/10.1145/2939672.2939785

Elahi, E., Khalid, Z., Tauni, M.Z., Zhang, H., and Lirong, X. (2022). Extreme weather events risk to crop-production and the adaptation of innovative management strategies to mitigate the risk: a retrospective survey of rural Punjab, Pakistan. *Technovation*, 117: 102255, https://doi.org/10.1016/j.technovation.2021.102255

Folberth, C., Elliott, J., Müller, C., Balkovič, J., Chryssanthacopoulos, J., Izaurralde, R.C., Jones, C.D., et al. (2019). Parameterization-induced uncertainties and impacts of crop management harmonization in a global gridded crop model ensemble. *Plos One*, 14(9): e0221862, https://doi.org/10.1371/journal.pone.0221862

Foster, A.J., Kakani, V.G., and Mosali, J. (2017). Estimation of bioenergy crop yield and N status by hyperspectral canopy reflectance and partial least square regression. *Precision Agriculture*, 18(2): 192–209, https://doi.org/10.1007/s11119-016-9455-8

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, JSTOR, 1189–1232.

Harmani, V.P., Himawan, B.M., Alhadi, M.A., Gunawan, A.A.S., and Anderies (2022). Systematic literature review: implementation of artificial intelligence in precision agriculture. In: *2022 5th international conference on information and communications technology (ICOIACT)*. IEEE, pp. 479–484, https://doi.org/10.1109/ICOIACT55506.2022.9971917

Ines, A.V.M., Das, N.N., Hansen, J.W., and Njoku, E.G. (2013). Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment*, 138: 149–164, https://doi.org/10.1016/j.rse.2013.07.018

Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., et al. (2016). Random forests for global and regional crop yield predictions. *Plos One*, 11(6): e0156571, https://doi.org/10.1371/journal.pone.0156571

Khaki, S. and Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10, https://doi.org/10.3389/fpls.2019.00621

Khaki, S., Wang, L., and Archontoulis, S.V. (2020). A CNN-RNN framework for crop yield prediction. *Frontiers in Plant Science*, 10, https://doi.org/10.3389/fpls.2019.01750

Khan, S.N., Khan, A.N., Tariq, A., Lu, L., Malik, N.A., Umair, M., Hatamleh, W.A., et al. (2023). County-level corn yield prediction using supervised machine learning. *European Journal of Remote Sensing*, 56(1), https://doi.org/10.1080/22797254.2023.2253985

Lesk, C., Coffel, E., Winter, J., Ray, D., Zscheischler, J., Seneviratne, S.I., and Horton, R. (2021). Stronger temperature–moisture couplings exacerbate the impact of climate warming on global crop yields. *Nature Food*, 2(9): 683–691, https://doi.org/10.1038/s43016-021-00341-6

MacCarthy, D.S., Adiku, S.G.K., Freduah, B.S., Gbefo, F., and Kamara, A.Y. (2017). Using CERES-maize and ENSO as decision support tools to evaluate climate-sensitive farm management practices for maize production in the Northern Regions of Ghana. *Frontiers in Plant Science*, 8, https://doi.org/10.3389/fpls.2017.00031

Maeda, Y., Goyodani, T., Nishiuchi, S., and Kita, E. (2018). Yield prediction of paddy rice with machine learning. In: *Proceedings of the international conference on parallel and distributed processing techniques and applications (PDPTA)*. The Steering Committee of The World Congress in Computer Science, Computer …, pp. 361–365.

Nyéki, A., Kerepesi, C., Daróczy, B., Benczúr, A., Milics, G., Kovács, A.J., and Neményil, M. (2019). Maize yield prediction based on artificial intelligence using spatio-temporal data. In: *Precision agriculture '19*. Wageningen Academic Publishers, The Netherlands, pp. 1011–1017, https://doi.org/10.3920/978-90-8686-888-9_124

Nyéki, A., Milics, G., Kovács, A.J., and Neményi, M. (2013). Improving yield advisory models for precision agriculture with special regards to soil compaction in maize production. In: *Precision agriculture '13*. Brill | Wageningen Academic, pp. 443–450, https://doi.org/10.3920/9789086867783_056

Nyéki, A., Teschner, G., Ambrus, B., Neményi, M., and Kovács, A.J. (2020). Architecting farmer-centric internet of things for precision crop production. *Hungarian Agricultural Engineering*, (38): 71–78, https://doi.org/10.17676/HAE.2020.38.71

Ramesh, D. and Vardhan, B.V. (2015). Analysis of crop yield prediction using data mining techniques. *International Journal of Research in Engineering and Technology*, 4(1): 47–473.

Ranum, P., Peña-Rosas, J.P., and Garcia-Casal, M.N. (2014). Global maize production, utilization, and consumption. *Annals of the New York Academy of Sciences*, 1312(1): 105–112, https://doi.org/10.1111/nyas.12396

Ren, J., Chen, Z., Zhou, Q., and Tang, H. (2008). Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *International Journal of Applied Earth Observation and Geoinformation*, 10(4): 403–413, https://doi.org/10.1016/j.jag.2007.11.003

Sakamoto, T. (2020). Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160: 208–228, https://doi.org/10.1016/j.isprsjprs.2019.12.012

Sinha, A., Shrivastava, G., and Kumar, P. (2019). Architecting user-centric internet of things for smart agriculture. *Sustainable Computing: Informatics and Systems*, 23: 88–102, https://doi.org/10.1016/j.suscom.2019.07.001

Sun, J., Di, L., Sun, Z., Shen, Y., and Lai, Z. (2019). County-level soybean yield prediction using deep CNN-LSTM model. *Sensors*, 19(20): 4363, https://doi.org/10.3390/s19204363

Verón, S.R., de Abelleyra, D., and Lobell, D.B. (2015). Impacts of precipitation and temperature on crop yields in the Pampas. *Climatic Change*, 130(2): 235–245, https://doi.org/10.1007/s10584-015-1350-1

Viscarra Rossel, R.A. and Bouma, J. (2016). Soil sensing: a new paradigm for agriculture. *Agricultural Systems*, 148: 71–74, https://doi.org/10.1016/j.agsy.2016.07.001

Vuran, M.C., Salam, A., Wong, R., and Irmak, S. (2018). Internet of underground things in precision agriculture: architecture and technology aspects. *Ad Hoc Networks*, 81: 160–173, https://doi.org/10.1016/j.adhoc.2018.07.017

Wang, L., Wang, P., Liang, S., Qi, X., Li, L., and Xu, L. (2019). Monitoring maize growth conditions by training a BP neural network with remotely sensed vegetation temperature condition index and leaf area index. *Computers and Electronics in Agriculture*, 160: 82–90, https://doi.org/10.1016/j.compag.2019.03.017

Wang, X., Huang, J., Feng, Q., and Yin, D. (2020). Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches. *Remote Sensing*, 12(11): 1744, https://doi.org/10.3390/rs12111744

Wei, Z., Meng, Y., Zhang, W., Peng, J., and Meng, L. (2019). Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau. *Remote Sensing of Environment*, 225: 30–44, https://doi.org/10.1016/j.rse.2019.02.022

Xie, Y., Wang, P., Bai, X., Khan, J., Zhang, S., Li, L., and Wang, L. (2017). Assimilation of the leaf area index and vegetation temperature condition index for winter wheat yield estimation using landsat imagery and the CERES-wheat model. *Agricultural and Forest Meteorology*, Elsevier, 246: 194–206, https://doi.org/10.1016/J.AGRFORMET.2017.06.015

Zelenák, A., Szabó, A., Nagy, J., and Nyéki, A. (2022). Using the CERES-Maize model to simulate crop yield in a long-term field experiment in Hungary. *Agronomy*, 12(4): 785, https://doi.org/10.3390/agronomy12040785