Smart Agricultural Technology 9 (2024) 100671

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Enhancing corn yield prediction: Optimizing data quality or model complexity?

Yuting Zhou [a,*], Shengfang Ma [b], Huihui Zhang [c,*], Sathyanarayanan Aakur [d]

[a] *Department of Geography, Oklahoma State University, Stillwater, OK, USA*
[b] *Oklahoma Water Resources Center, Oklahoma State University, Stillwater, OK, US*
[c] *Water Management and Systems Research Unit, United States Department of Agriculture, Agricultural Research Service, Fort Collins, CO, US*
[d] *Department of Computer Science and Software Engineering, Auburn University, Auburn, AL, USA*

ARTICLE INFO

Keywords:
UAV
Multispectral
RGB
Yield prediction
Deep learning
Vision transformer

ABSTRACT

Field-scale corn yield prediction before harvest can assist farmers in better organizing their resources. Machine learning-based pipelines for analyzing remote sensing imagery offer an efficient solution to this problem. However, the cost of data acquisition and training requirements for machine or deep learning models depend on various factors, such as equipment (multispectral vs. RGB sensors) and the ability to predict yield from observations across growth stages. In this study, we aim to provide a comprehensive analysis of the effectiveness of traditional ensemble learning methods (Random Forest and Gradient Boosting) and deep learning models (ResNet 18, ResNet34, and ViT) in predicting corn yield across deficit and fully irrigated fields using UAV-based RGB and multispectral imagery. The performance of these models was examined across early, middle, and late growth stages, considering both computational complexity and accuracy. We also developed a novel shallow CNN framework called SimRes, inspired by the ResNet framework but tailored for streamlined efficiency and simplicity for yield prediction. Extensive quantitative analysis demonstrated that the customized SimRes performed as well as deep learning baselines but with faster computing times, while traditional approaches, such as Random Forests and Gradient Boosting exhibited marginally smaller R-squared values. Models utilizing multispectral data outperformed models using RGB, albeit with variations across growth stages. Deep learning methods performed better than ensemble learning methods in the early and late growth stages using RGB, while performance became comparable in the middle stage. These results underscore the importance of additional information or more complex models to enhance prediction accuracy alongside a trade-off between computational complexity and accuracy. This research provides valuable insights for optimizing corn yield prediction across different growth stages, informing agricultural management and harvest planning decisions.

1. Introduction

As the world's population continues to increase, the demand for food escalates correspondingly [1]. Corn is a globally significant crop, with the United States (U.S.) leading in its production, accounting for 32 % of the global output in 2023 [2]. Accurate and timely crop yield prediction is essential to safeguarding global food security [3]. Effective corn yield prediction not only aids in meeting the surging food requirements but also facilitates optimal resource allocation and management strategies [[4]. As corn serves as a vital source of economic income for numerous farmers in the U.S., farm-level yield prediction is crucial in offering farmers invaluable insights to better prepare for harvests. Timely predictions empower farmers to proactively adjust their management practices, potentially enhancing yields and mitigating adverse impacts of environmental variability, especially in this climate change era. Thus, anticipating crop yields accurately and in advance is essential in fostering resilience within agricultural systems. Accurate and easy-to-get crop yield prediction can offer invaluable insights and information to farmers, aiding in informed decision-making and enhancing overall agricultural productivity.

Corn yield is affected by many factors [5], such as its growth state, environmental conditions during a growth season, soil conditions, management practices (e.g., irrigation and fertilizer), and seed varieties. Those factors need to be considered when predicting corn yield.

* Corresponding authors.
E-mail addresses: yuting.zhou@okstate.edu (Y. Zhou), Huihui.zhang@usda.gov (H. Zhang).

https://doi.org/10.1016/j.atech.2024.100671
Received 11 October 2024; Received in revised form 26 November 2024; Accepted 26 November 2024
Available online 28 November 2024

Conventional corn yield estimation relies on ground field survey data, which is costly and time-consuming. Thus, two more widely used yield prediction methods were developed: a physical process model and a statistical model. The physical process model for predicting corn yield dates back to the early 1960s [6], which uses physiology-based crop growth knowledge and dynamically simulates the crop growth situation to predict yield. The physical model has been successfully used to predict crop yields [7], especially after integrating it with remote sensing techniques [8]. However, the physical process model is complicated compared to statistical models, which require a large amount of input data and many parameters for calibration and are computationally expensive. The statistical models also have a long history, with early examples by [9,10]. Statistical modelling approaches have become increasingly prevalent in recent years [11], with an increasing abundance of data on crops, weather patterns, and soil conditions captured by various sensors, alongside advancements in machine learning techniques.

Among the statistical models, machine learning models stand out in crop yield estimation and become the mainstream methods [12,13]. Machine learning algorithms such as Linear Regression, Ensemble Learning, and deep learning are widely used to predict crop yield [14,15,16]. In recent years, deep-learning-based models have gained widespread adoption over traditional machine learning [17]. Deep learning models such as Feedforward Neural Networks (FNN) and Convolutional Neural Networks (CNNs) are better at modelling the complicated interactions between complex inputs and yield. When handling remote sensing images, CNNs have outperformed other models by extracting useful features for regression and classification [12,18,19]. Ensemble learning, such as Random Forest (RF) and Gradient Boosting (GB), is suitable for small sample data and widely used in crop yield estimation [14,15,20,21,22]. Specifically, RF has been used at large regional and field scales and has achieved accurate predictions [4,16,23]. Gradient Boosting extracts detailed features to make finer predictions in theory, but it is more computationally expensive. Higher spatial resolution data is required to train a machine learning model to get farm-level yield prediction. Ideally, a simple, robust model with minimal memory and computational requirements is more practical to integrate into farm monitoring devices.

Remote sensing has emerged as a viable alternative to a physical survey of the field for yield prediction using either the physical process model or the statistical models [11,17,19]. While proven to be popular, satellite remote sensing data is primarily useful for district-level and county-level yield prediction due to the low spatial and temporal resolution of satellite remote sensing data [24]. Unmanned Aerial Vehicles (UAV) provide a low-cost solution for collecting remote sensing data [25] at higher spatial resolution images along different spectral ranges such as red-green-blue (RGB) (or visible-light), multispectral (MS), or hyperspectral (HS), providing timely information about the variations for crop and soil conditions in precision agriculture [26].

While UAV-based imagery is effective in training machine learning models to predict crop yields [27,28,29,30], several key questions remain before its widespread adoption. First, the price of cameras, a major factor in UAV design and pricing, depends on the spectral ranges it can capture. A multispectral camera can cost ten times more than an RGB camera. In addition, multispectral cameras are heavier than RGB cameras and require a larger drone that can carry a high payload. It would make the technique more affordable if we could achieve comparable performance with a multispectral camera using an RGB camera. Geipel et al. [31] showcased a strong correlation between corn grain yield and three vegetation indices (VIs) derived from UAV RGB images. Second, machine learning is a quickly evolving field. New and more complex models are developed constantly. However, it is unclear how much is the real advantage of deep learning models to ensembled machine learning, when considering both prediction accuracy and computation time. Last but not least, most studies only use one epoch of data to predict corn yield, and few studies compared the performance of yield prediction at different growth stages. Early prediction of corn yield has tremendous meaning for farmers as it gives them time to adjust management practices to improve production and prepare for potential challenges such as market fluctuations or adverse weather conditions, ultimately maximizing their profitability and resilience in the agricultural sector. One recent study found that ML models can predict corn yield with an R^2 of 0.84 and 0.83 at vegetative (V6) and reproductive (R5) growth stages, respectively, using VIs calculated from multispectral imagery, though it is not clear a comparable performance could be achieved in an earlier or later stage [32]. Another study showed that an Ensemble Learning model was able to predict corn yield with an R^2 of 0.95 in the late vegetative growth stages (V10 and VT stages). However, it included <100 samples in both training and testing, indicating a possibility of overfitting [33]. To fill these gaps to make UAV and machine learning techniques more approachable for farmers, the objectives of this study are two-fold: (1) finding the most cost-effective machine learning model to predict corn yield when different bands were used and (2) determining the prediction accuracy at different growth stages. We anticipate findings from this study will guide farmers on which sensor to use (e.g., multispectral vs. RGB camera) and how early they can predict corn yield to adjust management practices and allocate their resources.

2. Materials and methods

2.1. Study Area

The study was conducted at the United States Department of Agriculture-Agricultural Research Service (USDA-ARS) Limited Irrigation Research Farm (LIRF) situated near Greeley, Colorado, USA. The farm is located at latitude 40.4463°N, longitude 104.6371°W, and an elevation of 1432 m above sea level (ASL). Pioneer 9998Q corn variety was planted on two adjacent rectangular-shaped fields on May 9, 2022 (Fig. 1), both equipped with sub-surface drip irrigation systems. The west field underwent frequent irrigation, aimed at maintaining the soil water content within the root zone at non-water stress levels. In contrast, the east field was managed as a deficit irrigated field, subject to water stress conditions. The irrigation regimes resulted in a total of seven events for the west field, totaling 509 mm of cumulative gross irrigation water depth, and six events for the east field, totaling 335 mm. The cumulative rainfall water depth during the growing season amounted to 116 mm based on measurements obtained from an onsite

Fig. 1. RGB image of the LIRF research site in northern Colorado, US. The study area is composed of fully irrigated (left) and deficit irrigated fields (right).

micrometeorological station (site name: GLY04, https://coagmet.colostate.edu/rawdata_form.php). Detailed descriptions of the experimental fields, farm management, data collection, and processing were also found in [34] (manuscript submitted and under review).

2.2. Datasets

This study utilized multispectral and RGB imagery captured by a DJI Spreading Wings S900 hexacopter (DJI Co., LTD, Shenzhen, China) equipped with a MicaSense RedEdge-MX camera (MicaSense Inc., Seattle, Washington, USA)[1] and a FLIR Duo Pro R camera (FLIR Systems, Inc., Wilsonville, OR). The multispectral camera acquired data in five bands including blue (475 nm, 32 nm bandwidth), green (560 nm, 27 nm bandwidth), red (668 nm, 16 nm bandwidth), RedEdge (717 nm, 12 nm bandwidth), and near-infrared (NIR) (842 nm, 57 nm bandwidth), while the RGB camera provided high-resolution 4k color imagery.

UAV flights were conducted at a height of 120 m with an overlap/sidelap of 88/70 %, resulting in a spatial resolution of 0.03 m for multispectral imagery and 0.01 m for RGB imagery. Image processing was performed using Agisoft Metashape software (Agisoft LLC, Saint Petersburg, Russia) to create orthomosaics and georectify the data.

Imagery was collected at three key growth stages: V12 (early, July 16), R1 (middle, August 2), and R6 (late, September 26). During the early stage (V12), maize tassel is developing rapidly but is not yet visible. In the middle stage (R1), the corn plant reaches full size, tassels are fully visible, and silks are emerging. Silking is one of the most critical stages in determining yield potential. The late stage (R6) is about 60 days after silking, physiological maturity is reached, and kernels have attained maximum dry weight at 30 to 35 % moisture.

The fully irrigated field was harvested on Oct 31 and Nov 1, 2022, and the deficit irrigated field was harvested on Oct 13–14, 2022, using a combine harvester, ALMACO RX 5FT Draper Head (ALMACO Inc., Nevada, IOWA, USA), equipped with a 2.5 cm accuracy GPS (NAV-900, Trimble, Inc., Westminster, CO, USA). The plot lengths for the fully and deficit irrigated fields were harvested in smaller blocks with a length and width of 4.57 m and 6.1 m, respectively (Fig. 2). Yield weight, field moisture content, and GPS location were recorded for each block. Yield weights were adjusted to 15.5 % moisture content and combined with geolocated polygons shapefiles of the smaller block using ArcGIS Pro version 3.0.0 (ESRI, Redlands, CA). The yield monitor on the combine harvester recorded 3990 data points for both fields. The yield's mean, minimum, maximum, and standard deviation were 11.74, 0.67, 18.75, and 3.02 ton/ha, respectively.

2.3. Data Preprocessing

The data processing phase involved the slicing of both high spatial resolution RGB and five-band multispectral imagery of the corn fields at three growth stages into image chips using block boundaries delineated during the corn harvesting process with GPS (Fig. 2). Each image chip corresponded to a yield value measuring 4.57 m in length and 6.1 m in width. These image chips, along with the associated yield data, constituted the training and testing datasets for the corn yield predictive model. The dataset was randomly divided, with 75 % allocated for training and validation and 25 % reserved for testing. Due to inherent random errors in GPS measurements, the boundaries were not uniform, resulting in image chips of slightly varying sizes. The multispectral image chips were approximately 20×80 pixels with a deviation of around ± 2 pixels, while RGB image chips were roughly 55×220 pixels with a deviation of approximately ± 5 pixels. To standardize the sizes of the image chips, a resizing operation was performed: all multispectral image chips were resized to 20×80 pixels, and RGB image chips were resized to 55×220 pixels. Furthermore, if there are missing values (NaN) in the imagery, a cubic interpolation method was applied to interpolate NaN pixels. All data processes were done in ArcGIS and Python.

2.4. Baselines: Ensemble learning

Ensemble learning is a powerful machine learning method. Ensemble learning uses a set of models that are combined to make a prediction [35]. Ensembles can improve the robustness and accuracy of prediction by mitigating the weaknesses of individual models. Two widely used ensemble learning methods in corn yield prediction are Random Forest (RF) and Gradient Boosting (GB).

Random Forest [36] is an ensemble learning algorithm that uses a decision tree as the basis learner with the principle of bagging [37]. The decision trees have been studied extensively in the field of machine learning since the 1960s. Decision trees partition the feature space into smaller regions based on a series of binary decisions, making them intuitive and easy to interpret. Bagging, also called Bootstrap Aggregating, developed by [37], involves training multiple instances of a base model on different subsets of the training data and subsequently aggregating their predictions through averaging or voting. This approach can reduce variance and enhance the robustness and generalization capabilities of the ensemble. Random Forest has found extensive application in crop yield prediction due to its ability to effectively handle complex and high-dimensional data, as well as their robust performance in predictive tasks.

Gradient Boosting [38] is another widely used ensemble learning that combines the predictions of multiple base learners to produce a strong learner with improved predictive performance. Unlike RF, it uses the Boosting ensemble method instead of bagging. Boosting algorithms use a forward stagewise process to transform weak learners into strong learners by increasing the weights of training samples that were mistakenly identified or wrongly calculated in a successive iteration [39]. The key idea behind boosting is to train each base learner to correct the errors made by the existing ensemble instead of training the base learner independently. It was treated as a powerful tool for corn yield prediction.

Feature Engineering: While ensemble learning excels in capturing complex non-linear relationships inherent in agricultural data and offering high predictive accuracy, it is challenging to extract features from images that are suitable for yield prediction. In traditional statistical models, spectral reflectance curves and vegetation indices are important. Here we also use the spectral information. We extracted the reflectance of the entire image and the histogram distribution of reflectance as features for prediction model input.

Model Optimization: To optimize the performance of Random Forest (RF) and Gradient Boosting (GB) models, we conducted hyperparameter tuning using grid search. For RF, we varied the number of trees and maximum tree depth, ultimately settling on 100 trees and a maximum depth of 25. For GB, we adjusted the number of trees, learning rate, maximum tree depth, and regularization parameters, selecting 250 trees, a maximum depth of 25, and appropriate regularization settings.

2.5. Baselines: deep learning

Deep learning, a branch of machine learning, utilizes artificial neural networks comprising multiple layers. Deep learning excels in feature extraction without the need for explicit programming. When handling image data, both Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) outperform other deep learning models.

Convolutional neural networks, or CNN, are especially prominent in modelling tasks where the input data is an image due to their ability to effectively capture spatial hierarchies and local patterns within the data [40]. CNNs leverage convolutional layers to automatically extract

[1] Mention of trademark, vendor, or proprietary product does not constitute a guarantee or warranty of the product by the USDA and does not imply its approval to the exclusion of other products that may also be suitable.

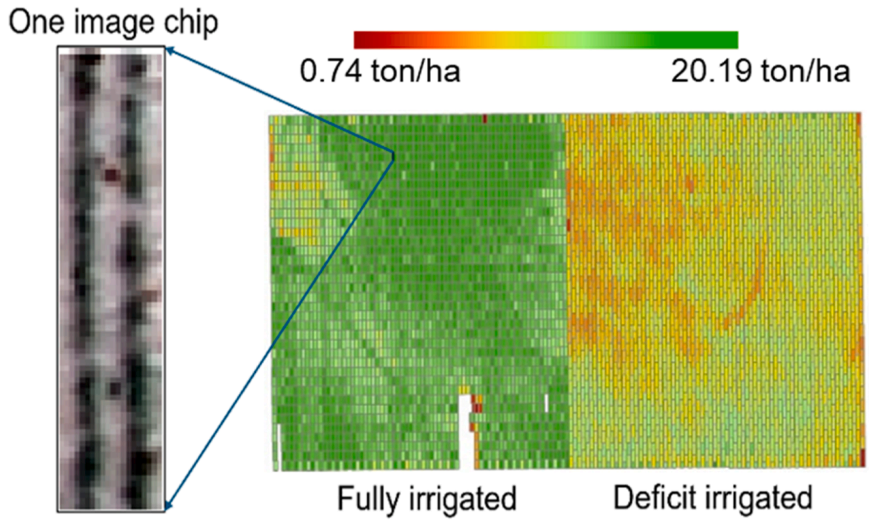

Fig. 2. Yield map. The color represents corn yield with green indicating higher and red indicating lower yield. The image chip is an example of how UAV imagery is extracted.

relevant features from images, preserving spatial information and reducing the computational burden when compared with feedforward neural networks. Various CNN architectures have been developed, including AlexNet, VGGNet, GoogLeNet, ResNet, and MobileNet. Each architecture has undergone evolution with different versions featuring varying numbers of parameters and layers tailored to specific application objectives. ResNet, in particular, has demonstrated exceptional performance across a wide range of computer vision tasks. Its potential in corn yield prediction has developed [18,40,41]. Additionally, a deeper CNN model has a stronger ability to extract features. However, there is always a trade-off between model complexity and accuracy. Deeper CNN models might perform better in predicting corn yield, but the computational cost will also be higher. It is unclear if a shallow CNN model can predict corn yield while minimizing computational cost. Here, we developed a customed Simple Resnet, SimRes, (see Fig. 3) constructed from Resnet Framework, Resnet18, and Resnet34 for corn yield prediction.

Vision Transformer (ViT) is a recent advancement in computer vision that adopts the Transformer architecture. Unlike traditional CNNs, which rely on convolutional layers for feature extraction, ViT processes images as sequences of fixed-size patches, leveraging self-attention mechanisms to capture global context and spatial relationships. This patch-based approach enhances ViT's ability to generalize across diverse datasets and scales seamlessly to images of varying resolutions. Additionally, ViT's transformer architecture facilitates parallel processing, enabling efficient training on modern hardware architectures, such as GPUs and TPUs. Moreover, ViT exhibits remarkable flexibility, allowing straightforward adaptation to various downstream tasks through fine-tuning or transfer learning. These advantages position ViT as a promising alternative to CNNs, particularly in tasks requiring a holistic understanding of image content, such as image classification, object detection, and semantic segmentation. While the ViT model has not been widely utilized in corn yield prediction [42], this paper explores ViT's potential in predicting corn yield from point-in-time images.

Hyperparameters for deep learning models: Hyperparameters such as learning rate, batch size, optimization algorithms, and optimizer coefficients, impact the performance of deep learning models. Adam Optimization algorithms with a small learning rate of 0.00075 with a

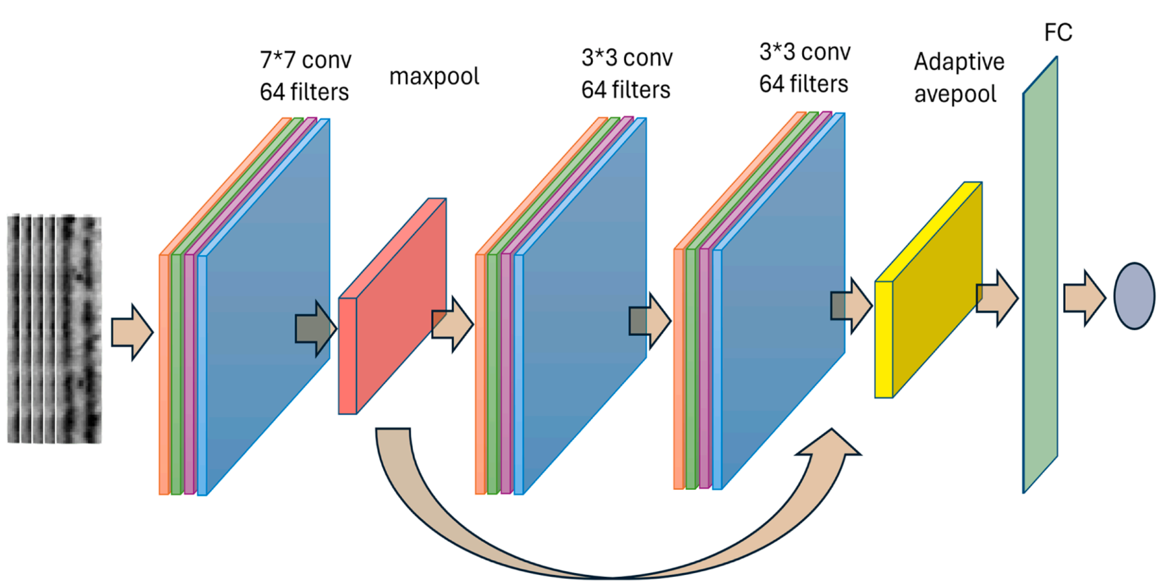

Fig. 3. Architecture of SimRes that consists of a single convolutional layer followed by a unique residual block. The residual block comprises two convolutional layers, with skip connections bypassing these layers.

batch size of 32 were used for all deep learning networks. Data augmentation techniques such as flipping, and rotating were applied to the input images to increase model robustness and generalization.

Several techniques were implemented to mitigate model overfitting in this study. First, data augmentation techniques, including horizontal and vertical flips, were applied to increase data diversity and reduce overfitting. Second, cross-validation was employed by splitting the training and validation dataset into four subsets. Three subsets were used for training while the remaining subset was used for validation, rotating this process to ensure robust performance. Additionally, transfer learning was utilized by initializing model parameters with those from pretrained models (ResNet and ViT), enhancing the training process. Other widely used techniques, such as regularization, dropout, and batch normalization, were also applied to further prevent overfitting and improve generalization.

2.6. Evaluation metrics

The 75 % data were used for training and validation and the 25 % data were only used to test the model performance. Three metrics including coefficient of determination (R-squared, R^2), RMSE, and mean absolute relative error (MARE) were used to indicate prediction accuracy while training and testing time were used to represent computation cost alongside model size. All these metrics were used to evaluate the overall performance of different models. The formulas for R^2, RMSE, and MARE are shown in the following equations.

$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-1}} \tag{2}$$

$$\mathrm{MARE} = \frac{1}{n-1}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{3}$$

where n denotes the number of data samples; y_i and $\hat{y}_i$ refer to the measured and predicted yield for a specific sample; $\bar{y}$ denotes the mean value of ground measured yield.

3. Results

3.1. Early stage yield prediction

3.1.1. RGB imagery

Overall, deep learning methods had better prediction accuracy than ensemble learning methods (Fig. 4a-b vs. c-f) when using RGB imagery as inputs at the early stage. However, the latter was much faster in terms of both training and testing time (Table 1). There were also slight differences among each type of method. Specifically, RF had a larger R^2 than GB, and with smaller RMSE and MARE. All the deep learning models had an R^2 larger than 0.8 and MARE less than 0.1, which indicated that they could predict corn yield well. The more complex methods did not always clearly increase the prediction accuracy. When increasing the layers in Resnet, which will increase the time cost, the prediction accuracy only increased marginally (Fig. 4c-e), while the training time was greatly increased. Notably, the more complex model did improve prediction accuracy for deficit irrigated corn (Fig. 4e). The relatively new ViT model, which is based on transformers, performed just as well as the SimRes.

3.1.2. Multispectral imagery

On the contrary, ensemble learning methods had similar prediction accuracy to deep learning methods (see Fig. 5) when using reflectance imagery as inputs in the early stage. All the models had an R^2 larger than 0.85. As was shown by using RGB imagery as inputs, RF had slightly

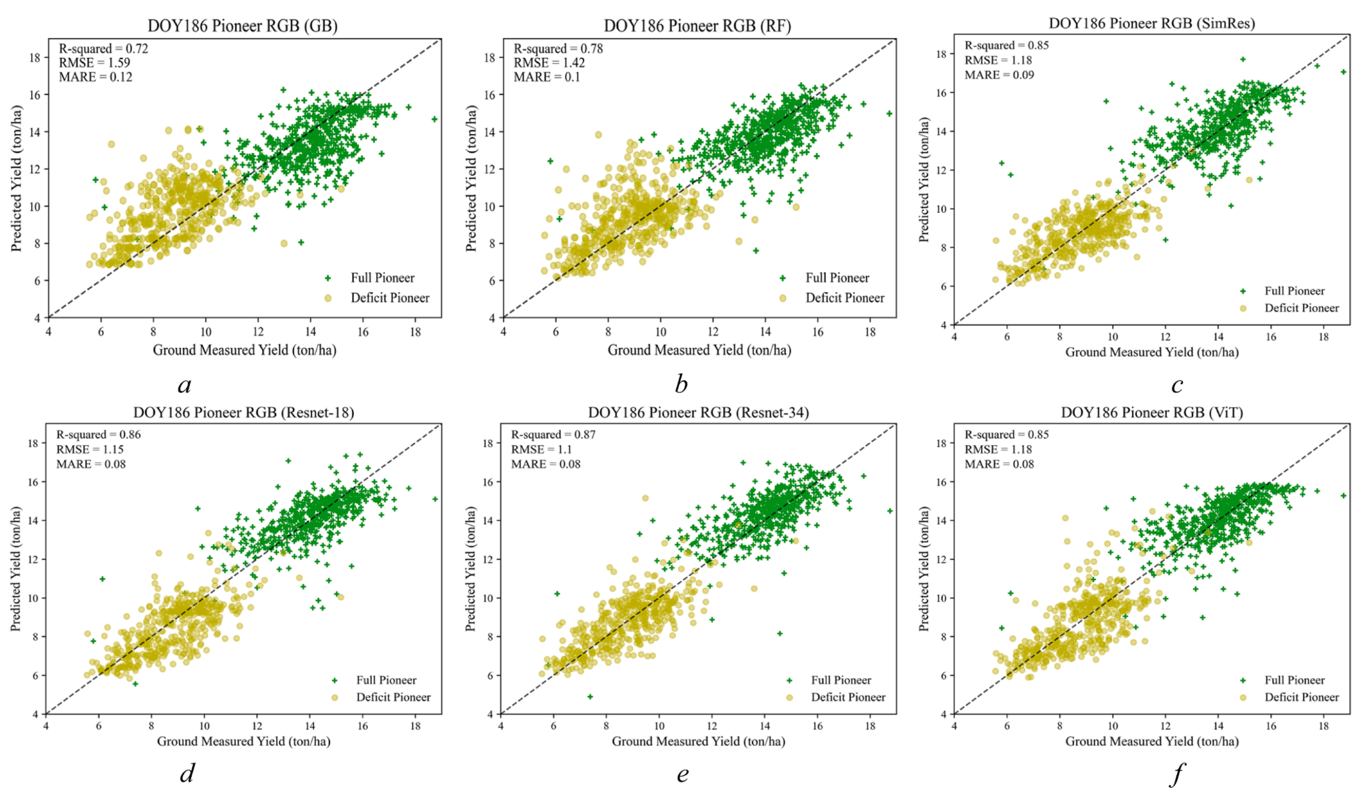

Fig. 4. Scatter plots of measured vs. predicted corn yield using UAV RGB with performance metrics included. (a) Gradient Boosting, (b) Random Forest, (c) SimRes, (d) Resnet-18, (e) Resnet-34, and (f) Vision Transformer (ViT).

Table 1
Comparison of the model performance (computational accuracy: R^2 and RMSE, and complexity: training and testing time) when using RGB or Multispectral imagery as inputs.

Model	R2		RMSE (ton/ha)		Training time (s)		Test time (s)	
	Ref	RGB	Ref	RGB	Ref	RGB	Ref	RGB
GB	0.87	0.72	1.1	1.59	~7	~1	0.08	0.04
RF	0.88	0.78	1.04	1.42	~13.7	~2.1	0.13	0.08
SimRes	0.88	0.85	1.04	1.18	~10,000	~8000	7	6
Resnet-18	0.87	0.86	1.08	1.15	~16,500	~14,800	17	17
Resnet-34	0.86	0.87	1.14	1.1	~20,904	~19,700	22	22
ViT	0.85	0.85	1.16	1.18	~20,004	~18,607	25	24

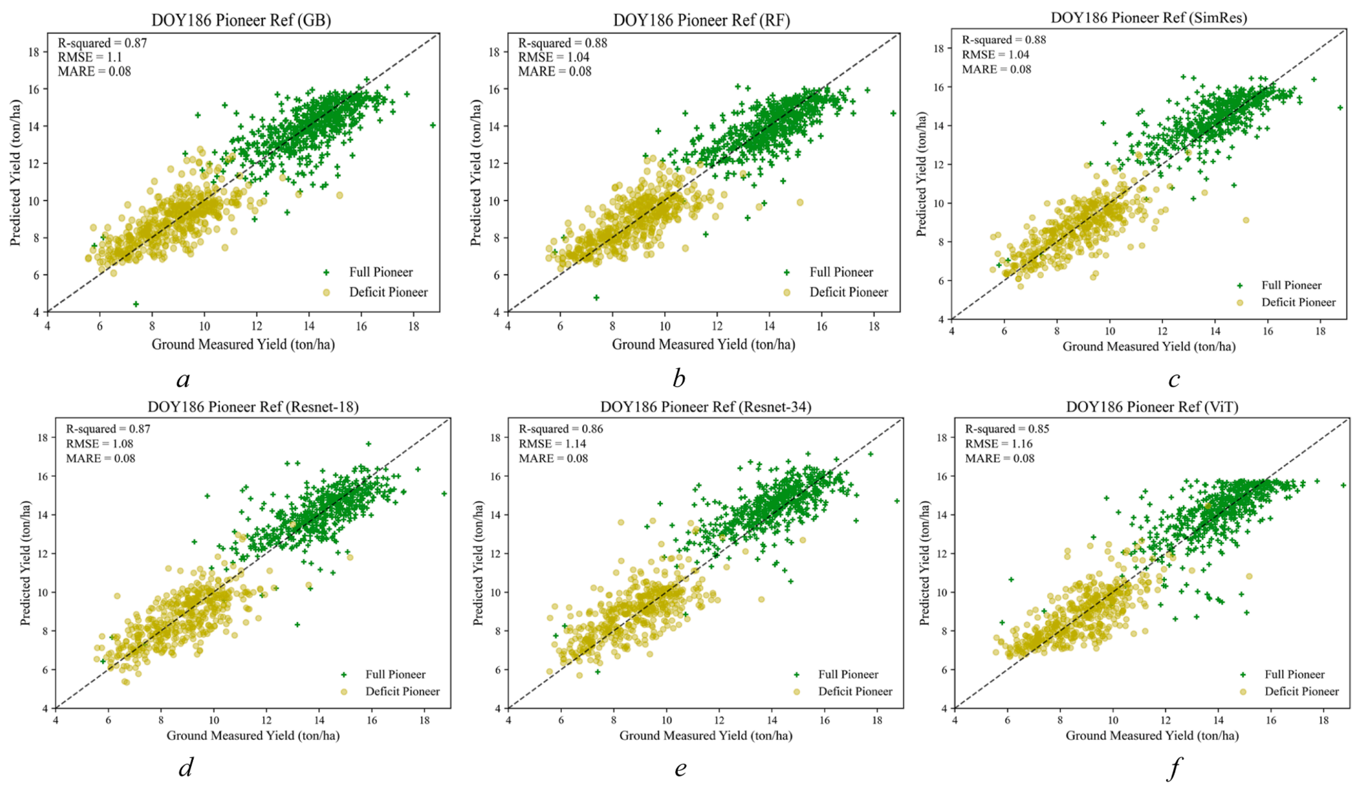

Fig. 5. Scatter plots of measured vs. predicted corn yield using reflectance from UAV multispectral imagery with performance metrics included. (a) Gradient Boosting, (b) Random Forest, (c) SimRes, (d) Resnet-18, (e) Resnet-34, and (f) Vision Transformer (ViT).

better prediction accuracy than GB (Fig. 5a and b). Also contradictory to the results in RGB imagery, the more complex deep learning models reduced R^2 values (Fig. 5c-e). The ViT had the smallest R^2 in all deep learning models. There is a clear ceiling for predicted yield in ViT results. Notably, the evaluation metrics were the same for RF and SimRes, while both can predict deficit irrigated corn yield well.

3.1.3. Multispectral and Rgb comparison

When comparing the results of multispectral to RGB imagery, ensemble learning methods performed much better here with an increase of 0.1 in R^2 values (Fig. 5a-b versus Fig. 4a-b). The results for deep learning models were similar using either RGB or reflectance as inputs (R^2 between 0.85 and 0.88). The ViT model had an even smaller R^2 in using reflectance data than RGB data. No matter whether the input data is RGB or multispectral imagery, the training time and testing time for deep learning models were significantly more than ensemble learning methods. The transformer-based model (ViT) did not outperform the ResNet models, whereas the size of the ViT model was much larger.

3.2. Middle stage yield prediction

When comparing ensemble learning and deep learning methods in the middle growth stage, we mostly focused on the prediction accuracy of the models, indicated by R^2 and RMSE, since the model time cost (training time and testing) and size were already compared and there was less change in them. As was shown in the early growth stage, deep learning models had larger R^2 values and smaller RMSE values than ensemble learning models when using RGB imagery as inputs (Fig. 6 red bars). RF still outperformed GB. Little difference was observed among deep learning models with R^2 values more than 0.85 for each deep learning model. When using reflectance imagery as inputs in the middle stage (Fig. 6 blue bars), all the models performed very well with R^2 close to or larger than 0.9. ViT had the best accuracy with an R^2 value of 0.92. No matter whether the input data was RGB or multispectral imagery, the SimRes performed as well as, if not better than, more complex models. As it was indicated in the early stage results (Fig. 4 and 5), it is clear that we need more complex models to achieve good accuracy when using RGB imagery as inputs and any model can perform well when using multispectral imagery as inputs.

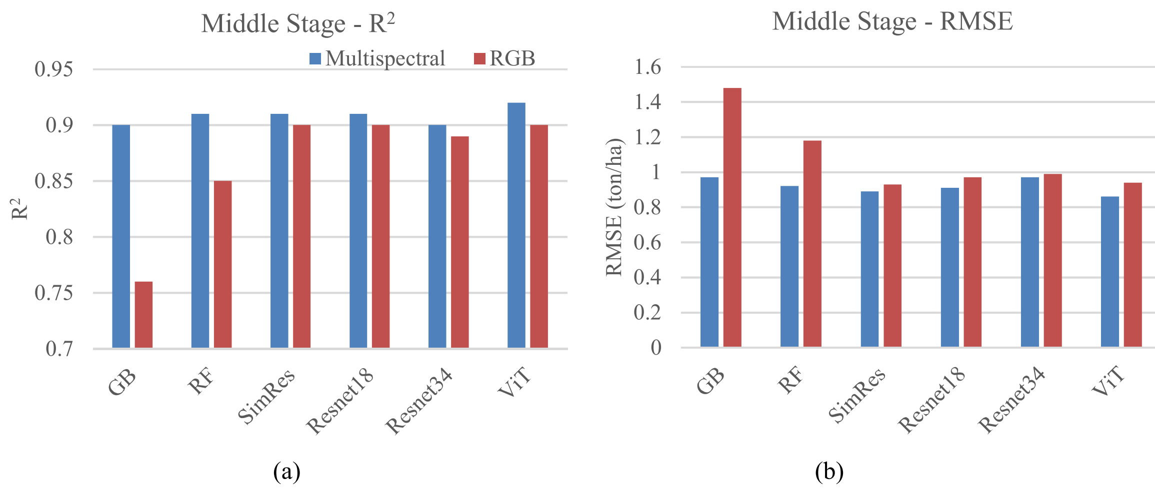

Fig. 6. Comparison of model prediction accuracy, indicated by (a) R^2, and (b) RMSE, in the middle growth stage.

3.3. Late stage yield prediction

When it comes to the late growth stage, deep learning methods still performed better than ensemble learning methods if using RGB imagery as inputs (Fig. 7 red bars), though all of them performed well with an R^2 of 0.85 or larger. In addition, the differences among models were less apparent than in the middle growth stage (Fig. 6 red bars). Unlike the very similar prediction accuracy in ensemble learning methods and deep learning methods in the middle growth stage when reflectance imagery as inputs, all deep learning models except ViT performed better than ensemble learning models, with an R^2 close to 0.9. Again, SimRes performed as well as more complex deep learning models when using either RGB or multispectral imagery as inputs.

4. Discussion and conclusion

This study evaluated the performance of ensemble learning models and deep learning models at predicting corn yield using RGB and multispectral imagery obtained by a UAV across different growth stages. While previous studies employed machine learning methods in predicting corn yield, they have not explicitly compared their performances in different growth stages using different data sources, namely RGB or multispectral imagery. Farmers are still not clear about the edge of a multispectral to an RGB camera, which is cheaper and requires less payload capacity. Furthermore, the real advantages of more complex deep learning models to ensemble learning models were not clear since a comprehensive comparison is missing.

We found that multispectral imagery worked better in predicting corn yield when using ensemble learning models in the early growth stage with an increase of 0.1 in R^2 values (Fig. 4a-b versus Fig. 5a-b). However, the prediction accuracy using deep learning models was very similar whether multispectral imagery or RGB imagery was used as inputs (Fig. 4b-e versus Fig. 5b-e). This indicates that more complex models were needed when the data quality is low, to further utilize the hidden signal in the data. In addition, the customized SimRes model had comparable if not better prediction accuracy than more complex models, while the computing cost (training/testing time and model size) was much lower. This implies that even if we need to use deep learning models, we do not need to use the most complex or newly developed models. Simplified models should be used to save computational resources if they can achieve prediction accuracy similar to those of more complex models. There were variations of prediction accuracy across different growth stages. Specifically, the middle stage had the highest prediction accuracy with an R^2 of 0.92 for the SimRes model when RGB imagery was used as input, demonstrating the best combination of the proposed SimRes with RGB information for corn yield prediction.

Numerous studies have aimed to predict corn yield at the field scale, given its pivotal role in the economy. However, before implementing these predictions, it is essential to compare various data sources and machine learning algorithms. One study used Random Forest to predict

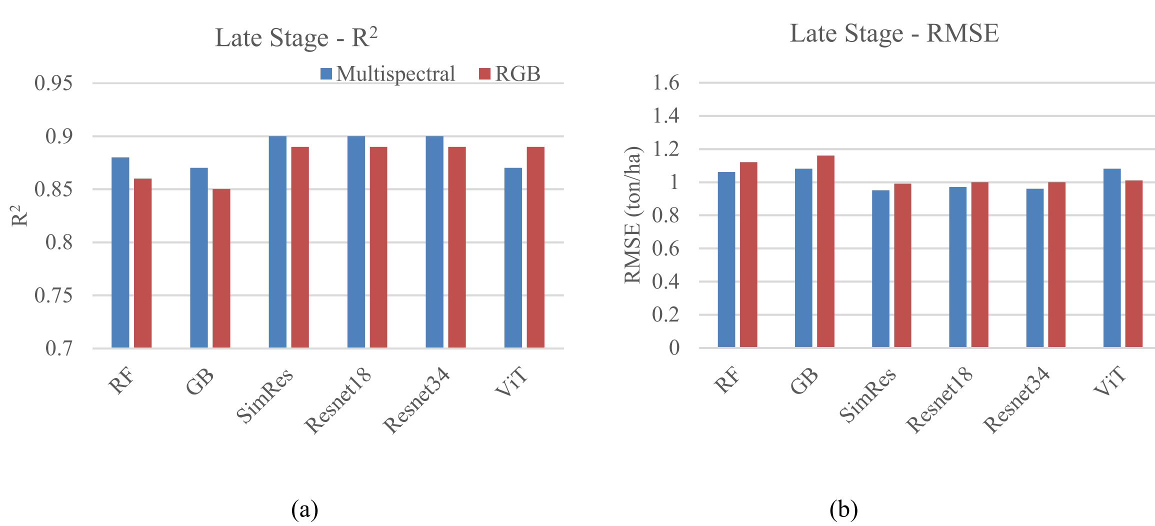

Fig. 7. Comparison of model prediction accuracy, indicated by (a) R^2, and (b) RMSE, in the late growth stage.

maize yield using vegetation indices extracted from multispectral imagery [21] and achieved a MARE of 0.78 kg ha^{-1}, which is comparable to our results in the early growth stage (Fig. 6b). However, this study did not explore the potential of deep learning models, nor did it elucidate the advantages of multispectral imagery over RGB imagery. Without such a comprehensive comparison, there is a risk of misallocation of instrumental or computational resources. Our findings underscore the importance of striking a balance between data quality and model complexity when predicting corn yield, which is the key to bringing the technology to farmers to benefit them in their farm operations.

This study comprehensively compared the performance of ensemble learning models and deep learning models to predict corn yield in different growth stages. However, other information such as weather data was not included in the models, which could potentially increase the prediction accuracy. A future study to include other information like weather data and soil data would be beneficial to further improve the model performance. Furthermore, it is not clear that the models trained here can be used in other corn fields for different scales. The transferability of the models developed in this study has to be tested in other areas and also in multiple years.

The combination of machine/deep learning and UAV-based imagery provides a great opportunity to collect high quality data and use complex algorithms to predict corn yield at the field scale to benefit farmers. However, a comprehensive examination of the technique, namely instrument and computation needs, is needed to make it approachable for farmers. Our findings provide strong evidence that we do not need to use the most advanced cameras and complex algorithms to predict corn yield. On the contrary, we should balance the data quality, which is greatly affected by camera cost, and model complexity, which is correlated with computational cost. Finding the sweet spot between data quality and model complexity will make the technical more affordable and applicable for farmers to bring real benefits to their farm operations. Future studies focusing on using more available data and the scalability of models can further help broaden the usages of machine/deep learning and UAV-based techniques in precision agriculture.

CRediT authorship contribution statement

Yuting Zhou: Writing – review & editing, Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Shengfang Ma:** Visualization, Methodology, Formal analysis, Data curation. **Huihui Zhang:** Writing – review & editing, Writing – original draft, Validation, Resources, Project administration, Investigation, Conceptualization. **Sathyanarayanan Aakur:** Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Ross Steward and Jon Altenhofen for managing the farm operation; Kevin Yemoto and Alex Olsen for collecting UAV images. This study was supported in part by research grants from the National Science Foundation (NSF) EPSCoR (OIA-1946093), the U.S. Geological Survey under Grant/Cooperative Agreement No G23AP00683, and the Bureau of Land Management Joint Fire Science Program (Project number: L23AC00388-00).

Data availability

Data will be made available on request.

References

[1] D. Tilman, C. Balzer, J. Hill, B.L. Befort, Global food demand and the sustainable intensification of agriculture, Proc. Natl. Acad. Sci. 108 (50) (2011) 20260–20264, https://doi.org/10.1073/pnas.1116437108.

[2] USDA Foreign Agricultural Service International Production Assessment Division Crop Explore. https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=0440000, (Accessed 22 August 2024).

[3] O. Marko, S. Brdar, M. Panic, P. Lugonja, V. Crnojevic, Soybean varieties portfolio optimization based on yield prediction, Comput. Electron. Agric. 127 (2016) 467–474, https://doi.org/10.1016/j.compag.2016.07.009.

[4] X. Yang, Z. Hua, L. Li, X. Huo, Z. Zhao, Multi-source information fusion-driven corn yield prediction using the Random Forest from the perspective of agricultural and forestry economic management, Sci. Rep. 14 (1) (2024) 4052, https://doi.org/10.1038/s41598-024-54354-9.

[5] Y. Ren, Q. Li, X. Du, Y. Zhang, H. Wang, G. Shi, M. Wei, Analysis of corn yield prediction potential at various growth phases using a process-based model and deep learning, Plants 12 (3) (2023) 446, https://doi.org/10.3390/plants12030446.

[6] C.T.de Wit, Photosynthesis of leaf canopies, Agric. Res. Rep 663 (1965) 57.

[7] X. Mo, S. Liu, Z. Lin, Y. Xu, Y. Xiang, T.R. McVicar, Prediction of crop yield, water consumption and water use efficiency with an SVAT-crop growth model using remotely sensed data on the North China Plain, Ecol. Modell. 183 (2) (2005) 301–322, https://doi.org/10.1016/j.ecolmodel.2004.07.032.

[8] J. Huang, J.L. Gómez-Dans, H. Huang, H. Ma, Q. Wu, P.E. Lewis, X. Xie, Assimilation of remote sensing into crop growth models: current status and perspectives, Agric. For. Meteorol 276-277 (2019) 107609, https://doi.org/10.1016/j.agrformet.2019.06.008.

[9] E.C.A. Runge, Effects of rainfall and temperature interactions during the growing season on corn yield, Agron. J. 60 (5) (1968) 503–507, https://doi.org/10.2134/agronj1968.00021962006000050018x.

[10] E.A. Thompson, The estimation of pairwise relationships, Ann. Hum. Genet. 39 (2) (1975) 173–188, https://doi.org/10.1111/j.1469-1809.1975.tb00120.x.

[11] D.T. Meshesha, M. Abeje, Developing crop yield forecasting models for four major Ethiopian agricultural commodities, RSASE 11 (2018) 83–93, https://doi.org/10.1016/j.rsase.2018.05.001.

[12] H. Jiang, H. Hu, R. Zhong, J. Xu, J. Xu, J. Huang, T. Lin, A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: a case study of the US Corn Belt at the county level, Glob. Chang. Biol. 26 (2019), https://doi.org/10.1111/gcb.14885.

[13] F. Babaie Sarijaloo, M. Porta, B. Taslimi, P.M. Pardalos, Yield performance estimation of corn hybrids using machine learning algorithms, AI Agric. 5 (2021) 82–89, https://doi.org/10.1016/j.aiia.2021.05.001.

[14] P. Charoen-Ung, P. Mittrapiyanuruk, Sugarcane yield grade prediction using random forest and gradient boosting tree techniques, in: The 15th JCSSE, 2018, pp. 1–6, https://doi.org/10.1109/JCSSE.2018.8457391.

[15] F. Huber, A. Yushchenko, B. Stratmann, V. Steinhage, Extreme gradient boosting for yield estimation compared with deep learning approaches, Comput. Electron. Agric. 202 (2022) 107346, https://doi.org/10.1016/j.compag.2022.107346.

[16] S.N. Khan, D. Li, M. Maimaitijiang, A Geographically weighted random forest approach to predict corn yield in the US Corn Belt, Remote Sens. 14 (12) (2022) 2843, https://doi.org/10.3390/rs14122843.

[17] A.M. Ali, M. Abouelghar, A.A. Belal, N. Saleh, M. Yones, A.I. Selim, I. Savin, Crop yield prediction using multi sensors remote sensing, The EJRS 25 (3) (2022) 711–716, https://doi.org/10.1016/j.ejrs.2022.04.006.

[18] P. Nevavuori, N. Narra, P. Linna, T. Lipping, Crop yield prediction using multitemporal UAV data and spatio-temporal deep learning models, Remote Sens. 12 (23) (2020) 4000, https://doi.org/10.3390/rs12234000.

[19] P. Muruganantham, S. Wibowo, S. Grandhi, N.H. Samrat, N. Islam, A systematic literature review on crop yield prediction with deep learning and remote sensing, Remote Sens. 14 (9) (2022) 1990, https://doi.org/10.3390/rs14091990.

[20] X. Dong, Z. Yu, W. Cao, Y. Shi, Q. Ma, A survey on ensemble learning, Front. Comput. Sci. 14 (2) (2020) 241–258, https://doi.org/10.1007/s11704-019-8208-z.

[21] A.P. Marques Ramos, L. Prado Osco, D. Elis Garcia Furuya, W. Nunes Gonçalves, D. Cordeiro Santana, L. Pereira Ribeiro Teodoro, H. Pistori, A random forest ranking approach to predict yield in maize with UAV-based vegetation spectral indices, Comput. Electron. Agric. 178 (2020) 105791, https://doi.org/10.1016/j.compag.2020.105791.

[22] M. Shahhosseini, G. Hu, S.V. Archontoulis, Forecasting corn yield with machine learning ensembles, Front. Plant Sci. 11 (2020), https://doi.org/10.3389/fpls.2020.01120.

[23] Y. Everingham, J. Sexton, D. Skocaj, G. Inman-Bamber, Accurate prediction of sugarcane yield using a random forest algorithm, Agron. Sustain. Dev. 36 (2) (2016) 27, https://doi.org/10.1007/s13593-016-0364-z.

[24] B. Peng, K. Guan, M. Pan, Y. Li, Benefits of seasonal climate prediction and satellite data for forecasting U.S. maize yield, Geophys. Res. Lett. 45 (18) (2018) 9662–9671, https://doi.org/10.1029/2018GL079291.

[25] I. Colomina, P. Molina, Unmanned aerial systems for photogrammetry and remote sensing: a review, ISPRS 92 (2014) 79–97, https://doi.org/10.1016/j.isprsjprs.2014.02.013.

[26] C. Zhang, J.M. Kovacs, The application of small unmanned aerial systems for precision agriculture: a review, Preci. Agric. 13 (6) (2012) 693–712, https://doi.org/10.1007/s11119-012-9274-5.

[27] H. García-Martínez, H. Flores-Magdaleno, R. Ascencio-Hernández, et al., Corn grain yield estimation from vegetation indices, canopy cover, plant density, and a neural network using multispectral and RGB images acquired with unmanned

aerial vehicles, Agric 10 (7) (2020) 277, https://doi.org/10.3390/agriculture10070277.
[28] P. Killeen, I. Kiringa, T. Yeap, P. Branco, Corn grain yield prediction using UAV-based high spatiotemporal resolution imagery, machine learning, and spatial cross-validation, Remote Sens. 16 (4) (2024) 683, https://doi.org/10.3390/rs16040683.
[29] I. Zualkernan, D.A. Abuhani, M.H. Hussain, J. Khan, M. ElMohandes, Machine learning for precision agriculture using imagery from unmanned aerial vehicles (UAVs): a survey, Drones 7 (6) (2023) 382, https://doi.org/10.3390/drones7060382.
[30] S. Sunoj, J. Cho, J. Guinness, J. van Aardt, K.J. Czymmek, Q.M. Ketterings, Corn grain yield prediction and mapping from unmanned aerial system (UAS) multispectral imagery, Remote Sens. 13 (19) (2021) 3948, https://doi.org/10.3390/rs13193948.
[31] J. Geipel, J. Link, W. Claupein, Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system, Remote Sens. 6 (11) (2014) 10335–10355, https://doi.org/10.3390/rs61110335.
[32] C. Kumar, P. Mubvumba, Y. Huang, J. Dhillon, K. Reddy, Multi-stage corn yield prediction using high-resolution UAV multispectral data and machine learning models, Agron 13 (5) (2023) 1277, https://doi.org/10.3390/agronomy13051277.
[33] R. Barzin, R. Pathak, H. Lotfi, J. Varco, G.C. Bora, Use of UAS multispectral imagery at different physiological stages for yield prediction and input resource optimization in corn, Remote Sens. 12 (15) (2020) 2392, https://doi.org/10.3390/rs12152392.
[34] Zhang, H., Zhou, Y., Ma, S., and Yemoto K. Optimizing corn yield prediction: integrating multi-temporal UAS data and machine learning. Submitted and under review, 2024.
[35] J. Moreira, C. Soares, A. Jorge, J. Sousa, Ensemble approaches for regression: a survey, ACM CSUR 45 (2012), https://doi.org/10.1145/2379776.2379786, 10:1-10:40.
[36] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[37] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140, https://doi.org/10.1007/BF00058655.
[38] J.H. Friedman, Stochastic gradient boosting, Comput. Stat. Data Anal. 38 (4) (2002) 367–378, https://doi.org/10.1016/S0167-9473(01)00065-2.
[39] Y. Zhang, J. Liu, W. Shen, A review of ensemble learning algorithms used in remote sensing applications, Appl. Sci. 12 (17) (2022) 8654, https://doi.org/10.3390/app12178654.
[40] W. Yang, T. Nigon, Z. Hao, G. Dias Paiao, F.G. Fernández, D. Mulla, C. Yang, Estimation of corn yield based on hyperspectral imagery and convolutional neural network, Comput. Electron. Agric. 184 (2021) 106092, https://doi.org/10.1016/j.compag.2021.106092.
[41] S. Khaki, L. Wang, S.V Archontoulis, A CNN-RNN framework for crop yield prediction, Front. Plant Sci. 10 (2020), https://doi.org/10.3389/fpls.2019.01750.
[42] Inderka, A., Huber, F., Steinhage, V. On convolutional vision transformers for yield prediction. ArXiv. 2024, abs/2402.05557. https://doi.org/10.48550/arXiv.2402.05557.