remote sensing

MDPI

Article

Multimodal Deep Learning Integration of Image, Weather, and Phenotypic Data Under Temporal Effects for Early Prediction of Maize Yield

Danial Shamsuddin [1], Monica F. Danilevicz [1], Hawlader A. Al-Mamun [1], Mohammed Bennamoun [2] and David Edwards [1,*]

[1] Centre for Applied Bioinformatics and School of Biological Sciences University of Western Australia, Perth, WA 6009, Australia; shamsuddin.danial@gmail.com (D.S.); monica.danilevicz@uwa.edu.au (M.F.D.); hawlader.almamun@uwa.edu.au (H.A.A.-M.)

[2] Department of Computer Science and Software Engineering, University of Western Australia, Perth, WA 6009, Australia; mohammed.bennamoun@uwa.edu.au

* Correspondence: dave.edwards@uwa.edu.au

Abstract: Maize (*Zea mays* L.) has been shown to be sensitive to temperature deviations, influencing its yield potential. The development of new maize hybrids resilient to unfavourable weather is a desirable aim for crop breeders. In this paper, we showcase the development of a multimodal deep learning model using RGB images, phenotypic, and weather data under temporal effects to predict the yield potential of maize before or during anthesis and silking stages. The main objective of this study was to assess if the inclusion of historical weather data, maize growth captured through imagery, and important phenotypic traits would improve the predictive power of an established multimodal deep learning model. Evaluation of the model performance when training from scratch showed its ability to accurately predict ~89% of hybrids with high-yield potential and demonstrated enhanced explanatory power compared with previously published models. Shapley Additive explanations (SHAP) analysis indicated the top influential features include plant density, hybrid placement in the field, date to anthesis, parental line, temperature, humidity, and solar radiation. Including weather historical data was important for model performance, significantly enhancing the predictive and explanatory power of the model. For future research, the use of the model can move beyond maize yield prediction by fine-tuning the model on other crop data, serving as a potential decision-making tool for crop breeders to determine high-performing individuals from diverse crop types.

Keywords: machine learning; high-throughput phenotyping; uncrewed aerial vehicles; computer vision; crop breeding

check for updates

Citation: Shamsuddin, D.; Danilevicz, M.F.; Al-Mamun, H.A.; Bennamoun, M.; Edwards, D. Multimodal Deep Learning Integration of Image, Weather, and Phenotypic Data Under Temporal Effects for Early Prediction of Maize Yield. *Remote Sens.* **2024**, *16*, 4043. https://doi.org/10.3390/rs16214043

Academic Editor: Liang Sun

Received: 7 August 2024
Revised: 15 October 2024
Accepted: 24 October 2024
Published: 30 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Crop breeding plays an important role in supporting global food production, with plant breeders and agricultural researchers developing new cultivars with desirable characteristics such as increased yield and stress tolerance [1–3]. As the global population is estimated to reach 10 billion individuals by 2050, increased pressure is placed on breeders to generate improved crop varieties to meet population demand [2,4,5]. However, crop breeders face several challenges in developing improved crop varieties as climate change impacts crop performance and influences growing practices [6]. The shift to unpredictable weather conditions increases the likelihood of volatile crop disease and pest trends and reductions in soil quality [6]. Additionally, the limited genetic diversity in domesticated crops, along with the time and costs associated with traditional breeding methods, poses another challenge for crop breeders [7,8]. As the second most produced crop globally, maize (*Zea mays* L.) has a pivotal role in these breeding programs, and projected yields are expected to increase at 1.6% annually [4,9,10]. Maize also represents 11% of the mean

annual intake of cereals globally, reflecting consumption rates in 161 countries [11]. Given the importance of maize, generating hybrid lines with desirable traits such as high yields is an important objective for crop breeders. Plant phenotyping is used by crop breeders to identify genetic variants that contribute to these desirable traits [12,13]. However, traditional approaches have several limitations, including requiring manual handling, being expensive, being destructive and invasive on samples, and being time-consuming [12,13]. With significant implications for food security, maize can be presented as a compelling case study for using state-of-the-art techniques such as high-throughput phenotyping (HTP) to further enhance current breeding practices.

High-throughput phenotyping can facilitate the precise prediction of genetic merit or phenotypes by using a range of imaging sensors deployed on ground platforms, uncrewed aerial vehicles (UAVs), or satellites that are employed for monitoring plant characteristics during growth under field or greenhouse conditions [14,15]. This capability expedites genetic progress in crop breeding by leveraging the large multivariate datasets generated from HTP platforms, streamlining the breeding process [14,16]. However, dealing with the substantial amount of data generated from HTP may pose a challenge for researchers in identifying and extracting measurable traits. The difficulty arises from the volume of data generated and manually identifying the specific features that may influence model prediction. To address this challenge, machine learning approaches such as deep learning and incorporating data from different sources have been used as powerful tools to improve the accuracy of prediction models [16,17].

Rising attention towards the use of multimodal data within deep learning frameworks mirrors the benefits that stem from combining complementary sources of data [14,17–20]. For example, tabular deep neural network (DNN) architectures have been shown to significantly improve in multimodal frameworks due to the ability of different modalities to share complementary information, known as co-learning [14,21]. This process involves integrating different forms of data—or modalities—such as images and text, into deep learning models [17]. Combining information from different modules provides increased contextual information compared with a single modality by itself [17]. Several studies have documented improved predictive power with multimodal deep learning models [14,18–20,22–24]. For example, Maimaitijiang et al. [22] reported multimodal data fusions from RGB, multispectral, and thermal sensors outperforming single modalities when predicting soybean grain yield, regardless of the machine learning method used, with intermediate-level feature fusion-based DNN displaying the highest performance (relative root mean square error of 15.9% and R^2 of 0.72). In addition, Danilevicz et al. [14] used a multimodal DNN architecture with individual specialised modules on multispectral images and tabular data to combine their information for early maize yield prediction and demonstrated superior predictive power (RMSE 1.07 t/ha, a relative RMSE of 7.60% of 16 t/ha, and R^2 score of 0.73) compared with single-modality models.

While several papers have reported the ability to accurately predict crop yield using deep learning models, there is still an opportunity for further improvement in their models' predictive power [14,18,25]. For example, despite outperforming individual modules, Danilevicz et al. [14] reported lower predicting power for individuals grown in one out of five environments, highlighting the model's difficulty in predicting crop performance under abnormal environmental conditions. It was suggested that incorporating additional temporal and weather information may be beneficial to improve model accuracy and address potential bias [14]. For example, crop yield can be adversely impacted by drought, especially during development periods that are vulnerable to short-term droughts lasting between one and three months [26]. As a result, the addition of weather information under temporal effects can allow prediction models to capture complex relationships between climate and yield potential.

Prior studies using deep learning models tend not to include these important variables due to budget restrictions, time limitations, or lack of data availability. For example, Yang et al. [25] used a convolutional neural network architecture to predict maize grain yield

using hyperspectral imagery only. Although they obtained a classification accuracy of 75.5%, integrating weather data under temporal effects might have further refined the model. Maize development is highly sensitive to environmental stressors such as extreme temperature deviations and frost damage [27]. The inclusion of this information would allow models to understand the effect of meteorological events on maize growth, which are not typically captured by images. In addition, the temporal effect typically captured by time series data (e.g., crop growth captured through imagery) allows models to learn the relationship between time and behaviour prediction [28]. For instance, Shook et al. [29] incorporated time series modelling into their machine learning models to capture weather variability over time and reported a significant role of minimum surface temperature in soybean yield prediction.

Previous machine learning approaches have shown high accuracy in predicting crop yield, but including complementary variables associated with yield could further improve a model's predictive power. More recent studies using multimodal deep learning models that combine image and weather data have shown significant potential, yet they overlook the importance of weather timing on maize yields, especially during the pregrowing seasons at the field trial level [18,20,30]. Incorporating important variables such as historical weather data and phenotypic information may bridge this knowledge gap and support the development of models with more predictive power. Our group has previously developed a multimodal deep learning model predicting maize crop yield at an early developmental stage, employing multispectral images and tabular data such as parental lines and seed source from publicly available datasets [14]. This study aims to further develop this model by incorporating additional images that capture early growth, climate historical data, and phenotypic data. These additional variables were integrated into the multimodal deep learning framework developed for the yield prediction of maize hybrids at an early developmental stage (the period before and during the onset of anthesis and silking stages) to identify more intricate feature patterns and improve the model's predictive power. The main objective of this study was to assess if the inclusion of historical weather data, maize growth captured through imagery, and important phenotypic traits would improve our predictive power. Our goal is to improve the interpretability of our model by using historical climate data to inform crop breeders of the importance of the magnitude and timing of each weather feature (e.g., temperature) on yield before sowing and throughout the development period of maize, which is rarely explored in the literature [18,20,30].

2. Materials and Methods

2.1. Experimental Conditions and Dataset Preparation

The dataset was made publicly available by Texas A&M University as part of the Genomes to Fields (G2F) initiative [31–35]. Detailed descriptions of the datasets can be obtained from [31–35]. In summary, these datasets contain tabular and image data corresponding to field trials conducted between 2017 and 2019 that were planted at College Station, Texas. The meteorological data were obtained from the 2022 G2F Genotype by Environment Prediction Competition compiled from the NASA Power website [36,37]. The NASA Power datasets are based on satellite data and can be generated using single-point coordinates [38,39]. For each year, hybrids were planted in ~1500 plots and arranged in two rows, with each row spanning 0.76 m at a length of 7 m. These trials used a randomised block design approach, consisting of at least two replicates of ~874 individual maize genotypes within three distinct environments per year. Commercial check hybrids were planted around the borders to separate each environment. These distinct environments were subject to different planting dates and fertiliser treatments. For 2017 and 2018, hybrids were sown during either the start of March (Optimal) or April (Delayed). Plants sown in 2017 were further treated with fertiliser at either 43 gallons per acre (Reduced) or 70 gallons per acre (Optimal) of a solution containing 32% nitrogen. In contrast, all 2019 hybrids were sown at the start of March and treated with fertiliser solution at 75 gallons per acre. These different environments were classified under distinct identifiers named G2LA, DG2F, G2

FE, G2F1, and G2FN. Further information on the sowing dates and fertiliser treatment for each environment is captured in Supplementary Table S1. Grain yield was collected ~3–4 months post sowing date using a plot combine harvester per plot. These yields were used as ground truth values during model training and development. A simplified overview of the development of the multimodal deep learning model is shown in Figure 1.

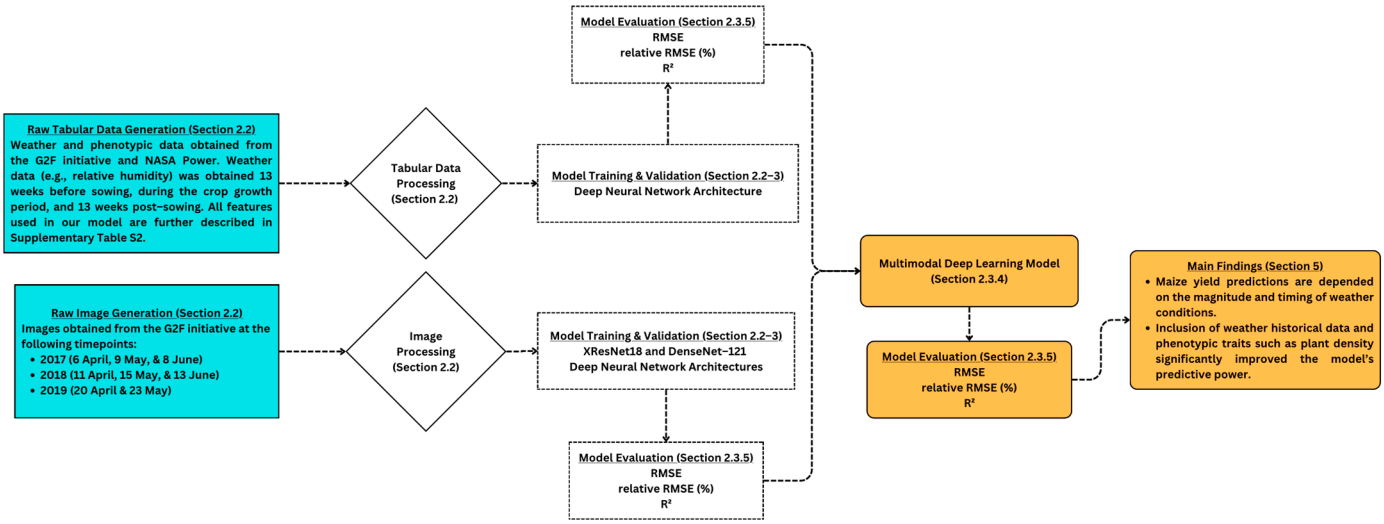


Figure 1. Schematic overview of the multimodal deep learning model development.

During 2017 and 2018, RGB images were collected using a DJI FC300X camera (DJI, Shenzhen, Guangdong, China), with a resolution and focal length of 4000 × 3000 and 3.61 mm, respectively. In contrast, a FC6310S camera was used to collect RGB images during 2019, working at a resolution of 5472 × 3648 and a focal length of 8.8 mm. The cameras were attached to a UAV platform and flown at an altitude of approximately 15–25 metres during 2017–18 and 85 metres during 2019 at midday.

2.2. Image and Tabular Preprocessing for Model Input

A simplified workflow of image preprocessing is shown in Figure 2. The images, taken one and two months after sowing date (at the beginning/onset of the crop's anthesis and silking stages), were joined to generate individual orthomosaics using Metashape (v2.0.1, Agisoft). Metashape allows users to generate orthomosaics by automatically aligning images, optimising their alignment, and building point clouds and digital elevation models using a specific coordinate system. The average ground sample distance was estimated to be 1.04 cm/pixel, 1.09 cm/pixel, and 2.26 cm/pixel for orthomosaics created using images from 2017, 2018, and 2019, respectively. Shapefiles containing information on the geographical location of individual plots were constructed using the plotshpcreate library available in R Studio (v2022.12.0 + 353, Posit Software) [40]. The resulting orthomosaics and shapefiles produced for each timepoint were visualised and manually aligned using QGIS (v3.28.7-Firenze, Free Software Foundation). Once aligned on QGIS, individual plot images across the two timepoints were trimmed from the orthomosaic, rotated horizontally, split into four quadrants, and stacked on top of each other using the custom Python script files, 1_imageclean.py, 2_stacktimepoints.py, and 3_converttojpg.py. All images were inspected before model input, with distorted images removed from the final analysis. This process generated a total of 8296 images of ~320 × 300 pixels, representing the hybrid phenotype throughout the two months and serving as a suitable input for the image-specific deep learning module. Before module input, our image dataset underwent several transformations including flipping, resizing to a width of 360 pixels, and normalising its numerical vectors using FastAI's Normalize function (v2.7.12) [41].

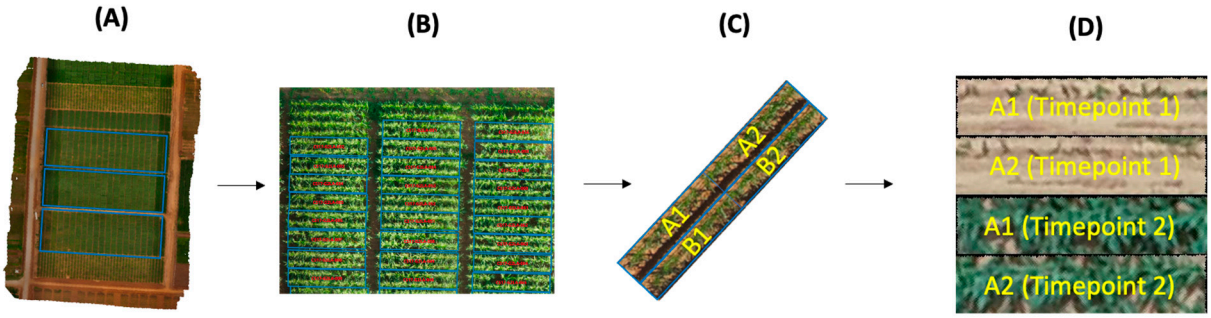


Figure 2. Overview schematic of the image preprocessing workflow: (**A**) Orthomosaic generation; Metashape (v2.0.1, Agisoft) was used to combine images taken from UAV platforms. (**B**) Visualisation of shapefiles on orthomosaics using QGIS (v3.28.7-Firenze, Free Software Foundation); shapefiles were generated using the plotshpcreate package on R software (v2022.12.0 + 353, Posit Software) and imported alongside their respective orthomosaic for each year in QGIS. (**C**) Cropped images generated from QGIS; individual plot images were cropped from aligned shapefiles using QGIS (v3.28.7-Firenze, Free Software Foundation) and were further rotated horizontally and split into four quadrants using a custom Python script. A1–2 and B1–2 represent replicates of the same hybrid lineage. (**D**) Final processed image; the quadrants corresponding to one replicate stacked on top of each other for all timepoints. Processed images will serve as inputs to the image-specific model, where the pixel values will used for feature extraction and model learning.

Parental lines, weather, and other environmental information were captured as tabular data for each year. The features were selected as input in the model, as shown in Supplementary Table S2 [42–45]. These data were further preprocessed to handle data integrity issues, including data gaps, erroneous values, and inappropriate formatting. Missing values in plots were input using values from their corresponding replicate, with plots removed from analysis if no replicate values were found. Daily climate data was available for all three years, with each weather input being either summed or averaged weekly, as described in Supplementary Table S2. Data during the 13 weeks before sowing, during the crop growth period, and 13 weeks post sowing were captured in the tabular input. Following quality control, a total of 636 plots (including replicates) were removed from analysis, with 8398 plots available for use as model input. Before model input, FastAI's Categorify and Normalize functions (v2.7.12) [41] were used to automatically convert categorical features into numerical vectors, and the continuous variables underwent mean subtraction and then division by standard deviation to ensure the dataset was normalised with uniform scaling.

Model validation and evaluation involved separating the processed dataset into a training, validation, and holdout set. The training set was used to train the model parameters, and the validation set was used to fine-tune the model hyperparameters. The holdout set was used to evaluate the final model performance. The final dataset contained a total of 8276 samples from all three years after alignment with the samples used within the tabular and image datasets. The holdout set was carefully selected to contain data unseen from the training or validation datasets, with 10% of samples chosen from the total dataset. The remaining dataset was further split into a training and validation set at a ratio of 70:30 using the RandomSplitter function available from the FastAI library (v2.7.12) [41].

The processed datasets served as inputs for their respective models, which had a batch size of 64. These models were developed using Python (v3.7.11) on Jupyter Notebooks (v6.5.7) with the aid of the FastAI library and Pawsey Supercomputing Resources.

2.3. Machine Learning Models

2.3.1. Model Hyperparameter Optimisation

To maximise the performance of each model, their hyperparameters were optimised using the Optuna Framework (v2.3) [46]. This package tests different combinations of the

hyperparameters and returns the combination with the best model performance. For both models, fifty trials were conducted to explore different hyperparameter combinations. This includes randomly choosing a learning rate value between 1×10^{-4} to 1×10^{-2}, weight decay values between 1×10^{-6} to 1×10^{-3}, and optimiser functions including "ranger", "Adam", "RAdam", and "QHAdam" [47–50].

2.3.2. Image Model Architecture

Several established architectures including XResNet and DenseNet are available in the FastAI and PyTorch libraries and serve as foundational models specific for image-based data for evaluation and to be further refined using Optuna [41,46,51–53]. In this study, we compared the model performance of XResNet with DenseNet, due to their respective strengths for computer vision tasks, especially in agriculture [54–57]. The ResNet architecture was employed as it was successfully integrated into the previous iteration of the multimodal deep learning model [14]. In addition, the ResNet architecture developed by He et al. [51] remains a popular framework for machine learning tasks, serving as a gold standard for numerous scientific publications as a baseline for comparison with more recently developed architectures [58]. ResNet's popularity also stems from its simple design, which uses a single identity shortcut to retain feature information by skipping residual blocks [55]. This simple design has drawbacks, as the use of the shortcut limits the model's learning capabilities [55]. DenseNet architectures have been extensively used in image classification tasks such as weed and crop identification and plant disease detection with high accuracies [54,56,57]. In DenseNet, the introduction of a dense connectivity between all layers ensures that each layer receives input from all previous layers which facilitates the efficient transfer of information across model layers [52,55]. DenseNet has been shown to outperform ResNet architectures by achieving lower error rates with fewer parameters [52,55]. Compared with ResNet, DenseNet requires more computational power and involves a longer training period [55]. Due to their strengths and limitations, DenseNet serves as an appropriate comparator with ResNet for maize yield estimation. Transfer learning was applied for each model being evaluated. The best-performing image–deep neural network (image-DNN) was integrated into the multimodal framework.

ResNet Model Architecture

A modified version of the ResNet architecture based on [59], coined 'XResNet' by the FastAI package [41], was used to process the RGB images, and its performance was evaluated. This modified architecture comprises 18 layers and uses optimised convolutional layers, with an integrated self-attention mechanism that allows the model to identify relevant regions of the input image for prediction [60]. The structure of the XResNet-18 architecture is captured in Figure 3.

DenseNet Model Architecture

The DenseNet framework was used to process our RGB images, and its performance was evaluated alongside ResNet [52]. The structure of the DenseNet-121 architecture is captured in Figure 3. The framework comprises 121 layers, with dense blocks connecting each layer [61]. Four dense blocks are used, with transition layers in between to aid in downsampling the feature maps and preparing them for subsequent layers [61].

2.3.3. Tabular Model Architecture

As previously mentioned, the performance of Tabular-DNN (tab-DNN) architectures in multimodal frameworks can be enhanced through co-learning [14,21]. As such, tab-DNN architectures were chosen over Random Forest and XGBoost algorithms for the multimodal framework used within this study. All features listed in Supplementary Table S2 were used to train the tab-DNN. The structure of the tab-DNN is captured in Figure 3.

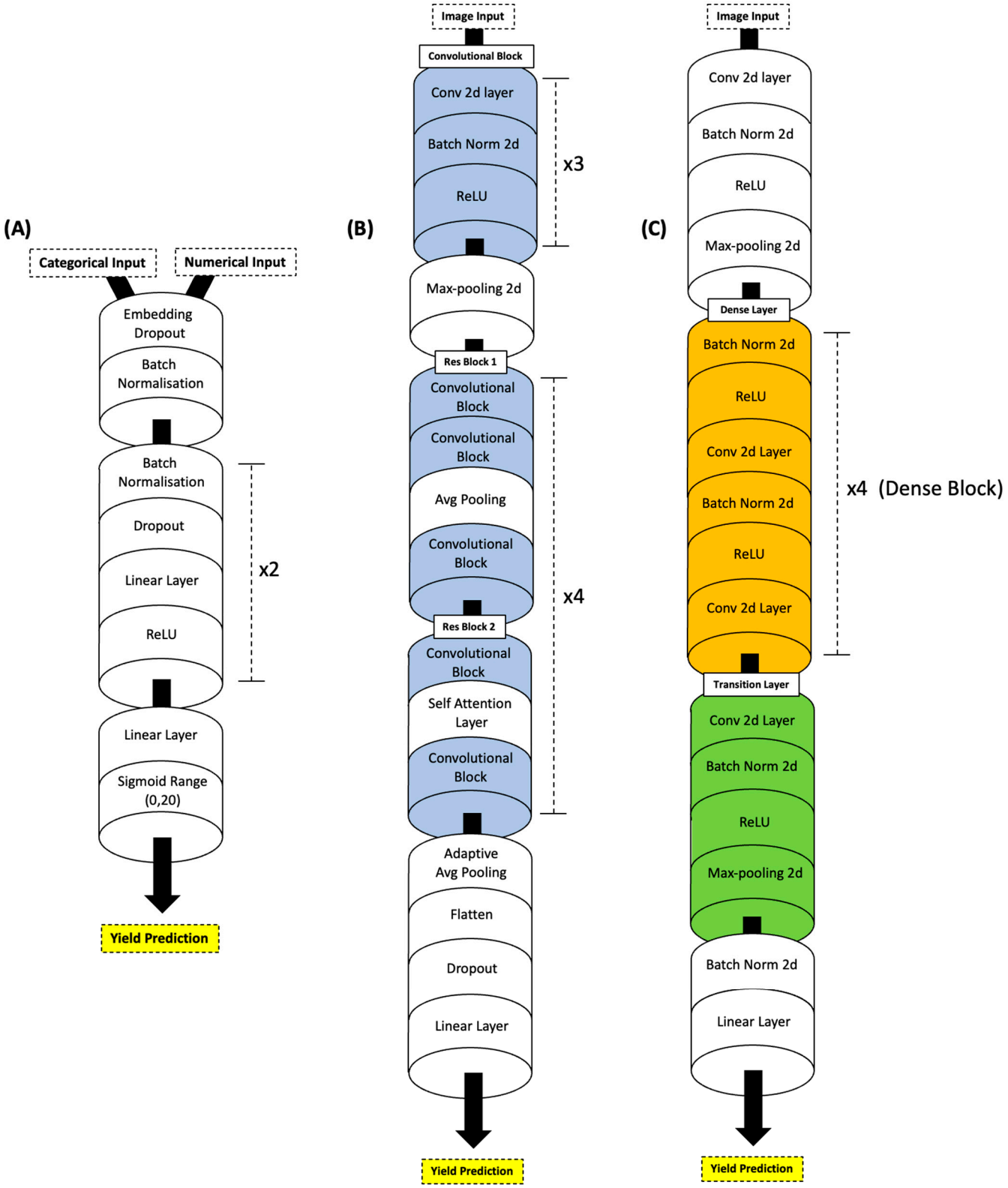


Figure 3. Overview of the layers used within each modality. All modules generate a target prediction. Repeated layers are visualised by a dotted line: (**A**) tab-DNN architecture with embedding layers for each categorical feature. (**B**) XResNet18 Image-DNN architecture; this framework uses a modified ResNet architecture with 18 layers, including a self-attention layer. (**C**) DenseNet-121 Image-DNN architecture; This framework uses four dense blocks, each with 6, 12, 24, 16 dense layers, respectively. Each dense block is connected by a transition layer.

2.3.4. Multimodal Architecture

The final Image-DNN module identifies image feature relationships that are important to the model predictions, which is reflected by their model weights. Similarly, the model weights from the top-performing tab-DNN module indicate the features (e.g., relative humidity) most relevant to model outcomes. The multimodal deep learning architecture combines the optimised Image-DNN and tab-DNN modules using a fusion module previously employed by Danilevicz et al. [14]. The final layer's weights from both individual modules are concatenated and are used as input to the fusion module for further processing, which consists of two linear layers and the activation function 'ReLU' [14], as shown in Figure 4. The combined input from both modules, allows the multimodal deep learning model to leverage complementary information, which may improve the accuracy of model predictions. Through this process, the combination of visual features (e.g., leaf density), phenotypic features (e.g., plant density), and meteorological features (e.g., temperature) over time provides the model with a more holistic perspective on plant health and the environmental factors influencing yield. The gradient-blending technique was employed and uses scaling weights to adjust the loss functions from multiple architectures based on their convergence behaviour [62]. The appropriate value for each scaling weight used for each modality was determined using Optuna. Target predictions are generated for each modality upon completion of model training for comparison and evaluation.

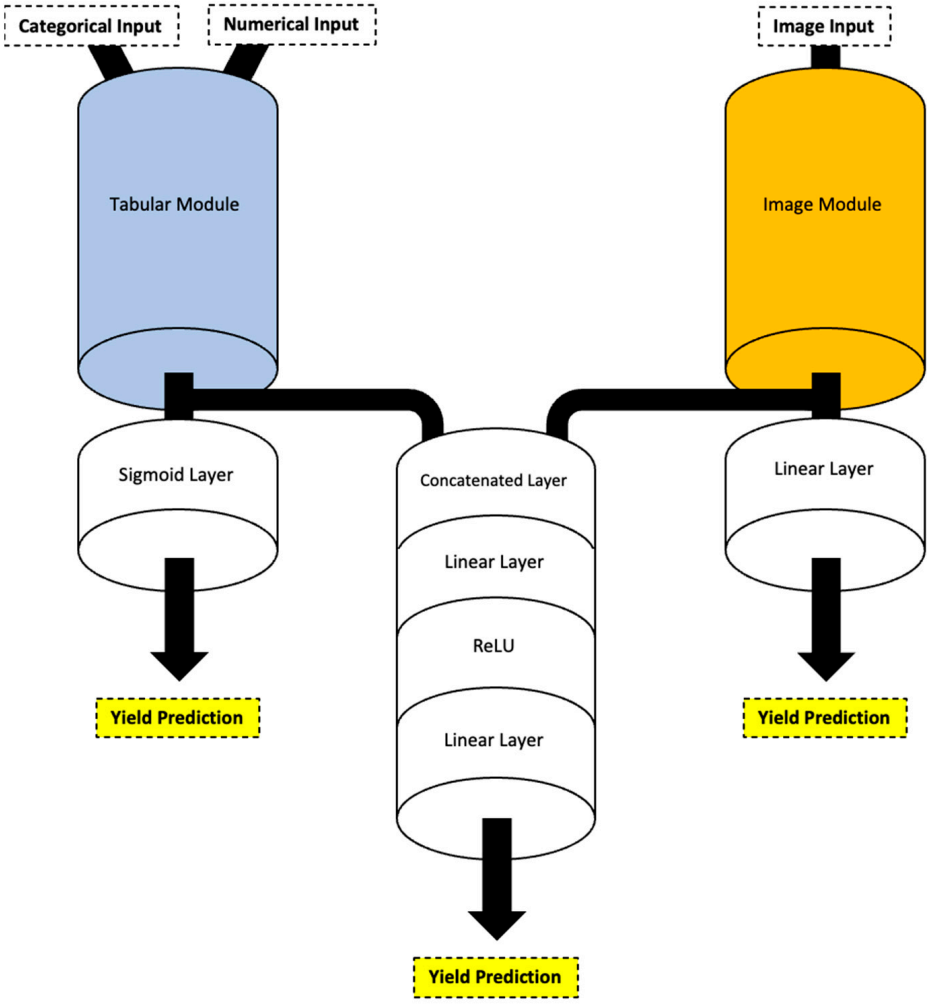


Figure 4. Overview of the fusion module used within the multimodal architecture. The last layer weights from each module are concatenated, then undergo linear and non-linear transformations before generating a yield prediction.

Upon finalising each module's architecture, several iterations of the multimodal deep learning model were evaluated. The multimodal framework involving pretrained modules incorporates fully trained weights during model training from both modalities using FastAI's SaveModelCallback function (v2.7.12), which monitors the valid loss at every epoch and only saves the model at the epoch with the lowest valid loss. In contrast, the multimodal framework involving training from scratch does not use these trained weights during model training. Multimodal model performance was evaluated through a comparison of training outcomes between training from scratch and with pretrained modules after training for 100 epochs. Additionally, the performance of the multimodal framework was compared with the individual modalities.

2.3.5. Model Evaluation

Fivefold stratified cross-validation is an evaluation method that splits up the dataset into five folds of equal sample numbers using the Scikit-learn library (v0.20.3) [63]. Initial model assessment was conducted by using four folds as the training set, leaving one to act as the validation set. Each fold contained an equal distribution of samples across all years. Model training iterates through each fold until all have been used as the validation set. FastAI's EarlyStoppingCallback function (v2.7.12) [41] was used during cross-validation, which monitors the valid loss for each epoch, with training halted if the valid loss does not increase by 0.01 every 20 epochs. Performance on the validation dataset was evaluated by averaging the evaluation metrics generated from all folds.

This set of evaluation metrics employed to evaluate each model's ability to predict the holdout set (i.e., final model performance) includes the root mean squared error (RMSE), relative RMSE (%), and coefficient of determination (R^2) using the Scikit-learn library (v0.20.3) [63], where the following equations were used:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(yi - \hat{y}\,i\right)^2}{n-1}} \quad (1)$$

$$RMSE\ \% = \frac{RMSE}{ymax - ymin} \times 100 \quad (2)$$

$$R^2 = 1 - \frac{\sum_i (yi - \hat{y}i)^2}{\sum_i (yi - \overline{y})^2} \quad (3)$$

where the target yield and predicted yield are symbolised as y and ŷ, respectively, for a given sample 'i'. N represents the total number of samples used. The values for ymax, ymin, ȳ were obtained by calculating the target yields' maximum, minimum, and average, respectively.

Feature importance analysis was conducted using the framework developed by Lundberg and Lee [64], called SHAP (Shapley additive explanations), which assigns an importance value per feature input for the model prediction. SHAP analysis was conducted using the fastinference package (v0.0.36) [65]. SHAP was selected for this study as it evaluates each feature individually, providing a more comprehensive evaluation of the model compared with drop column or permutation importance [66].

The model's ability to classify different yield ranges as classes was also evaluated using confusion matrices. These categories include high-yield (>10 t/ha), medium-yield (6 to 9.99 t/ha), and low-yield hybrids (<6 t/ha). The yield classes were defined previously by Danilevicz et al. [14] and were used in this study to enable performance comparison between models.

3. Results

3.1. Hybrid Characteristics

The ground truth yields corresponding to each hybrid are depicted in Figure 5. In our processed dataset, 874 distinct maize hybrids were sown throughout the three years, originating from the combination of 815 unique parental lines (Table 1). The hybrids sown in 2018 showed the greatest average yield ($p < 0.05$; 11.27 ± 2.04), followed by 2017 ($p < 0.05$; 9.21 ± 2.70) and 2019 ($p < 0.05$; 7.17 ± 1.18) (Figure 5A). When separating each yield by the environment, the 'G2FN' hybrids sown in 2018 possessed the greatest mean yield (11.86 ± 1.48), with 'G2LA' hybrids sown in 2017 possessing the lowest mean yield (5.83 ± 1.13) (Figure 5B). There is a trend for the mean yield to increase in 2017 and 2018, depending on the order of hybrids sown from each environment ($p < 0.05$; Figure 5B). For example, the mean yields sown in 2017 increased from G2LA, followed by DG2F and G2FE. In contrast, the mean yields sown in 2018 increased from G2LA, followed by G2FE and G2FN. Interestingly, the 'G2LA' hybrids showed a statistically significant relationship for yield to increase with year ($p < 0.05$; Figure 5B).

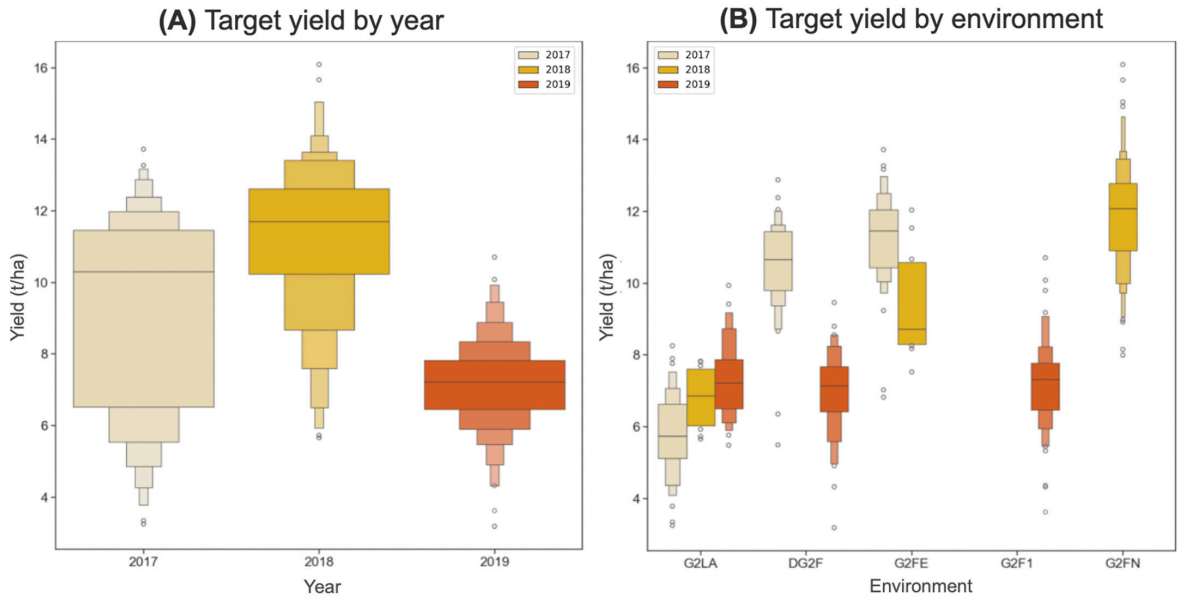


Figure 5. Maize hybrid and yield distribution throughout the years visualised using a boxplot portraying ground truth yields by year (**A**), boxplot portraying ground truth yields by environment and year (**B**), and a Venn diagram portraying the number of maize hybrids grown each year.

Table 1. Evaluation metrics for the tab-DNN model, per dataset. Fivefold cross-validation was used to generate the metric values for the validation dataset, represented as the mean and standard deviation.

Year(s)	No. of Unique Samples
2017	179
2018	17
2019	164
2017 and 2018 only	102
2017 and 2019 only	0
2018 and 2019 only	411
All Years	1
Total	874

3.2. Hyperparameter Optimisation

For the Image-DNN model, the DenseNet121 architecture returned the lowest valid loss compared with other architectures tested, as shown in Table 2. The optimised Image-DNN model comprised 121 layers, a learning rate of 0.000129, RAdam as the optimiser

function, and a weight decay value of 1.37×10^{-4}. The optimised Tabular-DNN model returned a valid loss of 0.968 using a learning rate of 0.000185, ranger as the optimiser function, and a weight decay value of 1.425×10^{-6} as its set of hyperparameters.

Table 2. The best hyperparameter combination for each Image-DNN architecture tested using Optuna. The model with the lowest valid loss is bolded.

Architecture	Learning Rate	Optimiser Function	Weight Decay	Valid Loss
XResNet18	0.000451	RAdam	1.31×10^{-4}	1.024
XResNet34	0.000155	ranger	3.82×10^{-6}	1.025
XResNet50	0.005622	Adam	1.13×10^{-5}	1.032
DenseNet121	**0.000129**	**RAdam**	$\mathbf{1.37 \times 10^{-4}}$	**1.012**
DenseNet169	0.000201	QHAdam	1.33×10^{-5}	1.018
DenseNet201	0.000167	RAdam	1.96×10^{-6}	1.020

3.3. Tab-DNN Results

The evaluation metrics generated for the tab-DNN model per dataset are shown in Table 3. For both datasets, the R^2 values indicate our model can explain ~77–78% of the total variation occurring in the dependent variables when predicting yield. The relative decrease in the RMSE and relative RMSE of ~22% between the holdout and validation datasets indicates the predictive performance of the validation dataset is superior compared with the holdout dataset. Assessing final model performance, year-by-year R^2 score analysis showed that 2017 displayed the greatest explainable variance at 80%, followed by 2018 at 52% and 2019 at −0.02% (Figure 6A). Using *t*-tests for statistical analysis revealed the mean absolute error for 2017 (0.299 ± 1.16) to be statistically different from other years ($p < 0.05$). Delving deeper, the G2FE treatment in 2017 was shown to possess the greatest average absolute error (0.82 ± 1.09) and was the only treatment group to be statistically different from all other treatment groups for each year ($p < 0.05$; Figure 6B). As shown in Figure 6C, the model accurately identified 92% of high-yielding hybrids, followed by medium-yield hybrids at 86%, and low-performing hybrids at 27%.

Table 3. Evaluation metrics for the tab-DNN model per dataset. Fivefold cross-validation was used to generate the metric values for the validation dataset, represented as the mean and standard deviation.

	RMSE	RMSE %	R^2
Validation Dataset	1.00 ± 0.02	7.28 ± 0.17	0.78 ± 0.005
Holdout Dataset	1.28	9.27	0.77

Feature importance analysis using SHAP values evaluated each feature's impact on the model prediction, with plant density showing the greatest impact (Figure 7A,B). The top ten features that displayed the greatest impact on model predictions include plant density (mean absolute SHAP value = 0.229), vertical arrangement of plot (mean absolute SHAP value = 0.226), days to anthesis (mean absolute SHAP value = 0.185), Parental Line 1 (mean absolute SHAP value = 0.166), Dew/Frost Point (Week 9; mean absolute SHAP value = 0.146), Max-Min Temperature Difference (Week 3; mean absolute SHAP value = 0.144), All Sky Surface Shortwave Downward Irradiance (Week −1; mean absolute SHAP value = 0.144), Growing degree-days (Week 4; mean absolute SHAP value = 0.143), relative humidity (Week 13; mean absolute SHAP value = 0.141), and Dew/Frost Point (Week 4; mean absolute SHAP value = 0.132) (Figure 7A). The directional impact of each feature was also captured in Figure 7B, where high values of plant density, days to anthesis, growing degree days (Week 4), and relative humidity (Week −13) were associated with positive model predictions. In contrast, low values of the vertical arrangement of the plot, Dew/Frost Point (Week 9), Max-Min Temperature Difference (Week 3), All Sky Surface

Shortwave Downward Irradiance (Week 1), and Dew/Frost Point (Week 4) were associated with positive model predictions (Figure 7B).

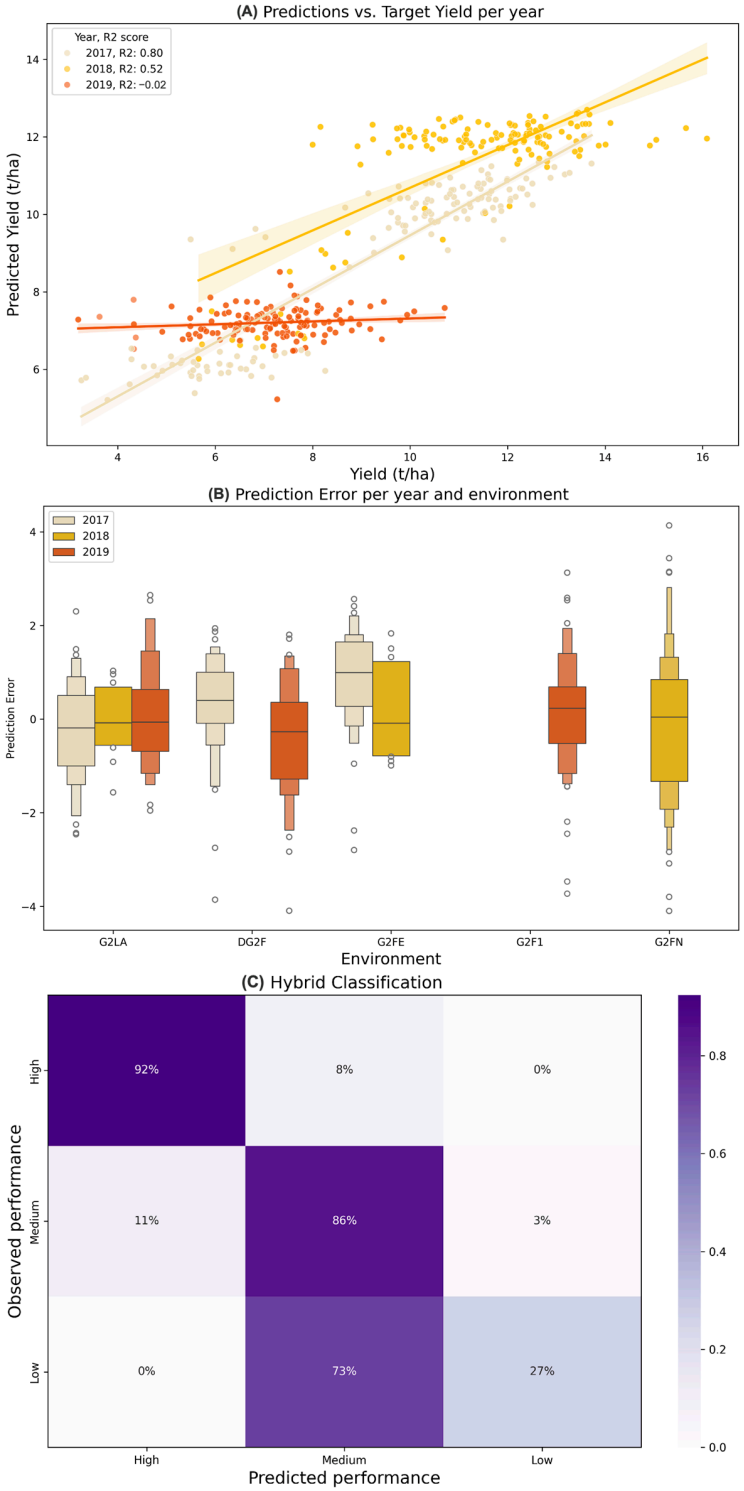


Figure 6. Tab-DNN model holdout results: (**A**) Comparison between predictions (*y*-axis) vs. ground truth yields (*x*-axis) for each sample per year, measured at tonne/ha. (**B**) Comparison between prediction error (*y*-axis) vs. environment (*x*-axis) for each sample per year. Calculated by subtracting the predicted yield from ground truth. (**C**) Confusion matrix showing the percentage of correctly predicted hybrids that were classified as either high (>10 tonne/ha), medium (6–9.99 tonne/ha), or low (<6 tonne/ha). Ground truth classifications are represented on the *y*-axis, with predicted performance on the *x*-axis.

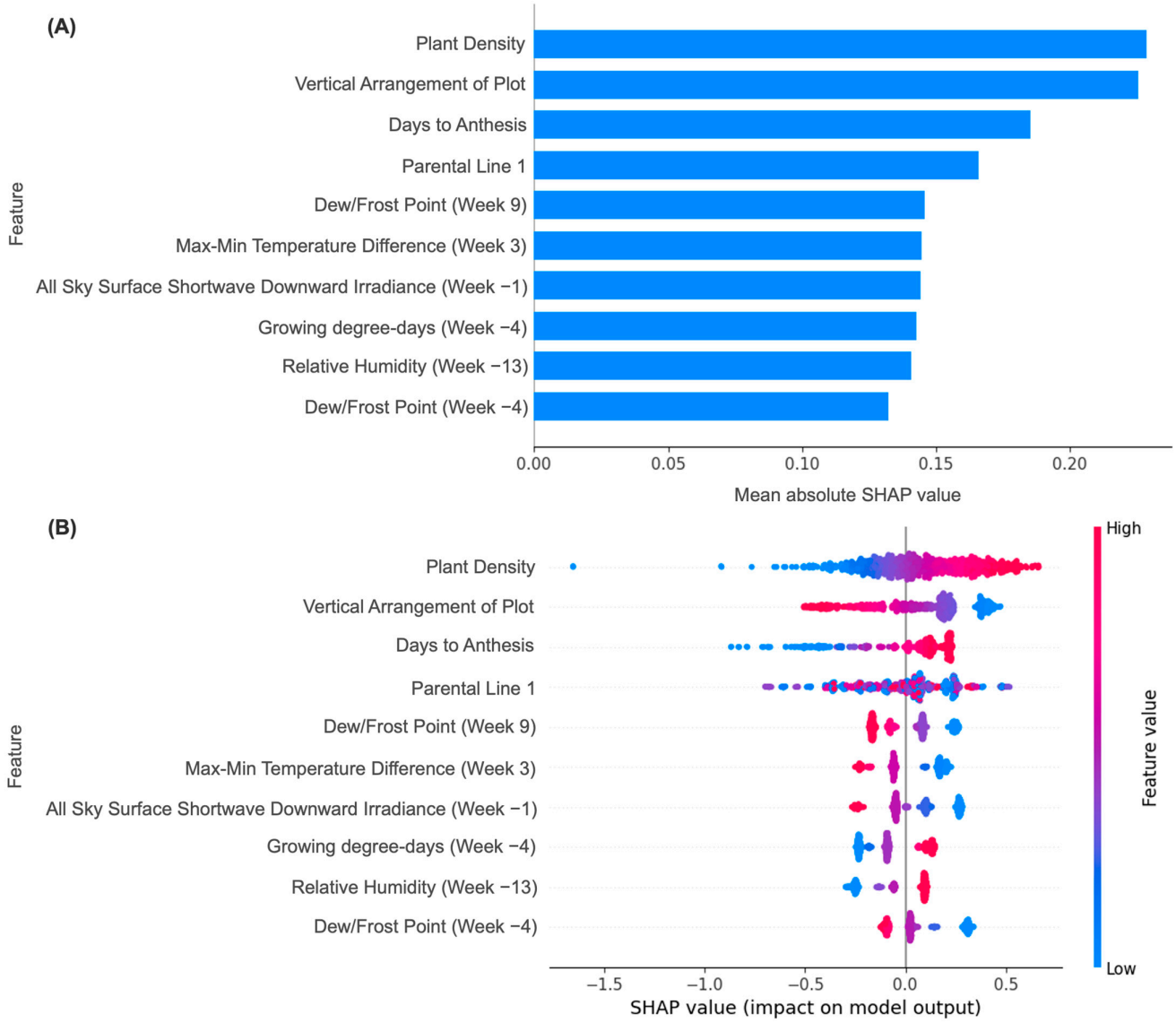


Figure 7. The top ten features based on SHAP values: (**A**) represents the top ten features and their average impact on tab-DNN model holdout predictions by mean absolute SHAP value. (**B**) represents the top ten features and their directional impact on tab-DNN model holdout predictions by SHAP value per sample. Features are further described in Supplementary Table S2.

3.4. Image-DNN Results

The evaluation metrics generated for the Image-DNN model per dataset are shown in Table 4. For both datasets, the R^2 values indicate our model can explain ~71–75% of the total variation occurring in the dependent variables when predicting yield. The relative decrease in the RMSE and relative RMSE of ~15% between the holdout and validation datasets indicate the predictive performance of the validation dataset is superior compared with the holdout dataset. Assessing final model performance, year-by-year R^2 score analysis showed that 2017 displayed the greatest explainable variance at 76%, followed by 2018 at 45% and 2019 at 15% (Figure 8D). In Figure 8E, the predicted and ground truth yields display two distinct hybrid groups. This trend is further apparent when analysing the predictive accuracy based on treatment and year (Figure 8A–C). Except for 2019, two treatment groups tend to separate into distinct clusters, with the third group sharing characteristics from both groups. Using *t*-tests for statistical analysis revealed the mean absolute error for 2019 (-0.278 ± 1.06) to be statistically different from other years ($p < 0.05$). Delving deeper, the G2FE treatment in 2017 was shown to possess the greatest average absolute error

(0.79 ± 1.26) and was statistically different from all treatment groups for each year, except for G2FN and G2FE treatments in 2018 ($p < 0.05$; Figure 8F). As previously discussed in Section 3.3, plots were categorised based on their performance to assess the model's capabilities to identify high-yielding hybrids. As shown in Figure 8G, the model accurately identified 87% of high-yielding hybrids, followed by medium-yield hybrids at 82% and low-performing hybrids at 39%.

Table 4. Evaluation metrics for the Image-DNN model, per dataset. Fivefold cross-validation was used to generate the metric values for the validation dataset, represented as the mean and standard deviation.

	RMSE	RMSE %	R^2
Validation Dataset	1.13 ± 0.04	8.23 ± 0.26	0.71 ± 0.02
Holdout Dataset	1.33	9.65	0.75

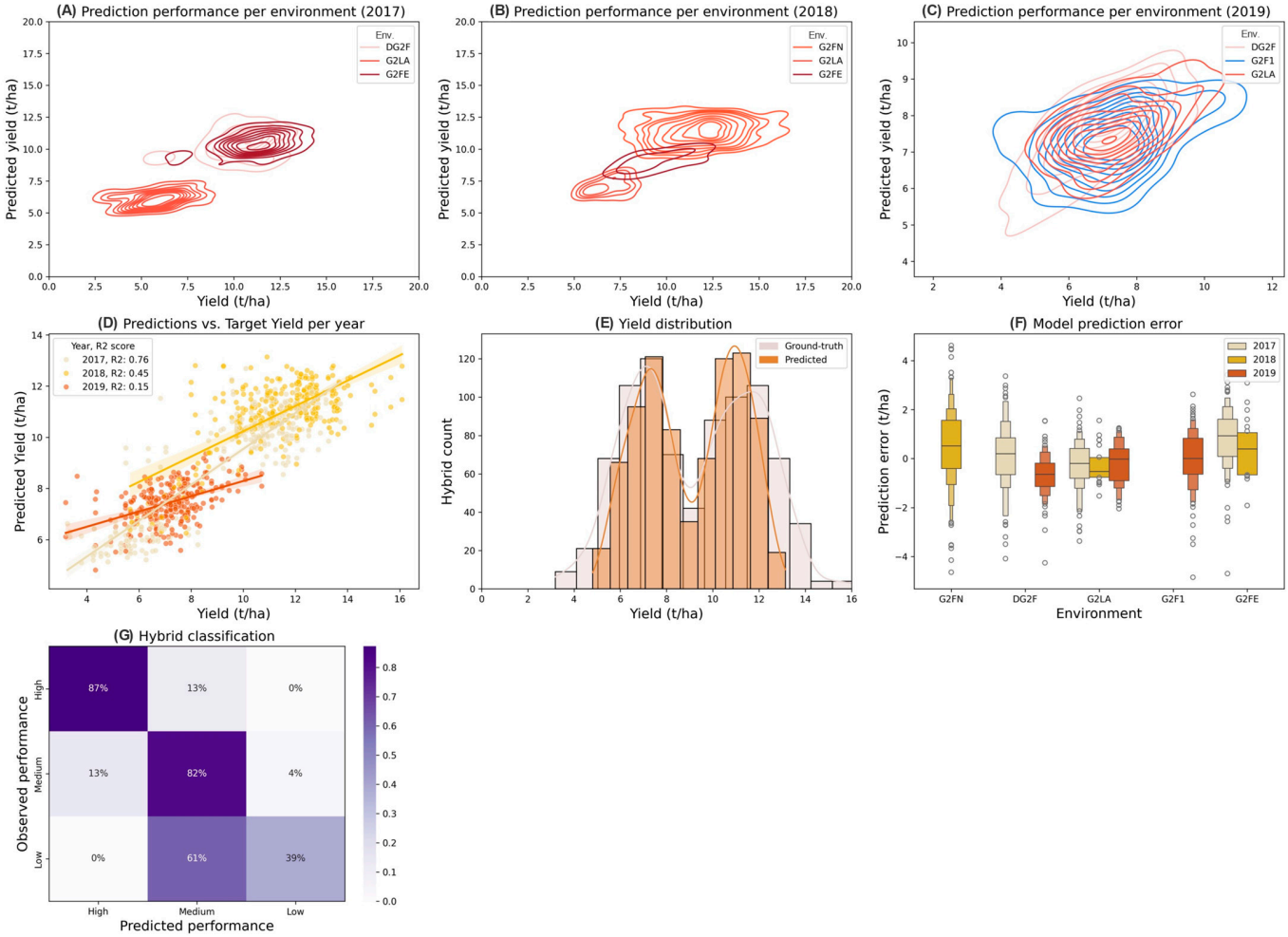


Figure 8. Image-DNN model holdout results: (**A**–**C**) Comparison between predictions (*y*-axis) vs. ground truth yields (*x*-axis) for each sample by environment and year, measured as tonne/ha. (**D**) Comparison between predictions (*y*-axis) vs. ground-true yields (*x*-axis) for each sample per year, measured as tonne/ha. (**E**) Comparison between hybrid count and yield distribution between ground truth and prediction values. (**F**) Comparison between prediction error (*y*-axis) vs. environment (*x*-axis) for each sample per year. Calculated by subtracting the predicted yield from ground truth. (**G**) Confusion matrix showing the percentage of correctly predicted hybrids that were classified as either high (>10 tonne/ha), medium (6–9.99 tonne/ha), or low (<6 tonne/ha). Ground truth classifications are represented on the *y*-axis, with predicted performance on the *x*-axis.

3.5. Multimodal Results

The different modalities were combined into a multimodal deep learning model, with its evaluation metrics presented in Tables 5–7. For each metric, the performance of our current model, when predicting for the validation dataset, outperforms the previous iteration. However, when predicting for the holdout dataset, its performance regarding RMSE and relative RMSE only outperforms the previous iteration within the tabular module (Tables 5 and 6). This trend is not apparent for R^2, as our current model substantially exceeds this metric for every modality compared with its previous iteration (Table 7). Specifically for the tabular component, the current version of the model improves the RMSE by ~37% for both RMSE and relative RMSE when training with pretrained modules. In comparison, the current version of the model improves the RMSE by ~12% for both RMSE and relative RMSE when training from scratch.

Table 5. Performance comparison between the current and previous iteration of the multimodal model based on RMSE. Fivefold cross-validation was used to generate the metric values for the validation dataset, represented as the mean and standard deviation. Weighted predictions were calculated by scaling the predictions produced by each modality based on their assigned weights and summing the predictions together. The best module is displayed in bold. Previous model values extracted from [14].

Multimodal Framework		Validation Dataset		Holdout Dataset	
		Previous Version (RMSE)	Current Version (RMSE)	Previous Version (RMSE)	Current Version (RMSE)
Feature Fusion	Tabular module	1.29 ± 0.20	**0.99 ± 0.03**	1.53	**1.32**
	Image module	1.27 ± 0.17	**1.11 ± 0.03**	**1.14**	1.33
	Fusion module	1.16 ± 0.05	**1.08 ± 0.03**	**1.17**	1.32
	Weighted Prediction	1.13 ± 0.04	**0.96 ± 0.03**	**1.07**	1.26
Feature Fusion with pretrained modules	Tabular module	1.29 ± 0.21	**0.92 ± 0.02**	2.16	**1.34**
	Image module	1.27 ± 0.18	**0.64 ± 0.01**	**1.19**	1.36
	Fusion module	1.14 ± 0.04	**0.63 ± 0.01**	**1.22**	1.36
	Weighted Prediction	1.12 ± 0.02	**0.65 ± 0.01**	**1.21**	1.30

Table 6. Performance comparison between the current and previous iteration of the multimodal model based on relative RMSE (%) of 13.76 t/ha. Fivefold cross-validation was used to generate the metric values for the validation dataset, represented as the mean and standard deviation. The best module is displayed in bold. Weighted predictions were calculated by scaling the predictions produced by each modality based on their assigned weights and summing the predictions together. Previous model values extracted from [14].

Multimodal Framework		Validation Dataset		Holdout Dataset	
		Previous Version (RMSE%)	Current Version (RMSE%)	Previous Version (RMSE%)	Current Version (RMSE%)
Feature Fusion	Tabular module	9.15 ± 1.40	**7.21 ± 0.20**	10.84	**9.60**
	Image module	9.02 ± 1.22	**8.09 ± 0.21**	**8.06**	9.65
	Fusion module	8.24 ± 0.38	**7.87 ± 0.21**	**8.27**	9.58
	Weighted Prediction	8.00 ± 0.31	**6.99 ± 0.20**	**7.60**	9.16
Feature Fusion with pretrained modules	Tabular module	9.10 ± 1.50	**6.67 ± 0.14**	15.31	**9.77**
	Image module	8.96 ± 1.31	**4.68 ± 0.09**	**8.44**	9.88
	Fusion module	8.09 ± 0.25	**4.59 ± 0.09**	**8.63**	9.86
	Weighted Prediction	7.90 ± 0.16	**4.69 ± 0.11**	**8.55**	9.47

Table 7. Performance comparison between the current and previous iteration of the multimodal model based on R^2. Fivefold cross-validation was used to generate the metric values for the validation dataset, represented as the mean and standard deviation. The best module is displayed in bold. Weighted predictions were calculated by scaling the predictions produced by each modality based on their assigned weights and summing the predictions together. Previous model values extracted from [14].

Multimodal Framework		Validation Dataset		Holdout Dataset	
		Previous Version (R^2)	Current Version (R^2)	Previous Version (R^2)	Current Version (R^2)
Feature Fusion	Tabular module	0.64 ± 0.11	**0.78 ± 0.009**	0.47	**0.75**
	Image module	0.65 ± 0.10	**0.72 ± 0.010**	0.71	**0.75**
	Fusion module	0.71 ± 0.02	**0.74 ± 0.010**	0.69	**0.76**
	Weighted Prediction	0.73 ± 0.016	**0.79 ± 0.009**	0.73	**0.78**
Feature Fusion with pretrained modules	Tabular module	0.64 ± 0.12	**0.81 ± 0.006**	−0.07	**0.75**
	Image module	0.65 ± 0.11	**0.91 ± 0.004**	0.67	**0.74**
	Fusion module	0.72 ± 0.02	**0.91 ± 0.004**	0.66	**0.74**
	Weighted Prediction	0.74 ± 0.01	**0.91 ± 0.004**	0.66	**0.76**

No significant differences existed between the predictions generated when training from scratch or using pretrained modules ($p > 0.05$; Figures 9 and 10). However, there were statistically significant differences in mean absolute error across different years and environments ($p < 0.05$). For example, using *t*-tests for statistical analysis revealed the mean absolute error for 2019 to be statistically different from other years ($p < 0.05$). When training from scratch, 2019 hybrids displayed the lowest mean absolute error (-0.08 ± 1.05) compared with 2017 (-0.45 ± 1.21) and 2018 (-0.35 ± 1.37). Similarly, 2019 hybrids displayed the lowest mean absolute error (-0.12 ± 1.06) compared with 2017 (-0.42 ± 1.24) and 2018 (-0.45 ± 1.42) when training using pretrained modules. Delving deeper, the G2FE treatment in 2017 was shown to possess the greatest average absolute error (-0.98 ± 1.15 and -0.95 ± 1.21 for both models, respectively) and was statistically different from all other treatment groups for each year ($p < 0.05$). As previously discussed in Section 3.3, plots were categorised based on their performance to assess the model's capabilities to identify high-yielding hybrids. As shown in Figures 9E and 10E, both models accurately identified 88–89% of high-yielding hybrids, followed by medium-yield hybrids at 83% and low-performing hybrids at 41%.

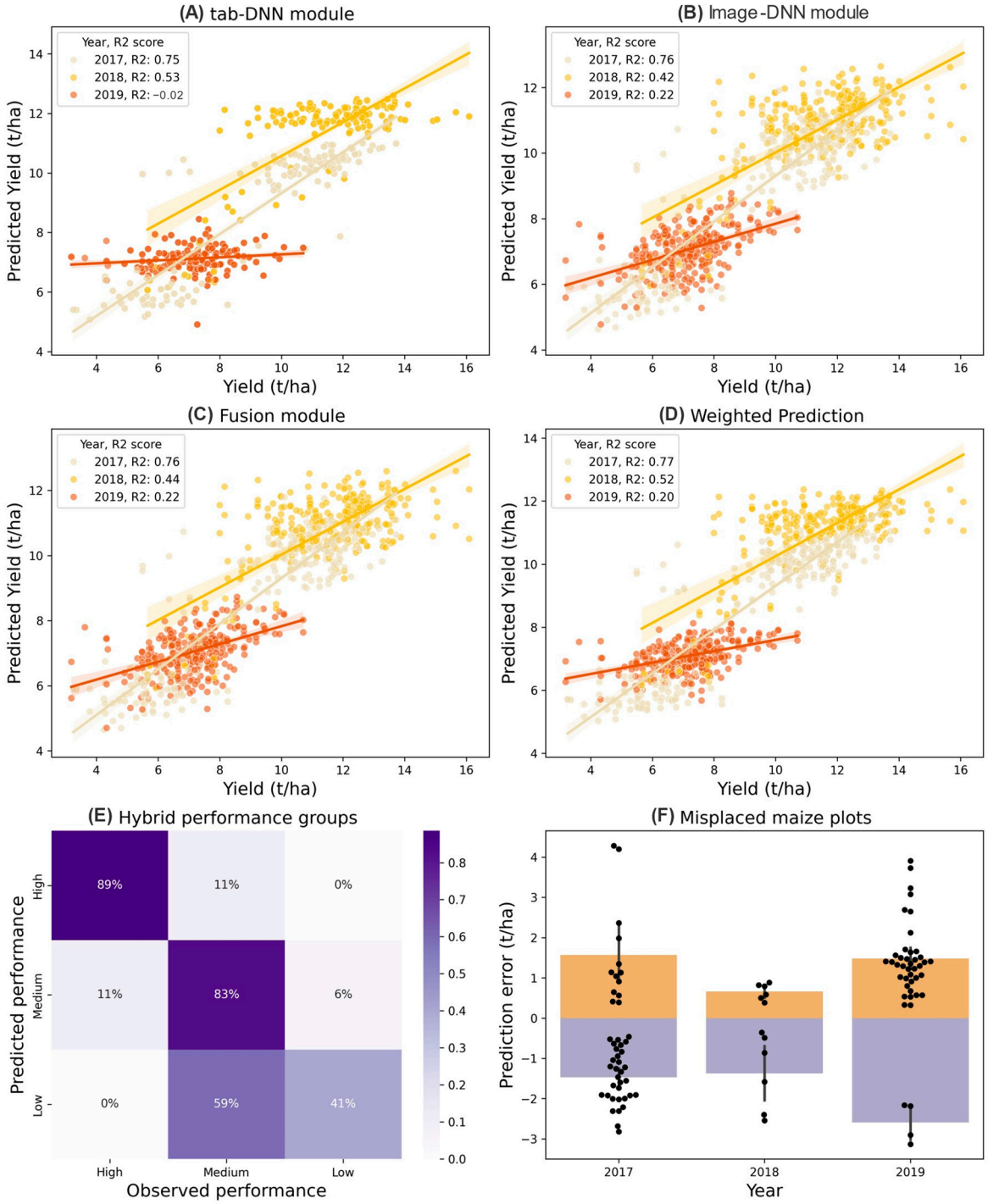


Figure 9. Yield predictions were generated from the multimodal architecture when trained from scratch. Comparison between predictions (*y*-axis) vs. ground truth yields (*x*-axis) for each sample by environment and year, measured at tonne/ha using tab-DNN predictions (**A**), Image-DNN predictions (**B**), fusion module predictions (**C**), and weighted predictions (**D**). (**E**) Confusion matrix showing the percentage of correctly predicted hybrids that were classified as either high (>10 tonne/ha), medium (6–9.99 tonne/ha), or low (<6 tonne/ha). Ground truth classifications are represented on the *y*-axis, with predicted performance on the *x*-axis. (**F**) Classification error plot showcasing hybrids predicted as medium yield but with a ground truth class of high yield (light purple), and hybrids predicted as medium yield but with a ground truth class of low yield (mustard).

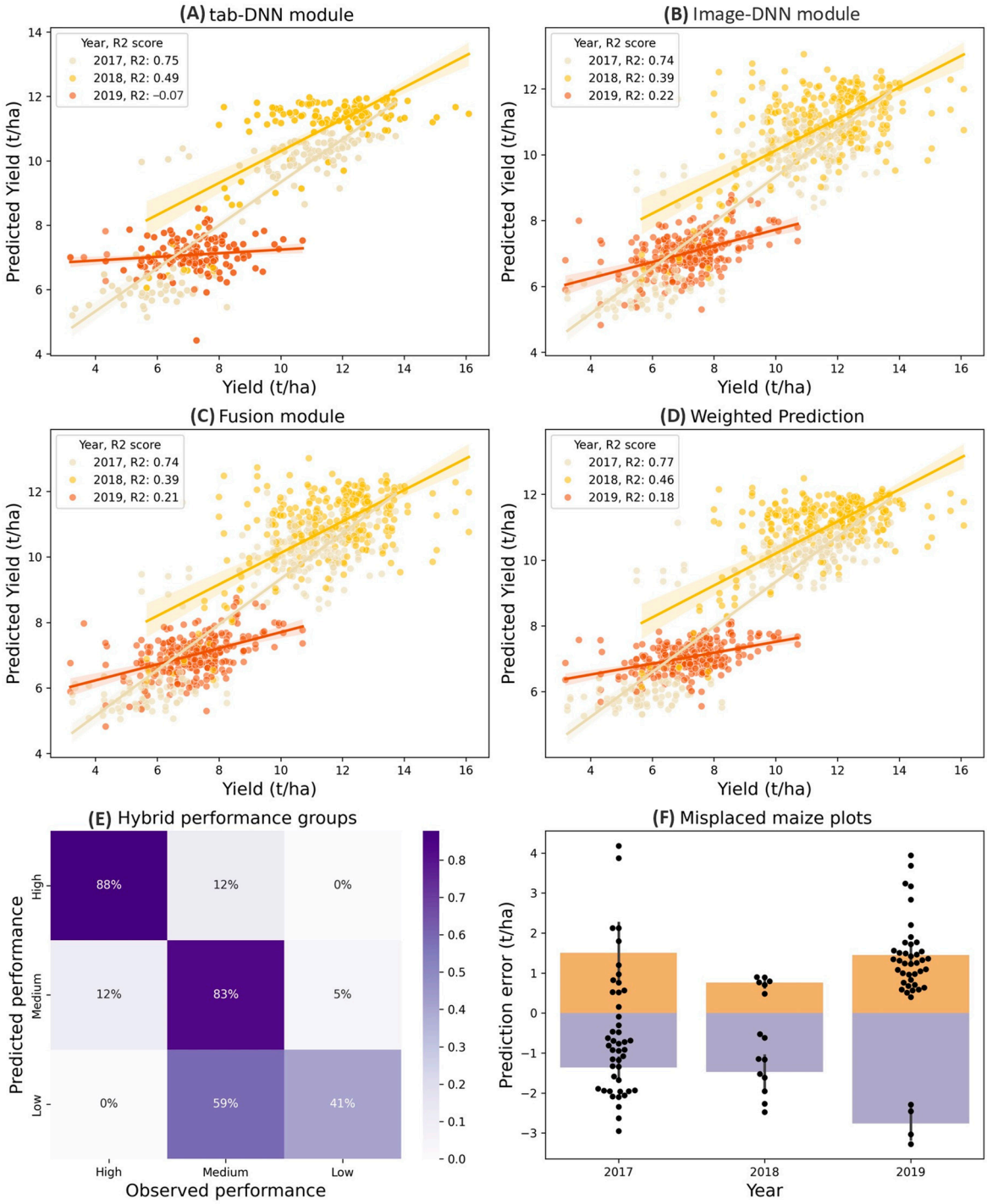


Figure 10. Yield predictions generated from the multimodal architecture when using pretrained modules. Comparison between predictions (*y*-axis) vs. ground truth yields (*x*-axis) for each sample by environment and year, measured at tonne/ha using tab-DNN predictions (**A**), Image-DNN predictions (**B**), fusion module predictions (**C**), and weighted predictions (**D**). (**E**) Confusion matrix showing the percentage of correctly predicted hybrids that were classified as either high (>10 tonne/ha), medium (6–9.99 tonne/ha), or low (<6 tonne/ha). Ground truth classifications represented on the *y*-axis, with predicted performance on the *x*-axis. (**F**) Classification error plot showcasing hybrids predicted as medium yield but with a ground truth class of high yield (light purple), and hybrids predicted as medium yield but with a ground truth class of low yield (mustard).

4. Discussion

Crop growth and yield are heavily dependent on meteorological factors. For example, weather variability has been shown to account for more than 75% of maize yield variability in the US central Corn Belt [67]. Investigating the yield data revealed variability between hybrids sown in different years, highlighting a complex interplay between various factors, such as weather influencing yield potential. Several meteorological factors, such as precipitation and air temperature, have been reported to significantly impact maize yield during the early stages of development [67–69]. Despite its importance, using weather variables as feature input in a multimodal approach is rarely explored, especially before the growing season and in a field-trial context where few of the same hybrids are grown [14,18,20,24,30]. In our study, implementing additional feature inputs, including time-dependent weather conditions, into our tab-DNN model significantly outperformed the previous iteration of the model [14]. The superior performance could potentially result from the significant enhancement of the tab-DNN's explanatory power (R^2 of 0.77 vs. 0.27; Table 3). With an improvement in the model's explanatory power, this modality demonstrated significantly reduced RMSE and relative RMSE (RMSE 1.28 t/ha and a relative RMSE 9.27% of 13.76 t/ha) compared with the previously published model (RMSE 1.77 t/ha and a relative RMSE 12.57% of 16 t/ha) when predicting for the holdout set [14]. The model presented in this study was able to correctly categorise ~19% more high-yielding hybrids than previously reported by Danilevicz et al. [14]. The tab-DNN module was also able to better distinguish hybrids of high- and low-yield potential, as the previous iteration could not correctly identify any hybrids with low yield [14]. Although the explanatory power of the model was significantly improved, there is still a trend for the R^2 score to decrease across years [14]. This observation highlights that the features used in the tabular module could not explain most of the yield variation in 2019. A plausible explanation is the similar environmental conditions (e.g., date sown) seen in 2019. For example, the hybrids sown in 2019 had the same plant date, whereas those sown in 2017 and 2018 were exposed to different planting dates (Supplementary Table S1). The limited variation in agricultural practices in 2019 may have contributed to the low explanatory power. Another potential factor may be the absence of other appropriate tabular features. Tabular machine learning models incorporating features such as vegetation indices have also been shown to significantly improve predictive power [70,71]. For example, Khan et al. [70] predicted maize crop yield by training several machine learning models, including Geographically Weighted Random Forest Regression (GWRFR), Random Forest Regression, and Partial Least Square Regression, using several features, including vegetation indices, gross primary production, climate data, and soil data. They reported the GWRFR model outperformed all other models (R^2 = 0.90 and RMSE = 0.764 MT/ha) [70]. Interestingly, their Random Forest Regression model outperformed our tab-DNN model, expressing an R^2 value of 0.828 and RMSE of 1.006 MT/ha [70]. As our tab-DNN model uses weather and crop-management-specific features only, the development of future models incorporating additional features, such as vegetation indices, may improve their predictive performance.

Regarding the Image-DNN module, the yield distribution of ground truth and predicted yields show a trend for two group clusters to form (Figure 8E). These group clusters showcase hybrids with high-yield potential compared with ones with lower yields, illustrating that the image modality can distinguish hybrids corresponding to either high or low yields. This observation is further supported when classifying hybrids based on their yield performance, with the Image-DNN module being able to correctly classify hybrids with high, medium, and low yields at a rate of 87%, 82%, and 39%, respectively. Compared with the tab-DNN module, there is greater explanatory power when using images to predict the yield of hybrids sown in 2019. Multiple factors may have contributed to this result, with one significant contributor being the use of RGB images capturing time dependences that can depict the growth conditions of the plant over time [14,72,73]. Including image features in our model enabled it to outperform our previously published tabular framework, providing a significantly higher R^2, RMSE, relative RMSE, and classification rate on

unseen data [14]. Despite this, the hybrids grown in 2019 displayed the lowest explanatory power compared with previous years. One possibility may be due to the quality of the images. The average ground sample distance of the 2019 images was considerably higher than captured in 2017 and 2018, probably associated with a change in camera and flying altitude in 2019. Image quality plays a key role in machine learning tasks [74,75]. For example, Roznik et al. [74] demonstrated that using higher-quality imagery from satellite data corresponded to enhanced explanatory power (R^2) for crop yield predictions. This underscores the importance for future studies to take into consideration the quality of their image datasets, especially when dealing with publicly available data. Capturing imagery through vegetation indices in tabular frameworks also seems to perform significantly better than using RGB images in our image framework [70,71]. For example, Guo et al. [71] captured UAV-based hyperspectral images using vegetation indices in a tabular format and reported an R^2 of 0.92 and RMSE of 0.53 t/ha, outperforming our Image-DNN results (R^2 of 0.75 and RMSE of 1.33 t/ha). The Image-DNN module was also able to demonstrate similar performance (RMSE of 1.33 t/ha, relative RMSE 9.65% of 13.76 t/ha, and R^2 of 0.75) to our previously published image model (RMSE of 1.27 t/ha, relative RMSE 8.99% of 16 t/ha, and R^2 of 0.63) that used multispectral images and vegetation indices for holdout predictions [14], with the exception of its enhanced explanatory power. This result highlights the use of capturing crop growth in RGB images, showing comparable performance to using multispectral features lacking time dependencies in image-specific modalities.

The multimodal framework was able to successfully forecast the yield potential of hybrids sown in different years. In 2019, the mean absolute error was the lowest compared with previous years, which is likely associated with a smaller yield range for 2019 compared with 2017 and 2018 (Supplementary Figure S1). The multimodal model shared similar performance to the individually trained modules on unseen data. This result illustrates the use of co-learning in our framework does not yield significant improvements to individual modules, as previously seen in [14]. In multimodal co-learning, an information-rich module can enhance the performance of another modality through the transfer of knowledge, such as their feature representations [76]. In this case, the modalities used in the multimodal framework demonstrate sufficient feature representation for co-learning to not be of much benefit. This is further supported as the tabular module in the multimodal framework contains a higher number of features than its previous counterpart [14]. The integration of these additional feature inputs also significantly enhanced the performance of the tab-DNN architecture. However, the Image-DNN module dragged down the overall performance of the multimodal framework, with the image, fusion, and weighted predictions generating suboptimal RMSE and relative RMSE values relative to our previously published model [14]. The lower relative RMSE of the tab-DNN module for holdout predictions indicates its superior generalisation abilities compared with its prior equivalent module [14], a pattern not seen in our Image-DNN module. Our model used a significantly greater number of tabular features compared with Danilevicz et al. [14], which might explain the difference in model generalisation capabilities. On the other hand, the performance of image models is highly dependent on image quality [74,75], and while the average ground resolution of the orthomosaics used in the previous model was 7.4 cm/pix [14], our study included variable image resolution. The disparity in performance over the evolution of the framework may also stem from differences in inputs for the Image-DNN module. For example, the use of multispectral images and vegetation indices was not available for this study, as only RGB images were available between timepoints. Danilevicz et al. [14] reported the inclusion of vegetation indices to have a greater impact on maize prediction than RGB or multispectral images alone. The usage of RGB data in the current iteration, where the previous version used time-independent multispectral features, highlights promising opportunities for model improvement and provides guidance for data collection in future studies.

The improvements shown in our framework demonstrate that the inclusion of environmental features such as weather and phenotypic traits play a role in influencing yield potential. The feature inputs that contributed to our model were explored using SHAP

values. Notably, six out of the ten most influential features of the model were related to meteorological factors, including temperature, humidity, and solar radiation at specific timepoints. For instance, these meteorological features include the difference between the maximum and minimum temperatures (MTD), All Sky Surface Shortwave Downward Irradiance (ASSSDI; a measure of solar radiation), Dew/Frost Point (indicator of absolute humidity), Growing degree–days (GDD; a measure of plant growth relative to temperature), and relative humidity (Figure 7A,B) [42–45,77]. The study presented here showcases that weather features—indicators of temperature, humidity, and solar radiation—have been shown to have a significant impact on model improvement. This result is not entirely surprising, especially since a considerable amount of the variability in maize yields can be explained by weather conditions in the US [67]. For example, the reproductive stage of maize is the most sensitive to temperature deviations. During this critical period, high temperatures can result in altered metabolic changes in maize crops, such as reductions in their CO_2 exchange rate [78]. Solar radiation is also a driver for photosynthesis and maturation of plant tissues, causing the plants' radiation use efficiency to be a determinant for final yield levels [78,79]. Under heat stress, humidity levels can also determine final yield levels, as increased humidity levels can reduce the internal temperature of crop canopies [80]. Extensive research has been conducted on these weather features and their role in influencing maize yield [78,79,81,82]. For example, Muchow et al. [81] reported optimal maize yields associated with lower temperatures and heightened solar radiation exposure. Subsequent studies have corroborated these findings [79,82]. In relation to this study, indicators of temperature (MTD and GDD) and solar radiation (ASSSDI) demonstrated directional impact in yield prediction, with high values of TDM and ASSSDI associated with lower predicted yields (Figure 7B). Interestingly, high values of GDD were associated with higher predicted yield values (Figure 7B). The results on the impact of solar radiation are consistent with previous findings [79,81,82]. In this study, the effect of temperature was measured through MTD to better capture temperature extremes and GDD at specific timepoints. The positive impact of MTD (3 weeks after sowing) and negative impact of GDD (4 weeks before sowing) on maize grain yield indicates that the impact of different measures of temperature on maize grain yield is not only determined by its magnitude but also its timing before and during the growth period. The effect of these features to influence maize yield before planting further demonstrates the importance of soil quality. For example, Li et al. [30] reported the importance of meteorological effects before planting in influencing nitrogen levels in soil, and subsequently, crop yield. Our study reports Dew/Frost Point (4 weeks before sowing and 9 weeks after sowing) and relative humidity (13 weeks before sowing) on yield prediction (Figure 7A), showing that lower values of Dew/Frost Point at these timepoints were associated with greater predicted yields (Figure 7B). Interestingly, higher values of relative humidity at 13 weeks before sowing were associated with greater predicted yields (Figure 7B). Zhang et al. [83] investigated the importance of relative humidity for crop growth. They reported a significant positive relationship between relative humidity and maize yield, with every percent increase in humidity boosting crop yield by 0.61% [83]. When using dew point temperatures to calculate vapor pressure deficit (VPD; a measure of humidity), they reported no significant relationship between maize yields and VPD [83]. Although our results on relative humidity are consistent with Zhang et al. [83], their findings on the impact of dew point temperatures diverge. This suggests that dew point temperatures may be a more reliable measure than VPD to assess the relationship between absolute humidity and maize yield. Overall, our results demonstrate the importance of these weather features on model improvement. Individually, these factors have been shown to contribute uniquely to maize growth. However, it is important to acknowledge that weather variables do not work in isolation, but rather interact in a complex way with crop growth. For example, heat stress tolerance is a measure of a plant's ability to resist temperature changes that may damage crop growth [84]. Genetic variation plays a significant role in determining a crop's tolerance to heat stress, with exposure to it inducing >5000 maize genes [84]. Heat stress tolerance is regulated by several weather variables

including humidity and solar radiation [84]. In the face of evolving climate conditions, the development of climate-resistant maize crops plays an important role in sustaining future global population growth.

In addition to weather variables, crop management-specific features were influential in the predictions generated from the multimodal framework. SHAP analysis showcased the top four most influential features of the model were plant density, the vertical placement of the hybrid in the field, days to anthesis, and parental line. Plant density was shown to be the leading feature driving model predictions, with a mean absolute SHAP value of 0.229 (Figure 7A). The directional impact of plant density was also evident, where high plant density values were associated with greater yield predictions (Figure 7B). Plant density has also been shown to play a significant effect on maize yield in other studies [85,86]. For example, Gozubenli et al. [86] demonstrated the trend for maize yield to increase with plant density, up to 90,000 plants/ha. Following plant density, the placement of the hybrid in the field seems to significantly contribute to our model predictions, indicating that hybrids positioned further to the left on the field map were generating greater yield predictions (Figure 7B). It was unexpected that the arrangement would impact maize yields, with further investigation revealing that certain regions of the field trial corresponded to higher yields (Supplementary Figure S2). The randomised block design employed by the study may have reduced the bias introduced to our model from the arrangement effect. Another key feature, days to anthesis, was also shown to contribute significantly to model outcomes (Figure 7A). High values of this feature were associated with higher yield predictions (Figure 7B). Anthesis is a critical developmental stage for maize, with studies using days to 50% anthesis as a metric to determine the maturity period for each hybrid and correlate it to their study endpoints [78,87]. For example, Badu-Apraku et al. [88] reported increased days to anthesis to be associated with higher yield gains under different stress conditions. Compared with the other influential features, parental line plays an important role in determining the genetic makeup of the maize hybrid, with 'Parental Line 1′ corresponding to the name of one of the parent plants that contributed to the hybrid lineage. Although prediction models using drone imagery have been shown to outperform genomic prediction methods, incorporating additional genetic markers such as SNPs into our model could further elucidate the importance of certain loci for predicting complex traits and improve the model's accuracy [89,90]. For example, Adak et al. [89] reported that genomic prediction methods using SNP markers perform similarly to prediction methods using UAV datasets when predicting complex traits such as yield. Crop breeders can utilise this information by selectively crossing parental lines harbouring influential loci, prioritising traits such as heat stress tolerance and yield potential.

5. Conclusions

Field-level integration of weather and phenotypic data into the multimodal framework addresses an important gap in the literature while enhancing the model's ability to identify high-yielding hybrids. The findings of our study address this knowledge gap by providing model developers and crop breeders information regarding the importance of the magnitude and timing of certain weather conditions on maize yield predictions. For example, low levels of Dew/Frost Point temperatures at 9 weeks after sowing were associated with greater yield predictions. This study demonstrated significant improvements in the predictive accuracy and explanatory power compared with previous iterations of the model. Specifically, the inclusion of historical weather data (e.g., relative humidity) and important phenotypic traits (e.g., plant density) was able to significantly improve the model predictions, as indicated by the improvement of R^2, RMSE, and relative RMSE compared with our previous version [14]. The contributions of these features to the model were also evaluated using SHAP values, and their directional impact was also highlighted. Through this analysis, the importance of several weather conditions (e.g., dew point temperatures 9 weeks after sowing) at specific timepoints was shown to significantly contribute to model predictions. Using our current image dataset showcased no significant improvements to

model predictions. Despite issues with data quality, capturing the early growth period of maize using RGB imagery performed similarly—albeit with an improvement in R^2—to multispectral images and vegetation indices lacking time dependencies. This highlights future directions for model improvement, such as capturing the early growth phases of maize using multispectral imagery, which may contribute more significantly to model predictions. Limited by location, our model can be further refined by the addition of data from different locations and environments. This study is also limited by differences in mean ground sample distance between years, which may explain the differences in the accuracy of model predictions. However, the model may become more refined with the inclusion of additional images with varying resolutions. This may improve the model's ability to generalise to different datasets, especially when crop breeders may face limitations in funding. This would make our findings more applicable to crop breeders facing challenges in accessing advanced UAV platforms with high-quality imaging equipment, and their time availability to capture close-range images. Despite challenges in data availability and variability in data quality, our findings reinforce the importance of harnessing multiple data sources for refining agricultural predictive models. The development of agricultural datasets for predictive models must consider the inclusion of weather, phenotypic information, and multispectral images at different timepoints. Moving forward, the incorporation of genetic information such as heat-stress SNP markers could provide more detailed information about the genetic makeup driving the yield potential of maize hybrids. The predictive capabilities of our model can potentially move beyond maize by enabling our pretrained weights to be used on other crop datasets, such as wheat and rice, serving as a single decision-making tool for breeders to identify high-yielding crops of diverse types.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/rs16214043/s1, Supplementary Table S1. Dates of images taken, sowing dates, and fertiliser application used for each environment per year; Supplementary Table S2. Features used for the Tabular DNN model. Descriptions extracted from the README.txt file of the G2F dataset; Supplementary Figure S1. Yield distribution for each year for (A) the ground truth yields during training and validation and (B) the holdout predictions generated from the multimodal deep learning model (training from scratch); Supplementary Figure S2. Yield variability between specific crop locations within environments per year. (A) represents G2LA hybrids sown in 2018. (B) represents G2FE hybrids sown in 2018. (C) represents G2FN hybrids sown in 2018. (D) represents G2LA hybrids sown in 2017. (E) represents G2FE hybrids sown in 2017. (F) represents DG2F hybrids sown in 2017.

Author Contributions: Conceptualisation, D.S., D.E., M.B., H.A.A.-M. and M.F.D.; methodology, D.S. and M.F.D.; software, D.S. and M.F.D.; validation, D.S.; formal analysis, D.S.; investigation, D.S.; data curation, D.S.; writing—original draft preparation, D.S.; writing—review and editing, D.S., D.E., H.A.A.-M., M.B. and M.F.D.; visualisation, D.S. and M.F.D.; supervision, D.E., H.A.A.-M., M.B. and M.F.D. All authors have read and agreed to the published version of the manuscript.

Funding: We gratefully acknowledge funding from the Australia Research Council (Projects DP210100296 and DP200100762). This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. Danial Shamsuddin was supported by the UWA Data Institute Summer Research Grant and the Jennifer Arnold Memorial Research Award.

Data Availability Statement: The datasets analysed in this study were made publicly available by the Genomes to Fields initiative (www.genomes2fields.org, accessed on 1 December 2023) and can be accessed from doi:10.25739/4ext-5e97, doi:10.25739/96zv-7164 and doi:10.25739/d9gt-hv94. The meteorological data were obtained from the NASA Langley Research Center (LaRC) POWER Project funded through the NASA Earth Science/Applied Science Program. This dataset was compiled by the Genomes to Fields initiative and can be accessed from doi:10.25739/tq5e-ak26. Access to the code employed for the development and evaluation of the multimodal framework has been made available at https://github.com/dshamsuddin/multimodal_maize_prediction.git, accessed on 27 May 2024.

Acknowledgments: The authors of this study would like to express their gratitude to Alper Adak for his commitment and time spent ensuring the datasets were in a suitable format before publishing them for public use. Extended appreciation to Scott Wilde, Seth Murray, and his research group for compiling and releasing the data applied in this study for public use and the Pawsey Supercomputing Resource Centre for providing the computational resources for running our models. Further acknowledgments are due to the UWA Data Institute and the Jennifer Arnold Memorial Research Award for their generous support of this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Govindaraj, M.; Vetriventhan, M.; Srinivasan, M. Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. *Genet. Res. Int.* **2015**, *2015*, 431487. [CrossRef] [PubMed]
2. Fess, T.L.; Kotcon, J.B.; Benedito, V.A. Crop Breeding for Low Input Agriculture: A Sustainable Response to Feed a Growing World Population. *Sustainability* **2011**, *3*, 1742–1772. [CrossRef]
3. Fan, J.; Zhou, J.; Wang, B.; de Leon, N.; Kaeppler, S.M.; Lima, D.C.; Zhang, Z. Estimation of Maize Yield and Flowering Time Using Multi-Temporal UAV-Based Hyperspectral Data. *Remote Sens.* **2022**, *14*, 3052. [CrossRef]
4. Ray, D.K.; Mueller, N.D.; West, P.C.; Foley, J.A. Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS ONE* **2013**, *8*, e66428. [CrossRef]
5. Beltran-Peña, A.; Rosa, L.; D'Odorico, P. Global food self-sufficiency in the 21st century under sustainable intensification of agriculture. *Environ. Res. Lett.* **2020**, *15*, 095004. [CrossRef]
6. Xiong, W.; Reynolds, M.; Xu, Y. Climate change challenges plant breeding. *Curr. Opin. Plant Biol.* **2022**, *70*, 102308. [CrossRef]
7. Wolter, F.; Schindele, P.; Puchta, H. Plant breeding at the speed of light: The power of CRISPR/Cas to generate directed genetic diversity at multiple sites. *BMC Plant Biol.* **2019**, *19*, 176. [CrossRef] [PubMed]
8. Watson, A.; Ghosh, S.; Williams, M.J.; Cuddy, W.S.; Simmonds, J.; Rey, M.-D.; Asyraf Md Hatta, M.; Hinchliffe, A.; Steed, A.; Reynolds, D.; et al. Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat. Plants* **2018**, *4*, 23–29. [CrossRef]
9. Haarhoff, S.J.; Swanepoel, P.A. Plant Population and Maize Grain Yield: A Global Systematic Review of Rainfed Trials. *Crop Sci.* **2018**, *58*, 1819–1829. [CrossRef]
10. Food and Agriculture Organization. FAOSTAT Home Page. Available online: https://www.fao.org/faostat/en/#home (accessed on 1 May 2024).
11. Erenstein, O.; Jaleta, M.; Sonder, K.; Mottaleb, K.; Prasanna, B.M. Global maize production, consumption and trade: Trends and R&D implications. *Food Secur.* **2022**, *14*, 1295–1319. [CrossRef]
12. Atefi, A.; Ge, Y.; Pitla, S.; Schnable, J. Robotic Technologies for High-Throughput Plant Phenotyping: Contemporary Reviews and Future Perspectives. *Front. Plant Sci.* **2021**, *12*, 611940. [CrossRef] [PubMed]
13. Ninomiya, S. High-throughput field crop phenotyping: Current status and challenges. *Breed. Sci.* **2022**, *72*, 3–18. [CrossRef] [PubMed]
14. Danilevicz, M.F.; Bayer, P.E.; Boussaid, F.; Bennamoun, M.; Edwards, D. Maize Yield Prediction at an Early Developmental Stage Using Multispectral Images and Genotype Data for Preliminary Hybrid Selection. *Remote Sens.* **2021**, *13*, 3976. [CrossRef]
15. Danilevicz, M.F.; Bayer, P.E.; Nestor, B.J.; Bennamoun, M.; Edwards, D. Resources for image-based high-throughput phenotyping in crops and data sharing challenges. *Plant Physiol.* **2021**, *187*, 699–715. [CrossRef]
16. Yang, W.; Feng, H.; Zhang, X.; Zhang, J.; Doonan, J.H.; Batchelor, W.D.; Xiong, L.; Yan, J. Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. *Mol. Plant* **2020**, *13*, 187–214. [CrossRef] [PubMed]
17. Gao, J.; Li, P.; Chen, Z.; Zhang, J. A Survey on Deep Learning for Multimodal Data Fusion. *Neural Comput.* **2020**, *32*, 829–864. [CrossRef]
18. Aviles Toledo, C.; Crawford, M.M.; Tuinstra, M.R. Integrating multi-modal remote sensing, deep learning, and attention mechanisms for yield prediction in plant breeding experiments. *Front. Plant Sci.* **2024**, *15*, 1408047. [CrossRef]
19. Li, L.; Liu, L.; Peng, Y.; Su, Y.; Hu, Y.; Zou, R. Integration of multimodal data for large-scale rapid agricultural land evaluation using machine learning and deep learning approaches. *Geoderma* **2023**, *439*, 116696. [CrossRef]
20. Lin, F.; Crawford, S.; Guillot, K.; Zhang, Y.; Chen, Y.; Yuan, X.; Chen, L.; Williams, S.; Minvielle, R.; Xiao, X.; et al. MMST-ViT: Climate Change-aware Crop Yield Prediction via Multi-Modal Spatial-Temporal Vision Transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023; pp. 5774–5784.
21. Zadeh, A.; Liang, P.P.; Morency, L.P. Foundations of multimodal co-learning. *Inf. Fusion* **2020**, *64*, 188–193. [CrossRef]
22. Maimaitijiang, M.; Sagan, V.; Sidike, P.; Hartling, S.; Esposito, F.; Fritschi, F.B. Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.* **2020**, *237*, 111599. [CrossRef]
23. Ramzan, Z.; Asif, H.M.S.; Yousuf, I.; Shahbaz, M. A Multimodal Data Fusion and Deep Neural Networks Based Technique for Tea Yield Estimation in Pakistan Using Satellite Imagery. *IEEE Access* **2023**, *11*, 42578–42594. [CrossRef]
24. Mia, M.S.; Tanabe, R.; Habibi, L.N.; Hashimoto, N.; Homma, K.; Maki, M.; Matsui, T.; Tanaka, T.S.T. Multimodal Deep Learning for Rice Yield Prediction Using UAV-Based Multispectral Imagery and Weather Data. *Remote Sens.* **2023**, *15*, 2511. [CrossRef]
25. Yang, W.; Nigon, T.; Hao, Z.; Paiao, G.D.; Fernández, F.G.; Mulla, D.; Yang, C. Estimation of corn yield based on hyperspectral imagery and convolutional neural network. *Comput. Electron. Agric.* **2021**, *184*, 106092. [CrossRef]

26. Zipper, S.C.; Qiu, J.; Kucharik, C.J. Drought effects on US maize and soybean production: Spatiotemporal patterns and historical changes. *Environ. Res. Lett.* **2016**, *11*, 94021. [CrossRef]
27. Sánchez, B.; Rasmussen, A.; Porter, J.R. Temperatures and the growth and development of maize and rice: A review. *Global Chang. Biol.* **2014**, *20*, 408–417. [CrossRef]
28. Magro, R.B.; Alves, S.A.M.; Gebler, L. Computational Models in Precision Fruit Growing: Reviewing the Impact of Temporal Variability on Perennial Crop Yield Assessment. *SN Comput. Sci.* **2023**, *4*, 554. [CrossRef]
29. Shook, J.; Gangopadhyay, T.; Wu, L.; Ganapathysubramanian, B.; Sarkar, S.; Singh, A.K. Crop yield prediction integrating genotype and weather variables using deep learning. *PLoS ONE* **2021**, *16*, e0252402. [CrossRef]
30. Li, Z.; Guan, K.; Zhou, W.; Peng, B.; Jin, Z.; Tang, J.; Grant, R.F.; Nafziger, E.D.; Margenot, A.J.; Gentry, L.E.; et al. Assessing the impacts of pre-growing-season weather conditions on soil nitrogen dynamics and corn productivity in the U.S. Midwest. *Field Crops Res.* **2022**, *284*, 108563. [CrossRef]
31. Lima, D.C.; Aviles, A.C.; Alpers, R.T.; McFarland, B.A.; Kaeppler, S.; Ertl, D.; Romay, M.C.; Gage, J.L.; Holland, J.; Beissinger, T.; et al. 2018–2019 field seasons of the Maize Genomes to Fields (G2F) G × E project. *BMC Genom Data* **2023**, *24*, 29. [CrossRef]
32. McFarland, B.A.; AlKhalifah, N.; Bohn, M.; Bubert, J.; Buckler, E.S.; Ciampitti, I.; Edwards, J.; Ertl, D.; Gage, J.L.; Falcon, C.M.; et al. Maize genomes to fields (G2F): 2014–2017 field seasons: Genotype, phenotype, climatic, soil, and inbred ear image datasets. *BMC Res. Notes* **2020**, *13*, 71. [CrossRef]
33. Cope, D.A.; Wilde, S.C. Maize G2F UAV Data, College Station, TX 2019. Available online: https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/CS19_G2F (accessed on 10 June 2023).
34. Wilde, S.C.; Cruzato, N.; Jung, J.; Chang, A.; Anderson, S.L.; Cope, D.; Popescu, S.; Lonesome, M.; Murray, S.C. CS18_G2F. Available online: https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/Murray_G2F_UAV_Maize_2018 (accessed on 10 June 2023).
35. Murray, S.C.; Lonesome, M.; Popescu, S.; Cope, D.; Anderson, S.L.; Chang, A.; Jung, J.; Cruzato, N.; Wilde, S.; Walls, R.L. G2F Maize UAV Data, College Station, Texas 2017. Available online: https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/tamu_corn_2017_CS17_G2F_20190305 (accessed on 10 June 2023).
36. Lima, D.C.; Washburn, J.D.; Varela, J.I.; Chen, Q.; Gage, J.L.; Romay, M.C.; Holland, J.; Ertl, D.; Lopez-Cruz, M.; Aguate, F.M.; et al. Genomes to Fields 2022 Maize genotype by Environment Prediction Competition. *BMC Res. Notes* **2023**, *16*, 148. [CrossRef] [PubMed]
37. Genomes to Fields. Genomes to Fields 2022 Maize Genotype by Environment Prediction Competition. Available online: https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/GenomesToFields_GenotypeByEnvironment_PredictionCompetition_2023 (accessed on 10 June 2023).
38. Rodrigues, G.C.; Braga, R.P. Evaluation of NASA POWER Reanalysis Products to Estimate Daily Weather Variables in a Hot Summer Mediterranean Climate. *Agronomy* **2021**, *11*, 1207. [CrossRef]
39. Duarte, Y.C.N.; Sentelhas, P.C. NASA/POWER and DailyGridded weather datasets—How good they are for estimating maize yields in Brazil? *Int. J. Biometeorol.* **2020**, *64*, 319–329. [CrossRef] [PubMed]
40. Anderson, S.L.; Murray, S.C. R/UAStools::plotshpcreate: Create Multi-Polygon Shapefiles for Extraction of Research Plot Scale Agriculture Remote Sensing Data. *Front. Plant Sci.* **2020**, *11*, 511768. [CrossRef]
41. Howard, J.; Gugger, S. Fastai: A Layered API for Deep Learning. *Information* **2020**, *11*, 108. [CrossRef]
42. Yu, Y.-C.; Shi, J.; Wang, T.; Letu, H.; Zhao, C. All-sky total and direct surface Shortwave Downward Radiation (SWDR) estimation from satellite: Applications to MODIS and Himawari-8. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102380. [CrossRef]
43. Parthasarathi, T.; Velu, G.; Jeyakumar, P. Impact of crop heat units on growth and developmental physiology of future crop production: A review. *J. Crop Sci. Technol.* **2013**, *2*, 2319–3395.
44. McMaster, G.S.; Wilhelm, W.W. Growing degree-days: One equation, two interpretations. *Agric. For. Meteorol.* **1997**, *87*, 291–300. [CrossRef]
45. Ramirez-Cabral, N.Y.Z.; Kumar, L.; Shabani, F. Global alterations in areas of suitability for maize production from climate change and using a mechanistic species distribution model (CLIMEX). *Sci. Rep.* **2017**, *7*, 5910. [CrossRef]
46. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631.
47. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015.
48. Wright, L. Ranger—A Synergistic Optimizer. Available online: https://github.com/lessw2020/Ranger-Deep-Learning-Optimizer (accessed on 1 November 2023).
49. Liu, L.; Jiang, H.; He, P.; Chen, W.; Liu, X.; Gao, J.; Han, J. On the variance of the adaptive learning rate and beyond. In Proceedings of the 8th International Conference on Learning Representations, Addis Ababa, Ethiopia, 30 April 2020.
50. Ma, J.; Yarats, D. Quasi-hyperbolic momentum and adam for deep learning. In Proceedings of the 7th International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
51. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

52. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
53. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. Pytorch: An imperative style, high-performance deep learning library. In Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS), Vancouver, BC, Canada, 8–14 December 2019; pp. 8024–8035.
54. Sharma, A.; Jain, A.; Gupta, P.; Chowdary, V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access* **2021**, *9*, 4843–4873. [CrossRef]
55. Zhang, C.; Benz, P.; Argaw, D.M.; Lee, S.; Kim, J.; Rameau, F.; Bazin, J.-C.; Kweon, I.S. Resnet or densenet? Introducing dense shortcuts to resnet. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2021; pp. 3550–3559.
56. Dhaka, V.S.; Kundu, N.; Rani, G.; Zumpano, E.; Vocaturo, E. Role of Internet of Things and Deep Learning Techniques in Plant Disease Detection and Classification: A Focused Review. *Sensors* **2023**, *23*, 7877. [CrossRef] [PubMed]
57. Mamat, N.; Othman, M.F.; Abdoulghafor, R.; Belhaouari, S.B.; Mamat, N.; Hussein, S.F.M. Advanced Technology in Agriculture Industry by Implementing Image Annotation Technique and Deep Learning Approach: A Review. *Agriculture* **2022**, *12*, 1033. [CrossRef]
58. Wightman, R.; Touvron, H.; Jégou, H. Resnet strikes back: An improved training procedure in timm. *arXiv* **2021**, arXiv:2110.00476. [CrossRef]
59. He, T.; Zhang, Z.; Zhang, H.; Zhang, Z.; Xie, J.; Li, M. Bag of tricks for image classification with convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 558–567.
60. Hu, G.; Qian, L.; Liang, D.; Wan, M. Self-adversarial training and attention for multi-task wheat phenotyping. *Appl. Eng. Agric.* **2019**, *35*, 1009–1014. [CrossRef]
61. Albelwi, S.A. Deep architecture based on DenseNet-121 model for weather image recognition. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 10. [CrossRef]
62. Wang, W.; Tran, D.; Feiszli, M. What makes training multi-modal classification networks hard? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 12695–12705.
63. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
64. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774.
65. Mueller, Z. Fastinference. Available online: https://github.com/muellerzr/fastinference (accessed on 1 May 2024).
66. Won, J.; Tutumluer, E.; Byun, Y.-H. Predicting permanent strain accumulation of unbound aggregates using machine learning algorithms. *Transp. Geotech.* **2023**, *42*, 101060. [CrossRef]
67. Joshi, V.R.; Kazula, M.J.; Coulter, J.A.; Naeve, S.L.; Garcia y Garcia, A. In-season weather data provide reliable yield estimates of maize and soybean in the US central Corn Belt. *Int. J. Biometeorol.* **2021**, *65*, 489–502. [CrossRef]
68. Chen, J.; Ren, B.-Z.; Bin, Z.; Liu, P.; Zhang, J.-W. Environment especially minimum temperature affecting summer maize grain yield through regulating ear differentiation and grain development. *J. Integr. Agric.* **2024**, *23*, 2227–2241. [CrossRef]
69. Wang, X.; Li, X.; Gu, J.; Shi, W.; Zhao, H.; Sun, C.; You, S. Drought and Waterlogging Status and Dominant Meteorological Factors Affecting Maize (*Zea mays* L.) in Different Growth and Development Stages in Northeast China. *Agronomy* **2023**, *13*, 374. [CrossRef]
70. Khan, S.N.; Li, D.; Maimaitijiang, M. A Geographically Weighted Random Forest Approach to Predict Corn Yield in the US Corn Belt. *Remote Sens.* **2022**, *14*, 2843. [CrossRef]
71. Guo, Y.; Xiao, Y.; Hao, F.; Zhang, X.; Chen, J.; de Beurs, K.; He, Y.; Fu, Y.H. Comparison of different machine learning algorithms for predicting maize grain yield using UAV-based hyperspectral images. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *124*, 103528. [CrossRef]
72. Näsi, R.; Viljanen, N.; Kaivosoja, J.; Alhonoja, K.; Hakala, T.; Markelin, L.; Honkavaara, E. Estimating biomass and nitrogen amount of barley and grass using UAV and aircraft based spectral and photogrammetric 3D features. *Remote Sens.* **2018**, *10*, 1082. [CrossRef]
73. Prabhakara, K.; Hively, W.D.; McCarty, G.W. Evaluating the relationship between biomass, percent groundcover and remote sensing indices across six winter cover crop fields in Maryland, United States. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *39*, 88–102. [CrossRef]
74. Roznik, M.; Boyd, M.; Porth, L. Improving crop yield estimation by applying higher resolution satellite NDVI imagery and high-resolution cropland masks. *Remote Sens. Appl. Soc. Environ.* **2022**, *25*, 100693. [CrossRef]
75. Fromm, M.; Schubert, M.; Castilla, G.; Linke, J.; McDermid, G. Automated Detection of Conifer Seedlings in Drone Imagery Using Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 2585. [CrossRef]
76. Rahate, A.; Walambe, R.; Ramanna, S.; Kotecha, K. Multimodal co-learning: Challenges, applications with datasets, recent advances and future directions. *Inf. Fusion* **2022**, *81*, 203–239. [CrossRef]
77. Denson, E.; Wasko, C.; Peel, M.C. Decreases in relative humidity across Australia. *Environ. Res. Lett.* **2021**, *16*, 074023. [CrossRef]
78. Waqas, M.A.; Wang, X.; Zafar, S.A.; Noor, M.A.; Hussain, H.A.; Nawaz, M.A.; Farooq, M. Thermal Stresses in Maize: Effects and Management Strategies. *Plants* **2021**, *10*, 293. [CrossRef]
79. Yang, Y.; Xu, W.; Hou, P.; Liu, G.; Liu, W.; Wang, Y.; Zhao, R.; Ming, B.; Xie, R.; Wang, K.; et al. Improving maize grain yield by matching maize growth and solar radiation. *Sci. Rep.* **2019**, *9*, 3635. [CrossRef] [PubMed]

80. Dong, X.; Li, B.; Yan, Z.; Guan, L.; Huang, S.; Li, S.; Qi, Z.; Tang, L.; Tian, H.; Fu, Z.; et al. Impacts of high temperature, relative air humidity, and vapor pressure deficit on the seed set of contrasting maize genotypes during flowering. *J. Integr. Agric.* **2024**, *23*, 2955–2969. [CrossRef]
81. Muchow, R.C.; Sinclair, T.R.; Bennett, J.M. Temperature and Solar Radiation Effects on Potential Maize Yield across Locations. *Agron. J.* **1990**, *82*, 338–343. [CrossRef]
82. Chen, C.; Baethgen, W.E.; Robertson, A. Contributions of individual variation in temperature, solar radiation and precipitation to crop yield in the North China Plain, 1961–2003. *Clim. Chang.* **2013**, *116*, 767–788. [CrossRef]
83. Zhang, P.; Zhang, J.; Chen, M. Economic impacts of climate change on agriculture: The importance of additional climatic variables other than temperature and precipitation. *J. Environ. Econ. Manag.* **2017**, *83*, 8–31. [CrossRef]
84. Djalovic, I.; Kundu, S.; Bahuguna, R.N.; Pareek, A.; Raza, A.; Singla-Pareek, S.L.; Prasad, P.V.V.; Varshney, R.K. Maize and heat stress: Physiological, genetic, and molecular insights. *Plant Genome* **2024**, *17*, e20378. [CrossRef]
85. Shafi, M.; Bakht, J.; Ali, S.; Khan, H.; Khan, M.A.; Sharif, M. Effect of planting density on phenology, growth and yield of maize (*Zea mays* L.). *Pak. J. Bot* **2012**, *44*, 691–696.
86. Gozubenli, H.; Sener, O.; Konuskan, O.; Kilinc, M. Effect of hybrid and plant density on grain yield and yield components of maize (*Zea mays*). *Indian J. Agron.* **2003**, *48*, 203–205. [CrossRef]
87. Ullah, Z.; Rahman, H.; Muhammad, N. Evaluation of maize hybrids for maturity and related traits. *Sarhad J. Agric.* **2017**, *33*, 624–629. [CrossRef]
88. Badu-Apraku, B.; Yallou, C.G.; Obeng-Antwi, K.; Alidu, H.; Talabi, A.O.; Annor, B.; Oyekunle, M.; Akaogu, I.C.; Aderounmu, M. Yield Gains in Extra-Early Maize Cultivars of Three Breeding Eras under Multiple Environments. *Agron. J.* **2017**, *109*, 418–431. [CrossRef]
89. Adak, A.; Murray, S.C.; Anderson, S.L. Temporal phenomic predictions from unoccupied aerial systems can outperform genomic predictions. *G3* **2022**, *13*, jkac294. [CrossRef] [PubMed]
90. Washburn, J.D.; Adak, A.; DeSalvio, A.J.; Arik, M.A.; Murray, S.C. High temporal resolution unoccupied aerial systems phenotyping provides unique information between flight dates. *Plant Phenome J.* **2024**, *7*, e20113. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.