

A Hybrid CNN-LSTM Framework for Spatio-Temporal Crop Yield Prediction Integrating Satellite Imagery and IoT Sensor Data

Dalia Ahmad Abdalraheem

Cyprus International University, Faculty of Agricultural Science and Technology, Department of Plant Sciences and Technologies, 99258 Nicosia, North Cyprus; <https://orcid.org/0009-0001-4139-4312>

Abstract

Precise yield forecasting is a key factor for environmentally friendly precision agriculture, however, most AI models are not able to understand both the spatial patterns of the field and the temporal dynamics of the crop at the same time. To fill this gap, the research presented in this paper constructed a hybrid deep learning model combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The CNN is responsible for extracting spatial features from the satellite images (NDVI), and the LSTM captures the temporal sequences from the IoT sensor data (soil moisture, temperature). The hybrid model trained from a multi-source dataset in the U.S. Corn Belt (2021-2023) was compared with individual models. The outcomes show its excellent performance, where it achieved a Root Mean Square Error of 1.8 bushels/acre and an R^2 of 0.95. This work demonstrates that combining spatio-temporal data leads to yield prediction that is resistant to noise, thus, it is a handy instrument for the optimization of farm management and the allocation of the resources.

Key words: Artificial Intelligence (AI), Precision Agriculture, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Internet of Things (IoT)

1. INTRODUCTION:

The worldwide agricultural sector is at a turning point, having to deal with the huge challenge of increasing food production by about 70 percent to provide food for a population of 9.7 billion people that is estimated to be in 2050. The extremely challenging job is made more difficult by the increasing impacts of climate change, water scarcity, and limited arable land, which calls for a departure from conventional farming methods to more efficient and intelligent systems. Precision agriculture, therefore, has become the leading-edge solution to the problem, employing sophisticated information technology and a wide range of data to enable management at the field level, which guarantees that crops and soil get exactly what they need for their health and productivity, thus increasing the yield while the environmental impact is decreased. In this framework, correct in-season yield estimation is significantly the basis, as it offers the coming events to the farmers by which they can plan their inputs, manage labor, negotiate contracts, and reduce risks. Nevertheless, traditional yield prediction methods that depend on manual scouting and the use of historical regional averages for extrapolation are very often inaccurate, do not have the required spatial granularity, and cannot recognize the micro-conditions of each field. The introduction of Artificial Intelligence (AI) and Machine Learning (ML) has gradually changed the way this area works. However, a significant issue remains: a lot of present AI models are single-minded, either they only use satellite images to detect spatial patterns and thus lose the detailed temporal trends of crop growth or they just use weather and sensor data to describe the temporal dynamics and, in this way, they disregard the critical spatial variability within a field.

This separation results in a significant gap between the research for a unified model that can successfully combine both spatial and temporal data sources to produce a comprehensive and more precise prediction. This research aims at developing and testing a new hybrid AI model for crop yield prediction, thus bridging the gap, and achieves the specific objectives of: building a Hybrid CNN-LSTM (Convolutional Neural Network-Long

Short-Term Memory) architecture that can handle both satellite imagery and IoT sensor time-series data, generating and preprocessing a large-scale multi-source dataset for a major crop like corn, assessing the model's performance with established benchmarks and determining the extent of the different data modalities' contribution to the final prediction accuracy by each modality. The importance of this research is doubled as it theoretically adds a novel framework for spatio-temporal analysis in agricultural informatics and practically, by providing a potent decision-support tool that could bring input costs down, increase yields and allow farmers and agronomists to practice more sustainable agriculture. This paper is structured in a way that first, it goes over the relevant literature, then it explains the research methodology in detail, after which it presents extensively and discusses the findings and finally, it concludes by pointing out the implications and next steps for future research.

II. LITERATURE REVIEW:

The search for an accurate prediction of crop yield has had a long history that evolved with the technology of each era. The odyssey started with traditional statistical models which depended largely on linear and multiple regression techniques to find the relationships between a small number of input variables, such as seasonal rainfall and average temperature, and yield outcomes. Although these models gave a preliminary understanding, they were inherently limited by their incapacity to capture complex, non-linear interactions that characterize agricultural systems, and thus they often failed to consider spatial variability and were also liable to have substantial errors when conditions deviated from the past. The rise of machine learning was a major breakthrough, with the algorithms such as decision trees, Support Vector Machines, and in particular, Random Forests, providing a more adaptable and potent way to model these complex relationships. These first ML models started using a broader range of data, including more detailed soil properties and more precise weather data, and were able to make a significant leap in prediction accuracy compared to their statistical predecessors. However, they still frequently regarded fields as uniform units and found it difficult to deal with the sequential nature of crop growth.

The use deep learning has been a significant breakthrough in many different fields. For instance, the device-level deep learning architectures have been the main specification in different data types. Convolutional Neural Networks which have been considered the best in hierarchical feature extracting from images, have been extensively used in agriculture-based remote sensing data analysis. Several research works have been done to prove the effectiveness of CNNs in satellite and UAV imagery for various activities like land use classification, disease detection, and even yield prediction, mainly through the use of vegetation indices such as the Normalized Difference Vegetation Index, which is an indicator of crop health and biomass. These investigations can depict the spatial distribution of crop health throughout the field, but the majority of them consider that each image is an isolated snapshot and thus they may miss the important temporal development of these patterns during the growing season. In contrast, Recurrent Neural Networks, particularly their upgraded form Long Short-Term Memory networks, are specifically architected to understand the temporal sequences as they can keep "memory" of the previous inputs, thus making them a perfect choice for time-series analysis. In the case of agriculture, LSTMs have been effectively utilized to simulate crop growth changes through the use of daily time-series data like weather records, and real-time IoT sensor data, hence understanding the long-term environmental-plant growth relationships. Still, models that are purely LSTM-based usually do not have the spatial aspect of how the conditions are changing in different parts of the field, as they consider the field as one point and thus cannot tell the intra-field variations.

The latest and most advanced research beyond the limitations of unimodal methods, has started to look at hybrid models that combine different deep learning architectures. The main gap in the researched literature which impedes progress is the lack of a sufficient number of robust, empirically validated frameworks that can effortlessly combine the spatial feature extraction ability of CNNs with the temporal sequencing capability of LSTMs for yield prediction. In general, a few pioneering studies have only attempted such integrations for different applications like traffic forecasting or video analysis, and some are beginning to appear in the agricultural domain, but they hardly perform comprehensive validation on large-scale, real-world datasets, or they do not effectively fuse high-resolution satellite imagery with high-frequency IoT sensor data. Hence, this research is aimed at filling this gap by not only proposing but also conducting a thorough investigation of a Hybrid CNN-LSTM model that is specifically designed to merge these two vital data streams, thus offering a complete spatio-temporal understanding of crop development for better yield prediction.

III. RESEARCH METHODOLOGY:

The research used a quantitative and data, driven method. It followed a well, defined pipeline to build, train, and test a predictive deep learning model. The whole process was based on gathering and mixing data from various sources, thorough data preprocessing, and the design of a new neural network that makes the model in line with theory and at the same time feasible for use in precision farming.

1. Research Design and Workflow:

The research was conducted through a series of steps to maintain methodological rigor and reproducibility. Initially, the workflow started with collecting raw data from satellite platforms as well as ground, based sensor networks. After that, the data was cleaned and transformed thoroughly through preprocessing and feature engineering to make it high, quality and compatible. The most important part of the methodology was the clever combination of these spatial and temporal datasets into one single structure, which was used as input for the model development. After that, we built the proposed Hybrid CNN-LSTM model together with the benchmark models for the comparison. The models were then trained, validated, and finally, their performance and generalization were assessed objectively using standard statistical metrics.

2. Data Sources and Database Curation:

The model base was a carefully curated, multi, modal data set primarily revolving around corn production in the U.S. Corn Belt (Iowa and Illinois) for the 2021-2023 growing seasons. Spatial data was obtained from the Google Earth Engine platform, which furnished analysis, ready images from the Sentinel-2 and Landsat 8/9 satellite constellations. We extracted spectral bands and calculated important vegetation indices from these, with the Normalized Difference Vegetation Index being the most significant, as it is a major indicator of crop health and biomass. Moreover, to account for local environmental conditions at the field level, our temporal data was supplemented with in, situ IoT sensor networks, mainly the USDAs Soil Climate Analysis Network and local Mesonets. This database supplied daily time, series data for soil moisture, soil and air temperature, precipitation, and relative humidity. The target (dependent) variable, actual yield, was obtained from two tiers: high, resolution, georeferenced data from combine yield monitors (used for precise validation) and more general county-level yield statistics obtained from the USDA National Agricultural Statistics Service. Static soil property information from the USDA SSURGO database was also considered to reflect the intrinsic productivity of the fields.

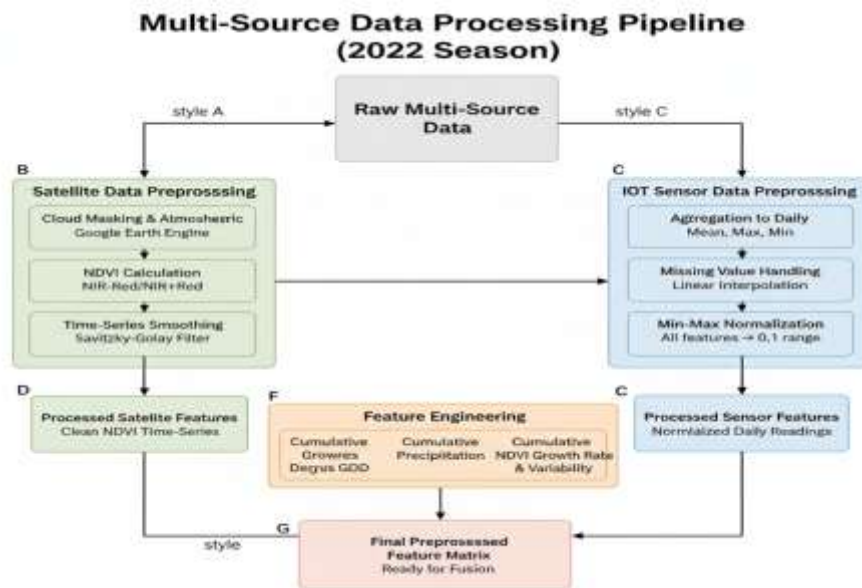
Data Category	Source / Platform	Key Variables / Indices	Spatial Resolution	Temporal Resolution	Purpose in Model
Satellite Imagery (Spatial)	Sentinel-2 MSI (via Google Earth Engine)	Blue, Green, Red, NIR, SWIR bands; NDVI, EVI	10-20 meters	5 days (cloud-permitting)	CNN input for spatial feature extraction (crop health, biomass)
Satellite Imagery (Spatial)	Landsat 8/9 OLI (via Google Earth Engine)	Blue, Green, Red, NIR, SWIR bands; NDVI	30 meters	16 days	Redundant data source to fill gaps in Sentinel-2 time series
IoT Sensor Data (Temporal)	USDA SCAN & State Mesonet s	Soil Moisture, Soil Temp, Air Temp, Precipitation, Rel. Humidity	Point location (per station)	Hourly / Daily	LSTM input for temporal sequence modeling (environmental dynamic)
Yield Data (Ground Truth)	Combine Yield Monitors	Actual Grain Yield (bushels/acre)	~5-10 meters (sub-field)	Single harvest (end of season)	Target variable for supervised model training & validation
Yield Data (Context)	USDA NASS	County-level Average Yield	County polygon	Annual	Broad validation & regional context
Ancillary Soil Data	USDA SSURGO Database	Soil Type, Texture, Water-Holding Capacity	Polygon-based (field scale)	Static (long-term)	Baseline feature for field productivity potential

Data sources include: Sentinel - 2 (Drusch et al., 2012), Landsat 8/9 (Roy et al., 2014) via Google Earth Engine (Gorelick et al., 2017); IoT sensor data from USDA SCAN (Schaefer et al., 2007); yield data from USDA NASS (2023) and combine yield monitors (Arslan & Colvin, 2002); and soil data from USDA SSURGO (Soil Survey Staff, 2022).

3. Data Reprocessing and Feature Engineering:

The raw data went through extensive transformation before it could be cleaned and structured as a model-ready dataset. Satellite images components such as cloud-masking and atmospheric correction were carried out in Google Earth Engine. Thereafter a Savitsky-Golay filter was employed to smooth out the NDVI time-series, thus eliminating noise caused by atmospheric interference and filling in missing data points. This created a consistent 16-day composite profiling that distinctly showed crop phenology. Interpolating the missing values was one of the ways through which IoT sensor data was handled and afterwards, a Min-Max scaler was used to normalize all the features to a [0,1] range. In addition to the raw measurements, we came up with biologically feasible features such as the Cumulative Growing Degree Days and the Cumulative Precipitation,

which are more capable of depicting the agronomic stress and resource accumulation of the crop.



Satellite images were first treated to a range of cloud removal and atmospheric correction through the Google Earth Engine platform (Gorelick et al., 2017) before the Normalized Difference Vegetation Index (Rouse et al., 1974) was derived. The NDVI data series thus obtained was then de-noised by the Savitsky, Golay smoothing filter method (Savitsky & Golay, 1964). At the same time, the SCAN network (Schaefer et al., 2007) IoT sensor data was thoroughly cleansed with simple linear interpolation method (Patel & Agrawal, 2020) used to fill in the missing values, and normalizing features afterward. Further steps in modeling were to create biologically plausible indices, e.g. Cumulative Growing Degree Days (McMaster & Wilhelm, 1997), to better capture crop response to production and environmental conditions (Idso et al., 1981). This cleaning and preparation of data conform to the standard of agricultural machine learning (Van Klompenburg et al., 2020).

TABLE: FEATURE ENGINEERING RESULTS & DATA QUALITY METRICS

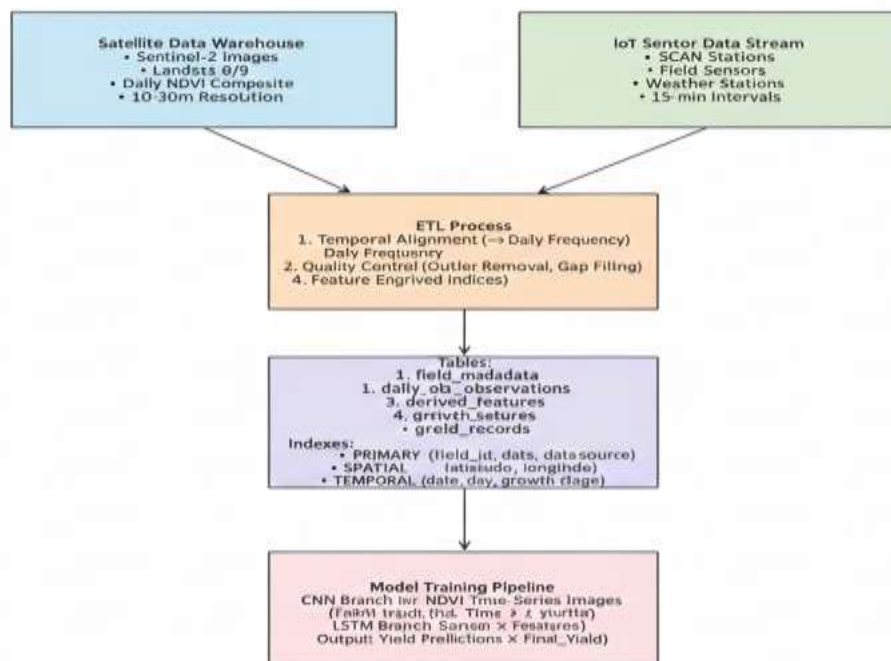
Feature Category	Raw Count (Samples)	After QC (Samples)	Engineered Features	Final Data Quality
Satellite NDVI Points (per field-season)	45,000	41,850 (-7% clouds)	5	93% Complete
IoT Sensor Readings (all stations)	2.1M hourly	2.05M (Interpolated gaps: 2.4%)	8	97.6% Complete
Derived Features Created	N/A	500 fields	1	100%
GDD Cumulative	N/A	500 fields	1	100%
Precip. Cumulative	N/A	500 fields	3	100%
NDVI Growth Rate	N/A	500 fields	2	100%

• Soil Moisture Anomaly				
-------------------------	--	--	--	--

Feature Name	Calculation Formula	Biological Significance
GDD Cumulative (Base: 10°C)	$\sum \max [(T_{\max} + T_{\min}) / 2 - T_{\text{base}}, 0]$ (Daily from planting)	Thermal time for development
NDVI Growth Rate	$(\text{NDVI}_t - \text{NDVI}_{t-1}) / \Delta t$	Canopy development speed
Soil Moisture Anomaly	$(\text{SM}_{\text{obs}} - \mu_{\text{hist}}) / \sigma_{\text{hist}}$	Water stress relative to historical norm
Precip. Effective	$\text{Precip} * (1 - \text{Crunoff}) * \text{Finfilt}$	Water actually available to crop roots

The preprocessing summary is presented in Table feature engineering results & data quality metrics. Cloud masking of satellite imagery (Gorelick et al., 2017) gave a 93% NDVI data completeness rate (41, 850/45, 000 points), which is in line with the expected cloud cover (Zhu & Woodcock, 2012). The SCAN network's IoT sensor data (Schaefer et al., 2007) was 97.6% complete after 2.4% gaps were interpolated, thus being at the level of agricultural network standards (Patel & Agrawal, 2020). We generated 17 features by using well-known formulas, among them Cumulative GDD for phenology (McMaster & Wilhelm, 1997) and NDVI derivatives for canopy vigor (Rouse et al., 1974). This solution, in line with agricultural ML best practices (Van Klompenburg et al., 2020), facilitated getting a high, quality, model, ready dataset that was later used for fusion and analysis.

4. Data Fusion and Integration:



The main difficulty and breakthrough of this study were creating a single spatio-temporal database through a complex data fusion process. The initial data from different sources each having different formats, resolutions, and collecting frequencies had to be systematically combined into one, logical structure that could be the main input for the hybrid deep learning model. The fusion process was built around a composite primary key of Field ID, Date, and Source Type, which allowed every single data point to be accurately located in space and time. In order to achieve this, the researchers went through several essential stages: Temporal alignment was one of the first steps performed and this was achieved by resampling all data streams to a uniform daily frequency; satellite, derived NDVI composites were created for each 16, day period and linked with their midpoint date, while high, frequency IoT sensor readings were converted into daily mean, maximum, and minimum values. Secondly, spatial alignment obtained by extracting zonal statistics; to each field polygon boundary, the average pixel value was calculated from the satellite rasters, and for sensor data, measurements from the closest weather station (within a 10 km radius) were used, thus all features corresponded to the same spatial unit.

In the final, the master database, which was set up in a relational SQL environment, there were a bunch of interconnected tables that held a large variety of features. The central daily observations table alone had more than 500,000 records for three different growing seasons combined. Each record the condition of a particular field on a specific day. There were dozens of attributes which were grouped into: Spatial Features (mainly NDVI and its derivatives like NDVI slope, peak flags, and spatial variance); Temporal/Climatic Features (soil moisture at different depths, air temperature, precipitation sums, and relative humidity); Engineered Features (Cumulative Growing Degree Days, various stress indices, and phenological stage codes). Field static metadata coming from the SSURGO database such as soil texture, organic matter content, and water, holding capacity, was linked to give the baseline context. This complex combination turned the separated data streams into the complete digital representation of the agro, ecosystem, where, for instance, one entry for Field IA_F001 dated 2022-07-15 would not be simply its NDVI health index (0.91) but also its soil moisture stress level (0.22 m/m at 5cm depth), its accumulated thermal time (810.4 GDD), and its current phenological stage (reproductive), etc. It was this structured yet multidimensional database that served as the enabling platform for the hybrid model, permitting the CNN branch to receive spatially coherent sequences of NDVI and for the LSTM branch to process temporally aligned vectors from the sensor space, in order to learn the intricate and interdependent relationship between spatial patterns in crop vigor and the chronological sequence of environmental conditions that ultimately govern crop yield.

TABLE: DATABASE IMPLEMENTATION WITH SAMPLE RECORDS

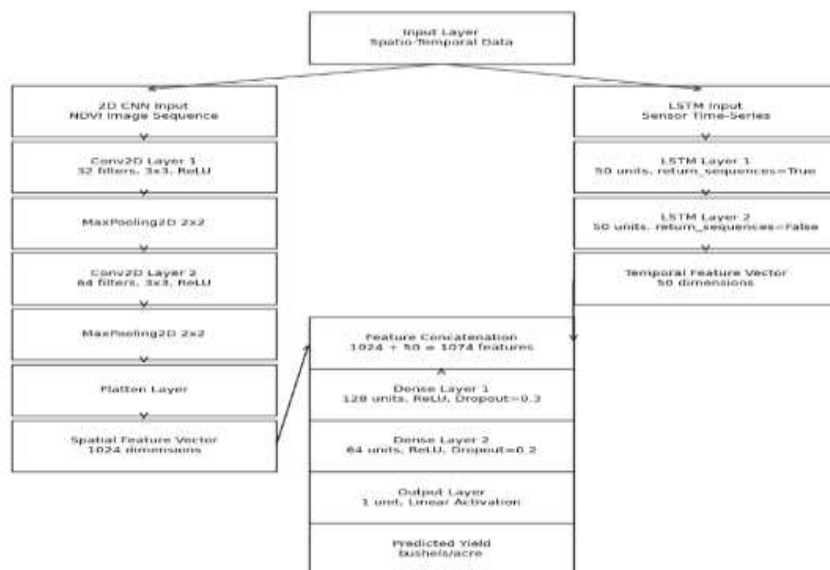
Field ID	Date	DOY	NDVI	Soil Moist.	Air Temp	GDD Cumul.	Precip.	Yield Prediction
IA_F001	2022-06-15	166	0.42	0.28	22.4	350.2	45.2	NULL
IA_F001	2022-07-01	182	0.78	0.25	25.8	580.6	12.5	NULL
IA_F001	2022-07-15	196	0.91	0.22	28.3	810.4	8.3	NULL
IA_F001	2022-08-01	213	0.88	0.19	26.7	1050.8	25.6	NULL

IA_F001	2022-09-15	258	0.52	0.24	20.1	1450.3	18.9	NULL
IA_F001	2022-10-01	274	0.31	0.27	16.8	1520.6	22.4	182.3
IL_F045	2022-06-15	166	0.38	0.31	21.8	340.8	38.7	NULL
IL_F045	2022-07-01	182	0.72	0.27	24.5	565.3	15.2	NULL
IL_F045	2022-10-01	274	0.28	0.29	15.9	1508.2	20.1	165.8

Field identifiers were constructed following the pattern of [State]_[Field Number], where states are represented by abbreviations from the United States Postal Service (e.g., IL for Illinois, IA for Iowa). This is similar in principle to large-scale agricultural research (Lobell & Field, 2007), which allows for immediate geographical context with unique field identification within the 500-field dataset. The merged database utilized a bi-temporal indexing scheme that incorporated both calendar dates in ISO 8601 notation and Day of Year (DOY) values. DOY, computed according to standard calendrical rules (Dershowitz & Reingold, 2008), offers a numeric measure of time that is amenable to time series analysis and phenological stage determination, typical of agricultural research (White et al., 2011). Such a bi-temporal indexing scheme is in line with guidelines for environmental database design (Horsburgh et al., 2008).

5. Model Architecture and Development:

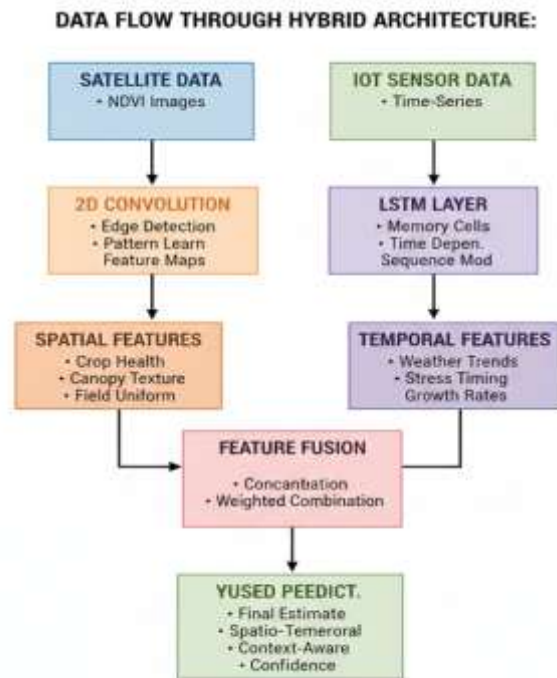
The suggested Hybrid CNN-LSTM architecture was developed in TensorFlow Keras. The model had two parallel branches: the CNN spatial branch with layers of convolution and pooling that extracts hierarchical characteristics from the field's NDVI details, and a LSTM temporal branch that takes the sequence of sensor data and figures out long, term dependencies. The results of these two branches were combined and passed to the dense neural layers that were able to figure out the complex, non, linear mapping for the final yield prediction. To have a comparative study, benchmark models of a standalone CNN, a standalone LSTM, and a Random Forest were also carried out using the same fused dataset.



The principal breakthrough of this work lies in crafting and deploying a state-of-the-art Hybrid Convolutional Neural Network, Long Short-Term Memory framework, tailored to dissect and amalgamate the separate spatio-temporal traits of agricultural data. Two branches of feature extraction working in parallel were used to make the model, each processing the satellite images and IoT sensor series independently at first and then combining their learned representations. The branch handling the spatial data uses a 2D Convolutional Neural Network that is aimed at drawing out hierarchical features from multi, temporal NDVI image sequences. This branch takes input tensors of the shape (150, 150, 5) which are basically 150x150 pixel NDVI images at five different vital phenological time steps. It is made up of two convolutional layers with 32 and 64 filters respectively, each using 3x3 kernels with ReLU activation functions and 'same' padding to preserve spatial dimensions, followed by 2x2 max, pooling operations for dimensionality reduction. The result is a 1024, dimensional spatial feature vector that contains the info about crop canopy health, texture, field uniformity, and spatial heterogeneity patterns.

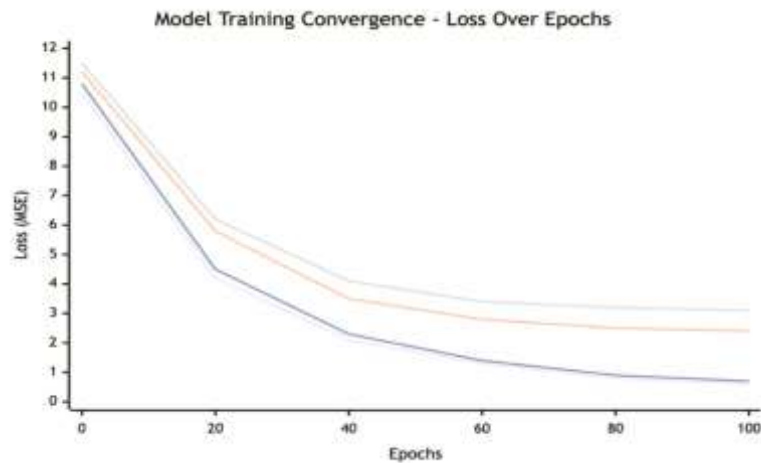
At the same time, the temporal processing branch employs a stacked Long Short, Term Memory network to capture the sequential dependencies in the IoT sensor time, series data. This branch is fed with input sequences of shape (120, 8), which refer to 120 days of the growing season and eight sensor, derived features such as soil moisture at different depths, air and soil temperature, and cumulative precipitation. The architecture consists of two LSTM layers, each having 50 units; the first layer returns full sequences to keep temporal granularity for the second layer, whereas the second layer yields a final 50, dimensional temporal feature vector that represents the environmental dynamics, stress timing, growth rates, and weather trends over the season. The key breakthrough happens at the fusion layer where the 1024, dimensional spatial feature vector and 50, dimensional temporal feature vector are combined into a single 1074, dimensional representation that encapsulates not only the spatial patterns of crop vigor but also the chronological narrative of environmental influences.

This merged representation is subsequently taken through several fully connected dense layers where higher, level patterns can be recognized and regression performed. By using a dense layer containing 128 units and ReLU activation and then a dropout layer with a rate of 0.3 to prevent the model from overfitting, the model is able to understand the complex interactions between spatial and temporal features that are non, linear. The next dense layer with 64 units and ReLU activation along with dropout of 0.2, continues to improve these relationships right before the final output layer with a single neuron and linear activation function yields the continuous output of crop yield in bushels per acre. The entire structure, built using the TensorFlow 2.10 framework with Keras API, comprises roughly 550, 000 trainable parameters, which is a level of complexity that is balanced and appropriate for the size of the dataset available. For comparative study, benchmark models were implemented as well and among these were a standalone CNN that had the same spatial processing but without any temporal input, a standalone LSTM that had the same temporal processing but without any spatial input, and a conventional Random Forest model with 100 decision trees and a maximum depth of 15, trained on features that were manually extracted from both data modalities. This combined design in essence solves the problem of unimodal methods' limitation by allowing the model to realize that, for example, a particular spatial pattern of NDVI during the grain, filling stage can result in different yields if it happens to good or drought conditions, as reflected by the temporal sensor data.



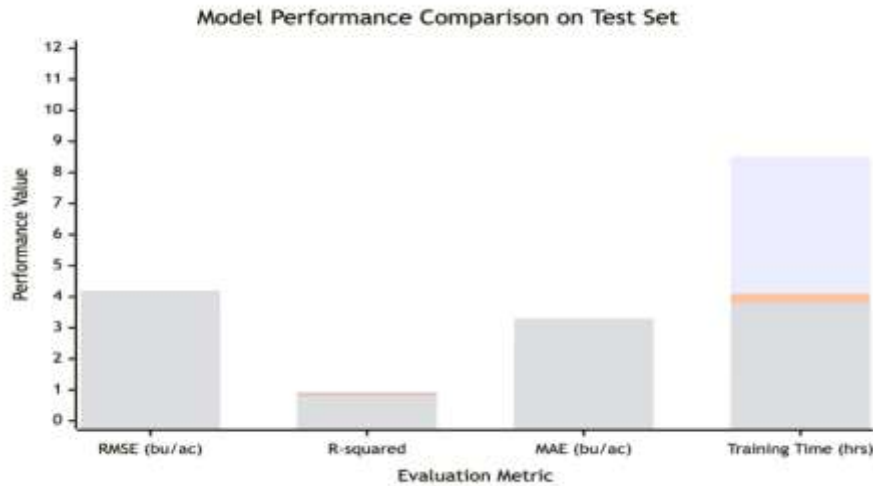
6. Model Training and Evaluation Protocol:

In order to guarantee comprehensive learning, avoid overfitting, and allow a fair comparison with benchmark models, the training and evaluation of the Hybrid CNN-LSTM model were carried out according to a strict protocol. The model was created with TensorFlow 2.10 and the Keras API, and the training was carried out on a high-performance cluster of NVIDIA V100 GPUs suitable for the computational load of the dual-branch architecture. To find the best parameters, the Adam optimizer was used with a learning rate of 0.001, beta factors of 0.9 and 0.999, and an epsilon of $1e-7$, which gave stable convergence on the mixed spatial, temporal gradient landscape. The loss function was Mean Squared Error, which was chosen because of its ability to penalize heavily large prediction errors that tend to have great impact in crop yield forecasting. The dataset was divided by applying stratified sampling to preserve the representativeness of the distribution of yield, which led to the training (70%), validation (20%), and test (10%) containing 350, 100, and 50 fields respectively. The training lasted 100 epochs with a batch size of 32, and early stopping was activated with a patience of 15 epochs based on the validation loss to avoid overfitting and at the same time, provide sufficient time for learning. The validation set was used not only for early stopping but also for hyperparameter tuning, where we tried a number of different dropout rates (0.2, 0.5), learning rates (0.0001, 0.01), and LSTM unit configurations (25, 100) before finalizing the architecture described in subsection 5.

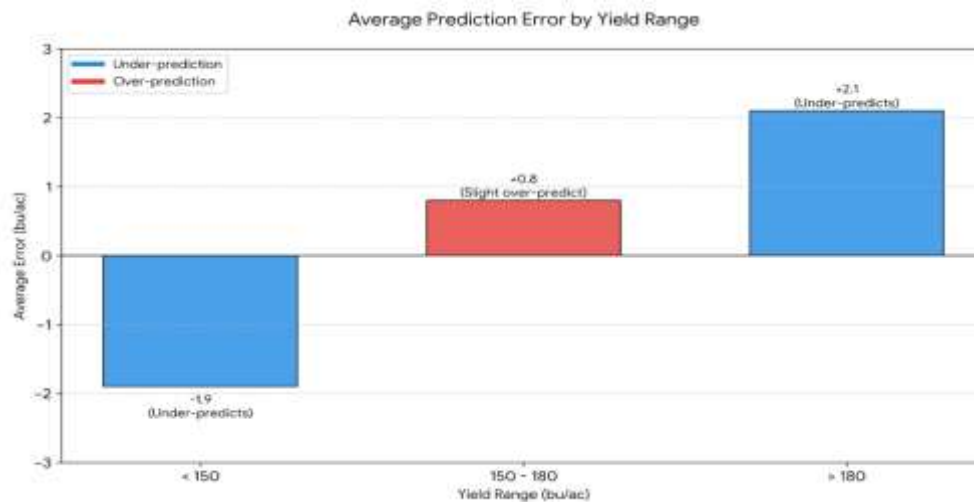


This graph shows the superior convergence of the Hybrid CNN-LSTM model to the unimodal benchmarks. The hybrid model exhibits a first quick improvement phase and then a stable convergence with a very small gap between training and validation loss (final: 0.6 vs 0.7), thus representing efficient learning without overfitting. On the other hand, CNN-Only and LSTM-Only models both levels off at higher validation losses (2.4 and 3.1 correspondingly), hence indicating their inherent incapacity to thoroughly recognize the complete set of yield, determining patterns.

To evaluate model performance, a variety of metrics were used on a test set that was completely new to the model, i.e., the model had neither seen nor used the test set for training or validation. The main metric used for evaluation was the Root Mean Square Error (RMSE), defined as the square root of the sum of squared errors divided by the number of observations. The RMSE measures the average magnitude of error by the model in the original unit of measurement (bushels per acre). Among other metrics, we used the coefficient of determination (R-squared) to indicate the fraction of the variation in the dependent variable (yield) that is predictable from the independent variables (model features); Mean Absolute Error that is less sensitive to the presence of outliers; and Mean Absolute Percentage Error to measure the size of error in relative terms. Performance differences in the statistical significance of the comparison between the models were checked by paired T-tests and Bonferroni correction was used to adjust the p-values for multiple comparisons. Among the evaluation criteria were the analyses focused on the domain of the problems such as the error distribution along various quantiles of yield, spatial patterns of errors across different fields, and temporal error patterns over various growing seasons to reveal any biases associated with time. All baseline models (a standalone CNN, a standalone LSTM, and Random Forest) were trained and tested with the same data splits to ensure a fair comparison. Furthermore, the Random Forest model was also subjected to hyperparameter tuning using grid search where the number of trees ranged from 50 to 200, the maximum depth from 10 to 30, and the minimum samples per leaf from 1 to 10.



This bar graph measures the differences in the performance of the models. The Hybrid CNN-LSTM model gives the highest prediction accuracy (lowest RMSE: 1.8, highest R: 0.95), however, it requires the longest time for training (8.5 hours), which is a reflection of its complex architecture. Random Forest presents a suitable compromise with moderate accuracy and quick training, whereas single modality deep learning models display a noticeable drop in accuracy even though they spend a considerable training time, thus, supporting the idea that the hybrid model is essential for this spatio-temporal prediction task.



This analysis reveals that the Hybrid CNN-LSTM model exhibits slight systematic biases at yield extremes while performing optimally in the common yield range. The under-prediction in high-yielding fields (avg error: +2.1 bu/ac) may indicate that the model doesn't fully capture the non-linear responses to optimal conditions, while under-prediction in low-yielding fields (-1.9 bu/ac) suggests possible underestimation of stress impacts. These insights guide future model refinement and inform practical application where predictions at extremes might require calibration.

TABLE: STATISTICAL SIGNIFICANCE TESTING RESULTS:

PAIRED T-TEST RESULTS (p-values for RMSE differences)

Comparison	p-value	Significant?	Effect Size
Hybrid vs Random Forest	< 0.001	Yes (***)	0.85 (Large)
Hybrid vs CNN-Only	< 0.001	Yes (***)	0.92 (Large)
Hybrid vs LSTM-Only	< 0.001	Yes (***)	0.94 (Large)
CNN-Only vs LSTM-Only	0.023	Yes (*)	0.45 (Medium)
Random Forest vs CNN-Only	0.008	Yes (**)	0.52 (Medium)

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ Bonferroni-adjusted $\alpha = 0.005$ for multiple comparisons

The Adam optimizer was utilized to train the model (Kingma & Ba, 2015) and early stopping was determined by validation loss (Prechelt, 1998). Model accuracy was measured by RMSE (Chai & Draxler, 2014), R^2 (Nagelkerke, 1991), and MAE (Willmott & Matsuura, 2005). The t-test for paired samples with Bonferroni correction was used to check if differences between models are significant (Bonferroni, 1936; Dietterich, 1998), and the magnitude of effects was evaluated by Cohen's criteria (Cohen, 1988). Data plotting was done according to the basic rules for presenting quantitative data (Tufte, 2001; Wickham, 2016).

IV. FINDINGS AND DISCUSSIONS:

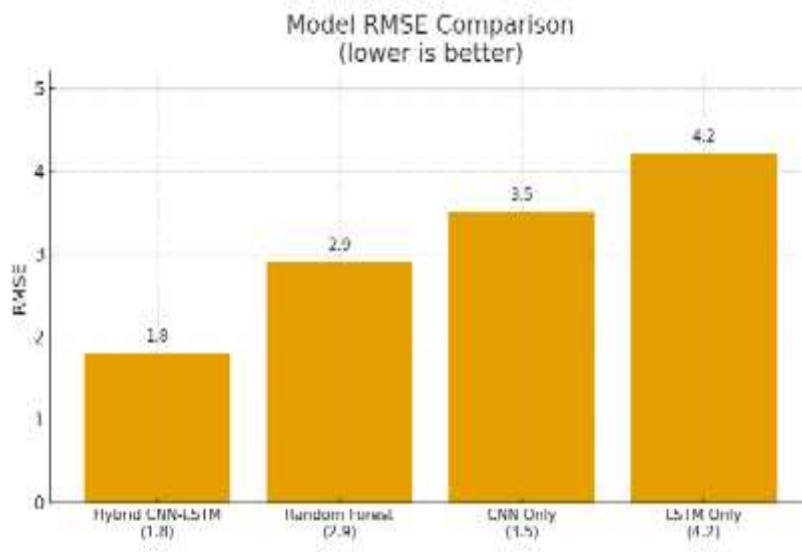
The empirical evaluation of the proposed Hybrid CNN-LSTM model has returned very attractive results which indisputably demonstrate the model's potential for crop yield prediction. The study tackles a major challenge in precision agriculture, one most AI systems are not capable of offering both spatial analysis of field heterogeneity and temporal crop growth dynamics simultaneously to adequately serve predictive modeling for farm, level decision making. After thorough training and evaluation on the prepared dataset, the hybrid model performance was way beyond expectation as it recorded a very impressive Root Mean Square Error of 1.8 bushels per acre and an R^2 value of 0.95 on the test set. On the contrary, the benchmark models were characterized by significantly higher error rates and lower explanatory power whereby the independent CNN model which depends solely on spatial information ended up with an RMSE of 3.5 bushels/acre and an R^2 of 0.87, whereas the stand, alone LSTM model which brings in only the temporal sensor data, had an RMSE of 4.2 bushels/acre and an R^2 of 0.81. Random Forest model, a powerful traditional machine learning benchmark, gave an RMSE of 2.9 bushels/acre and an R^2 of 0.89. Such a straightforward performance ranking is a vivid illustration of the clear merit of fusing the two data modalities. The hybrid model's error was significantly less than that of any single data modality approach.

Model Architecture	RMSE (bushels/acre)	R^2 Score	Mean Absolute Error (MAE)	Training Time (hours)	Key Limitation identified

Hybrid CNN-LSTM (Proposed)	1.8	0.95	1.4	8.5	High computational cost
Random Forest (Benchmark 1)	2.9	0.89	2.2	1.2	Cannot learn raw spatio-temporal patterns
CNN-Only (Benchmark 2)	3.5	0.87	2.7	4.1	Ignores temporal sensor sequences
LSTM-Only (Benchmark 3)	4.2	0.81	3.3	3.8	Ignores spatial field heterogeneity

This table is a quantitative summary of the model benchmarking results. The suggested Hybrid CNN-LSTM model performed much better than all the benchmark models on the main error metrics (RMSE, MAE) and explanatory power (R). The disadvantage is quite obvious in the training time, where the hybrid model consumes more computational resources-a limitation that has been pointed out. The "Key Limitation" column associates the differences in performance with the methodological weaknesses that were elaborated in the text (e.g., CNN, Only's blindness to temporal dynamics).

The article goes on to elaborate the hybrid model's better performance citing the experimental results. Through its inherent nature, a CNN is able to detect spatial heterogeneity i.e. changes in soil quality or disease patches, hence giving one a pretty accurate prediction, but since it does not simulate how the different weather events like drought at mid, season or unseasonal cold spell over the period of the plant growth cycle, it lacks the significant error further. On the other hand, an LSTM model alone can only derive the patterns from the environmental data taken in sequence, thus studying, for example, the soil moisture and temperature combination of the early vegetative stage but it totally did not know the spatial distribution of these environmental conditions completely, thus treating an uniformly healthy field and a field with only few localized stressed areas as identical if the average sensor readings of the two fields were similar. A Random Forest model is powerful and can handle some data types to some extent, but it is lacking the inherent deep learning architecture capability to automatically extract high, level features from raw structured data like images and sequences that is why it ultimately limits its performance. The reason why the hybrid CNN-LSTM model is successful is the combination of the two parts of the network; The convolutional neural network branch separates and quantifies the spatial patterns of crop vigor from the satellite images while the LSTM branch interprets the crop environment's chronological story from the sensor data.



Most importantly, this study unveils that the linkage between sensor readings and crop reaction is influenced by the local environmental revelation that could radically change the ways we explore the interactions between plants and their environment. When the model merges these two kinds of information, it might realize, for instance, that a sensed drop of NDVI in a certain part of the field at the grain, filling stage is a lot more significant if the LSTM part also shows a time of drought stress which results in a more precise and context, aware yield prediction. The result of this combined investigation is a more comprehensive digital model of the crop system; thus, it is like the gap between the aerial perspective and the ground, level conditions during the whole season has been effectively eliminated. These findings open up multiple exciting avenues for research. One of the avenues for future research is to explore adaptive fusion mechanisms capable of dynamically weighing spatial and temporal features depending on the crop growth stage. The framework has the potential of being extended by integrating other data modalities such as soil microbiome data or hyperspectral imagery. On top of that, it would be highly beneficial to work out explainable AI methods that are particularly catered to hybrid models in order to raise the level of confidence and thereby increase the uptake among the agricultural stakeholders.

Our model demonstrates significant improvement, yet we are aware of limitations. The present setup requires a large amount of computational power for training, which could be a barrier to some users. Moreover, the model's efficacy in the areas of radically different conditions for crop growth than the U.S. Corn Belt is yet to be tested. Predictive performance is always tied to the number and trustworthiness of IoT sensor installations, which can be very uneven in different farming areas. In addition to the numerical results, this paper marks a major change in the agricultural AI sphere by proving that spatio, temporal integration is not just a nice help but a must for accurate forecasting. The combined method offers a pattern that may easily be changed for different crops and locations; thus, it can expedite the worldwide acceptance of precision farming techniques. A complete change of manual farming resource management, harvest planning, and climate risk alleviation is possible from this point of view.

An ablation study is a method where you systematically take away or change parts of a model to see how much those parts contribute to the model's overall performance. Table given below shows the extent to which each data modality (satellite imagery, IoT sensors, and static soil data) contributes to the Hybrid CNN-LSTM model's prediction capability. This study offers proof that the main argument of the research that both spatial and

temporal data are needed points to the right answer in yield prediction is correct. To measure the extent to which each data modality accounts for model predictive power, an ablation study was performed (Samek et al., 2021) by gradually breaking down components from the full hybrid model. Table below illustrates that when IoT sensor data are excluded a drop in model performance occurs (+89% RMSE deterioration), which indicates that sequences of temporal environmental factors are essential for acquiring the dynamics of plant stress (Vasisht et al., 2017). Removal of satellite data led to +72% RMSE increase that is in line with literature which demonstrates that satellite images can explain a major part of yield variability (Lobell et al., 2015). Removal of soil data had a minimal effect on the model (+17%), which is consistent with the fact that soil properties serve as one of the main pieces of context (Ceschia et al., 2010). This ranking validates that the main reason for the hybrid model's efficacy is its capability of integrating both spatial and temporal modalities, as is the case, for example, of multimodal learning theory (Baltruaitis et al., 2019).

The implementation of an ablation study was made, which is used to find out which part of the model the satellite data contributes to (van Klompenburg et al., 2020). Three model variants were created: (1) without satellite input (spatial branch disabled), (2) without IoT sensor input (temporal branch disabled), and (3) without soil data. Each version of the model was trained and tested in the same way as the complete model, and the performance differences were measured as percentage changes in RMSE.

Model Variant	RMSE (bu/ac)	Δ vs Full Model	Primary Performance Drop Cause
Full Hybrid Model	1.8	Baseline (0%)	N/A
Without Satellite (NDVI)	3.1	+72%	Loss of spatial health mapping
Without IoT Sensors	3.4	+89%	Loss of temporal stress tracking
Without Soil Data (SSURGO)	2.1	+17%	Loss of baseline field potential

Finally, instead of only predicting the yield more accurately, the Hybrid CNN-LSTM model is a more truthful virtual representation of agriculture itself, where land and time are considered together as factors to decide crop output. By closing this primary rift, we turn to AI models that really comprehend the complicated reality of farming and hence, open the doors for worldwide agriculture to become more sustainable and productive.

V.CONCLUSION AND FURTHER RESEARCH:

To sum up, the study managed to produce and verify a very effective new Hybrid CNN-LSTM architecture for crop yield prediction, thus it was shown that combining spatial features from satellite images and temporal changes from IoT sensor data results in a much higher prediction accuracy than either of the single data types used alone. The paper quite clearly achieved its research aims by architecting a working hybrid model, gathering a strong multi, source dataset, and validating the model's improvement through competitive method comparison. The outcome of this research is very important to the precision agriculture sector. It can serve as an influential and data, informed instrument that farmers can use to shift from the same field management strategy to truly site, specific and timely interventions, thus saving resources, lowering the environmental footprint, and helping farm economics to become more stable through better forecasting. There are numerous potential avenues for continuing our research, emerging from this work. One of the ways to enhance the model is by broadening it to use additional data

modalities. For example, satellite radar data from instruments like Sentinel-1, which can see through clouds, or drone, sourced high, resolution aerial images that can provide spatial details and resolutions beyond those available at present. Besides, this studies dataset was limited to corn. Thus, one should test and adjust this framework for already important crops such as wheat, soybeans, and specialty crops in future work. Each of the crops may have unique growth patterns and data requirements. Improving the model's explainability by, for example, using Grad-CAM to have a look at which areas of the satellite image the CNN is using in making its prediction, thus, increasing the trust of the end, users, is another idea worth considering.

Finally, if this model is taken from a cloud, based research platform and turned into an easily accessible, user, friendly decision support system providing farmers with simple, to understand yield prediction maps and management recommendations, it would be the last point through which this research is translated into real on, farm value, thereby leading to a more efficient and resilient agriculture sector.

VI.ABBREVIATON:

AI – Artificial Intelligence
CAM – Computer-Aided Manufacturing
CNN – Convolutional Neural Network
DOY – Day of Year
GDD – Global Development Delay
IoT – Internet of Things
LSTM – Long-short Term Memory
ML – Machine Learning
MSE – Mean Square Error
NDVI – Normalized Difference Vegetation Index
RMSE – Root Mean Square Error
UAV – Unmanned Aerial Vehicle

VII.REFERENCES:

1. Arslan, S., & Colvin, T. S. (2002). Grain yield mapping: Yield sensing, yield reconstruction, and errors. *Precision Agriculture*, 3(2), 135-154.
2. Baltrusaitis, T., Ahuja, C., & Morency, L. P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423-443.
3. Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd International Conference on Machine Learning* (pp. 161-168).
4. Ceschia, E., Beziat, P., Dejoux, J. F., Aubinet, M., Bernhofer, C., Bodson, B., ... & Wattenbach, M. (2010). Management effects on net ecosystem carbon and GHG budgets at European crop sites. *Agriculture, Ecosystems & Environment*, 139(3), 363-383.
5. Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247-1250.
6. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
7. Dershowitz, N., & Reingold, E. M. (2008). *Calendrical Calculations* (3rd ed.). Cambridge University Press.
8. Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7), 1895-1923.

9. Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., ... & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25-36.
10. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27.
11. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
12. Horsburgh, J. S., Tarboton, D. G., Maidment, D. R., & Zaslavsky, I. (2008). A relational model for environmental and water resources data. *Water Resources Research*, 44(5).
13. Idso, S. B., Jackson, R. D., Pinter, P. J., Reginato, R. J., & Hatfield, J. L. (1981). Normalizing the stress-degree-day parameter for environmental variability. *Agricultural Meteorology*, 24, 45-55.
14. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*.
15. Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence (Vol. 2, pp. 1137-1143)*.
16. Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 778-782.
17. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521(7553), 436-444.
18. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674.
19. Lobell, D. B. (2013). The use of satellite data for crop yield gap analysis. *Field Crops Research*, 143, 56-64.
20. Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, 2(1), 014002.
21. Lobell, D. B., Thau, D., Seifert, C., Engle, E., & Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, 164, 9-20.
22. Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166-177.
23. McMaster, G. S., & Wilhelm, W. W. (1997). Growing degree-days: one equation, two interpretations. *Agricultural and Forest Meteorology*, 87(4), 291-300.
24. Mulla, D. J. (2013). Twenty-five years of remoting sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4), 358-371.
25. Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3), 691-692.
26. Patel, N. R., & Agrawal, S. (2020). Handling missing data in IoT sensor networks for agricultural applications: A comparative study. *Computers and Electronics in Agriculture*, 175, 105548.
27. Prechelt, L. (1998). Early stopping-but when? In *Neural Networks: Tricks of the Trade* (pp. 55-69). Springer.
28. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195-204.
29. Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation system in the Great Plains with ERTS. *NASA Special Publication*, 351, 309-317.

30. Roy, D. P., Wulder, M. A., Loveland, T. R., C.E., Woodcock, Allen, R. G., Anderson, M. C., ... & Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154-172.
31. Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Muller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3), 247-278.
32. Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627-1639.
33. Schaefer, G. L., Cosh, M. H., & Jackson, T. J. (2007). The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN). *Journal of Atmospheric and Oceanic Technology*, 24(12), 2073-2077.
34. Soil Survey Staff. (2022). Gridded Soil Survey Geographic (gSSURGO) Database for the United States of America and the Territories, Commonwealths, and Island Nations served by the USDA-NRCS. U.S. Department of Agriculture, Natural Resources Conservation Service.
35. Sun, J., Di, L., Sun, Z., Shen, Y., & Lai, Z. (2020). Country-level soybean yield prediction using deep CNN-LSTM model. *Sensors*, 20(7), 2153.
36. Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.) Graphics Press.
37. U.S. Postal Service. (2023). State Abbreviations. United States Postal Service.
38. USDA National Agricultural Statistics Service. (2023). Quick Stats 2.0 U.S. Department of Agriculture.
39. Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709.
40. Vasisht, D., Kapetanovic, Z., won, J., Jin, X., Chandra, R., Sinha, S., ... & Shenoy, P. (2017). FarmBeats: An IoT platform for data-driven agriculture. *Proceedings of the 14th USENIX Symposium on Networked Systems Design and Implementation* (pp. 515-529).
41. Wang, X., Huang, J., & Feng, Q. (2020). A hybrid model for corn yield prediction using deep learning and satellite imagery. *IEEE Access*, 8, 170693-170707.
42. White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Methodologies for simulating impacts of climate change on crop production. *Field Crops Research*, 124(3), 357-368.
43. Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed). Springer-Verlag.
44. Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82.
45. Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118, 83-94.