International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

A Robust Hybrid AI Framework for Predicting Date Palm Yields from Multiple Sources Amidst Environmental Change

Ziyad Alhazmi[1], Muhannad Alahmadi[1], Sheikh Mohammed[1], Mostafa Elgayar[1]
Department of Computer Science, Arab East Colleges, Riyadh 11583, Saudi Arabia
Email address: mmalgayar@arabeast.edu.sa

Abstract—*This paper proposes a robust hybrid artificial intelligence framework for predicting date palm yields by integrating multi-source agricultural data including climatic variables, soil physicochemical properties, satellite-derived vegetation indices (NDVI), and historical yield records within a unified CNN-LSTM-XGBoost architecture enhanced with SHAP-based explain ability. The framework is designed for cross-regional validation across desert and semi-desert agroecological zones, addressing a critical gap in computational modeling for date palm systems.*

Keywords— *Date Palm; Hybrid AI; CNN-LSTM; XGBoost; Yield Prediction; SHAP; Remote Sensing; Precision Agriculture.*

I. INTRODUCTION

The date sector is one of the most important strategic agricultural sectors in many desert and semi-desert countries, with global annual production exceeding 9.5 million metric tons across more than 40 countries [8]. It represents a cornerstone of food security, a vital economic resource, and a key component in agricultural value chains and processing industries [19]. Date cultivation is gaining particular importance with the expansion of agricultural development programs, increased investment in modern technologies, and progress toward smart and sustainable agriculture [7].

However, date production faces increasing challenges due to the interplay of environmental, climatic, and economic factors, most notably: rainfall variability, rising temperatures, worsening droughts, water scarcity, varying soil characteristics, the spread of agricultural pests, and rising production costs [18]. Climate change has further increased the uncertainty surrounding production seasons,

with the FAO estimating that approximately 22% of global agricultural yield losses are directly attributable to climate-related events [8], making accurate production forecasting more complex than ever before.

In this context, agricultural production forecasting models have become essential for precision agriculture systems, supporting decision-making for farmers and regulators, improving water resource management, harvesting, storage, and marketing processes, reducing losses, and enhancing economic and environmental sustainability [11].

Despite significant progress in applying artificial intelligence (AI) technologies to agriculture, most previous studies have relied on single-source models or individual techniques such as artificial neural networks, decision trees, or regression models [19, 17]. While important, these approaches fail to capture the complex, non-linear, and time-varying nature of date production, which is influenced by the multidimensional interaction of climatic factors, soil characteristics, remote sensing data, irrigation patterns, and agricultural practices [5, 14].

Relying on a single model or data source often leads to lower forecast accuracy, limited generalizability across geographic regions, and performance uncertainty when environmental conditions change [9, 16]. This highlights a clear research gap: the absence of an integrated hybrid model that combines multiple AI technologies and integrates diverse data sources within a unified framework capable of representing the complex and dynamic relationships affecting date production [2].

Therefore, the problem addressed by this study is the need to develop a hybrid AI model capable of integrating diverse multi-source agricultural data, representing nonlinear and time-varying relationships, improving the accuracy of crop production forecasts, achieving stability and generalizability across varying climatic and environmental conditions, and supporting smart farming systems and data-driven decision-making. Solving this problem will increase agricultural planning efficiency, improve water resource management, reduce production risks, and make data-driven decisions more reliable, in line with the digital transformation in agriculture [20].

Based on the above, this study seeks to bridge this gap by designing, developing, and evaluating a hybrid artificial intelligence model capable of providing accurate and stable predictions for date production, thereby supporting agricultural sustainability and enhancing the efficiency of the Date production system in desert and semi-desert environments.

This study makes several original contributions to the field of agricultural AI and date palm yield prediction. First, it proposes the first comprehensive hybrid CNN-LSTM-XGBoost framework specifically designed and validated for date palm yield prediction, addressing a critical gap in the existing literature where hybrid frameworks have been predominantly applied to staple grain crops. Second, the framework integrates four heterogeneous data streams climatic variables, soil physicochemical properties, satellite-derived vegetation indices (NDVI), and historical yield records within a single unified architecture, enabling the model to capture both spatial and temporal dimensions of date production dynamics

Ziyad Alhazmi, Muhannad Alahmadi, Sheikh Mohammed, and Mostafa Elgayar, "A Robust Hybrid AI Framework for Predicting Date Palm Yields from Multiple Sources Amidst Environmental Change," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 9, Issue 3, pp. 113-118, August 2026.

simultaneously. Third, the study incorporates SHAP-based explainable AI (XAI) to provide transparent, interpretable predictions that translate directly into actionable recommendations for farmers, water resource managers, and agricultural policymakers [9]. Fourth, the framework is designed for cross-regional validation across multiple agroecological zones in desert and semi-desert environments, rigorously testing model generalizability under diverse climatic and soil conditions [17, 16]. Fifth, by combining ensemble stacking with attention mechanisms, the proposed architecture is capable of adaptively weighting input features according to their predictive relevance across different growth stages, improving both accuracy and robustness under climate uncertainty [18, 12].

II. Literature Review

The following section reviews prior research on date palm productivity forecasting, hybrid artificial intelligence modeling, remote sensing integration, and climate-adapted agricultural forecasting.

Rahmanian et al. [18] systematically reviewed the vulnerability of date palm cultivation to projected climate change scenarios, documenting evidence that rising temperatures above 32°C during pollination and fruit development stages significantly reduce fruit set and quality across major producing regions. Their analysis identified temperature extremes, shifting precipitation regimes, and increased drought frequency as the primary climate-driven threats to date yields, and evaluated adaptation strategies implemented in Saudi Arabia, Iran, and North Africa.

Alhammadi & Kurup [1] conducted a comprehensive review of the impacts of salinity stress on Phoenix dactylifera L., quantifying yield reductions of 15–40% across varying salinity thresholds and examining how soil electrical conductivity interacts with temperature and irrigation management to determine final productivity. The review also highlighted the scarcity of integrated predictive models that jointly account for soil and climatic stressors in date palm systems a gap the current study explicitly addresses.

Dehghanisanij et al. [6] developed a hybrid intelligent model combining an Anti-Coronavirus Optimization algorithm (ACVO) with an Adaptive Neuro-Fuzzy Inference System (ANFIS) to estimate date palm crop yield and water productivity under three levels of subsurface drip irrigation. Their model demonstrated that hybrid optimization-inference architectures reduced prediction RMSE by 23% compared to standalone ANFIS models and identified irrigation quantity and solar radiation as the dominant predictors of yield variability.

Rashid et al. [19] conducted a landmark comprehensive review of machine learning applications in palm oil yield prediction, systematically evaluating algorithms including Linear Regression, Random Forest, ANN, and SVM across more than 60 studies. Their meta-analysis revealed that ensemble methods consistently outperformed single-algorithm approaches, and that combining meteorological and biometric variables produced the largest accuracy improvements.

Ang et al. [2] built a multi-source predictive system for palm yield estimation at block level, integrating satellite-derived NDVI, WDRVI, land surface temperature (LST), precipitation, weather station records, and field survey data. Comparing XGBoost, SVR, Random Forest, and Deep Neural Networks, they demonstrated that multi-source integration improved R^2 by 12–18% over single-source models, with XGBoost achieving the highest performance.

Mohamad Radzi et al. [13] evaluated 17 machine learning and deep learning models for palm yield prediction using a comprehensive agronomic dataset spanning over a decade of plantation records. Their extensive comparison revealed that no single algorithm universally dominated across all metrics, directly reinforcing the rationale for hybrid model design.

Cao et al. [5] developed a large-scale rice yield prediction framework for China, integrating MODIS satellite imagery, daily meteorological observations, and historical yield statistics across diverse climatic zones. Their hybrid approach combined Random Forest for feature selection with deep neural networks for final prediction, consistently achieving R^2

> 0.85 across validation regions and outperforming single-model baselines by margins of 15–22%.

Paudel et al. [17] evaluated ML algorithms for large-scale wheat yield forecasting across European zones, with particular attention to cross-regional generalizability. Their key finding that LSTM networks outperformed Gradient Boosted Decision Trees in capturing inter-annual yield variability provides direct empirical support for including LSTM components in the temporal modeling layer of the proposed framework.

Wang et al. [22] developed a hybrid seasonal prediction system for winter wheat yield in northern China by coupling a dynamical climate model with ML post-processing. Results showed that physics-data hybrid combinations achieved substantially higher forecast skill than either component alone, and that ensemble averaging of multiple ML algorithms reduced prediction variance by 30% compared to any individual model.

Oikonomidis et al. [16] conducted a systematic review of deep learning for crop yield prediction, analyzing 44 primary studies. Their synthesis found that the CNN-LSTM hybrid architecture consistently outperformed standalone CNN or LSTM models when processing multi-temporal remote sensing data. Crucially, they identified the lack of validated multi-source datasets as the primary constraint on current model performance.

Muruganantham et al. [14] systematically reviewed deep learning and remote sensing integration for crop yield prediction across 44 studies spanning 2010 to 2022. Their analysis confirmed that CNN and LSTM architectures fed with Sentinel-2 and MODIS time-series NDVI achieved the highest accuracy, and established that multi-temporal data fusion was the single most impactful methodological factor.

Bouras et al. [4] developed ML-based yield forecasting models for cereal crops in Morocco using satellite drought indices, regional climate indicators, and weather data. Results showed drought-based remote sensing indices were superior predictors of yield variability compared to precipitation data alone, achieving RMSE reductions of 18–27% over baseline statistical models.

Naghdyzadegan Jahromi et al. [15] developed ML models

Ziyad Alhazmi, Muhannad Alahmadi, Sheikh Mohammed, and Mostafa Elgayar, "A Robust Hybrid AI Framework for Predicting Date Palm Yields from Multiple Sources Amidst Environmental Change," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 9, Issue 3, pp. 113-118, August 2026.

for wheat yield prediction integrating ground-measured data, satellite actual evapotranspiration (ET), and multiple vegetation indices. Their study demonstrated that combining ET with NDVI and EVI produced substantially more accurate predictions than vegetation indices alone, with an R^2 improvement of 0.09.

Aviles Toledo et al. [3] developed a multi-modal deep learning architecture fusing hyperspectral UAV imagery, LiDAR point clouds, and environmental sensor data through late-fusion LSTM networks with attention mechanisms for maize yield prediction. Their study demonstrated that attention mechanisms improved model interpretability by highlighting biologically meaningful growth stage features, and that late fusion of heterogeneous data modalities outperformed early fusion.

Hu et al. [9] critically examined the risks of black-box ML models in assessing climate change impacts on crop yields. Integrating SHAP-based explainable AI with ensemble ML, they identified temperature and vapor pressure deficit as the dominant climate-yield relationship drivers globally, providing an important template for the interpretability component of the current study.

Iqbal et al. [10] evaluated ML algorithms for wheat yield prediction under CMIP6 climate change scenarios, demonstrating that ensemble ML models maintained robust predictive accuracy where process-based crop models showed substantial degradation under future climate projections.

Saleem et al. [20] reviewed ML, DL, ensemble learning, and XAI techniques for crop yield prediction under abnormal climate conditions spanning 2020 to 2024. Their meta-analysis found that stacking-based ensemble methods achieved the highest overall performance across diverse climate conditions, directly informing the ensemble design of the current framework.

Liu et al. [12] assessed a CNN-LSTM-Attention hybrid for multi-crop yield prediction in Northeast China using multi-source data from 2014 to 2020. Their model outperformed standalone CNN, LSTM, RF, and XGBoost across all crop types, with R^2 improvements of 8–15% attributable to the attention mechanism weighting critical growth stage features.

TABLE I. Comparison of Related Studies on Crop Yield Prediction Using AI and Remote Sensing

S. No.	Comparison of Previous Studies		
	Search title and study year	**Methodology/technique used**	**Innovative tool/proposal**
1	Rahmanian et al. (2023) [18]	Climate scenario review; temp and drought impact on date palm	Justifies climatic variables as core predictors in the model
2	Alhammadi & Kurup (2022) [1]	Salinity stress review; soil EC and irrigation interaction	Informs multi-variable soil input design
3	Dehghanisanij et al. (2023) [6]	ACVO-ANFIS hybrid model; subsurface drip irrigation levels	Direct hybrid AI precedent on date palm; 23% RMSE reduction
4	Rashid et al. (2021) [19]	LR, RF, ANN, SVM meta-analysis across 60+ palm oil studies	Ensemble methods outperform single algorithms consistently
5	Ang et al. (2022) [2]	XGBoost, SVR, RF, DNN; NDVI, WDRVI, LST, precipitation fusion	Multi-source integration improved R^2 by 12–18%
6	Mohamad Radzi et al. (2024) [13]	17 ML/DL models; decade-long agronomic plantation dataset	No single algorithm dominates; reinforces hybrid model rationale
7	Cao et al. (2021) [5]	RF + DNN; MODIS imagery, meteorological and historical yield data	$R^2 > 0.85$; outperforms single-model baselines by 15–22%
8	Paudel et al. (2021) [17]	LSTM vs. GBDT; large-scale European wheat yield forecasting	LSTM outperforms GBDT for inter-annual yield variability
9	Wang et al. (2022) [22]	Dynamical climate model coupled with ML post-processing ensemble	Ensemble averaging reduces prediction variance by 30%
10	Oikonomidis et al. (2023) [16]	Systematic review of 44 DL studies for crop yield prediction	CNN-LSTM hybrid outperforms standalone models consistently
11	Muruganantham et al. (2022) [14]	Systematic review; Sentinel-2 and MODIS NDVI time-series	Multi-temporal fusion is the most impactful methodological factor
12	Bouras et al. (2021) [4]	ML with drought remote sensing indices; Morocco cereal yields	RMSE reduced 18–27% over baselines in arid environments
13	Naghdyzadegan et al. (2023) [15]	ML models integrating ground data, ET, NDVI, and EVI	ET + NDVI combination improves R^2 by 0.09 over NDVI alone
14	Aviles Toledo et al. (2024) [3]	Late-fusion LSTM with attention; UAV hyperspectral and LiDAR data	Attention mechanism improves interpretability and fusion design
15	Hu et al. (2023) [9]	SHAP + ensemble ML; climate change impact on global crop yields	SHAP identifies temperature as dominant driver; XAI template
16	Iqbal et al. (2024) [10]	Ensemble ML under CMIP6 future climate scenarios for wheat	Ensemble ML robust under future climate projection uncertainty
17	Saleem et al. (2025) [20]	Meta-analysis: ML, DL, stacking, XAI under abnormal climate	Stacking achieves highest performance across diverse climates
18	Liu et al. (2024) [12]	CNN-LSTM-Attention; multi-source data; Northeast China 2014–2020	R^2 improved 8–15% with attention; closest architectural precedent
19	Srivastava et al. (2022) [21]	CNN with environmental and phenological remote sensing data	R^2 0.82–0.91; phenological features highly informative
20	Jabed & Murad (2024) [11]	Review of 115 ML/DL studies (2018–2023); multi-source integration	Hybrid $R^2 > 0.85$; most comprehensive benchmark for this study

Srivastava et al. [21] developed a CNN-based yield prediction for winter wheat using integrated environmental and phenological data from remote sensing and climate reanalysis, achieving R^2 of 0.82–0.91 across validation regions. Phenological timing features combined with climatic variables were among the most informative predictors.

Jabed & Murad [11] reviewed ML and DL applications in crop yield prediction across 115 studies spanning 2018 to 2023. Their synthesis established that hybrid approaches consistently achieved $R^2 > 0.85$, representing 15–30% error reductions over statistical baselines, and identified multi-source data integration as the single most critical methodological factor in prediction performance.

Previous studies reveal a clear pattern: hybrid AI models that integrate multi-source data outperform single-model, single-source approaches across diverse crop types and geographic locations [5, 19, 11]. Despite this demonstrated superiority, three interconnected gaps remain unaddressed in the existing literature.

First, the systematic application of hybrid frameworks to date palm yield prediction specifically is virtually absent from the scientific literature, as most relevant studies have focused on staple grain crops, resulting in a clear neglect of date palm in the computational modeling literature [19, 6]. Second, the few date palm-specific studies that do exist have been characterized by reliance on a single data source, limited geographic scope, and relatively simple model architectures limitations that preclude adequate representation of the multi-dimensional, time-varying nature of date production systems [6, 2]. Third, the cross-regional generalizability of AI-driven crop yield prediction models in arid and semi-arid environments a prerequisite for any large-scale operational deployment has rarely been subjected to rigorous systematic testing [17, 16].

The current study addresses all three gaps simultaneously, through its exclusive focus on date palm cultivation systems, the integration of four heterogeneous data streams within a unified CNN-LSTM-XGBoost framework, and the validation of model performance across multiple agroecological zones in desert and semi-desert environments. In doing so, this study represents the first comprehensive hybrid AI approach purpose-built for cross-regional date palm yield prediction.

III. Methodology

This study proposes a hybrid artificial intelligence framework designed to predict date palm yield by integrating multi-source agricultural data within a unified deep learning and ensemble learning pipeline. The framework combines Convolutional Neural Networks (CNN) for spatial feature extraction, Long Short-Term Memory (LSTM) networks for temporal sequence modeling, and Extreme Gradient Boosting (XGBoost) as a meta-learner for final yield estimation. The overall design is motivated by the demonstrated limitations of single-model approaches in capturing the multi-dimensional and time-varying nature of date palm production systems [5, 11].

A. Dataset Description

The proposed framework will be evaluated using a multi-source dataset compiled from four complementary data streams. The first stream consists of climatic variables including daily maximum and minimum temperature (°C), total precipitation (mm), relative humidity (%), and solar radiation (MJ/m^2) obtained from regional meteorological stations and ERA5 reanalysis products across selected desert and semi-desert agroecological zones in Saudi Arabia. The second stream comprises soil physicochemical measurements, specifically electrical conductivity (dS/m), pH, volumetric soil moisture content (%), and exchangeable sodium percentage (ESP), which have been identified as primary determinants of date palm productivity under arid conditions [1]. The third stream incorporates satellite-derived Normalized Difference Vegetation Index (NDVI) time series extracted from Sentinel-2 and MODIS imagery at 10-day intervals across each growing season, capturing canopy development dynamics throughout the phenological cycle. The fourth stream consists of historical annual yield records (kg/palm/year) sourced from regional agricultural registries spanning a minimum of ten growing seasons. All four data streams are temporally aligned to the agricultural calendar of the date palm (March–October) and spatially matched to plantation-level geographic coordinates, yielding an integrated multi-source dataset suitable for supervised learning.

B. Data Preprocessing

Prior to model training, all data streams undergo a standardized preprocessing pipeline. Missing values in climatic and soil records are imputed using linear interpolation for short gaps ($\leq$ 3 consecutive observations) and seasonal decomposition-based imputation for longer gaps, following the approach validated by Naghdyzadegan Jahromi et al. [15]. NDVI time series are smoothed using a Savitzky–Golay filter (window size = 7, polynomial order = 2) to reduce noise introduced by cloud contamination and atmospheric variability. All continuous features are normalized to the [0, 1] range using Min-Max scaling, applied independently to training, validation, and test partitions to prevent data leakage. Temporal alignment across data streams is enforced by resampling all inputs to a uniform weekly resolution using forward-filling for instantaneous measurements and cumulative aggregation for flux variables such as precipitation and radiation. A variance inflation factor (VIF) analysis is subsequently applied to identify and remove multicollinear features, retaining only variables with VIF < 5 for inclusion in the final input feature set.

C. Proposed Framework Architecture

The proposed hybrid framework operates as a three-stage pipeline, as illustrated in Fig. 1. In the first stage, a one-dimensional CNN processes spatially structured input features specifically the multi-band NDVI time series and soil property maps through two successive convolutional blocks, each consisting of 64 and 128 filters of kernel size 3, respectively, followed by batch normalization and ReLU activation. Max-pooling layers with a stride of 2 reduce the spatial resolution between blocks, and the output is flattened into a compact feature vector representing learned spatial representations of each input field.

In the second stage, the flattened CNN output is

Ziyad Alhazmi, Muhannad Alahmadi, Sheikh Mohammed, and Mostafa Elgayar, "A Robust Hybrid AI Framework for Predicting Date Palm Yields from Multiple Sources Amidst Environmental Change," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 9, Issue 3, pp. 113-118, August 2026.

concatenated with the sequential climatic feature vector and fed into a two-layer LSTM network comprising 128 units in the first layer and 64 units in the second, with a dropout rate of 0.3 applied between layers to mitigate overfitting. The LSTM layers capture inter-annual variability and within-season temporal dependencies in the climatic inputs patterns that are critical to date palm phenology but cannot be modeled by convolutional operations alone [8, 17]. A temporal attention mechanism is applied to the LSTM output sequence, enabling the model to adaptively assign higher weight to growth-stage intervals that are most predictive of final yield, consistent with the architectural design validated by Liu et al. [12].

In the third stage, the attended output vector from the LSTM and the CNN feature representation are concatenated and used as input to an XGBoost meta-learner, which performs the final regression to yield estimate. XGBoost is selected for this role because of its demonstrated robustness to feature interactions and its capacity to correct residual prediction errors from the deep learning components [2, 5]. The complete CNN and LSTM components are trained end-to-end, while the XGBoost stage is trained on the held-out validation set outputs to avoid target leakage.

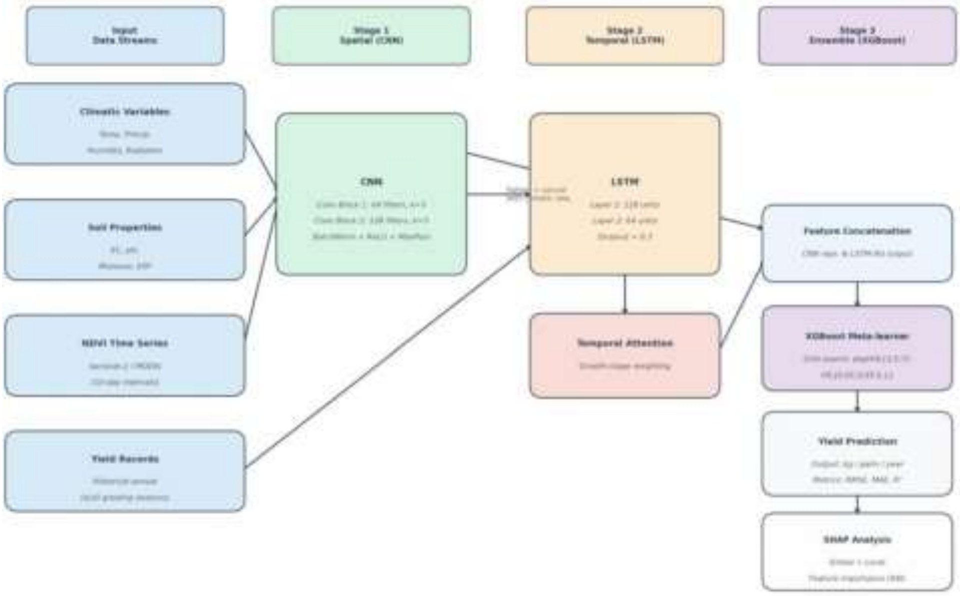

Fig. 1. Architecture of the proposed hybrid CNN-LSTM-XGBoost framework for date palm yield prediction. The pipeline integrates spatial feature extraction (CNN), temporal sequence modeling (LSTM with attention), and ensemble regression (XGBoost) across four heterogeneous input data streams.

D. *Training Protocol*

The integrated dataset is partitioned into training (70%), validation (15%), and test (15%) subsets using a chronological split that preserves temporal ordering, thereby preventing future data from informing past predictions. The CNN and LSTM components are trained jointly using the Adam optimizer with an initial learning rate of 1×10^{-3} and a weight decay of 1×10^{-4}. A cosine annealing learning rate scheduler is applied over a maximum of 200 epochs, with early stopping triggered when validation loss fails to decrease for 20 consecutive epochs. Batch size is set to 32. The XGBoost meta-learner is subsequently trained on the validation-set predictions of the CNN-LSTM sub-network using a randomized grid search over the following hyperparameter space: number of estimators $\in \{100, 200, 300\}$, maximum tree depth $\in \{3, 5, 7\}$, learning rate $\in \{0.01, 0.05, 0.1\}$, and subsampling ratio $\in \{0.7, 0.85, 1.0\}$. Five-fold cross-validation is applied within the training partition to select the optimal XGBoost configuration.

E. *Evaluation Metrics*

Model performance is quantified using three complementary metrics: Root Mean Square Error (RMSE), which penalizes large prediction deviations; Mean Absolute Error (MAE), which provides an interpretable measure of average prediction magnitude; and the coefficient of determination (R^2), which indicates the proportion of yield variance explained by the model. All metrics are computed on the held-out test set and reported separately for each agroecological zone to assess cross-regional generalizability. The proposed framework is additionally benchmarked against four baseline models standalone LSTM, standalone XGBoost, a CNN-LSTM without attention, and the ACVO-ANFIS model of Dehghanisanij et al. [6] to quantify the marginal contribution of each architectural component.

F. *Explainability Analysis*

To ensure that the framework's predictions are interpretable and actionable for agricultural stakeholders, SHAP (SHapley Additive exPlanations) values are computed for the XGBoost meta-learner output across the entire test partition. SHAP provides a theoretically grounded decomposition of each prediction into additive contributions from individual input features, enabling the identification of the most influential climatic, soil, and remote sensing variables for each growing season and geographic zone [9]. Global feature importance rankings are derived by averaging absolute SHAP values across all test samples, while local explanations are used to illustrate

Ziyad Alhazmi, Muhannad Alahmadi, Sheikh Mohammed, and Mostafa Elgayar, “A Robust Hybrid AI Framework for Predicting Date Palm Yields from Multiple Sources Amidst Environmental Change,” *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 9, Issue 3, pp. 113-118, August 2026.

how feature contributions shift across contrasting yield outcomes (high-yield vs. low-yield seasons). This analysis is intended to move the framework beyond black-box prediction toward decision-support, providing farmers and policymakers with evidence-based guidance on the primary drivers of date palm yield variability under changing environmental conditions.

IV. CONCLUSION

This paper presented a hybrid artificial intelligence framework for date palm yield prediction that integrates Convolutional Neural Networks, Long Short-Term Memory networks, and Extreme Gradient Boosting within a unified multi-source pipeline. The proposed architecture addresses three interconnected gaps identified in the existing literature: the near-complete absence of hybrid deep learning frameworks purpose-built for date palm systems, the reliance of prior date palm studies on single data sources and simplified model architectures, and the limited cross-regional validation of AI-driven yield prediction models in arid and semi-arid environments. The framework processes four heterogeneous data streams — climatic variables, soil physicochemical properties, satellite-derived NDVI time series, and historical yield records — through a three-stage pipeline in which CNN extracts spatial features, LSTM with temporal attention captures within-season and inter-annual dynamics, and XGBoost performs the final yield regression as a meta-learner. This design enables the model to exploit cross-modal information that neither purely spatial nor purely temporal architectures can access independently. The integration of SHAP-based explain ability further ensures that predictions are interpretable and directly actionable for farmers, water resource managers, and agricultural policymakers, addressing a critical limitation of black-box approaches documented in the literature [9]. The framework is designed for cross-regional validation across multiple agro ecological zones in desert and semi-desert environments, with performance benchmarked against four baseline models to quantify the marginal contribution of each architectural component. By targeting date palm systems specifically and operating within the memory constraints of standard computing infrastructure, the proposed approach is intended to be both scientifically rigorous and practically deployable in the agricultural contexts where it is most needed. Future work will focus on the collection and curation of the multi-source dataset described in this study, the empirical validation of the proposed architecture, and the extension of the framework to incorporate additional phenological variables and real-time

REFERENCES

[1] Alhammadi, M. S., & Kurup, S. S. (2022). Impact of salinity stress on date palm Phoenix dactylifera L.—A review. Agronomy, 12(10), 2522.
[2] Ang, Y., Shafri, H. Z. M., Lee, Y. P., et al. (2022). Oil palm yield prediction across blocks from multi-source data using machine learning and deep learning. Earth Science Informatics, 15, 2349–2367.
[3] Aviles Toledo, C., Crawford, M. M., & Tuinstra, M. R. (2024). Integrating multi-modal remote sensing, deep learning, and attention mechanisms for yield prediction. Frontiers in Plant Science, 15, 1408047.
[4] Bouras, E. H., Jarlan, L., Er-Raki, S., et al. (2021). Cereal yield forecasting with satellite drought-based indices using machine learning in Morocco. Remote Sensing, 13(16), 3101.
[5] Cao, J., Zhang, Z., Tao, F., et al. (2021). Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches. Agricultural and Forest Meteorology, 297, 108275.
[6] Dehghanisanij, H., Salamati, N., Emami, S., et al. (2023). An intelligent approach to improve date palm crop yield and water productivity under different irrigation and climate scenarios. Applied Water Science, 13, 56.
[7] Dhanaraju, M., Chenniappan, P., Ramalingam, K., et al. (2022). Smart farming: Internet of Things (IoT)-based sustainable agriculture. Agriculture, 12(11), 1745.
[8] FAO. (2023). The state of food and agriculture 2023. Food and Agriculture Organization of the United Nations.
[9] Hu, T., Zhang, X., Bohrer, G., et al. (2023). Crop yield prediction via explainable AI and interpretable machine learning. Agricultural and Forest Meteorology, 336, 109458.
[10] Iqbal, N., Shahzad, M. U., Sherif, E. S. M., et al. (2024). Analysis of wheat-yield prediction using machine learning models under climate change scenarios. Sustainability, 16(16), 6976.
[11] Jabed, M. A., & Murad, M. A. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches. Heliyon, 10(24), e40836.
[12] Liu, B., et al. (2024). Deep learning for multi-source data-driven crop yield prediction in Northeast China. Agriculture, 14(6), 794.
[13] Mohamad Radzi, N. A., et al. (2024). Predicting oil palm yield using a comprehensive agronomy dataset and 17 machine learning and deep learning models. Ecological Informatics, 80, 102595.
[14] Muruganantham, P., Wibowo, S., Grandhi, S., et al. (2022). A systematic literature review on crop yield prediction with deep learning and remote sensing. Remote Sensing, 14(9), 1990.
[15] Naghdyzadegan Jahromi, M., et al. (2023). Developing machine learning models for wheat yield prediction using ground-based data and vegetation indices. European Journal of Agronomy, 146, 126820.
[16] Oikonomidis, A., Catal, C., & Kassahun, A. (2023). Deep learning for crop yield prediction: A systematic literature review. New Zealand Journal of Crop and Horticultural Science, 51(1), 1–26.
[17] Paudel, D., Boogaard, H., de Wit, A., et al. (2021). Machine learning for large-scale crop yield forecasting. Agricultural Systems, 187, 103016.
[18] Rahmanian, S., et al. (2023). Date palm (Phoenix dactylifera) cultivation under climate change: Challenges and adaptation strategies. Sustainability, 15(4), 3549.
[19] Rashid, M., Bari, B. S., Yusup, Y., et al. (2021). A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. IEEE Access, 9, 63406–63439.
[20] Saleem, M. H., et al. (2025). Recent trends in machine learning, deep learning, ensemble learning, and explainable AI for evaluating crop yields under abnormal climate conditions. Applied Sciences, 15(2), 615.
[21] Srivastava, A. K., Safaei, N., Khaki, S., et al. (2022). Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. Scientific Reports, 12, 3231.
[22] Wang, J., Cao, J., Wang, H., et al. (2022). Improving the forecasting of winter wheat yields with machine learning-dynamical hybrid ensemble prediction. Remote Sensing, 14(7), 1707.

Ziyad Alhazmi, Muhannad Alahmadi, Sheikh Mohammed, and Mostafa Elgayar, "A Robust Hybrid AI Framework for Predicting Date Palm Yields from Multiple Sources Amidst Environmental Change," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 9, Issue 3, pp. 113-118, August 2026.