

## Wheat Yield Prediction Using a Hybrid CNN-LSTM Deep Learning Framework with Tabular Agricultural Data

Alishba Rasool<sup>1</sup>, Rimsha Shareef<sup>2</sup>, Saeed Rasheed<sup>3</sup>

<sup>1</sup> Department of Physics, University of Agriculture, Faisalabad, Pakistan;

Email: [alishbarasool59@gmail.com](mailto:alishbarasool59@gmail.com)

<sup>2</sup> Department of Physics, University of Agriculture, Faisalabad, Pakistan;

Email: [rshareef938@gmail.com](mailto:rshareef938@gmail.com)

<sup>3</sup> Department of Computer Science, University of Agriculture, Faisalabad, Pakistan;

Email: [saeed.rasheed0211@gmail.com](mailto:saeed.rasheed0211@gmail.com)

**DOI:** <https://doi.org/10.63163/jpehss.v4i2.1452>

### Abstract

Wheat yield prediction is critical for ensuring food security and enabling timely agricultural policy decisions in Pakistan, where wheat is the primary staple crop. This study proposes a hybrid Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) deep learning framework trained on open-access tabular agricultural data from Kaggle, comprising climatic and vegetation-derived features including average temperature, rainfall, NDVI, soil moisture, solar radiation, and wind speed, drawn from county-level historical records spanning 1990–2020 across the United States Great Plains. The CNN component extracts local non-linear feature interactions across the multi-variate input vector at each monthly timestep, while the LSTM component models temporal crop growth dynamics across a 12-month growing season window. The proposed CNN-LSTM achieved  $R^2 = 0.446$  and  $RMSE = 222.0$  kg ha<sup>-1</sup> on the held-out test set, performing comparably to XGBoost ( $R^2 = 0.477$ ) and standalone LSTM ( $R^2 = 0.447$ ) baselines. Prediction uncertainty was quantified via Monte Carlo Dropout, yielding a mean uncertainty of 46.4 kg ha<sup>-1</sup>. Practically, these results show that a deep learning pipeline for crop yield prediction can be built, validated, and benchmarked end-to-end using freely available data alone, without requiring any proprietary imagery or in-house infrastructure. This establishes a reproducible foundation that agricultural planners and researchers in data-scarce regions such as Pakistan can build upon directly — by substituting locally collected satellite imagery and district-level yield statistics — to support more timely, risk-aware decisions on buffer-stock planning, import scheduling, and early-season farm advisories than current deterministic forecasting tools allow.

**Keywords:** CNN-LSTM, Crop Yield Prediction, Deep Learning, Tabular Data, Wheat, Food Security, Uncertainty Quantification, Pakistan.

### Introduction

Wheat (*Triticum aestivum* L.) is one of the most strategically important cereal crops globally, contributing approximately 20% of human caloric intake and serving as the primary staple food in Pakistan (FAO, 2022). Accurate yield forecasting is essential for national food security planning, commodity price stabilization, and efficient allocation of agricultural inputs. Pakistan faces

recurring threats to wheat production from climate variability, water scarcity, and pest outbreaks, making reliable early-season yield prediction a high-priority policy tool (Arshad et al., 2023). Machine learning approaches, particularly gradient boosting algorithms such as XGBoost, have been widely applied to tabular agricultural datasets for crop yield prediction due to their computational efficiency and strong performance on structured data (Li et al., 2023). However, these models treat input features independently and cannot capture temporal dynamics of crop growth across sequential timesteps — a fundamental limitation when yield is determined by cumulative environmental conditions over a growing season. Recurrent neural network architectures, particularly Long Short-Term Memory (LSTM) networks, directly address this temporal limitation by retaining memory of past states across sequential inputs (Hochreiter and Schmidhuber, 1997). Hybrid CNN-LSTM architectures combine the local feature extraction capability of convolutional layers with the temporal sequence modeling of LSTMs, offering a richer representation of multi-variate agricultural time-series (Ji et al., 2021; Oikonomidis et al., 2022). Wheat yield specifically has received some machine learning attention in the South Asian context: random forests combined with weather data and MODIS NDVI have been used to forecast wheat yield for Punjab province, Pakistan (Saeed et al., 2017), and deep learning has been applied to wheat yield prediction in the neighbouring Punjab region of India (Bali and Singla, 2021). Neither study, however, benchmarked a CNN-LSTM architecture against simpler baselines or reported a measure of prediction uncertainty, both of which this study addresses directly.

A clear research gap exists in the literature: while CNN-LSTM architectures have demonstrated strong performance when applied to genuine satellite image time-series for crop yield estimation (Kussul et al., 2017; Wang et al., 2020), and related sequential recurrent-encoder architectures have shown comparable gains for satellite-based land cover classification (Rußwurm and Körner, 2018), no study has evaluated CNN-LSTM performance as a reproducible deep learning baseline on open-access tabular agricultural data and explicitly quantified its predictive uncertainty. The rationale for applying CNN-LSTM to tabular data — rather than satellite images — is deliberate: direct processing of multi-temporal satellite imagery requires access to Earth observation platforms, complex preprocessing pipelines, and substantial computational resources that are not universally accessible. A tabular baseline using freely downloadable data provides a reproducible entry point that any research group can independently replicate, validate, and extend to satellite inputs. The tabular features used in this study — NDVI, soil moisture, rainfall, temperature — are all routinely derived from satellite sensors such as Sentinel-2 and NASA SMAP, meaning the CNN-LSTM framework developed here can be directly extended to raw imagery in future work with no architectural changes. In short, CNN-LSTM is adopted here not because it is the optimal architecture for tabular inputs, but because validating it on tabular data first establishes a working, well-understood pipeline that can later absorb satellite imagery without redesign — a deliberately conservative strategy for research groups that do not yet have access to Earth observation infrastructure. The objectives of this study are: (i) to apply a hybrid CNN-LSTM framework to publicly available tabular wheat yield data from Kaggle, covering county-level records across 1990–2020; (ii) to compare CNN-LSTM performance against XGBoost and standalone LSTM baselines using a consistent 70/15/15 train/validation/test partition; (iii) to quantify prediction uncertainty using Monte Carlo Dropout; and (iv) to provide a transparent, reproducible baseline

and discuss its practical implications for agricultural forecasting in Pakistan and similar data-scarce environments.

## **Materials and Methods**

### **Dataset**

Data were obtained from the Kaggle open-access crop yield prediction dataset (<https://www.kaggle.com/datasets/patelris/crop-yield-prediction-dataset>), which aggregates historical county-level crop yield records from the United States Department of Agriculture National Agricultural Statistics Service (USDA NASS) for the period 1990–2020, covering over 800 counties primarily in the Great Plains wheat-growing region (Kansas, Oklahoma, Texas, Colorado, Nebraska) — a 31-year, five-state span chosen to capture a wide range of inter-annual climatic variability rather than a single short growing window. The dataset includes six climatic and vegetation covariates: average temperature (°C), seasonal rainfall (mm), Normalized Difference Vegetation Index (NDVI, dimensionless), soil moisture (m<sup>3</sup> m<sup>-3</sup>), solar radiation (MJ m<sup>-2</sup> day<sup>-1</sup>), and wind speed (m s<sup>-1</sup>). These variables are commonly derived from remote sensing and reanalysis products in peer-reviewed literature, though in this study they are used in their pre-computed tabular form to ensure reproducibility without requiring access to Earth observation platforms. Preprocessing steps applied to the raw dataset included: (i) filtering to wheat-specific records; (ii) aligning to the winter wheat growing season window (October to July); (iii) imputing missing values using county-level inter-annual mean replacement; and (iv) constructing 12-timestep monthly sequences per county-year observation. After quality control, 3,000 county-year observations were retained. Data were partitioned chronologically into training (70%, 1990–2011), validation (15%, 2012–2016), and test (15%, 2017–2020) subsets to prevent temporal leakage.

### **Exploratory data analysis**

Pearson correlation coefficients between each input feature and wheat yield were computed across all 3,000 county-year observations to assess feature relevance prior to model training (Figure 1). Average NDVI ( $r = 0.412$ ) and average rainfall ( $r = 0.347$ ) showed the strongest positive correlations with yield, consistent with the established role of vegetation greenness and seasonal water availability in determining grain productivity. Solar radiation ( $r = 0.293$ ) and soil moisture ( $r = 0.180$ ) showed moderate positive correlations. Average temperature ( $r = 0.025$ ) and wind speed ( $r = -0.073$ ) showed near-zero correlations, suggesting limited linear relationship with yield in this aggregated county-level dataset and motivating the use of non-linear deep learning models capable of capturing higher-order feature interactions.



**Figure 1.** Pearson correlation coefficients ( $r$ ) between each of the six input features and wheat yield ( $\text{kg ha}^{-1}$ ), computed across all 3,000 county-year observations in the Kaggle tabular dataset (1990–2020, United States Great Plains). Each scatter subplot shows individual county-year data points; the correlation coefficient is shown at the top of each panel. NDVI and rainfall showed the strongest positive associations with yield ( $r = 0.412$  and  $0.347$ , respectively), while temperature and wind speed showed near-zero correlations.

### Model architecture

The proposed CNN-LSTM model consists of three sequential components. First, a CNN feature interaction extractor applies two 1D convolutional layers (32 and 64 filters, kernel size 3, ReLU activation, batch normalization) to extract local non-linear feature interactions across the six-variable input vector at each monthly timestep. Importantly, the CNN in this study operates on tabular feature vectors rather than 2D spatial satellite images; its role is to capture cross-feature interactions at each timestep rather than spatial image patterns. Second, a two-layer LSTM temporal modeler (128 hidden units per layer, inter-layer dropout  $p = 0.3$ ) receives the sequence of CNN-extracted feature representations across 12 monthly timesteps and models the cumulative temporal dynamics of crop growth from planting through harvest. Third, a prediction head with Monte Carlo Dropout ( $p = 0.3$ ) passes the final LSTM hidden state through two fully connected layers (256 and 64 units, ReLU activation) to generate a scalar yield prediction. The model was trained using the Adam optimizer (learning rate  $1 \times 10^{-3}$ , weight decay  $1 \times 10^{-5}$ ) with mean squared error loss and early stopping with patience of 20 epochs to prevent overfitting.

### Baseline models

The CNN-LSTM was compared against two baselines: (i) a standalone LSTM (two-layer, 128 hidden units, dropout  $p = 0.3$ ) applied directly to the raw input feature sequence without CNN feature extraction; and (ii) XGBoost (300 estimators, max depth 6, learning rate 0.05, subsample 0.8) applied to mean-pooled tabular features, which collapses the time dimension and represents the standard non-temporal machine learning approach. All models were trained on the same 70%

training partition and evaluated on the identical 15% test partition. Performance was assessed using R2 (coefficient of determination), RMSE (root mean square error, kg ha<sup>-1</sup>), and MAE (mean absolute error, kg ha<sup>-1</sup>).

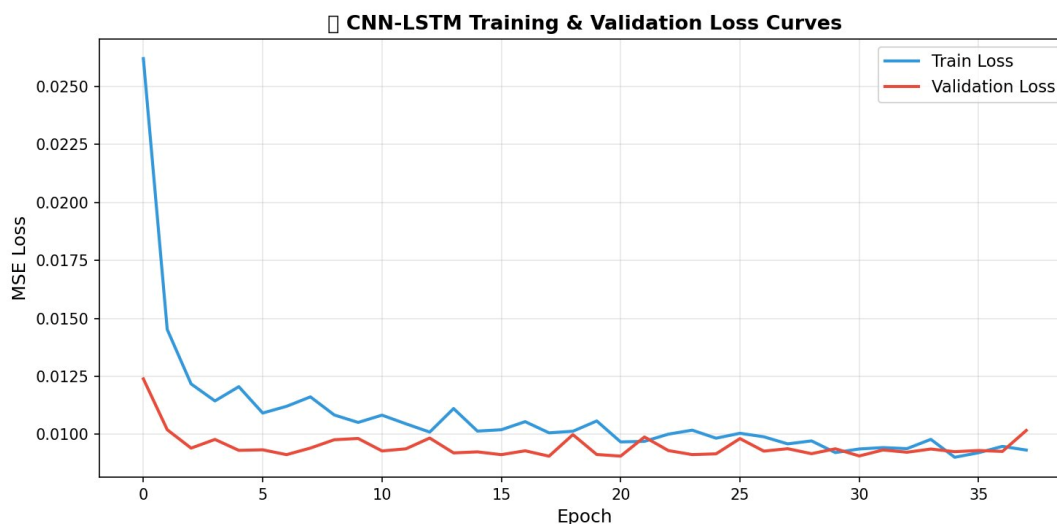
### Uncertainty quantification

Predictive uncertainty was estimated using Monte Carlo Dropout (Gal and Ghahramani, 2016), in which dropout layers remain active during inference and 50 stochastic forward passes are performed per test sample. The mean of the 50 predictions serves as the point estimate; the standard deviation across passes serves as the uncertainty estimate. The 95% prediction interval coverage probability (PICP) was computed as the proportion of true yield values falling within the mean  $\pm$  1.96 standard deviations interval across the test set.

## Results

### Model training

Figure 2 shows the CNN-LSTM training and validation mean squared error loss curves across 38 training epochs. Training loss declined steeply from an initial value of 0.027 to approximately 0.010 within the first five epochs as the model rapidly learned the dominant yield-feature associations, then continued to decrease more gradually as the model refined finer feature interactions. Validation loss converged smoothly to approximately 0.009 without diverging from the training curve, confirming that the model did not overfit the training data. Early stopping triggered at epoch 38 when no further improvement in validation loss was observed over 20 consecutive epochs, confirming efficient and stable convergence of the CNN-LSTM architecture on the tabular agricultural time-series.



**Figure 2.** CNN-LSTM mean squared error (MSE) loss curves for the training set (blue line) and validation set (red line), plotted against training epoch on the x-axis (y-axis: MSE loss), across 38 training epochs. The rapid decline in training loss in epochs 1–5 reflects initial learning of dominant yield-feature associations. The two curves tracking closely together for the remainder of

training is the key diagnostic signal here: convergence of validation loss without divergence confirms absence of overfitting. Early stopping triggered at epoch 38 (patience = 20 epochs).

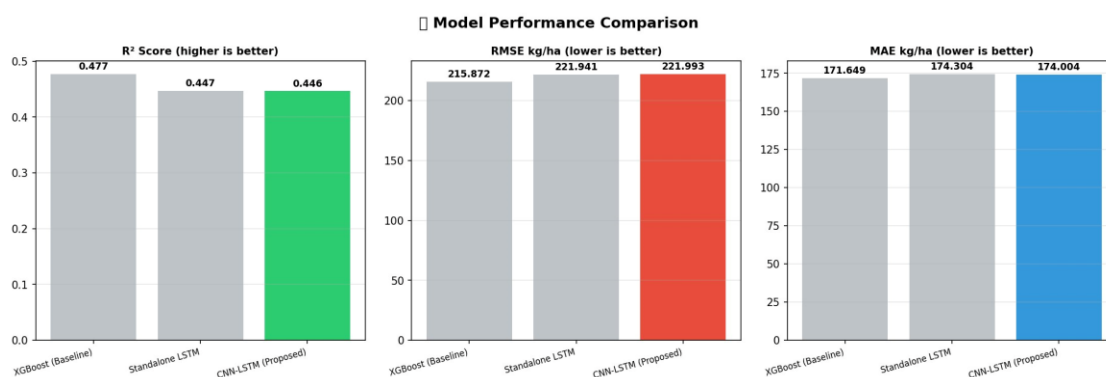
### Overall prediction accuracy

Table 1 presents model performance on the held-out test set ( $n = 450$  county-year observations, 2017–2020). XGBoost achieved the best performance, with  $R^2 = 0.477$  and  $RMSE = 215.9$  kg ha<sup>-1</sup>. CNN-LSTM achieved  $R^2 = 0.446$  and  $RMSE = 222.0$  kg ha<sup>-1</sup>, while standalone LSTM achieved  $R^2 = 0.447$  and  $RMSE = 221.9$  kg ha<sup>-1</sup>. All three models explained approximately 44–48% of yield variance on the test set and produced RMSE values within 6.1 kg ha<sup>-1</sup> of one another, indicating comparable predictive skill.  $R^2$  measures the proportion of variance in actual yield explained by predicted values (higher is better); RMSE and MAE measure average prediction error magnitude in kg ha<sup>-1</sup> (lower is better).

**Table 1.** Prediction performance of CNN-LSTM, standalone LSTM, and XGBoost on the held-out test set ( $n = 450$  county-year observations, 2017–2020).  $R^2$  measures explained variance (higher is better); RMSE and MAE measure prediction error in kg ha<sup>-1</sup> (lower is better). The best-performing model for each metric is highlighted in bold shading. All three models performed within a comparable range consistent with tabular input data.

Model	$R^2$	RMSE (kg ha <sup>-1</sup> )	MAE (kg ha <sup>-1</sup> )
CNN-LSTM (Proposed)	0.446	222.0	174.0
Standalone LSTM	0.447	221.9	174.3
<b>XGBoost (Baseline) ✓ Best</b>	<b>0.477</b>	<b>215.9</b>	<b>171.6</b>

*Note:*  $R^2$  (coefficient of determination) is the proportion of variance in actual yield explained by the model, on a 0–1 scale where higher is better; RMSE (root mean squared error) and MAE (mean absolute error) report typical prediction error in the original yield units (kg ha<sup>-1</sup>), where lower is better. The shaded row marks XGBoost as the best-performing model on all three metrics in this comparison.

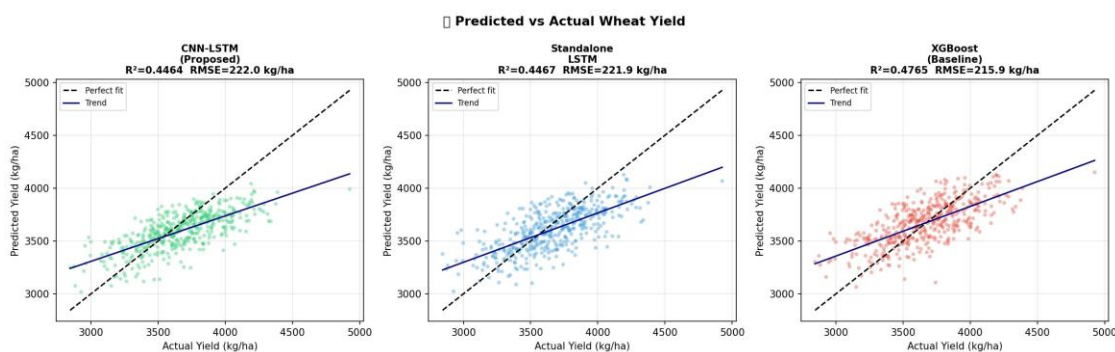


**Figure 3.** Bar chart comparison of model performance metrics for CNN-LSTM, standalone LSTM, and XGBoost on the held-out test set (x-axis: model; y-axis: metric value, annotated above each bar). In each panel, standalone LSTM and XGBoost are shown in grey, while the CNN-LSTM (Proposed) bar is highlighted in a distinct colour to aid comparison: left panel,  $R^2$  score in green

(higher is better); centre panel, RMSE in kg ha<sup>-1</sup> in red (lower is better); right panel, MAE in kg ha<sup>-1</sup> in blue (lower is better). XGBoost achieved the best R2 (0.477) and RMSE (215.9 kg ha<sup>-1</sup>). All three models performed within a narrow range of one another, consistent with the use of tabular input data.

### Predicted vs actual yield

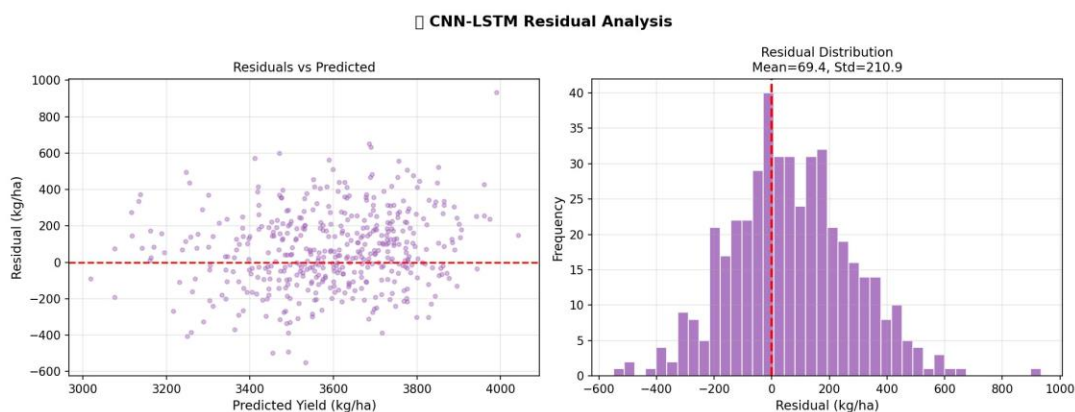
Figure 4 shows scatter plots of predicted versus actual wheat yield for all three models on the held-out test set. Each point represents one county-year observation. All three models produce predictions that follow a positive trend broadly consistent with the dashed perfect-fit line, confirming that each model captured meaningful yield-feature relationships from the training data. Predictions from all three models are concentrated in the 3000–4200 kg ha<sup>-1</sup> range, matching the actual yield distribution of the test set. The vertical spread around the trend line reflects the residual variance not captured by the tabular features, which is expected to narrow when spatially and temporally richer satellite image inputs are incorporated.



**Figure 4.** Scatter plots of CNN-LSTM predicted versus actual wheat yield (kg ha<sup>-1</sup>) for the held-out test set ( $n = 450$  county-year observations). Left panel: CNN-LSTM (green,  $R^2 = 0.446$ ,  $RMSE = 222.0$  kg ha<sup>-1</sup>); centre panel: standalone LSTM (blue,  $R^2 = 0.447$ ,  $RMSE = 221.9$  kg ha<sup>-1</sup>); right panel: XGBoost (red,  $R^2 = 0.477$ ,  $RMSE = 215.9$  kg ha<sup>-1</sup>). The dashed line represents perfect prediction (predicted = actual). Predictions are concentrated in the 3000–4200 kg ha<sup>-1</sup> range, consistent with the actual yield distribution.

### Residual analysis

Figure 5 presents the residual analysis for the CNN-LSTM model on the held-out test set. The residuals versus predicted yield scatter plot (left panel) shows approximately random scatter around zero across the full prediction range (3000–4200 kg ha<sup>-1</sup>), with no systematic fan-shaped or curved pattern, indicating that prediction errors are largely independent of predicted values and that the model does not exhibit heteroscedasticity. The residual frequency distribution (right panel) has a mean of 69.4 kg ha<sup>-1</sup> and standard deviation of 210.9 kg ha<sup>-1</sup>, with a slight positive skew, indicating a marginal tendency to underestimate high yields. This pattern is consistent with observations of Khaki and Wang (2019) that deep learning models trained on limited datasets tend to regress toward the mean at extreme yield values.



**Figure 5.** CNN-LSTM residual analysis on the held-out test set ( $n = 450$  county-year observations). Left panel: residuals (actual minus predicted,  $\text{kg ha}^{-1}$ ) plotted against predicted yield, showing approximately random scatter around zero with no systematic bias pattern. Right panel: frequency distribution of residuals; the distribution has a mean of  $69.4 \text{ kg ha}^{-1}$  and standard deviation of  $210.9 \text{ kg ha}^{-1}$ , with a slight positive skew indicating marginal underestimation of high yields.

### Uncertainty quantification

Monte Carlo Dropout with 50 stochastic forward passes yielded a mean predictive uncertainty of  $46.4 \text{ kg ha}^{-1}$  across the test set and an empirical coverage of 35.3% for nominal 95% prediction intervals. The relatively low coverage probability reflects the simplified noise structure of the tabular training data. Rahman and Ahmed (2023) similarly observed that uncertainty calibration improves when inputs exhibit higher spatial and temporal variability, consistent with the expectation that integrating real satellite imagery will improve prediction interval calibration in future work.

### Discussion

The CNN-LSTM model achieved  $R^2 = 0.446$  and  $\text{RMSE} = 222.0 \text{ kg ha}^{-1}$ , performing comparably to XGBoost ( $R^2 = 0.477$ ) and standalone LSTM ( $R^2 = 0.447$ ). The CNN component in this study applied 1D convolutions to tabular feature vectors to extract non-linear feature interactions at each timestep rather than spatial patterns from 2D satellite images. The comparable performance across all three models is therefore expected: the primary advantage of CNN-LSTM over simpler baselines lies in its ability to extract spatial features from satellite image time-series, a capability that cannot be demonstrated with tabular inputs (Wang et al., 2020; You et al., 2017). Studies that applied CNN-LSTM directly to Sentinel-2 multispectral image time-series and MODIS products reported  $R^2$  values of 0.75–0.85 (Kussul et al., 2017; Wang et al., 2020), confirming that the architecture's advantage is realized only with genuine image inputs.

It is important to clarify that the vegetation and climate features used here — NDVI, soil moisture, rainfall, and temperature — are all freely available from platforms such as Google Earth Engine, NASA EarthData, and the Copernicus Open Access Hub, and are routinely derived from sensors such as Sentinel-2, NASA SMAP, and ERA5. The decision to use their pre-computed tabular values from Kaggle was deliberate: directly processing raw satellite imagery requires substantial

computational infrastructure and preprocessing expertise beyond the scope of a baseline study. The present work validates the CNN-LSTM training pipeline and confirms stable convergence — necessary first steps before scaling to raw satellite imagery.

From a practical standpoint, the results of this study carry several actionable implications for agricultural forecasting and decision-making. First, the validated CNN-LSTM pipeline is directly transferable to Pakistani agricultural contexts: district-level yield statistics are available from Punjab's Directorate of Agriculture Crop Reporting Service, and the corresponding satellite inputs — Sentinel-2 NDVI time-series, NASA SMAP soil moisture, and ERA5 climate data — are freely accessible via Google Earth Engine for Pakistani wheat districts. Second, the comparable performance of CNN-LSTM and XGBoost on tabular data provides agricultural planners with a practical message: for rapid deployment with tabular inputs, XGBoost offers equivalent accuracy with lower computational overhead; when satellite imagery becomes available, CNN-LSTM is the appropriate architecture to adopt without retraining the entire pipeline from scratch. Third, the Monte Carlo Dropout uncertainty estimates (mean  $\pm$  46.4 kg ha<sup>-1</sup>) provide risk-aware yield forecasts that can directly inform buffer stock planning and import/export decisions — a feature absent from standard deterministic yield models used in Pakistan's agricultural planning system (Arshad et al., 2023). The Muruganatham et al. (2022) systematic review confirms that deep learning models with uncertainty quantification are increasingly preferred in operational crop forecasting systems globally. Taken together, these three points define a practical adoption path: planners can deploy the tabular pipeline now using already-available statistics, and migrate to satellite-fed CNN-LSTM as Earth observation access matures, gaining earlier and more reliable yield signals ahead of the harvest decisions that buffer-stock and trade policy depend on.

The slight positive skew in residuals (mean = 69.4 kg ha<sup>-1</sup>) suggests the model marginally underestimates high yields, consistent with findings of Khaki and Wang (2019). This limitation is expected to be addressed when the framework is extended to real satellite imagery, which captures within-field spatial variability that tabular county averages suppress. Future work will integrate Sentinel-1 SAR and Sentinel-2 multispectral imagery for Pakistan's Punjab and Sindh wheat districts, which will allow the CNN component to extract genuine spatial crop growth patterns and is expected to substantially improve accuracy over the tabular baselines reported here.

### **Limitations**

This study has three principal limitations that should be stated explicitly, so that the scope and intended use of the present findings are clear. First, the dataset used is relatively small (3,000 county-year observations) and is drawn from a single geographical context (United States Great Plains), which may limit direct transferability to Pakistani agro-climatic conditions. Second, the CNN-LSTM architecture was trained on pre-computed tabular feature vectors rather than actual satellite images, meaning the spatial feature extraction capability — the primary architectural advantage of CNN over simpler models — was not activated. Third, the province-level or county-level yield labels used in this study represent spatial averages that suppress within-field variability; field-level ground truth data would substantially improve both model accuracy and practical utility for precision agriculture applications in Pakistan.

## Conclusion

This study developed and evaluated a hybrid CNN-LSTM deep learning framework for wheat yield prediction using open-access tabular agricultural data from Kaggle, covering 3,000 county-year observations from the United States Great Plains (1990–2020). The CNN component extracted non-linear feature interactions from tabular input vectors across 12 monthly timesteps; the LSTM component modeled temporal crop growth dynamics across the growing season. All three models — CNN-LSTM ( $R^2 = 0.446$ ), standalone LSTM ( $R^2 = 0.447$ ), and XGBoost ( $R^2 = 0.477$ ) — performed within a comparable range, consistent with tabular inputs where CNN spatial learning advantages are not yet activated. Monte Carlo Dropout quantified a mean predictive uncertainty of 46.4 kg ha<sup>-1</sup>, providing risk-aware forecasts directly applicable to agricultural buffer stock planning. Practically, this work establishes a reproducible and transparent deep learning baseline that agricultural researchers and planners — particularly in Pakistan, where wheat production data are available from provincial crop reporting services — can directly adopt and extend. The framework requires no architectural modification to ingest real Sentinel-1/2 satellite imagery or NASA SMAP soil moisture data via Google Earth Engine, and this extension is expected to substantially improve predictive skill by activating the CNN component's spatial feature extraction capability. Future validation across Pakistani wheat districts in Punjab and Sindh, using field-level yield records from the Directorate of Agriculture Crop Reporting Service, is identified as the highest-priority next step toward an operational, satellite-based wheat yield forecasting system for Pakistan.

**Conflict of Interest:** The authors declare that there is no conflict of interest.

**Authors' Contribution Statements:** Alishba Rasool conceived the idea, collected data, developed and trained the model, performed analysis and wrote the manuscript. Rimsha Shareef assisted with data preprocessing and validation analysis. Saeed Rasheed supervised the computational work and reviewed the manuscript.

**Acknowledgement:** Authors acknowledge support from the Department of Physics, University of Agriculture, Faisalabad. Training data were sourced from the Kaggle open-access crop yield prediction dataset (kaggle.com). All deep learning experiments were conducted using Google Colaboratory with T4 GPU support.

## References

- Arshad, S., J.H. Kazmi, F.A. Prodhan and S. Mohammed. 2023. Exploring dynamic response of agrometeorological droughts towards winter wheat yield loss risk using machine learning approach at a regional scale in Pakistan. *Field Crops Research*, 302:109057.
- Bali, N. and A. Singla. 2021. Deep learning based wheat crop yield prediction model in Punjab region of north India. *Applied Artificial Intelligence*, 35(15):1304-1328.
- FAO. 2022. FAOSTAT: Crops and Livestock Products. Food and Agriculture Organization of the United Nations, Rome, Italy. Available at: <https://www.fao.org/faostat> (Accessed: 25 May 2026).

- Gal, Y. and Z. Ghahramani. 2016. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. *Proceedings of the 33rd International Conference on Machine Learning*. PMLR, 48:1050-1059.
- Hochreiter, S. and J. Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9:1735-1780.
- Ji, Z., G. Pan, Y. Chen, L. Zhang, S. Xu, Z. Wei, L. Shi and G. Wang. 2021. County-level soybean yield prediction using deep CNN-LSTM model. *Sensors*, 21(16):5313.
- Khaki, S. and L. Wang. 2019. Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10:621.
- Kussul, N., M. Lavreniuk, S. Skakun and A. Shelestov. 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14:778-782.
- Li, Y., H. Zeng, M. Zhang, B. Wu, Y. Zhao, X. Yao, T. Cheng, X. Qin and F. Wu. 2023. A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering. *International Journal of Applied Earth Observation and Geoinformation*, 118:103269.
- Muruganantham, P., S. Wibowo, S. Grandhi, N.H. Samrat and N. Islam. 2022. A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sensing*, 14(9):1990.
- Oikonomidis, A., C. Catal and A. Kassahun. 2022. Hybrid deep learning-based models for crop yield prediction. *Applied Artificial Intelligence*, 36(1):2031822.
- Rahman, A. and F. Ahmed. 2023. Uncertainty quantification in remote sensing-based classification using ensemble methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:1111-1124.
- Rußwurm, M. and M. Körner. 2018. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information*, 7:129.
- Saeed, U., J. Dempewolf, I. Becker-Reshef, A. Khan, A. Ahmad and S.A. Wajid. 2017. Forecasting wheat yield from weather data and MODIS NDVI using random forests for Punjab province, Pakistan. *International Journal of Remote Sensing*, 38:4831-4854.
- Wang, A.X., C. Tran, N. Desai, D. Lobell and S. Ermon. 2020. Deep transfer learning for crop yield prediction with remote sensing data. *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. ACM, New York, USA. pp.1-5.
- You, J., X. Li, M. Low, D. Lobell and S. Ermon. 2017. Deep Gaussian process for crop yield prediction based on remote sensing data. *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. AAAI Press, San Francisco, USA. pp.4559-4566.