

Crop Yield Forecasting using Hybrid CNN-LSTM Neural Networks

Soumya Rai

Research Scholar, M. Tech CSE, Shri Ramswaroop Memorial University, Lucknow, UP, India

Neeraj Kumar

Assistant Professor, Shri Ramswaroop Memorial University, Lucknow, UP, India

Abstract - Crop yield forecasting helps in enhancing agricultural planning and providing food security. In this paper an accurate agricultural yield forecasting is produced by using hybrid CNN-LSTM neural network model. The model is analysing data that are based on time, like weather, using Long Short-Term Memory (LSTM) and takes out features from satellite images using Convolutional Neural Networks (CNN). The combination of CNN-LSTM is making it easier to understand the spatial and temporal factors that affects crop development.

The study uses number of data sources, like weather data, satellite images, and previous crop yield records. The outcome tells us that the hybrid model performs better than individual and traditional machine learning models as it provides more accurate and trustworthy predictions. While studying this model there were other challenges that were noticed, like poor data quality and requirements of high processes. After considering all the things, the method holds significant potential for improving crop output forecasts and it also supports effective agricultural decision-making.

Keywords: Crop yields forecasting, CNN-LSTM hybrid models, deep learning, agriculture, neural networks, Prediction based on time series, modern farming, and food security

1. INTRODUCTION

Agriculture is very much important for the economy and supply of food in many countries. Farmers depend on agricultural production for their income, and people who are not farmer, they depend on it for daily nutrition. There are number of factors that affect crop production, like temperature, rainfall, soil quality, weather, and farming practices. The above factors are considered to be a fluctuating condition that makes it difficult for farmers to predict how much crop will be produced in a given season.

In past few years, Agriculture has started getting help from technology in a variety of ways. One of the most important discoveries is the use of data and machine learning techniques to predict agricultural production. The model provides correct prediction which can help farmers make better decisions about planting, irrigation, and harvesting. It can also provide great help to governments in planning the food supply and managing resources efficiently. Whereas, traditional prediction methods had overlooked many complex issue in agricultural data.

Promising results have been shown by deep learning models in handling such complex problems. Whereas Convolutional Neural Networks (CNN) is considered to be very good in extracting area details from data, like satellite photographs, Long Short-Term Memory (LSTM) networks are useful for understanding patterns that are based on time, like variations of weather over time. Both the techniques offer benefits, but they also have disadvantages, if used alone.

So, in order to overcome such limitations, researchers are now focusing on using models that can combine the benefits of both CNN and LSTM. Since a temporal and spatial data can be integrated by hybrid CNN-LSTM model, it is more suitable to use it for agricultural yield forecasting. This method can provide predictions that are more accurate than traditional and single-model methods.

The aim of this research is to develop a hybrid CNN-LSTM neural network crop yield forecasting model. The objective of this research is to provide accurate prediction by using both temporal and spatial data. The study also explains that how this method can improve production of agricultural, planning for farmers and also provides effective decision-making opportunity.

2. LITERATURE REVIEW

The last ten years of research on deep learning-based agricultural yield prediction—specifically, hybrid CNN–LSTM models—are covered in this review. We explain how CNNs and LSTMs work as well as the importance of yield projections. The model properties, inputs, training, accuracy (using metrics like RMSE, MAE, and R2), and data sources (satellite pictures, weather, and soil) are compared in a table that summarizes the main study. Overall, hybrid CNN–LSTM models often improve accuracy by

capturing both temporal (time) and spatial (image) features. Difficulties (data gaps, limited samples, model "black box") and common data processes (NDVI/EVI computation, smoothing, masking) are discussed. We identify issues (including interpretability and real-time issues) and suggest future directions, such as explainable AI, transfer learning, and the utilization of more data sources [1][2] [3][4].

2.1 Background: Why Forecast Crop Yields

Crop yields tell how much food a farm or region can produce. This helps the government in providing food supply, managing crop pricing and handling disaster relief by accurate prediction of crop production. It also helps farmers in planting, irrigation, and harvesting plans. It is very important as weather, soil, and farming practices can all change significantly from year to year. For example, a little rainfall can lower output. The complex data patterns are often overlooked in Traditional methods, like statistics or simple models. Now, researchers have started using machine learning and deep learning to increase their prediction accuracy [1][3].

Remote sensing, or satellite imagery, is very useful as it collects the information about the earth from a distance. Over large areas, satellites can track moisture, vegetation, and climate. For example, scientists use special indicators like NDVI (Normalized Difference Vegetation Index) or EVI (Enhanced Vegetation Index) as it helps measure how green and healthy plants are. Because these indicators change over the course of the growing season, they can be used to predict the final yield. These are combined with soil and weather data (rain, temperature) to create a complete picture. Prediction is challenging, though, because numerous data types must be aligned in time and place, and data may be noisy (cloud cover, missing values) [1][3][4].

2.2 CNN, LSTM, and Hybrid CNN–LSTM

- **Convolutional Neural Networks (CNN):** Similar to the human eye, CNNs are a kind of deep learning model that excels in extracting spatial features from images. They scan an image using filters to discover elements at various scales, such as edges, textures, and shapes. CNNs may use satellite data (such as vegetation maps and multi-band pixels) in crop forecasting to determine which visual patterns—such as plant density and green areas—are associated with yield [1].
- **Long Short-Term Memory (LSTM):** A special kind of recurrent neural network that can process data sequences across time is the long short-term memory (LSTM). The LSTM have a memory that can learn how past weather or soil conditions affect future outcomes. Because of this, LSTMs are useful for spotting temporal patterns or trends (such an extended dry spell or a string of hot days) [1].
- **Why Hybrid?** None of the model was able to capture all the pertinent data by its own. CNNs just examines each time snapshot and ignores how the conditions changes. Despite their emphasis on time, LSTMs usually have trouble with spatial data, including images. The CNN component extracts feature from images (spatial patterns in fields) and the LSTM component learns from the temporal evolution of these features across time in a hybrid CNN–LSTM model. In practice, research often feeds sequences of satellite-based pictures (like vegetation or soil moisture maps spanning weeks or months) through CNN layers before piping the features into LSTM layers. In this way, the model can learn "space-time" correlations that improve yield estimations. Numerous studies have shown that CNN-LSTM hybrids perform better on agricultural data than CNN-only or LSTM-only models [1].

2.3 Review of Recent Studies

Study (Year)	Crop / Region	Data Sources	Model (Architecture)	Input Features	Training / Eval	Metrics & Results
Sun et al. (2019)[1]	Soybean (US counties)	MODIS satellite (surface reflectance bands 1–7, LST); Daymet weather; USDA yield labels	CNN–LSTM (2D CNN + LSTM)	Multi-spectral reflectance (NDVI etc.), LST, weather series; masked to soybean areas[1]	Google Earth Engine framework; 2003–2015 data; 5-fold/annual validation	End-of-season RMSE \approx 329 kg/ha (CNN–LSTM) vs 359 kg/ha (CNN alone)[1]; $R^2 \approx$ 0.69–0.81 across years[1] (CNN–LSTM outperformed CNN/LSTM)
Taremw a et al.	Maize (Uganda)	Climate (rainfall, temperature);	CNN–LSTM	Meteorologic al variables + satellite-	2018–2020 data; SMOGN regression	RMSE = 0.327 t/ha (327 kg/ha), MAE = 0.267 t; R^2

Study (Year)	Crop / Region	Data Sources	Model (Architecture)	Input Features	Training / Eval	Metrics & Results
(2026)[2]		MODIS vegetation index (e.g. GPP, NDVI); ZARDI yield records		derived vegetation indices (biannual)	oversampling; feature selection; hyperparameter tuning	$R^2 = 0.783$ (CNN-LSTM). This beat CNN+RF ensemble ($R^2=0.722$, RMSE=0.370 t) and far outperformed CNN or RF alone[2].
Song et al. (2025)[5]	Wheat (China)	Climate (temperature, precipitation); Socio-economic data (machinery power, output value, land area, disasters)	Parallel CNN + LSTM + Attention (TPCLA) with transfer learning	Multivariate inputs: both direct (weather) and indirect (economics) factors	1993–2024 data; pre-train on similar regions, fine-tune on Shandong data; PSO for hyperparameters	$R^2 = 0.904$, 18.4% lower RMSE and 12.6% lower MAE than baseline LSTM+Attention[5]. Outperformed other DL models (RNN, LSTM-Attn). (Improvement credited to transfer learning from related regions.)
Zhang et al. (2024)[4]	Winter wheat (China)	Remote sensing: SIF (sun-induced fluorescence), LAI, EVI (MODIS); Climate (ERA5 temp, precip, etc)	BO-CNN-BiLSTM (Bayes-optimized CNN + bidirectional LSTM)	County-level 8-day composites of SIF, EVI, LAI and climate variables[4]	2011–2020 (Henan Province); Bayesian optimization for hyperparameters	$R^2 = 0.81$, RMSE = 617.0 kg/ha (best with SIF+EVI+climate). BCBL model outperformed RF, XGBoost, and LSTM (single) in all cases. Achieved stable estimates ~25 days before harvest. [4]

These examples show how spatiotemporal data-based hybrid models generally improve accuracy. For instance, Sun et al. found that CNN-LSTM decreased RMSE by about 8–9% when compared to CNN or LSTM alone. Similarly, Zhang et al. found that their CNN-BiLSTM model performed better in terms of R^2 and RMSE than simpler models. Taremwa et al. found an increase ($R^2 0.783$) when CNN layers were introduced to capture spatial context in maize fields. A wide range of data are also included in this research, such as weather data, soil or socioeconomic inputs, and remote-sensing indices (NDVI/EVI or raw reflectance bands) [1] [2][4][5].

2.4 Strengths and Limitations

- Strengths:** Hybrid CNN-LSTM models are capable of capturing intricate patterns that single models are unable to. By combining spatial (image) and temporal qualities, they learn how vegetation (seen by satellite) and changing weather mutually affect yield. These hybrids improve accuracy, according to numerous research; Sun et al. and Zhang et al., for example, report better RMSE and R^2 than pure CNN or LSTM models. They work well with multi-source data (satellite, climate, soil) and can handle the non-linear interactions that crop processes have [1][4].
- Limitations Hybrid:** DL models are more complex and demand a lot of high-quality data. In actuality, performance is greatly influenced by crop type, geography, and data quality. Typical issues include cloud contamination in satellite images, missing values in time series, and small sample sizes in ground yields. For example, Sun et al. found that unusual conditions and a paucity of training data caused their model to perform badly in 2012, an extreme drought year. Moreover, deep

models are "black boxes," which makes it difficult to determine what factors affect predictions. Another issue is computational cost: training CNN-LSTM on large satellite datasets (multi-year, high resolution) can be expensive, requiring the use of tools like Google Earth Engine (as Sun et al) [1] [3].

2.5 Common Data Preprocessing and Features

Researchers prepare data using similar methods. Typical behaviors include:

- **Vegetation indices:** Numerous studies use satellite bands to determine NDVI, EVI, or similar indices because of their association with plant health. CNN inputs are typically either indices over time or raw reflectance bands. [1]
- **Cloud/noise filtering:** Unprocessed satellite time series are noisy. For example, Zhang et al. used an 8-day maximum-value composite to reduce clouds and applied a Savitzky–Golay smoothing filter to LAI/EVI time series. Sun et al. selected cloud-free pixels using high-quality data [4].
- **Masking and alignment:** Data is hidden using crop regions (e.g., using the Cropland Data Layer to maintain only soybean fields). Crop phenology-corresponding 8- or 16-day periods are commonly used to transform date-aligned meteorological and satellite data [4].
- **Normalization and augmentation:** Datasets are regularly scaled (for instance, by standardizing reflectance values) and sporadically added. For instance, Taremwa et al. used SMOGN oversampling to balance constrained yield data. Others use window-based augmentation techniques or data fusion [2].

2.6 Implementation Challenges

- **Data availability:** Detailed yield records are difficult to find in many places. This limits how models can be trained. In order to resolve this limitation, some studies use data augmentation or transfer learning (pre-training on data-rich locations). [2] [5]
- **Spatial/temporal resolution:** Data received from the Satellite might be very imprecise. For example, MODIS gives 500m pixel; finer patterns in the fields can get missed. Also, the weather stations and remote sensing might not be able to synchronize perfectly in time or might not cover every area. [6]
- **Overfitting:** The complex models may overfit in the cases when samples are small. To handle small samples and avoid overfitting, Song et al. used cross-region transfer learning. To improve generalization, some uses cross-validation or Bayesian hyperparameter. [5]
- **Interpretability:** Hybrid networks have many parameters which is the reason that makes them hard to understand. To identify which inputs are significant and which not, the researchers use attention layers or feature importance tests. Even then it is tough to understand "why" the model makes a prediction.
- **Computation:** It requires a lot of processing power to handle imageries of years to handle, as Sun's MODIS time series from 2003 to 2015. Many of the researches uses cloud system like HPC and Google Earth Engine to manage massive volumes of data. [1]

2.7 Research Gaps and Future Directions

Despite of advancements, there are several gaps that still exist in the model:

- **Data fusion and modalities:** Several promising new sources of data comprise of satellite-based fluorescence such as SIF, high-resolution drones, and soil moisture satellites. Using additional data types like weather, management and socioeconomics, into unified models is one of the future directions. [4]
- **Scalability and transferability:** While more work is needed to enable models to automatically adapt across climates and crops, transfer learning and cross-regional training (as in Song et al.) show promise. Models taught in one site might not function in another. [5]
- **Real-time forecasting:** Mostly the studies focus on yield at the end of the season when crops are yielded. Continuous or in-season predictions that means using the most recent data as soon as it becomes available, is still a challenging task. The clear goal is the integration with a monitoring system for the real-time forecasting. [4]
- **Model interpretability and simplicity:** Models in the future might use AI approaches that are explainable. To highlight key components, Song et al. and others provide an interpretable layers and attention techniques. Further study is required in order to understand about what the models learn, like telling real yield factors from false correlations. [5]
- **Advanced architectures:** For the spatial and temporal data new networks such as Transformers or graph neural networks could be studied. According to Taremwa et al., in the future transformers and explainable layers will be used in order to improve the prediction. [2]
- **Benchmarking and diverse crops:** There are many studies that focuses on major crops like wheat, maize and soybeans etc, in specific regions. If the test were performed on smaller crops or in various area then wider application of results could

be possible. Evaluating approaches that uses common standards which includes common datasets and metrics would also simplify the process.

By learning from large amount of spatial and temporal data, hybrid CNN-LSTM models have a great deal of potential for better yield estimates. More accuracy is seen in recent studies as compared to the previous methods. However, they also need to deal with the quality and complexity of the data. As there will be further research on data fusion, scalability and interpretability, there will be a rise in the utility of these tools for farmers and policymakers. [1][3][4] [5]

3. METHODOLOGY

The main objective of this study's is to develop such a model that can forecast the agricultural yield accurately by using a hybrid CNN-LSTM neural network. The procedure is organized and simplified in such a way that both temporal and spatial information can be used effectively for the prediction of crop yield.

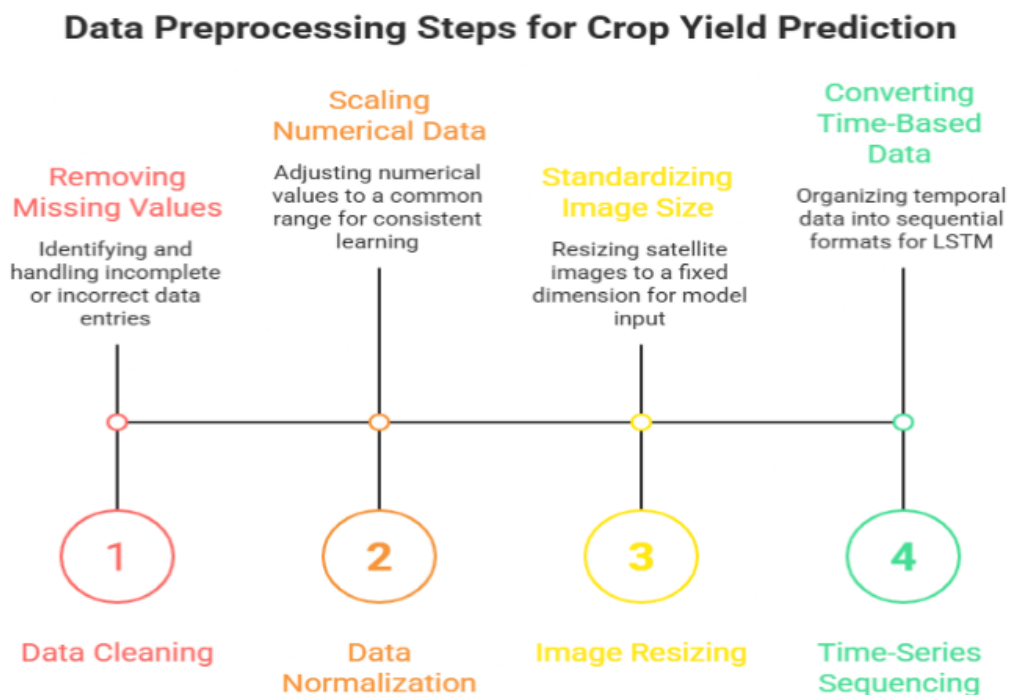
3.1. Data Collection

Firstly, we have collected a variety of data from reliable sources. The dataset consists of:

- Satellite images of the farming lands
- Weather information, including the humidity, temperature, and amount of rainfall in that location
- Fertility and moisture information about the soil
- Previous data about the crop yield

These data sources are important because the productivity of the agriculture is affected by environmental and geographic factors. Using many data sets and sources makes the model more precise and accurate.

3.2. Data Preprocessing



The data is cleaned and processed well before used in the model. This process consists of:

- Eliminating wrong or the missing values
- Normalizing numerical data so that every number falls within a comparable range of data
- Deciding a fixed size for the satellite images
- Creating a proper sequence from the time-based data

This phase is important because the raw data is sometime unreliable and noisy. The model learns and performs more precisely and accurately when preprocessing of the data is done correctly.

3.3. Feature Extraction using CNN

Convolutional Neural Networks (CNNs) are used in this step in order to extract key features of the satellite images.

CNN assists in spotting trends like:

- Crop well-being
- Density of vegetation
- Patterns of land usage

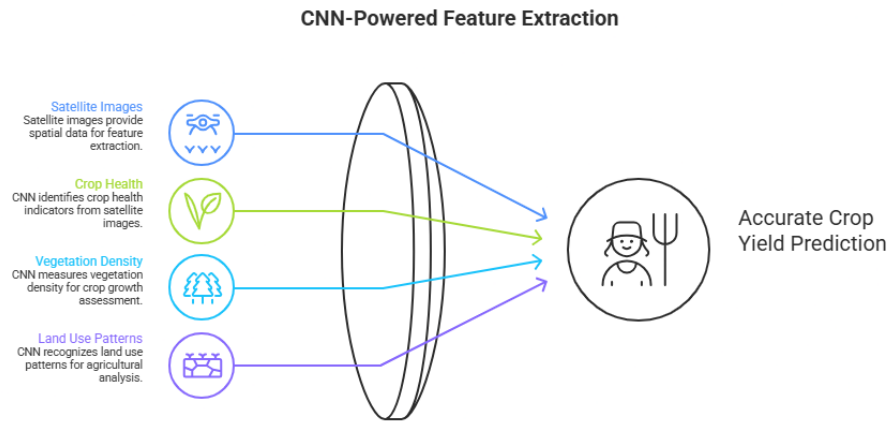


Fig 3.4.1 CNN Powered Feature Extraction

CNN automatically extracts useful features from photographs rather than choosing them by hand. Performance is enhanced and human effort is decreased as a result.

3.4. Temporal Modelling using LSTM

The next stage is to examine how spatial features evolve over time after they have been extracted. Long Short-Term Memory (LSTM) is employed for this.

LSTM is beneficial because:

- Time-series data can be handled by it.
- It retains historical data for a longer amount of time.
- It records weather-related and seasonal variations.

This aids in the model's comprehension of how crop conditions change during the growing season.

3.5. Hybrid CNN-LSTM Model Design

In this study, CNN and LSTM are combined to form a hybrid model. The model operates as follows:

- After processing the input images, CNN extracts spatial characteristics.
- After that, the LSTM network receives these features.
- LSTM examines the feature sequence over time.
- Lastly, crop yield is predicted using a fully connected layer.

This combination increases prediction accuracy by enabling the model to use both temporal and spatial information.

3.6. Model Training

Data from the past is used to train the model. While undergoing training:

- There are training and testing sets in the dataset.
- The training data helps the model identify trends.
- Error is measured using loss functions like Mean Squared Error (MSE).
- Model weights are updated using optimization methods such as the Adam optimizer.

Until the model's performance increases, training is performed several times (epochs).

3.7. Model Evaluation

Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) score are some of the evaluation metrics that are frequently used to assess how well the model performs after training using unseen data.

3.8. Implementation Tools

The following well-known tools and libraries are used to implement the model:

- Python is a programming language.
- Keras with TensorFlow for deep learning.
- Pandas and NumPy for managing data
- OpenCV for processing images.

These technologies facilitate the efficient construction and training of the model.

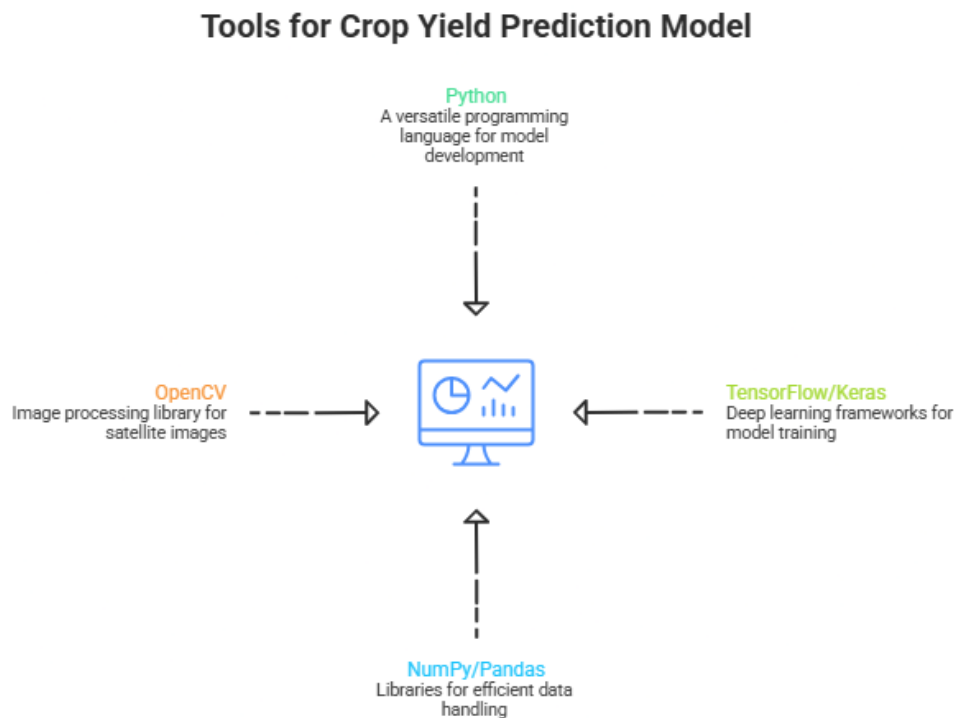


Fig.3.9.1 Tools for crop prediction Model

3.9. Workflow Summary

The research's whole workflow can be summed up as follows:

Data Collection → Data Preprocessing → Feature Extraction (CNN) → Temporal Analysis (LSTM) → Model Training → Evaluation → Prediction

3.10. Conclusion of Methodology

A hybrid CNN-LSTM model was used in this method to predict the agricultural yield in a clear and orderly manner. Using the time-based and image-based data together, the model becomes more robust and precise. The method of Deep learning allows the system to handle complex agricultural data and produces projections more precisely.

4. RESULTS

This section presents the performance of the proposed hybrid CNN-LSTM model. The main goal of this study was to find out how effective is this model in predicting crop output using both time-based and image-based data.

4.1 Model Performance

This model was firstly trained on the gathered data set and then it was tested on new data to examine its prediction capability. The model was very efficient in forecasting agricultural yield. In most of the situation the predicted values were very close to the actual yield.

The model analysed and learned from both spatial pattern from satellite image and temporal pattern from meteorological data. It therefore stand out to be better than the models that uses only one type of data.

4.2 Evaluation Metrics

In order to check the performance of the model, several evaluation metrics were used:

- Mean Absolute Error (MAE): The MAE value of the model shows that there was a slight variation in the expected values and actual yields.
- Root Mean Squared Error (RMSE): The low value of RMSE measurements indicated that there is a decrease in large errors.
- R-squared (R^2) Score: The high R^2 score of the model suggests that it can be capable in explaining most of the variation in crop yield data.

From this result we can clearly see that the Hybrid model produces the accurate forecast.

4.3 Comparison with Other Models

To further understand the effectiveness of the proposed model, it was compared to other popular models, such as:

- Conventional regression models.
- Machine learning models (Random Forest, SVM)
- Distinct deep learning models (just CNN and LSTM)

The results showed that:

- Traditional models had greater error rates.
- Although machine learning models performed better, they were still not very accurate.
- Time patterns could not be captured by CNN-only models, but spatial data could.

While capturing time patterns, LSTM-only models disregarded picture characteristics. Because it includes the advantages of both CNN and LSTM, the hybrid CNN-LSTM model fared better than any of these methods.

4.4 Visualization of Results

The actual and predicted crop yield figures were plotted. The graph, which showed that most of the predicted points were quite close to the actual values line, displayed high accuracy. Loss curves were also plotted during training. The model was learning appropriately and avoiding overfitting, as these figures showed a steady decrease in training and validation loss.

4.5 Impact of Different Data Inputs

The study also looked that how the different input data affected the performance of the model:

- The accuracy was decreased when only weather data was considered.
- The accuracy was improved but in a limited amount when only satellite images was used.
- Highest accuracy and best result were produced on combining both of the two data sets.

This shows that crop productivity is affected by various factors and combining these factors enhances the forecasting.

4.6 Observations

Some important conclusion that are drawn from the data are as:

- The functioning of the model was good when the high-quality of the data was provided.
- LST component captured the seasonal patterns successfully.

- CNN was successful in extracting useful features from images without human assistance.

However, the accuracy of forecast was reduced in a number of cases, which was due to noisy or missing data.

4.7 Limitations Observed

Despite the strong performance of the model, there were a few shortcomings that were identified:

- Quality of the data affected the performance significantly.
- A lot of processing power is required to train the model.
- Time taken by this model was longer as compared to the simpler models.
- From these limitations we can see that further improvements can be made in future work.

5. CONCLUSION

In order to improve the projection of crop yield, this study used a hybrid CNN-LSTM neural network model. The main goal of this study was to combine the benefits of the CNN with the benefits of LSTM in such a way that the model could handle spatial and temporal input both efficiently. The results clearly showed that our approach gave a better outcome together than both individual and traditional machine learning methods.

The study showed that the health of the crop, quality of the soil, and the weather in which they are grown, all of these factors directly impact the productivity of the agriculture. By combining the time-based data, including the temperature and rainfall details, with the satellite images, the model was able to better understand these factors. The CNN component of the model helped in the extraction of useful features from the images, while the component of LSTM helped in understanding how these features changes over time. Using this combination had resulted in an increased overall accuracy of the forecast.

The outcome of the study resulted in reduced prediction error and gave more reliable result. The performance of CNN-LSTM model was better than both traditional methods and individual deep learning models. This clearly shows that using both spatial and temporal information is required for predicting the crop yield accurately.

There were some more challenges that has appeared during the inquiry. The efficiency and the performance of the model was affected by the availability and the quality of the data. The prediction accuracy of the model was sometime affected by the Noisy or missing data. The model might not be effective or accurate for the small-scale farming as because for small scale farming, more time to train the model and more processing power is required.

Larger and more diverse datasets can be used by the researches in future on this topic. In order to further improve the performance, techniques like transformer-based models or attention mechanisms can be researched. Furthermore, more efforts can be made in order to develop models that are simpler and easier for agricultural expert and specially for farmers to use.

The study result show that the hybrid CNN-LSTM model has a potential as a crop production forecasting technique. This model gave a very accurate and reliable way of predicting the crop productivity, which can help in better decision-making, better planning of agriculture, and greater food security.

6. REFERENCES

- [1] County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model - PMC
<https://pmc.ncbi.nlm.nih.gov/articles/PMC6832950/>
- [2] Prediction of maize yield in Uganda using CNN-LSTM architecture on a multimodal climate and remote sensing dataset | Discover Artificial Intelligence | Springer Nature Link
<https://link.springer.com/article/10.1007/s44163-026-00855-7>
- [3] In-season crop yield prediction: State of the art and future research direction - ScienceDirect
<https://www.sciencedirect.com/science/article/pii/S1569843226000452>
- [4] Frontiers | BO-CNN-BiLSTM deep learning model integrating multisource remote sensing data for improving winter wheat yield estimation
<https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2024.1500499/full>
- [5] Wheat Yield Prediction Based on Parallel CNN-LSTM-Attention with Transfer Learning Model
<https://www.mdpi.com/2077-0472/15/23/2519>
- [6] A hybrid CNN-LSTM deep learning framework for enhanced crop yield prediction using spatial-temporal agricultural data
<https://www.mathsjournal.com/pdf/2025/vol10issue12S/PartA/S-10-11-3-466.pdf>