INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

**Received:** 15-02-2024

**Accepted:** 27-03-2024

**Published:** 24-04-2024

**Citation:** Linkesh M, Ghorpade M, Prasad P (2024) Jowar and Wheat Yield Prediction using a Wavelet based Fusion of Landsat and Sentinel Data with Meteorological Parameters. Indian Journal of Science and Technology 17(17): 1791-1799. https://doi.org/10.17485/IJST/v17i17.413

* **Corresponding author**.

monisha.linkesh@tcetmumbai.in

**Funding:** None

**Competing Interests:** None



Published By Indian Society for Education and Environment (iSee)

**ISSN**
Print: 0974-6846
Electronic: 0974-5645

# Jowar and Wheat Yield Prediction using a Wavelet based Fusion of Landsat and Sentinel Data with Meteorological Parameters

**Monisha Linkesh[1]*, Minakshi Ghorpade[1], Pratibha Prasad[1]**

**1** Assistant Professor, Department of Information Technology, Thakur College of Engineering and Technology, Mumbai, Maharashtra, India

## Abstract

**Objectives:** The objective of this study is to improve the accuracy of crop yield prediction models, specifically focusing on wheat and jowar crops in Maharashtra during the Rabi season, by integrating Landsat and Sentinel satellite data with meteorological parameters. **Methods:** The study utilizes Landsat 8 and Sentinel satellite datasets covering Maharashtra State. Atmospheric correction is applied to extract surface properties, followed by wavelet-based fusion to combine the images. Normalized Difference Vegetation Index (NDVI) is calculated and combined with meteorological parameters using ensemble learning techniques, including Random Forest and Ada-Boost algorithms. Comparative analysis is conducted against existing models, considering parameters such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). **Findings:** Significant findings reveal that the proposed methodology outperforms existing models, achieving lower MAE, MSE, and RMSE values for wheat and jowar yield predictions. Additionally, our research highlights the superiority of wheat production over jowar in the Rabi season, based on comprehensive analysis of crop yield predictions. **Novelty:** This study introduces a novel approach that integrates multiple data sources and employs ensemble learning techniques to enhance crop yield prediction accuracy. By combining Landsat and Sentinel satellite data with meteorological parameters, our methodology provides a more comprehensive understanding of crop growth dynamics, leading to more reliable predictions compared to existing methods.

**Keywords:** Satellite imagery; Machine learning; Normalized Difference Vegetation Index; Fusion; Ensemble learning

## 1 Introduction

Wheat cultivation plays a pivotal role in global food security, with India ranking second in production. Despite its significance, accurately predicting wheat yield remains a challenge due to various environmental factors. Previous studies have utilized Sentinel data and meteorological parameters to address this issue. In contrast, this research introduces a novel approach by integrating Landsat and Sentinel data with meteorological parameters to enhance prediction accuracy. This study aims to bridge existing research gaps by leveraging multi-source data fusion and ensemble learning techniques to predict wheat crop yield in Maharashtra during the Rabi season. By combining Landsat and Sentinel satellite imagery with meteorological data, including rainfall, temperature, and humidity, we seek to improve the precision of yield estimation models. However, there is a lack of research that thoroughly integrates both types of data and uses advanced machine learning methods to improve predictive accuracy. Our study seeks to fill this gap by proposing a novel approach that combines Landsat and Sentinel satellite data with meteorological parameters to predict wheat crop yield. By leveraging ensemble learning techniques such as Random Forest and Ada-Boost, we aim to improve the predictive performance compared to existing models. Additionally, our research addresses the need for more accurate and timely yield forecasts to support agricultural decision-making in the region.

The primary motivation behind this research is to offer farmers reliable insights into crop yield forecasts, aiding in decision-making processes related to agricultural practices. By considering the seven developmental stages of wheat: it starts from March to June. The Green-up (beginning of March) is the first stage, followed by the Jointing Stage (end of March), Elongation Stage (beginning of April), Booting (middle of April), Heading Stage (end of April), Anthesis Stage (beginning to middle May), and finally maturity stage (beginning of June) and incorporating comprehensive datasets, our methodology aims to provide a more holistic understanding of crop growth dynamics.

The major area of application for this work lies in agricultural management and decision support systems. Through accurate yield predictions, farmers can optimize resource allocation, mitigate risks, and enhance overall productivity. This study addresses a well-posed problem by integrating advanced data analytics techniques with domain-specific knowledge to improve crop yield forecasting accuracy.

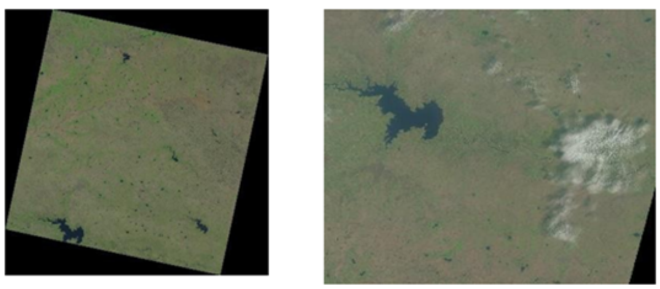

**Fig 1. Example Landsat and Sentinel Image**

Cheng Enhui et al. [1,2] proposed a study that compares various data-driven approaches for predicting winter wheat yields in China using Sentinel-2 and ZY-1 02D data. The LSTM model outperforms RF, GBDT, and SVR methods, with an RMSE of 0.201 t/ha. Hyperspectral data from ZY-1 02D yields more accurate estimates than 30-m Sentinel-2 data, with a 5% improvement in RMSE. Additionally, 10-m Sentinel-2 data show even better performance, highlighting the potential of deep learning methods like LSTM for accurate yield estimation.

To estimate the crop's yield, Rohit Ravi et al. [3] created a model utilizing the XG Boost algorithm. Predicting agricultural yields using data on cultivated area, rainfall, and maximum and lowest temperatures is the primary goal of this study. It will assist our Indian farmers in forecasting agricultural yields based on environmental factors. The crop that is considered an input is rice. This study compares the XG Boost method against decision trees, random forests, support vector regression, and linear regression. R2, minimum square error, and minimum absolute error are the metrics used to compare these algorithms. The data for the years 2000–2014 has been generated using the data.gov.in website. Since all four of these southern Indian states—Andhra Pradesh, Karnataka, Tamil Nadu, and Kerala have similar climates, data for these states are included. Compared to other models, the model suggested in this paper which is based on XG Boost is performing significantly better. Comparing the XG Boost $R^2$ to other models, it is the best at 0.9391.

Rohit Ravi et al. [4] proposed a yield prediction model for various crops in India using neural network regression modeling. A dataset for different seasons and years from 1998 to 2014 was gathered from an Indian government website. The study targets Maharashtra state and uses Python Pandas and Pandas Profiling tools for data filtering. Initially, a Multilayer Perceptron neural network with an RMS prop optimizer achieved 45% accuracy, which was later improved to 90% by adjusting network architecture, weights, and biases, and changing the optimizer to Adam. The developed model, utilizing a 3-layer neural network with Rectified Linear Activation Unit (ReLU) function, establishes relationships between various input parameters (e.g., cultivation area, crop, state, district, season, year) and crop yield. Backward and forward propagation techniques are utilized to train the model, demonstrating its effectiveness in predicting crop yields accurately.

Gohar Ghazaryan et al.[2] study aimed to assess the performance of various algorithms and remotely sensed time-series datasets for yield estimation at both county and field scales in the United States. For county-level analysis, MODIS-based surface reflectance, land surface temperature, and evapotranspiration time series were utilized, while field-level analysis employed NASA's Harmonized Landsat Sentinel-2 (HLS) product. Convolutional Neural Network (CNN) and CNN, followed by Long Short-Term Memory (LSTM) models, were applied, with the CNN-LSTM model demonstrating the highest accuracy for county-level analysis, achieving mean percentage errors of 10.3% for maize and 9.6% for soybean. Notably, robust results were obtained for the drought year of 2012. At the field level, all models produced accurate results, with R-squared values exceeding 0.8 when utilizing mid-growing season data.

Saeed Nosratabadi et al.[5] suggested new agricultural yield prediction models using hybrid machine learning methodologies. It encompasses crops such as wheat, barley, potatoes, and sugar beets. The current study focuses on farms near the Iranian city of "Kerman". Kerman is Iran's largest province (183,285 km2), accounting for 11% of Iran's total land area. To achieve the study's purpose, two types of data were collected: 1) agriculture production, and 2) information about the weather. This study compared the performance of artificial neural networks-imperialist competitive algorithm (ANN-ICA) and artificial neural networks-gray wolf optimizer (ANN-GWO) models for agricultural yield prediction. In contrast, in this study, a wide range of criteria are used to assess the models' performance. In order to accomplish this, the ideal attribute set for each technique was first identified from among the various attributes. According to the results, the ANN-GWO technique performed better in crop yield prediction, with an R of 0.48, an RMSE of 3.19, and a MEA of 26.65. Because a different set of qualities affects the performance of the model, it is advised that future study examines a new set of attributes and compare the results.

Potnuru Sai Nishant et al.[6] predicts the yield of almost all kinds of crops that are planted in India. Simple factors like state, district, season, and area are used in this method to create novelty, and the user can anticipate the crop's production in whatever year they want. The study predicts the yield using regression techniques such as Kernel Ridge, Lasso, and ENet algorithms. It also applies the idea of stacking regression to improve the algorithms' performance and provide a more accurate prediction.

## 2 Methodology

### 2.1 Dataset Details

The data set was obtained from the USGS (United States Geographic System) website. This study focused on the State of Maharashtra. Landsat 8 OLI/TIRS c2 L1 and sentinel images were collected for Maharashtra state. The path and row of this dataset are (146, 046). The latitude and longitude of this study area are 19 45'05" North latitude, 075 42'50" East longitude respectively. The cloud cover range is below 5%. The dataset years were collected from 2013 to 2021. The meteorological data was collected from the weather website of Maharashtra. It consists of minimum and maximum temperature, precipitation, dew point, wind speed, visibility, humidity, pressure, NDVI, condition and production.

The basic process is as follows:

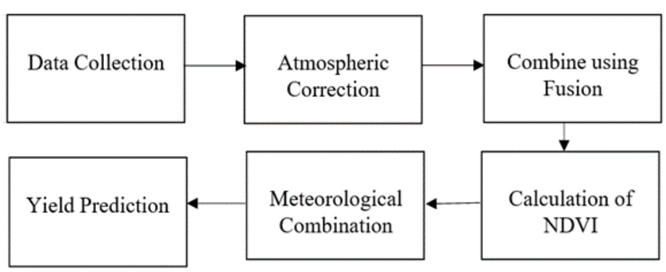


**Fig 2. The Workflow of Yield Prediction Model**

### 2.2 Atmospheric Correction

The main objective of doing atmospheric correction is to extract the surface properties from the satellite image. This method improves the accuracy of classification. The atmospherically corrected images will significantly improve accuracy when using multiple date images. It is a significantly necessary process when calculating the vegetation indices.

This process can be calculated using two phases:

1. DN to Radiance Conversion $L = G^* DN + B$, where DN: Digital Number G and B are gain and bias values for a particular band;
2. Radiance to TOA conversion P=Π * L* D 2 / ESUN * Cos⊠

All the values used in the formula are available in the metadata file of the image file.

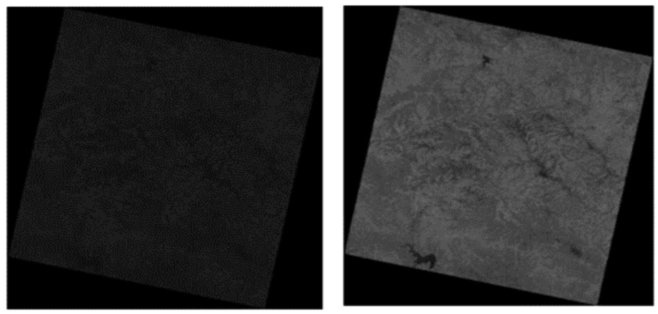

**Fig 3. 29/12/15: atmospheric correction of Landsat data**

## 2.3 Fusion

Image fusion is a technique for combining images. It combines two images and the resultant image is the fused one. The fused image contains more information than a single image. This fusion technique is not only needed for reducing the amount of data but it also helps to restore an image from more than one degraded image and for mixing images. There are various fusion techniques available for combining images. Some of the examples are; feature-based, pixel-based, and decision-based fusion. The fusion technique is of two types: spatial and frequency domain. The spatial domain consists of High Pass Filter, Intensity Hue Saturation, Brovey algorithm and Principal component substitution method. The frequency domain consists of pyramid-based algorithms, discrete cosine transforms, curvelet based and discrete wavelet transforms. This proposed work uses a wavelet based fusion technique. It is a mathematical tool for decomposing the images, and it provides efficient localization in both the spatial and frequency domains. It is an extension idea of the high pass filtering and it produces enhanced accuracy compared to other fusion methods. These wavelets are generated using High and low pass filters. The signal S is passed through these filters and down sampled by two. It produces low and high-frequency signals called an approximation and detailed coefficient respectively. The atmospherically corrected images were fed into the image registration step. This method converts the two different sets of images into one format. It is used in the remote sensing field to align different satellite images. Then resampling was done. It is the same as resizing, but it changes the physical number of pixels[7]. This proposed work uses the down sampling method rather than up sampling because it preserves most pixels information and enhances the quality of the image. The 30m resolution band was reduced to 10 m, the same as in the sentinel band.

## 2.4 Generation of Wavelets

Wavelet theory overcomes the limitations of Fourier transform and short time Fourier transform. The wavelet decomposition of a signal s(t) based on the multiresolution theory given by Mallet and Meyer can be done using digital FIR filters.

In one level decomposition, the signal S is passed through high and low pass filters and down sampled by two. It produces low and high-frequency signals called an approximation and detailed coefficient respectively.

The Wavelet transformation based fusion technique was used in this proposed work. It has two forms discrete and continuous wavelet transforms. This research work uses the discrete wavelet transform technique.

- **Steps followed in wavelet transform technique**

1. Get the input image.
2. Apply the wavelet transformation to two images to get the decomposition of two images. This function decomposes the two images into 4 coefficients. NDVI = NIR- R / NIR+ R.
3. Find the detail and approximation coefficients of two input images. CA is the approximation coefficient of image, and CH, CV, and CD are the detailed coefficients of input images. Here CA1 is the approximation coefficient of image1; CH1

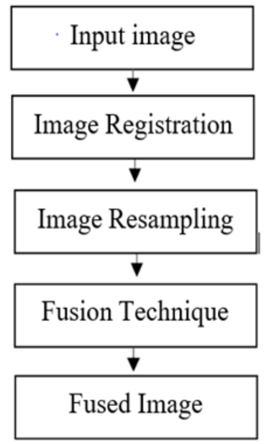


**Fig 4. Fusion Steps**

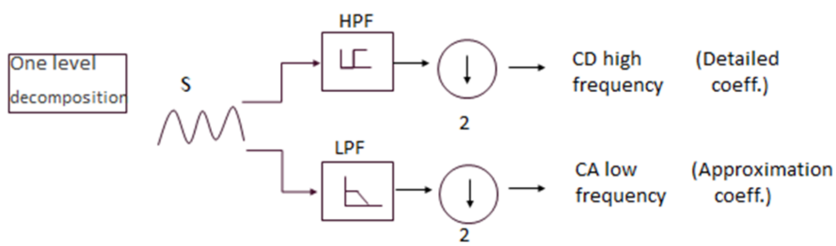


**Fig 5. Generation of wavelets using filters**

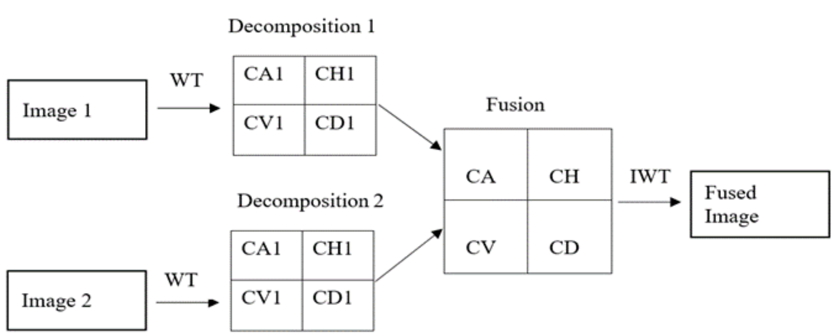


**Fig 6. Working of wavelets in fusion process**

is the detailed horizontal coefficient of image1; CV1 is the detailed vertical coefficient of image1. CD1 is the detailed dimensional coefficient of image1.

4. Merge the coefficient of two images by using the rule CA=fusion1(CA1,CA2), CH=fusion2(CH1,CH2), CV=fusion2(CV1,CV2), CD=fusion2(CD1,CD2). Fusion 1 and 2 are two mathematical operations. Fusion 1 is for the approximation coefficient and fusion 2 is for the detailed coefficient. These operations can be mean, max, or min. so nine combinations of operations are possible.
5. Apply inverse transform on the merged coefficient
6. Get the final fused image as output.

## 2.5 Calculation of NDVI

Normalized Difference vegetation Index (NDVI) estimates the green portion by calculating the ratio between the discrepancy between the near-infrared (NIR) and red bands (R) and the summation between the near-infrared and red bands.

This NDVI value fluctuates between minus one to plus one at all times. Water bodies are represented by a negative attribute. The value near to zero indicates, there is no vegetation portion, and the value near to positive one indicates a dense leaf portion.

In Landsat8 image, NIR is the 5th band and RED is the 4th band and in sentinel 2 image, 8 and 8a bands are NIR and RED is the 4th band as shown in Table 1.

The NDVI value was calculated from the atmospheric corrected image to get the value of vegetation and other parameters as shown in Figure 7.

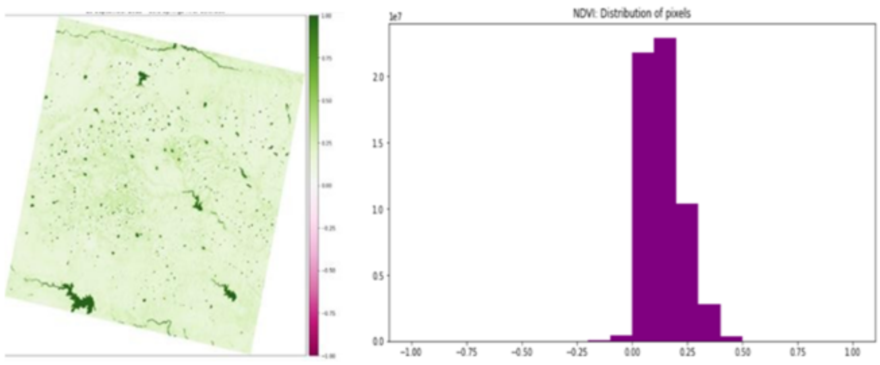


**Fig 7. Distribution of NDVI value**

It was observed that compared to the single satellite image the fusion of images increased the accuracy of the estimation model and when comparing wheat and jowar crops in rabi season, the wheat crop will give the maximum yield than the jowar crop as shown in Figure 7.

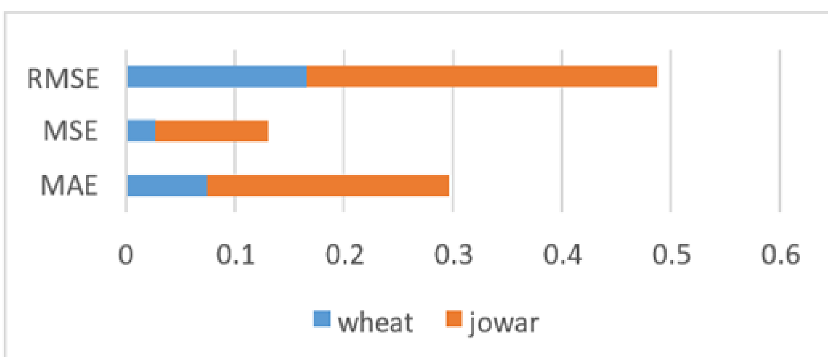


**Fig 8. Accuracy comparison of Landsat, Sentinel and Fused image**

## 2.6 Meteorological combination

Meteorological parameters such as temperature, precipitation, dew point, humidity, pressure, and production factors are combined with NDVI value to find the final yield of the crop using ensemble learning. This ensemble learning is a combination of the Random Forest and Ada-Boost algorithm.

### *2.6.1 **Working of Random Forest***

Step1: N number of the decision tree was built for the various subset of the given dataset.

Step 2: Then take the average of all decision trees to improve the accuracy of the prediction model. Advantages of using Random Forest Model:

1. It's Fast, Accurate and doesn't overfit.
2. It can be used to solve both classification and regression problems.
3. Work better with the high dimensional data point.
4. It handles well missing values.
5. Versatile in nature.

#### *2.6.2* ***Working of Ada-boost***

Step 1: First decision tree model is made from the dataset.

Step 2: The misclassified items from the first model only given as input to the second model.

Step 3: These steps will continue up to the specified number of base learners.

Advantages of using Ada-boost model:

1. It can be used with any model to improve its efficiency.
2. Can be used to solve a decision tree problem.
3. It learns from the model's prior error.

**Table 1. Landsat8 and Sentinel resolution band informations**

| Band No | Landsat Information | | Sentinel Information | |
|---|---|---|---|---|
| I | Ultra-blue | 30m | Coastal | 60m |
| II | Blue | 30m | Blue | 10m |
| III | Green | 30m | Green | 10m |
| IV | Red | 30m | Red | 10m |
| V | Near infrared | 30m | Vegetation Red Edge | 20m |
| VI | Cirrus | 30m | Vegetation Red Edge | 20m |
| VII | Short wave infrared | 30m | Vegetation Red Edge | 20m |
| VIII | Short wave infrared | 30m | Near Infrared | 10m |
| | | | a. Near Infrared | 20m |
| IX | Panchromatic | 15m | Water Vapor | 60m |
| X | Thermal infrared | 100m | Cirus | 60m |
| XI | Thermal infrared | 100m | Short Wave infrared | 20m |
| XII | - | - | Short Wave Infrared | 20m |

## 3 Results and Discussion

The performance of the yield prediction model was estimated using Mean Absolute Error, Mean squared Error, and Root Mean Squared Error metrics.

$$Mean\ Square\ Error = \sum_{i=1}^{N} \frac{(predicted\ value - actual\ value)^2}{N}$$

Table 2 shows that an ensemble algorithm shows better MAE results. Hence, the ensemble algorithm of random forest and Ada boost with ensemble of SVM and Linear Regression was compared and showed better results as shown in Table 2.

Table 3 shows wheat has less values in all three metrics which indicates the higher accuracy of production. Compared to jowar, wheat will give maximum production and jowar will give less production in rabi season as it needs different weather conditions for growing.

The satellite data were collected, and then atmospheric correction was done to get the surface properties of the land and then the images were fused and calculated NDVI value as shown in Figure 7. The calculated value was combined with meteorological parameters comprising minimum and maximum temperature, precipitation, dew point, wind speed, visibility, humidity, pressure, NDVI, crop condition, and production. Evaluation measures including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) were employed to assess predictive accuracy. The accuracy was increased when fusing two images.

**Table 2. Comparative Analysis of Existing and Proposed work**

| | Dataset Used | Evaluation Measures used / Techniques Used | MAE | MSE | RMSE |
|---|---|---|---|---|---|
| Proposed Work | minimum and maximum temperature, precipitation, dew point, wind speed, visibility, humidity, pressure, NDVI, condition and production | Random Forest | 0.1772 | 0.0865 | 0.2941 |
| | | Ada-Boost | 0.1493 | 0.0648 | 0.2547 |
| | | Ensemble Learning (Random Forest and Ada-Boost) | **0.0737** | **0.0276** | **0.1661** |
| | | SVM and Linear Regression | 0.1794 | 0.0562 | 0.2372 |
| Existing Work | MODIS-based surface reflectance, Land Surface Temperature, and Evapotranspiration time series [2] crop species, irrigation, crop yield rainfall, solar radiation, and temperatures [5], | Hybrid Models | **0.2665** | | **0.3190** |

**Table 3. Comparison of wheat and jowar yield prediction using ensemble of Random Forest and Ada Boost**

| Metrics Crops | MAE | MSE | RMSE |
|---|---|---|---|
| Wheat | 0.0737 | 0.0276 | 0.1661 |
| Jowar | 0.2222 | 0.1029 | 0.3209 |

This proposed work increases the precision of the yield forecasting model compared to utilizing the single satellite image [2]. Also, compared to the existing work [5], which used hybrid models for prediction, this proposed work used many meteorological parameters and all growing stages of wheat and improved the accuracy as shown in Table 1.

Comparing our proposed methods, Random Forest achieved an MAE of 0.1772, MSE of 0.0865, and RMSE of 0.2941. Ada-Boost exhibited improved performance with an MAE of 0.1493, MSE of 0.0648, and RMSE of 0.2547. The ensemble learning approach combining Random Forest and Ada-Boost yielded the lowest errors, with an MAE of 0.0737, MSE of 0.0276, and RMSE of 0.1661.

In contrast, existing works relying on MODIS-based surface reflectance and hybrid models reported with an MAE of 0.2665 and MSE of 0.3190.

These results demonstrate the superior predictive capability of our proposed methodology, leveraging ensemble learning techniques to achieve higher accuracy in wheat yield prediction.

Comparing our results with existing reports, we observe a significant improvement in predictive accuracy. Our methodology, integrating various meteorological parameters and employing ensemble learning techniques, outperforms previous models. For instance, while MODIS-based surface reflectance and hybrid models have been utilized in past studies, our approach demonstrates lower errors and higher precision in wheat yield prediction. One of the key contributions of our study lies in the comprehensive consideration of meteorological factors alongside satellite imagery. By incorporating parameters such as minimum and maximum temperature, precipitation, dew point, wind speed, visibility, and humidity, our model provides a more holistic understanding of crop yield dynamics. This approach enhances the predictive capability of our system, addressing a notable research gap in the field.

Furthermore, the utilization of ensemble learning algorithms, namely Random Forest and Ada-Boost, sets our study apart from existing works. This combination allows for the mitigation of bias and variance, leading to more robust predictions. The comparative analysis highlights the superiority of our model over traditional techniques, emphasizing its novelty and efficacy.

## 4 Conclusion

In conclusion, this study has contributed significantly to the field of crop yield prediction by introducing a novel approach that combines Landsat and Sentinel satellite data with meteorological parameters to enhance the accuracy of wheat yield forecasts. Our research has demonstrated a marked improvement in yield prediction accuracy compared to existing methods. By leveraging ensemble learning techniques such as Random Forest and Ada-Boost, coupled with advanced data fusion methods, we have achieved a substantial reduction in mean absolute error (MAE) and root mean squared error (RMSE) metrics. Specifically, our ensemble model, integrating Landsat and Sentinel data, exhibited an MAE of 0.0737 and an RMSE of 0.1661, outperforming previous studies that relied solely on single-source satellite imagery.

Strengths of our research lie in its comprehensive approach, which considers multiple factors influencing wheat growth and integrates diverse datasets to capture complex interactions. By incorporating meteorological parameters alongside satellite imagery, our model offers a more holistic understanding of crop dynamics, empowering farmers with actionable insights to optimize agricultural practices.

However, certain limitations persist, primarily related to data availability and processing complexities. While our study focused on Maharashtra state, broader geographical coverage could enhance the generalizability of our findings. Additionally, further refinement of data fusion techniques and ensemble learning algorithms may yield even greater predictive accuracy.

Moving forward, future research endeavors could explore the integration of additional variables, such as soil quality and pest infestation data, to enhance the robustness of yield prediction models. Furthermore, advancements in remote sensing technology and machine learning algorithms offer promising avenues for continued innovation in agricultural forecasting.

## References

1) Enhui C, Bing Z, Dailiang P, Liheng Z, Le Y, Yao L, et al. Wheat yield estimation using remote sensing data based on machine learning approaches. *Frontiers in Plant Science*. 2022;13:1–16. Available from: https://doi.org/10.3389/fpls.2022.1090970.
2) Ghazaryan G, Skakun S, König S, Rezaei EE, Siebert S, Dubovyk O. Crop Yield Estimation Using Multi-Source Satellite Image Series and Deep Learning. In: IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium. 2021. Available from: https://doi.org/10.1109/IGARSS39084.2020.9324027.
3) Ravi R, Baranidharan B. Crop Yield Prediction using XG Boost Algorithm. *International Journal of Recent Technology and Engineering*. 2020;8(5):3516–3520. Available from: https://www.ijrte.org/wp-content/uploads/papers/v8i5/D9547118419.pdf.
4) Patil PS, and SSK. A Machine Learning Approach to Predict Crop Yield and Success Rate. In: 2019 IEEE Pune Section International Conference (PuneCon). 2020. Available from: https://doi.org/10.1109/PuneCon46936.2019.9105741.
5) Nosratabadi S, Szell K, Beszedes B, Imre F, Ardabili S, Mosavi A. Comparative Analysis of ANN-ICA and ANN-GWO for Crop Yield Prediction. In: 2020 RIVF International Conference on Computing and Communication Technologies (RIVF). IEEE. 2020. Available from: https://doi.org/10.1109/RIVF48685.2020.9140786.
6) Nishant PS, Venkat PS, Avinash BL, Jabber B. Crop Yield Prediction based on Indian Agriculture using Machine Learning. In: 2020 International Conference for Emerging Technology (INCET). IEEE. 2020. Available from: https://doi.org/10.1109/INCET49848.2020.9154036.
7) Li C, Wang Y, Ma C, Ding F, Li Y, Chen W, et al. Hyperspectral Estimation of Winter Wheat Leaf Area Index Based on Continuous Wavelet Transform and Fractional Order Differentiation. *Sensors*. 2021;21(24):1–19. Available from: https://dx.doi.org/10.3390/s21248497.