Enhancing Crop Yield Estimation from Remote Sensing Data: A Comparative Study of the Quartile Clean Image Method and Vision Transformer

Manan Thakkar
mdt01@ganpatuniversity.ac.in

Ganpat University
Rakeshkumar Vanzara
Ganpat University

Research Article

Keywords: Machine Learning, Deep Learning, Quartile Clean Image, MODIS, Remote Sensing, Data Pre-processing, Noisy Feature Handling, Vision Transformer

Posted Date: September 27th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4969512/v1

Additional Declarations: No competing interests reported.

Enhancing Crop Yield Estimation from Remote Sensing Data: A Comparative Study of the Quartile Clean Image Method and Vision Transformer

Manan Thakkar[1*] and Rakeshkumar Vanzara[2†]

[1*] Computer Science and Information Technology, Ganpat University, Ganpat Vidyanagar, Gujarat, India
[2] Information Technology, Ganpat University, Ganpat Vidyanagar, Gujarat, India

*Corresponding author. E-mail: mdt01@ganpatuniversity.ac.in;
Contributing author: rakesh.vanzara@ganpatuniversity.ac.in;
†The author contributed equally to this work.

Abstract. The use of high-altitude remote sensing (RS) data from aerial and satellite platforms presents considerable challenges for agricultural monitoring and crop yield estimation due to the presence of noise caused by atmospheric interference, sensor anomalies, and outlier pixel values. This paper introduces a "Quartile Clean Image" pre-processing technique to address these data issues by analyzing quartile pixel values in local neighborhoods to identify and adjust outliers. Applying this technique to 20,946 Moderate Resolution Imaging Spectroradiometer (MODIS) images from 2003 to 2015 improved the mean peak signal-to-noise ratio (PSNR) to 40.91 dB. Integrating Quartile Clean data with Convolutional Neural Networks (CNN) models with exponential decay learning rate scheduling achieved RMSE improvements up to 5.88% for soybeans and 21.85% for corn, while Long Short-Term Memory (LSTM) models demonstrated RMSE reductions up to 11.52% for soybeans and 29.92% for corn using exponential decay learning rates. To compare the proposed method with state-of-the-art techniques, we introduce the Vision Transformer (ViT) model for crop yield estimation. The ViT model, applied to the same dataset, achieves remarkable performance without explicit pre-processing, with R^2 scores ranging from 0.9752 to 0.9875 for soybean and 0.9540 to 0.9888 for corn yield estimation. The RMSE values range from 7.75086 to 9.76838 for soybean and 26.25265 to 34.20382 for corn, demonstrating the ViT model's robustness. This research contributes by (1) introducing the Quartile Clean Image method for enhancing RS data quality and improving crop yield estimation accuracy, and (2) comparing it with the state-of-the-art ViT model. The results demonstrate the effectiveness of the proposed approach and highlight the potential of the ViT model for crop yield estimation, representing a valuable advancement in processing high-altitude imagery for precision agriculture applications.

Keywords: Machine Learning, Deep Learning, Quartile Clean Image, MODIS, Remote Sensing, Data Pre-processing, Noisy Feature Handling, Vision Transformer.

Declaration

- Funding: Not Applicable
- Conflict of interest/Competing interests: Not Applicable
- Data availability
  The code for data collection and processing used in this study is available at https://github.com/mananthakkar24/RemoteSensingBlobDetection. By using the Data Collection related code provided in this repository, researchers can extract the Remote Sensing dataset and store it in their Google Drive.
  A sample dataset used in this study is available at: https://drive.google.com/drive/folders/18Z3hcqRf0nnE5vjqDat99qh3o-OnP5cE?usp=sharing
- Code availability
  The code used in this study is openly available on GitHub at https://github.com/mananthakkar24/RemoteSensingBlobDetection. This repository contains all the necessary code for data collection, processing, and analysis as described in this paper. Users can clone or download the repository to access and run the code. Entire code also will be available on the same GitHub link.
- Research Involving Human and /or Animals: Not Applicable
- Author Contribution
  Conceptualization: Manan Thakkar; Methodology: Manan Thakkar; Formal anaysis and investigation: Manan Thakkar; Writing - original draft preparation: Manan Thakkar; Writing - review and editing: Manan Thakkar, Rakeshkumar Vanzara; Resources: Manan Thakkar; Supervision: Rakeshkumar Vanzara.

1 Introduction

Addressing the critical need for accurate crop yield estimation is paramount in the broader context of global food security challenges. With an estimated 795 million people living without sufficient food supply, and a projection of two billion additional people to feed by 2050, the urgency for precise crop output forecasts is more crucial than ever [1][2][3][4]. These forecasts are vital for effective agricultural production planning, securing food supply chains, and shaping policy decisions to prevent food crises [5][1]. The dynamic interplay of environmental factors, genetic traits, and agricultural practices significantly influences crop output, particularly in arid and semi-arid regions where variables like soil salinity can drastically impact yields [6][7][8][9].

The role of RS data in agricultural applications has become increasingly prominent. The advancements in satellite technology, such as the launch of Sentinel A/B and Landsat-8, have enabled the acquisition of multi-temporal, multi-altitude, and multispectral data, crucial for applications in climate change, urban planning, and agriculture [9][10][11][12][13]. In crop yield prediction, RS parameters like Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Indices (EVI), and Normalized Difference Water Index (NDWI) are frequently utilized, falling into the semi-empirical approach for yield estimation [14][15][16].

However, RS data obtained from high-altitude perspectives often present challenges such as low resolution and distortions from cloud cover, shadows, and sensor errors, introducing random noise into the dataset [17]. This noise can significantly impair the feature extraction capabilities of CNN models, often leading to overfitting and inaccurate yield predictions [18][19][20].

The need for pre-processing RS data is, therefore, critical. ML and Deep Learning (DL) algorithms, such as CNN and LSTM, have been extensively applied for yield estimation, demonstrating their potential in managing complex data and enhancing prediction accuracy [5][21][22][23][24]. However, the efficacy of these models is heavily reliant on the quality of the input data, highlighting the importance of effective data pre-processing.

To address these challenges, our research introduces the "Quartile Clean Image" technique, a novel data pre-processing method inspired by the Robust Scaler, which efficiently handles noisy features and outlier pixel values in remote sensing datasets [1][6]. By enhancing the quality of the data before inputting it into ML models such as CNN and LSTM, the Quartile Clean Image technique paves the way for more accurate and reliable crop yield estimations. To further assess the effectiveness of our proposed method, we compare its performance with a state-of-the-art DL model, the ViT, which has shown remarkable results in various computer vision tasks. Through this comparative analysis, we aim to demonstrate the significance of data pre-processing in improving crop yield predictions, even when using advanced models like ViT. Our approach not only tackles the immediate challenges associated with RS data quality but also contributes to the overarching goal of ensuring global food security through improved agricultural predictions. Figure 1 provides a concise visual summary of our research work, encapsulating the entire process flow for crop yield estimation, with a special emphasis on the application of the Quartile Clean technique in pre-processing RS image data, a key aspect of our methodology.

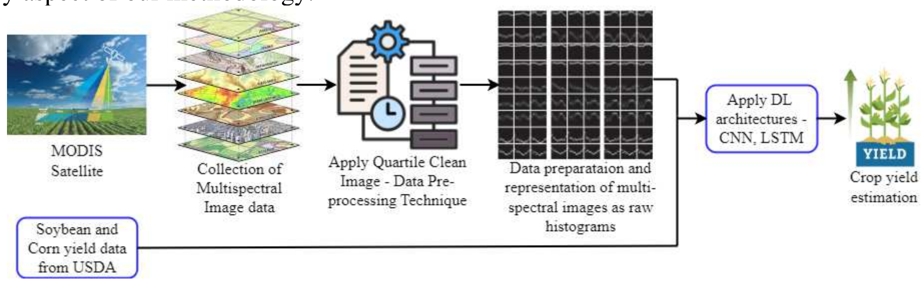

Fig. 1. Visual Representation of the Research Work Process Flow

1.1 Research Contributions

The following are this paper's main points:

1. **Development of Quartile Clean Image Technique:** Formulating a novel data pre-processing method specifically designed to tackle images with noisy features and outlier pixel values in high-altitude RS data.

2. **Enhancement of ML Model Performance:** Demonstrating significant improvements in the accuracy of CNN and LSTM models for agricultural yield prediction after applying the Quartile Clean Image technique.
3. **Comprehensive Analysis of Image Quality Improvements:** Detailed assessment of image quality enhancement using the Quartile Clean Image technique, observing quantifiable improvements in PSNR values in a large dataset of MODIS images.
4. **Introduction of ViT for Crop Yield Estimation:** Adapting the state-of-the-art ViT architecture for crop yield estimation from RS data and comparing its performance with the proposed Quartile Clean Image method.
5. **Comparative Study of Quartile Clean Image Method and ViT:** Conducting a comprehensive comparative analysis of the proposed Quartile Clean Image method and the ViT model in terms of crop yield estimation accuracy, robustness, and efficiency.

2 Related Work

In recent years, the use of RS data and advanced ML techniques for crop yield estimation has gained significant attention. Researchers have explored various approaches to improve the accuracy and reliability of yield predictions by leveraging diverse data sources and innovative methodologies. Table 1 summarizes the key findings and contributions of selected recent studies.

Table 1. Summary of recent studies on crop yield estimation using RS data and ML techniques

Crop	Features Used	Sample Characteristics	Algorithms	Metrics Evaluated	Best Performing Model & Accuracy
Winter Wheat [5]	GLASS LAI, AgERA Data, WOFOST Model, Weather data, Field observations	Chinese Data, Henan Survey Yearbook Yields	WOFOST, MCMC, EnKF	RMSE, MAPE	EnKF better accuracy than open loop WOFOST, RMSE reduced from 987 to 688 kg/ha
Corn, Soybean, Wheat [4]	EVI, soil moisture, weather data, Vegetation Optical Depth (VOD)	Multi-source RS datasets from Continental US (CONUS)	LSTM Neural Network	R^2 Score, RMSE	For Corn: R^2 Score: 0.61, RMSE: 1.32 For Soybean: R^2 Score: 0.69, RMSE: 1.28 For Wheat: R^2 Score: 0.66, RMSE: 1.25
Jute, Maize, Rice, Sugarcane, Wheat [25]	Fuzzy neighborhood spatial filtering, scaling, flipping, shearing, zooming on pooled images	Augmented imagery datasets from public FAO and World Bank data	Ensemble method, Gradient Boosting, Decision Trees, Random Forest, Naive Bayes	Accuracy, AUC, F1-Score, Precision, Recall	Ensemble method outperformed others by avg 13% over Gradient Boosting and 24% over Decision Trees. Gradient Boosting best for 2017-2021 yield prediction.

Winter Wheat [26]	VTCI, LAI, FPAR from remote sensing data	County-level data from Guanzhong Plain, China (2011-2020)	CNN-GRU, GRU	R^2, RMSE, MRE	CNN-GRU outperformed GRU: R^2 = 0.64, RMSE = 462.56 kg/ha, MRE = 8.90%. Leave-one-year-out cross-validation showed good stability and generalization.
Corn [27]	MODIS VIs, temperature, precipitation data	Multi-year county-level US dataset	proposed MMPD, DNN, DANN, MDAN, M3SDA	RMSE, MARE	Lowest error achieved by MMPD model for validation year – 2016 RMSE: 1.41
Early Winter Wheat [28]	Sentinel-2 derived domain adaptation techniques and vegetation indices	Data from Southwestern of the City of London, Ontario, Canada	MLR	RMSE	Average RMSE: 604.48 kg/ha
Wheat, Barley [29]	APAR, T, Kt·Kst derived from multispectral satellite imagery and meteorological data	Commercial fields in Spain under different water and nutrient conditions	MYRS model integrating RS data into classical CGM approaches	R^2, RMSE, d index	MYRS model performed well at field and within-field scales. Kt·Kst approach best for wheat (R^2=0.94, RMSE=71 g/m^2, d=0.98). Good performance for barley (R^2=0.81, RMSE=62 g/m^2, d=0.94).

Huang et al. (2023) [5] proposed an improved winter wheat yield estimation method by assimilating GLASS LAI into a crop growth model using a Bayesian posterior-based ensemble Kalman filter (EnKF), reducing the RMSE from 987 to 688 kg/ha compared to the open loop WOFOST model. Similarly, Mateo-Sanchis et al. (2023) [4] developed an LSTM neural network approach for predicting corn, soybean, and wheat yields using multi-source RS datasets from the Continental US, achieving promising accuracy with R^2 Scores ranging from 0.61 to 0.69 and RMSE values between 1.25 and 1.32. These studies highlight the effectiveness of integrating remote sensing data with advanced data assimilation techniques and DL models for accurate crop yield estimation.

Several studies have focused on developing novel architectures and techniques to further improve crop yield prediction. Ilyas et al. (2023) [25] proposed a fuzzy hybrid ensemble method that enhances pooled images with various transformations and identifies optimal weights of candidate classifiers using a bagging strategy, outperforming individual classifiers by an average of 13% to 24%. Wang et al. (2023) [26] introduced a CNN-GRU model that combines the feature extraction capability of CNNs with the time series memory advantage of GRUs to estimate county-level winter wheat yields in China, demonstrating good stability and generalization over different crop years. Ma et al. (2023) [27] proposed an adaptive domain adversarial neural network (ADANN) model for corn yield prediction using a multi-year county-level US dataset, achieving the lowest error for the validation year 2016 with an RMSE of 1.41. These studies showcase the potential of innovative architectures and techniques in improving crop yield estimation accuracy.

Other studies have explored the use of specific RS data sources and techniques for early-stage yield forecasting and within-field variability assessment. Liao et al. (2023) [28] employed domain adaptation techniques and vegetation indices (VIs) derived from Sentinel-2 imagery for early-stage winter wheat yield forecasting, achieving an average RMSE of 604.48 kg/ha. Campoy et al. (2023) [29] proposed the MYRS model, which integrates time series of satellite multispectral imagery and meteorological data into classical approaches for estimating key variables related to yield components, demonstrating good performance in reproducing final yield at field and within-field scales for wheat and barley crops under various conditions. These studies highlight the importance of utilizing specific RS data sources and techniques for targeted applications in precision agriculture.

Despite the significant advancements in crop yield estimation using RS data and ML, challenges such as the presence of noise, outliers, and the need for effective pre-processing techniques remain prevalent. Our research aims to address these challenges by introducing the Quartile Clean Image technique, a novel pre-processing method designed to enhance the quality of RS data for crop yield estimation. By comparing our approach with state-of-the-art models like the ViT, we contribute to the ongoing efforts in developing accurate and reliable crop yield estimation methods, providing valuable insights for future research in this domain.

3 Dataset Collection and Preparation

This section outlines the process of compiling the dataset and sources utilized for evaluating the Quartile Clean pre-processing technique through the yield estimation of soybean and corn crops.

3.1 Dataset Description

Using the Earth Engine's computational capability, we used MODIS products MOD09A1 v6 for Surface Reflectance, MOD11A2 v6 for Land Surface Temperature, and MCD12Q1 for Land Cover type. Covering a span of 14 years, from 2002 to 2015, the dataset includes multispectral RS data for 31 states in the United States. In order to provide ground truth data corresponding to the satellite imagery, bushels per acre of soybean and corn yield data for the aforementioned 31 states were obtained from the United States Department of Agriculture (USDA).

To compile a diverse multispectral RS dataset, we put to use the capabilities of the MODIS instrument aboard NASA's Terra and Aqua satellites. In particular, we utilized the MOD09A1 v6 product to collect surface reflectance data across seven different wavelength bands, namely visible light (Band 1: 620-670 nm) to shortwave infrared (Band 7: 2105-2155 nm). This procedure produced a robust dataset of 644 images per year over a period of 14 years, resulting in a vast collection of 1,833 image sets covering 31 U.S. states with resolution as precise as 500m for each state.

To further enrich our dataset, land surface temperature data was incorporated from the same image collection using the MOD11A2 v6 MODIS product, which contains two temperature bands. As a result, a dataset with the structure 50 * 50 * 1288 was produced. Furthermore, using the MCD12Q1 product, we focused on pixels recognized as cultivated cropland to extract land cover information. By only including pixels where a minimum of 60% of the area was classified as farmed cropland, we created a 50 x 50 x 14 land cover mask dataset to suit the needs of agricultural applications. Along with their spectral and spatial resolutions, these collated datasets provide a robust foundation to evaluate the Quartile Clean pre-processing technique. Table 2 summarizes the compiled datasets in detail, outlining their key characteristics.

Table 2. Summary of Collected MODIS Satellite Image Data

	Surface Reflectance	**Land Surface Temperature**	**Land Cover (Mask Image)**
Source	MOD09A1.006 Terra Surface Reflectance	MOD11A2.006 Terra Land Surface Temperature	MCD12Q1.006 MODIS Land Cover Type
Bands	Seven – 0, 1, 2, 3, 4, 5, and 6	Two – 0 and 4	One – 0
Image collection	644	644	14
Image size	50 * 50 * 4508	50 * 50 * 1288	50 * 50 * 14
Total count of Images to cover the area of all 31 states	1833	1833	1833

3.2 Data Processing

After compiling the complete 14-year image dataset, the first step in data preparation involved separating it into 14 distinct annual different subsets comprising temperature, mask, and surface reflectance images. This reduced the dimensionality to 50 x 50 x 92, 50 x 50 x 1, and 50 x 50 x 322 respectively. Merging these divided images into a single, unified dataset for analysis was the next crucial step. By combining the annual temperature and reflectance images, a 14 x 50 x 50 x 414 multi-dimensional array was produced. The dataset was subsequently enhanced by adding important land cover data with the crop mask data.

In order to tackle the limitations imposed by sparse data and redundant features, we utilized a dimensionality reduction method that was proposed by You et al. [1]. This method entailed reducing each image to its distinct 32 * 32 pixel values. Essentially, before reduction, the Quartile Clean Image method was applied to refine data quality by adjusting outlier pixels. Images can be converted into a histogram of pixel counts without losing any information thanks to this technique [1], which capitalizes on pixel variety rather than raw high-dimensionality. The final dataset was structured as a compressed NumPy array histogram with embedded metadata including year, location, and yield values. Our study's Figure 2 explains this data transformation journey, showing how it progresses from the first step of data gathering towards the last phase of data processing.

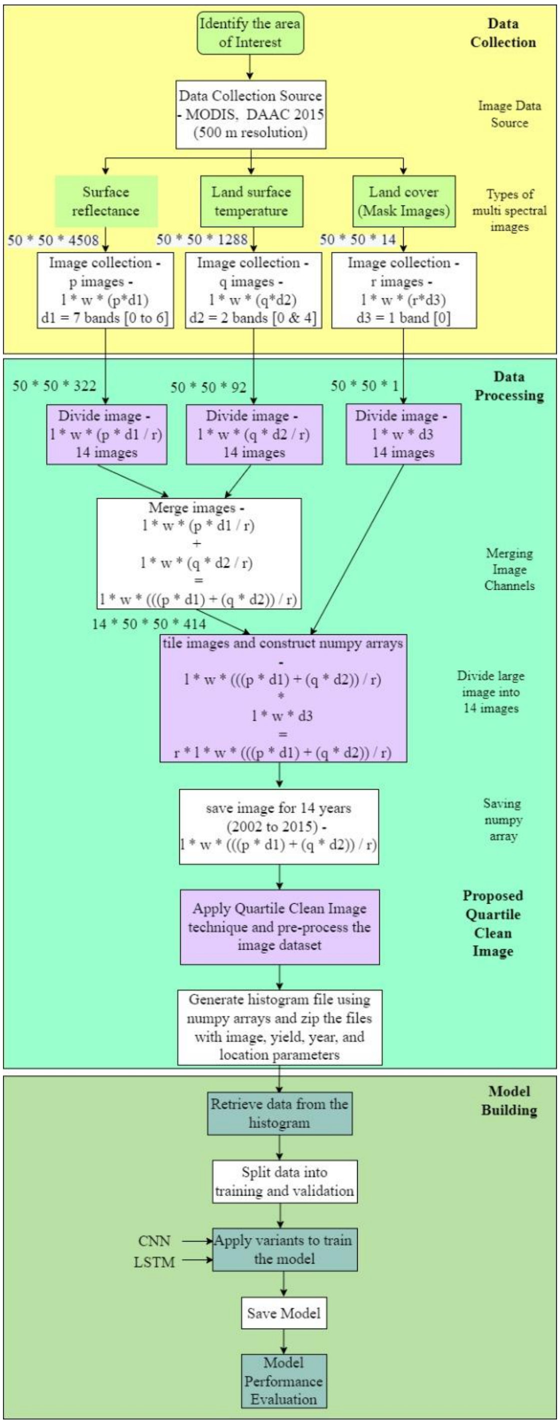

Fig. 2. Overall flow of MODIS data collection, data processing, and model building

3.3 Dataset Splitting

After finalizing the zipped NumPy histogram file containing image, location, year, and yield data, the dataset was split for model training and validation. Separate data for each year from 2011 to 2015 for soybean and corn crops was kept as reserved for validation purpose, while data of all preceding years from the year in scrutiny for validation was utilized for training purpose, ranging from 2003 up to one year prior. This temporal splitting approach enables rigorous validation of model performance on unseen recent data, while allowing robust training on a wide range of historical cycles and conditions for each crop. Tables 3 and 4 detail the distribution of these training and validation data points, ensuring a comprehensive evaluation of the model's performance.

Table 3. Number of Data Points for Training and Validation for Soybean

Validation Year	Training Samples Count	Validation Samples Count
2011	6321	770
2012	7091	754
2013	7845	712
2014	8557	726
2015	9283	667

Table 4. Number of Data Points for Training and Validation for Corn

Validation Year	Training Samples Count	Validation Samples Count
2011	6252	735
2012	6987	732
2013	7719	662
2014	8381	686
2015	9067	635

4 Architectural Overview and Methodology

This section describes the architectures and approaches used, including their essential components and implementation details. We delineate the structure of DL architectures - CNN and LSTM networks - as reconstructed based on the methodology developed by You et al. [1].

4.1 Existing CNN and LSTM architectures

The CNN model includes a three-block architecture, where each block contains convolutional layers for feature extraction using kernel filters. Block 1 has two convolutional layers with 128 kernels each. Block 2 also has two convolutional layers but with 256 kernels in each. Block 3 increases complexity with three convolutional layers, each utilizing 512 kernels. The convolution operation in each layer, defined mathematically in Equation 1, is crucial for feature extraction. With this structured three-block design, the model can extract low-level features in initial blocks before identifying higher-level features in subsequent blocks through the application of more complex kernel filters.

$$O(o_h, o_w) = \sum_j \sum_k k[j, k].\ I[o_h - j, o_w - k] \quad (1)$$

where the kernel matrix k traverses the input matrix I to produce the output matrix O. The size (height oh and width ow) of extracted output feature matrix is:

$$o_h = \frac{m_h - k_h + 2p}{s} + 1$$
$$o_w = \frac{m_w - k_w + 2p}{s} + 1$$

This output matrix is further refined by the Rectified Linear Unit (ReLU) activation function, generating the activated output as shown in Equation (2).

$$\hat{O} = f\big(O(o_h, o_w)\big) = \max\big(0, O(o_h, o_w)\big) \quad (2)$$

Following each convolutional layer, the architecture incorporates a batch normalization process to stabilize learning by normalizing layer inputs, as depicted in Equation (3). This is paired with a dropout layer set at a rate of 0.25 to prevent overfitting by randomly ignoring a subset of neurons during training.

$$\overline{O_i} = \frac{\widehat{O_i} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (3)$$

Parallelly, the LSTM model, a variant of Recurrent Neural Networks (RNN), is adopted for its proficiency in processing sequences, making it suitable for time-series prediction like crop yield estimation. Our LSTM architecture consists of four blocks with varying filter sizes: 16, 32, 64, and 128. Each block is equipped with a 3x3 kernel and is sequentially followed by a max pooling layer to reduce spatial dimensions, and a batch normalization layer to ensure efficient training dynamics.

4.2 Simulation Setup for CNN and LSTM

The experiments were carried out over Google Colab Pro with GPU specifications as 1xTesla K80 (4692 cores), P100 (3584 cores), T4 (2560 Tensor Cores), and 32GB GDDR5 vRAM with 2 x vCPU. The CNN and LSTM have been implemented by setting up the hyperparameters listed below in Table 5:

Table 5. Fine-Tuned Parameters in Implementation of CNN and LSTM

Parameter	Value
Optimizer	Adam
Activation (Hidden Layer)	Relu
Activation (Output Layer)	Linear
Loss function	Mean Squared Error (mse)
Epochs	200
Batch Size	12
Dropout (In CNN)	0.25
Batch Normalization	True
For LR Scheduler – Exponential Decay	
initial_learning_rate	0.005
decay_steps	100000
decay_rate	0.96

For LR Scheduler – Cosine Decay	
initial_learning_rate	1e-3
first_decay_steps	1000
t_mul	2.0
m_mul	1.0
alpha	0.0

4.3 Vision Transformer (ViT) Architecture

ViT is a neural network architecture that applies the transformer architecture to image classification tasks. In this study, we adapt ViT for crop yield estimation by modifying the transformer-based classifier to perform regression.

The ViT architecture consists of two main components: a feature extractor and a transformer-based classifier. The feature extractor converts an input image x with width W and height H into a sequence of patches X = [x1, x2, ..., xN], where each patch xi represents a local region of size p × p, and N is the number of patches given by:

$$N = \frac{W * H}{p^2}$$

The transformer-based classifier then processes the sequence of patches X to estimate the crop yield. It consists of multiple transformer blocks, each incorporating multi-head self-attention and feed-forward layers. The multi-head self-attention layer computes the weighted sum of the patches based on their similarity:

$$Z = softmax\left(\frac{\left(X_q * W_q\right) * \ (X_k * W_k)^T}{\sqrt{d}}\right) * (X_v * W_v)$$

where X_q, X_k, and X_v are the query, key, and value matrices, W_q, W_k, and W_v are learnable weight matrices, d is the dimensionality of the query and key vectors, and softmax normalizes the attention scores across all patches.

The feed-forward layer applies a non-linear transformation to the output of the multi-head self-attention layer:

$$F = max(0, Z * W1 + b1) * W2 + b2$$

where W1 and W2 are learnable weight matrices, b1 and b2 are learnable bias vectors, and max(0, x) is the ReLU activation function.

ViT stacks multiple transformer blocks to form the classifier, and the output of the final transformer block is passed through a linear layer to predict the crop yield.

4.4 Quartile Clean Image: Data Pre-processing Technique

The architectural overview of the Quartile Clean Image data pre-processing technique is specially engineered to boost the quality of high-resolution RS images, notably those used in agricultural yield estimation. This technique is crucial for reducing the effects of environmental factors that may distort image data, such as climate conditions, cloud cover, and shadows, which feature commonly in datasets collected from high altitudes.

In essence, the Quartile Clean Image approach works on the idea that related objects in an image cluster around the median value of their nearby pixels, with some fluctuation. We use this understanding to identify and handle outlier pixel values that may be caused due to noise or data capture errors. The process begins with the identification of the quartile values (q1, q2, and q3) for each pixel, taking into account its neighbouring pixels within a specified window. This establishes a local pixel value threshold for considering a 'normal' pixel value, becoming a yard-stick to correctly determine whether a specific pixel is normal or an outlier. Any pixel value that does not fall within the range of q1 to q3 is flagged as an outlier. Once these outliers are detected, the Quartile Clean Image technique replaces these values with a median value (q2) of the identified range. This intentional replacement removes anomalous values that could possibly hinder analysis while preserving image integrity.

An illustrative representation of this pre-processing step is provided in Figure 3. The figure showcases a zero-padding applied to the edges of the actual image pixel values, preparing them for the quartile-based cleaning process. The zero-padding ensures that edge pixels receive a neighbourhood of values to facilitate quartile calculations, maintaining consistency across the image. After this initial phase, quartile ranges are generated and outliers are converted to the median value with a view to creating a pre-processed image with optimum noise reduction for a better portrayal of the visual. This architectural approach allows for subsequent image analysis phases, such as feature extraction and pattern recognition, to operate on data that more accurately reflects the actual conditions of the observed area, thereby helping to considerably enhance the performance of CNN and LSTM models in estimating the crop yields accurately.

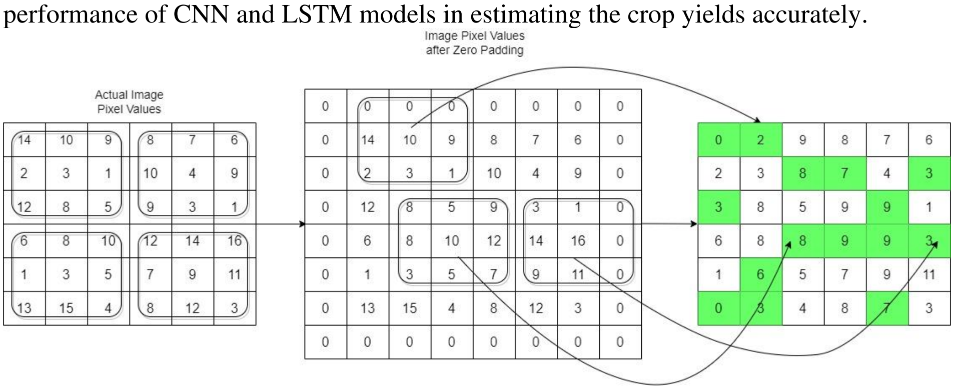

Fig. 3. Image Pre-process: Quartile Clean Image Data

4.5 Pseudo code for Quartile Clean Image

The pseudo code for generating the quartile clean image is as below. The Zero-Padding function serves as an initial step, preparing the image for quartile analysis. It accepts an image and a pad width k as inputs, creating zero matrices (z0 and z1) to pad the original image (Ii). The PSNR function is used to comparatively assess the quality of the processed image with the original. The function calculates the mean squared error (MSE) between the original and processed images and then uses the obtained value to calculate the PSNR. A higher PSNR indicates better quality, with 100 denoting no noise. The core of the technique is encapsulated in the QuartileClean function, which iterates over the image collection. For each image Ii, it applies zero padding and initializes a matrix for the cleaned image. It then proceeds to process each channel of the image, applying a moving window based on the pool size (pj, pk) and stride (s) to compute the quartile values for each pixel neighborhood. If the pixel value is within the interquartile range (q1 to q3), it is retained; otherwise, it is replaced with the median value (q2). This selective filtering effectively reduces noise and corrects outliers in the image data.

Function ZeroPadding(Image I_i , pad_width):

Parameter:
padding = pad_width = k
Return:
Array after padding with zeros: padded_image
begin function
1. Generate zeros matrix z0 of shape (k, I_i.shape[1], I_i.shape[2])
2. Generate zeros matrix z1 of shape (I_i.shape[0] + k * 2, k, I_i.shape[2])
3. Concatenate I_i, z0, and z1 to generate padded matrix as:
   padded_image = Concatenate((z1, concatenate((z0, I_i, z0), axis = 0), z1), axis=1)

end function

Function PSNR(Image I_i, Image $\widehat{I_i}$):

Parameter:
max_pixel = 255.0
Return:
psnr
begin function
1. Compute mean squared error(mse): mean((Image Ii − Image $\widehat{I_i}$)^2)
2. **if** mse == 0 **then**
3. return 100 #No noise present in the signal
4. psnr = 20 * ($log_{10}(\frac{\text{max\_pixel}}{\sqrt{mse}})$)

end function

Function QuartileClean(Image_collection):

Parameters:
pool_size = (p_h, p_w)
stride = s
Returns:
Quartile Clean Image ($\widehat{I_i}$)
PSNR(I_i , $\widehat{I_i}$)

```
begin function
  for I_i in image_collection do
     I_i_zero = ZeroPadding(I_i, pad_width)
     Î_i = zeros_matrix(I_i.shape[0] – p_h + 1, I_i.shape[1] – p_w+1, I_i.shape[2])
     for each channel m in I_i do
       X1 = I_i[:, :, m]
       for n in range (0, Î_i .shape[0]) do
          for o in range (0, Î_i .shape[1]) do
                compute Quartile values q1, q2, q3 from X1[i:i+p_h, j:j+p_w]:
                if q1 < X1[n][o] < q3 then
                   Î_i [n][o][m] = X1[n][o]
                else:
                   Î_i [n][o][m] = q2
                end if
          end for
       end for
     end for
  end for
end function
```

4.6 Assessment of Quartile Clean Image on Metric PSNR

To evaluate the Quartile Clean Image technique, we assessed 20,946 MODIS satellite images from 2003-2015 using PSNR. PSNR quantifies image quality, with typical values of 30-50 dB for 8-bit images; higher PSNR indicates lower noise. The mean PSNR post-Quartile Clean processing was 40.91 dB, demonstrating substantial noise reduction. The median PSNR was 33.76 dB, further evidencing improved quality. Figure 4 shows the PSNR distribution, concentrated favourably within the target range and peaked at frequently observed values. This analysis verifies the Quartile Clean technique successfully enhances MODIS image clarity by suppressing noise and distortions, as quantified by the elevated PSNR metrics within optimal bounds across a large satellite image dataset.

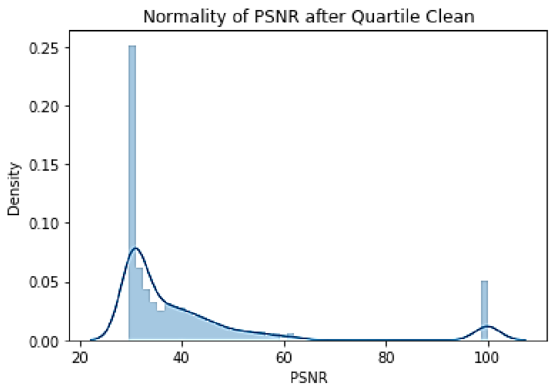

Fig. 4. Distribution Plot for PSNR Values

5 Results Analysis

A comparison of the CNN and LSTM models in estimation of soybean and corn crop yields is presented in Tables 6 and 7, with an emphasis on important metrics including R^2 score, adjusted R^2 score, RMSE, and RMSLE. The well-established model architectures demonstrate substantially improved accuracy on these metrics when using the pre-processed imagery.

5.1 Assessment of CNN and LSTM on Original Data

1. Performance Analysis on Metrics R^2 Score and Adjusted R^2 Score:

The R^2 score and adjusted R^2 score are essential metrics for evaluating the performance of regression models, particularly in the context of high noise levels that may lead to overfitting. In crop yield estimation, an R^2 score closer to 1 indicates a better fit, while a lower score suggests that the model may not be capturing the variability in the data adequately. Our analysis applies these metrics to assess CNN and LSTM models' ability to estimate soybean and corn crop yields.

For soybeans, the CNN model with an exponential learning rate (LR) scheduler exhibits R^2 scores ranging from 0.5023 to 0.8352 (Table 6), indicating that the model explains between 50.23% and 83.52% of the variance in the data. The LSTM model for soybeans shows slightly better performance, with R^2 scores spanning from 0.6989 to 0.8649. These results suggest that both models capture a significant portion of the variability in soybean yield, with the LSTM model having a slight edge over the CNN model.

In the case of corn, the CNN model with different LR schedulers achieves R^2 scores ranging from 0.6779 to 0.8893 (Table 7). However, the negative R^2 score of -2.107 in the validation year 2013 (exponential LR) indicates severe overfitting, suggesting that the model may be learning noise in the data rather than the underlying patterns.

The adjusted R^2 score, which considers the number of predictors, highlights potential overestimation in some validation years. For soybeans, the adjusted R^2 scores exceed 1 in certain instances, such as 2.4996 for the CNN model without pre-processing in the validation year 2015 (Table 6). These values greater than 1 are mathematically impossible for the adjusted R^2 score and indicate the presence of outlier variables that heavily influence the model's performance. Moreover, for corn, negative and undefined (nan) adjusted R^2 values are observed in several validation years for both CNN and LSTM models (Table 7). The presence of negative R^2 and Adjusted R^2 values, particularly for corn yield prediction, raises concern about the reliability of the model's performance and the potential for overfitting. These anomalies suggest that the CNN and LSTM models, when trained on the original data without pre-processing, may be learning noise and outliers rather than the underlying patterns in the data. This highlights the need for effective data pre-processing techniques to mitigate these issues and improve the model's generalization ability.

2. Performance Analysis on Parameters RMSE and RMSLE:

RMSE and RMSLE are widely used metrics for evaluating the performance of regression models, particularly in the context of crop yield prediction. RMSE measures the average magnitude of the errors between the predicted and actual values, providing an intuitive interpretation of the model's accuracy in the original units of the target variable. On the other hand, RMSLE is more sensitive to relative errors and is less affected by outliers compared to RMSE, making it a useful metric when the dataset contains extreme values or when the focus is on relative performance.

In our study, we examine the RMSE values across CNN and LSTM models for both soybean and corn yield prediction. For soybeans, the RMSE values range from 7.9862 to 11.3344 for the CNN model and from 7.47865 to 9.12522 for the LSTM model (Table 6). The LSTM model exhibits slightly better performance, with its RMSE values concentrated in a lower range compared to the CNN model.

However, for corn yield prediction, the RMSE ranges are noticeably wider, spanning from 23.94368 to 41.17015 (Table 7). These higher RMSE values suggest that the models face greater challenges in accurately predicting corn yields compared to soybeans. The wider ranges also indicate the models' sensitivity to outliers in the corn dataset, which may be attributed to the presence of extreme values or greater variability in the data.

To gain further insights into the models' performance, we also consider the RMSLE metric, which is particularly relevant when outliers significantly affect RMSE. Our findings show that CNN models with exponential decay schedulers produced RMSLE values ranging from 0.128 to 0.3172 for corn, while LSTM models with cosine decay schedulers resulted in RMSLE values from 0.1378 to 0.3425 (Table 7). These RMSLE values provide a different perspective on the models' performance, focusing on the relative errors rather than the absolute differences.

To improve the models' performance, particularly for corn yield prediction, several strategies can be explored. These include data pre-processing techniques to handle outliers and normalize the data, feature engineering to create more informative input variables, and model regularization methods to mitigate overfitting.

5.2 Assessment of CNN and LSTM with Quartile Clean RS Data

The integration of the Quartile Clean Image pre-processing technique in our comparative analysis of CNN and LSTM models for soybean and corn yield prediction has yielded remarkable improvements in model performance. This technique effectively addresses the presence of outliers and noise in the remote sensing data, which can adversely affect the accuracy of crop yield predictions. By applying the Quartile Clean Image method, we have observed consistent enhancements in the models' predictive capabilities, as evidenced by the significant reductions in RMSE values across different validation years and learning rate schedulers (Tables 6 and 7).

A detailed comparative analysis of CNN and LSTM model performance for soybean and corn yield prediction was conducted, integrating the Quartile Clean Image pre-processing technique. As shown in Tables 6 and 7, CNN models paired with exponential decay LR scheduling and Quartile Clean data achieved RMSE improvements up to

5.88% for soybeans. Similar enhancements up to 8.54% were attained using cosine decay LR scheduling. For LSTM models, sizable RMSE reductions ranged from 1.75% to 8.34% with exponential decay LR, and up to 11.52% with cosine decay LR, on soybeans. For corn yield prediction, CNN models exhibited RMSE decreases up to 21.85% using exponential decay LR. LSTM models demonstrated an even wider RMSE improvement range from 29.86%. Cosine decay LR scheduling also led to RMSE boosts from 1.02% to 4.83% in CNNs and 0.85% to 16.40% in LSTMs.

For soybean yield prediction, the application of Quartile Clean data in conjunction with CNN models and exponential decay LR scheduling has resulted in RMSE improvements ranging from 3.38% to 19.28%. These improvements are particularly notable in validation year such as 2015, where the RMSE values decreased from 11.05045 to 10.67621. Similarly, when using cosine decay LR scheduling, the CNN models exhibited RMSE reductions between 2.19% and 5.92%, with the most significant improvement observed in validation year 2015, where the RMSE decreased from 8.79026 to 8.27365 (5.92% improvement).

The impact of Quartile Clean data on LSTM models for soybean yield prediction is equally impressive. With exponential decay LR scheduling, the LSTM models achieved RMSE reductions ranging from 2.23% to 8.34%, with the highest improvement seen in validation year 2014, where the RMSE decreased from 8.91761 to 8.17308 (8.34% improvement). When using cosine decay LR scheduling, the LSTM models demonstrated even more substantial RMSE improvements, ranging from 1.75% to 11.52%. The most notable enhancement was observed in validation year 2014, where the RMSE reduced from 7.68459 to 6.79858 (11.52% improvement).

Regarding corn yield prediction, the integration of Quartile Clean data has led to even more pronounced RMSE reductions. CNN models with exponential decay LR scheduling achieved RMSE improvements ranging from 2.17% to 21.85%, with the most significant decrease observed in validation year 2013, where the RMSE dropped from 33.3056 to 26.027 (21.85% improvement). Similarly, LSTM models with exponential decay LR showcased RMSE reductions between 0.97% and 29.92%, with the highest improvement seen in validation year 2014, where the RMSE decreased from 29.03402 to 20.36222 (29.92% improvement).

The application of cosine decay LR scheduling in combination with Quartile Clean data also yielded substantial RMSE improvements for both CNN and LSTM models in corn yield prediction. CNN models exhibited RMSE reductions ranging from 1.02% to 4.83%, with the most significant improvement observed in validation year 2012, where the RMSE decreased from 38.86628 to 36.98624 (4.83% improvement). LSTM models, on the other hand, demonstrated RMSE improvements between 0.85% and 13.17%, with the highest reduction seen in validation year 2013, where the RMSE dropped from 30.80472 to 26.70376 (13.31% improvement).

The consistency of the RMSE improvements across different validation years and learning rate schedulers highlights the robustness and effectiveness of the Quartile Clean Image pre-processing technique. By removing outliers and noise from the RS data, this method enables the CNN and LSTM models to capture the underlying patterns and relationships more accurately, leading to enhanced crop yield predictions.

Despite the significant improvements in RMSE values after applying the Quartile Clean Image pre-processing technique, some instances of negative R^2 and Adjusted R^2 values persist, particularly for corn yield prediction in certain validation years (Table 7). While these negative values are less prevalent compared to the models trained on the original data, they still warrant further investigation. Potential strategies to address this issue include exploring additional data pre-processing techniques, feature selection and engineering, regularization methods, and ensemble approaches to enhance the model's robustness and generalization ability.

5.3 Comparison with State-of-the-Art: Vision Transformer

The ViT model, a state-of-the-art architecture for image classification tasks, has been adapted for crop yield estimation in this study. By modifying the transformer-based classifier to perform regression, we compare the performance of the ViT model with the proposed Quartile Clean Image method and assess its effectiveness in estimating soybean and corn yields without explicit pre-processing.

1. **Performance Analysis on Metrics R^2 Score and Adjusted R^2 Score:**

For soybean yield estimation, the ViT model with the Adabelief optimizer achieves remarkable R^2 scores ranging from 0.9752 to 0.9875 (Table 6), indicating that the model explains between 97.52% and 98.75% of the variance in the data. Similarly, the ViT model with the Adam optimizer yields R^2 scores between 0.9754 and 0.9881, demonstrating its ability to capture a substantial portion of the variability in soybean yield without the need for explicit pre-processing.

In the case of corn yield estimation, the ViT model with the Adabelief optimizer attains R^2 scores ranging from 0.954 to 0.9888 (Table 7), while the Adam optimizer achieves R^2 scores between 0.954 and 0.9887. These high R^2 scores highlight the ViT model's robustness in estimating corn yields, even in the presence of potential noise and outliers in the remote sensing data.

The adjusted R^2 scores for the ViT model in both soybean and corn yield estimation are consistently high and within the expected range of 0 to 1. This indicates that the model's performance is not significantly influenced by the number of predictors, and the high R^2 scores are not a result of overfitting.

2. **Performance Analysis on Metrics RMSE and RMSLE:**

For soybean yield estimation, the ViT model with both Adabelief and Adam optimizers achieves RMSE values ranging from 7.75086 to 9.76838 (Table 6). These RMSE values are indeed lower than those obtained by CNN and LSTM models without the Quartile Clean Image pre-processing technique. However, when compared to CNN and LSTM models with the Quartile Clean Image pre-processing, the ViT model's RMSE values are generally higher. For example, the LSTM model with Quartile Clean Image data and exponential decay LR achieves RMSE values as low as 7.31592, which is slightly better than the ViT model's best RMSE of 7.75086.

In the case of corn yield estimation, the ViT model with both Adabelief and Adam optimizers achieves RMSE values ranging from 26.24478 to 39.47279 (Table 7). These RMSE values are comparable to or slightly lower than those obtained by CNN and

LSTM models without the Quartile Clean Image pre-processing. However, when compared to CNN and LSTM models with the Quartile Clean Image pre-processing, the ViT model's RMSE values are generally higher. For instance, the LSTM model with Quartile Clean Image data and exponential decay LR achieves an RMSE value of 20.36222, which is considerably lower than the ViT model's best RMSE of 26.24478.

The RMSLE values for the ViT model in soybean yield estimation range from 0.1797 to 0.2461 (Table 6), while for corn yield estimation, they range from 0.1735 to 0.3885 (Table 7). These RMSLE values are generally comparable to those obtained by CNN and LSTM models, both with and without the Quartile Clean Image pre-processing technique.

Upon reviewing the values in Tables 6 and 7, it is evident that the ViT model's performance in terms of RMSE is not consistently superior to CNN and LSTM models, particularly when compared to those with the Quartile Clean Image pre-processing technique applied. However, the ViT model still demonstrates competitive performance, especially considering that it does not require explicit pre-processing of the remote sensing data.

The ViT model's ability to capture long-range dependencies and spatial relationships through the self-attention mechanism remains a notable advantage, enabling it to extract meaningful features and patterns for crop yield estimation. While the ViT model's performance may not be universally superior to CNN and LSTM models with pre-processing, its competitive results without the need for explicit pre-processing highlight its potential as a powerful tool for precision agriculture applications.

Figures 5-8 visually depict the RMSE and R^2 score improvements pre- and post-Quartile Clean application for both crops.

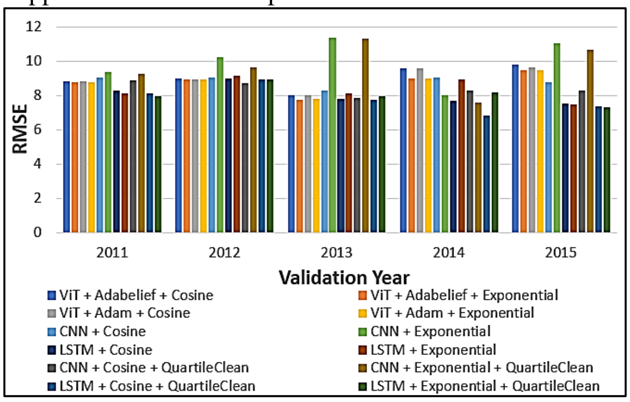

Fig. 5. Comparative Analysis of CNN and LSTM before and after Applying Quartile Clean Pre-Processing along with SOTA ViT on Metric RMSE for Soybean

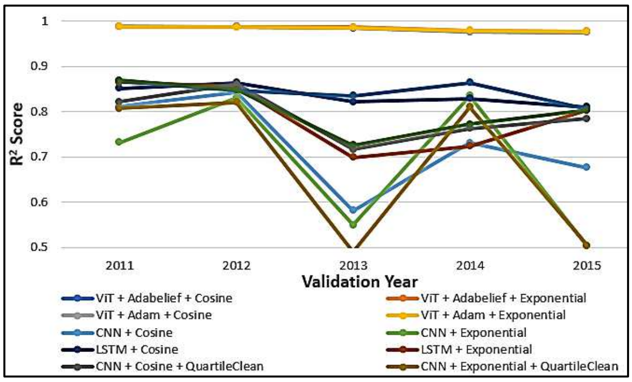

Fig. 6. Comparative Analysis of CNN and LSTM before and after Applying Quartile Clean Pre-Processing along with SOTA ViT on Metric R^2 Score for Soybean

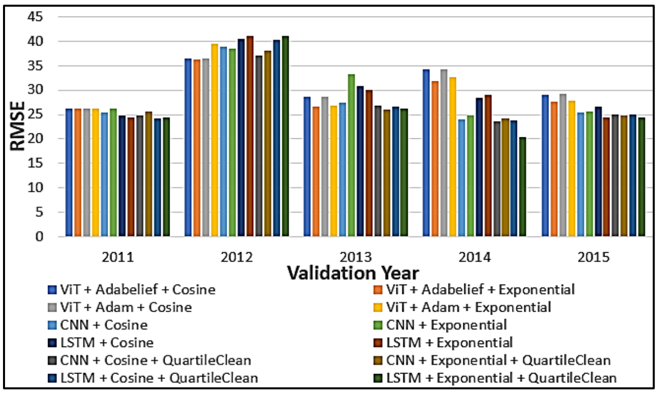

Fig. 7. Comparative Analysis of CNN and LSTM before and after Applying Quartile Clean Pre-Processing along with SOTA ViT on Metric RMSE for Corn

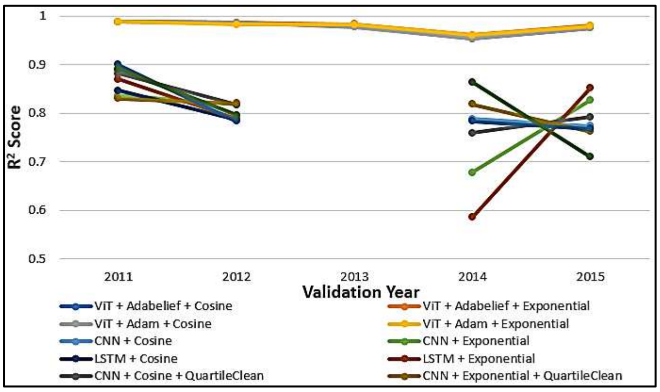

Fig. 8. Comparative Analysis of CNN and LSTM before and after Applying Quartile Clean Pre-Processing along with SOTA ViT on Metric R^2 Score for Corn*

***NOTE:** Negative R^2 scores have been omitted from the figure for clarity. The presence of negative R^2 scores in certain validation years for CNN and LSTM models, particularly without pre-processing, indicates poor model performance and potential overfitting. These negative values suggest that the models may be learning noise and outliers rather than the underlying patterns in the data.

Table 6. Assessment of CNN and LSTM model before and after Pre-processing using Quartile Clean Technique for Soybean Crop Yield

Algorithm	**Validation Year**	**RMSE**		**RMSLE**		**R^2 Score**		**Adjusted R^2 Score**	
		Cosine Decay LR	Exponential LR	Cosine Decay LR	Exponential LR	Cosine Decay LR	Exponential LR	Cosine Decay LR	Exponential LR
CNN with Quartile Clean Image data	2011	8.88736	9.26102	0.1834	0.1927	0.8218	0.8068	0.8595	0.7921
	2012	8.72154	9.61863	0.1972	0.2241	0.8609	0.8205	0.8155	0.8276
	2013	7.83369	11.32481	0.1469	0.2294	0.7158	0.4898	18.2683	9.4752
	2014	8.28044	7.59598	0.1485	0.1358	0.7629	0.8104	0.7527	0.7664
	2015	8.27365	10.67621	0.1425	0.1897	0.7847	0.5034	0.7882	2.1587
LSTM with Quartile Clean Image data	2011	8.1047	7.9375	0.1634	0.1598	0.868	0.8687	0.8029	0.761
	2012	8.92371	8.90538	0.1915	0.1949	0.8469	0.8483	0.9287	0.8552
	2013	7.74811	7.97643	0.1466	0.1515	0.8345	0.7257	1.1641	11.2351
	2014	6.79858	8.17308	0.1144	0.1448	0.8641	0.772	0.8334	0.824
	2015	7.38282	7.31592	0.1249	0.1239	0.805	0.8032	0.9024	0.9261
CNN	2011	9.0229	9.33824	0.1929	0.1907	0.8112	0.7321	0.8046	0.7077
	2012	9.02973	10.21969	0.2007	0.2275	0.8426	0.8307	0.8636	0.7513
	2013	8.25432	11.3344	0.1573	0.1776	0.5813	0.5491	48.7197	26.2899
	2014	9.05366	7.9862	0.1629	0.1397	0.7307	**0.8352**	0.7192	0.8096
	2015	8.79026	11.05045	0.1407	0.1971	0.6763	**0.5023**	1.3919	2.4996
LSTM	2011	8.2714	8.11766	0.1672	0.1631	0.8516	**0.8649**	0.8006	0.7273
	2012	8.97318	9.12522	0.1907	0.2004	0.8641	0.856	0.8449	0.8462
	2013	7.80117	8.11917	0.1433	0.1549	0.8219	**0.6989**	2.9425	11.0177
	2014	7.68459	8.91761	0.1336	0.1609	0.8282	0.7235	0.8073	0.7942
	2015	7.51185	7.47865	0.1253	0.1252	0.8101	0.8032	0.936	0.9488
ViT + Adabelief	2011	8.81379	8.77692	0.2245	0.2237	0.987	0.9875	0.9746	0.9757
	2012	8.98373	8.9441	0.2461	0.2451	0.9873	0.9875	0.974	0.9745
	2013	8.02821	7.75086	0.1861	0.1797	0.9845	0.9865	0.9682	0.9723
	2014	9.54976	9.00565	0.2068	0.1943	0.9766	0.9787	0.9525	0.9568
	2015	9.76838	9.46421	0.2036	0.1968	0.9752	0.9771	0.9498	0.9535
ViT + Adam	2011	8.81552	8.77641	0.2245	0.2237	0.9881	0.9871	0.9767	0.9749
	2012	8.94466	8.95077	0.2451	0.2453	0.9878	0.9869	0.9751	0.9732
	2013	8.02762	7.81578	0.1861	0.1812	0.9854	0.9859	0.9701	0.9711
	2014	9.55191	8.95676	0.2069	0.1932	0.9766	0.9796	0.9525	0.9584
	2015	9.62273	9.48338	0.2003	0.1972	0.9754	0.977	0.9502	0.9533

Table 7. Assessment of CNN and LSTM model before and after Pre-processing using Quartile Clean Technique for Corn Crop Yield

Algorithm	Validation Year	RMSE		RMSLE		R^2 Score		Adjusted R^2 Score	
		Cosine Decay LR	Exponential LR	Cosine Decay LR	Exponential LR	Cosine Decay LR	Exponential LR	Cosine Decay LR	Exponential LR
CNN with Quartile Clean Image data	2011	24.81149	25.53795	0.1555	0.1604	0.8818	0.8301	nan	nan
	2012	36.98624	38.13069	0.3175	0.317	0.8176	0.8202	0.785	0.7737
	2013	26.80136	26.027	0.1489	0.1474	-1.863	-2.5691	-2551.66	-4962.4619
	2014	23.69743	24.11044	0.117	0.1175	0.7593	0.8191	0.0671	0.5403
	2015	25.10863	24.73492	0.1277	0.126	0.792	0.7624	0.9007	1.0496
LSTM with Quartile Clean Image data	2011	24.13071	24.31677	0.1502	0.1498	0.8998	0.8912	nan	nan
	2012	40.22044	41.10841	0.3407	0.3343	0.7839	0.7959	0.7765	0.7441
	2013	26.70376	26.17598	0.1505	0.1426	-0.514	-0.6165	-1005.21	-1404.2379
	2014	23.8334	20.36222	0.1175	0.098	0.7839	0.8645	0.4176	-0.1868
	2015	24.95029	24.39694	0.1287	0.1234	0.768	0.7094	1.0865	1.7904
CNN	2011	25.3862	26.24564	0.1535	0.1649	**0.8893**	0.8369	nan	nan
	2012	38.86628	38.51291	0.3357	0.3172	0.7961	0.8174	0.8316	0.7883
	2013	27.45831	33.3056	0.1456	0.1675	-0.286	-2.107	-770.265	-3032.9072
	2014	23.94368	24.78605	0.1178	0.128	0.7882	**0.6779**	-0.4873	-6.4438
	2015	25.4786	25.66778	0.1287	0.1289	0.7735	0.8278	0.9466	0.8392
LSTM	2011	24.74114	24.50401	0.1613	0.1521	0.8469	0.8704	nan	nan
	2012	40.56584	41.17015	0.3425	0.3456	0.7869	0.7882	0.7723	0.7692
	2013	30.80472	30.04143	0.1789	0.1766	-1.330	-1.602	-2048.43	-3092.8167
	2014	28.51151	29.03402	0.1496	0.1493	0.4371	0.5855	-30.3967	-4.06
	2015	26.57907	24.43664	0.1378	0.1596	0.6633	0.8529	1.7563	nan
ViT + AdaBelief	2011	26.25277	26.25265	0.2155	0.2155	0.9888	**0.9888**	0.977	**0.9771**
	2012	36.40415	36.3189	0.3648	0.3641	0.9869	**0.987**	0.9732	**0.9735**
	2013	28.68751	26.62948	0.1949	0.1807	0.9789	**0.9838**	0.957	**0.9668**
	2014	34.20382	31.88006	0.2127	0.1968	0.954	0.9614	0.9089	0.9226
	2015	29.00538	27.56208	0.1833	0.1735	0.9765	0.9805	0.9525	0.9603
ViT + Adam	2011	26.25276	26.24478	0.2155	0.2154	0.9887	0.9885	0.9768	0.9765
	2012	36.38152	39.47279	0.3646	0.3885	0.9871	0.9826	0.9737	0.9646
	2013	28.57183	26.85423	0.1941	0.1823	0.9782	0.9832	0.9557	0.9656
	2014	34.19587	32.73254	0.2127	0.2026	0.954	0.9618	0.9089	0.9234
	2015	29.2871	27.74993	0.1852	0.1748	0.9763	0.9801	0.952	0.9595

6 Conclusion and Future Work

This research demonstrates the remarkable effectiveness of the Quartile Clean Image technique in processing remote sensing data for accurate crop yield estimation, a critical component in addressing global food security challenges. By pre-processing 14 years of MODIS multispectral imagery covering diverse regions in the United States, the technique significantly reduced noise and distortions inherent in high-altitude data. The integration of Quartile Clean data with CNN and LSTM models led to substantial improvements in soybean and corn yield prediction accuracy, with RMSE reductions ranging from 3.38% to 19.28% for soybeans and 2.17% to 29.92% for corn. These findings highlight the importance of data quality in yield estimation and set a new standard for leveraging remote sensing data in precision agriculture.

The significant improvements in RMSE reduction and R^2 score enhancement achieved by applying the Quartile Clean Image technique as a pre-processing step demonstrate its superiority over methods that do not employ pre-processing. Moreover, the comparison with the state-of-the-art ViT model highlights the importance of the Quartile Clean Image method in enhancing the performance of CNN and LSTM models, even when compared to advanced architectures that can handle raw data effectively. The ViT model, adapted for crop yield estimation in this study, demonstrated remarkable performance in terms of R^2 scores and robustness to negative values, showcasing its potential as a powerful tool for precision agriculture applications. However, the Quartile Clean Image pre-processing technique remains crucial for enhancing the performance of CNN and LSTM models, as evidenced by the lower RMSE values achieved with pre-processed data. This underscores the importance of effective data pre-processing in conjunction with advanced machine learning models for accurate crop yield estimation.

Future research should focus on exploring the scalability and applicability of the Quartile Clean Image technique across different geographical regions, crop types, and remote sensing platforms. Combining this pre-processing method with emerging ML approaches, such as the ViT, presents exciting opportunities for further improving yield prediction accuracy and gaining a deeper understanding of the intricate relationships between crops, environmental factors, and management practices. Moreover, investigating the potential benefits of integrating advanced pre-processing techniques, like the Quartile Clean Image method, with state-of-the-art models like ViT could lead to the development of even more precise and reliable crop yield estimation systems. By building upon the findings of this research and exploring these possibilities, future studies can contribute to the ongoing advancement of precision agriculture and its applications in ensuring global food security.

References

1. J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep Gaussian process for crop yield prediction based on remote sensing data," 31st AAAI Conf. Artif. Intell. AAAI 2017, pp. 4559–4565, 2017, doi: 10.1609/aaai.v31i1.11172.

2. M. Qiao et al., “Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks,” Int. J. Appl. Earth Obs. Geoinf., vol. 102, no. July, p. 102436, 2021, doi: 10.1016/j.jag.2021.102436.
3. A. Tripathi, R. K. Tiwari, and S. P. Tiwari, “A deep learning multi-layer perceptron and remote sensing approach for soil health based crop yield estimation,” Int. J. Appl. Earth Obs. Geoinf., vol. 113, no. August, p. 102959, 2022, doi: 10.1016/j.jag.2022.102959.
4. A. Mateo-Sanchis, J. E. Adsuara, M. Piles, J. Munoz-Mari, A. Perez-Suay, and G. Camps-Valls, “Interpretable Long Short-Term Memory Networks for Crop Yield Estimation,” IEEE Geosci. Remote Sens. Lett., vol. 20, 2023, doi: 10.1109/LGRS.2023.3244064.
5. H. Huang et al., “The Improved Winter Wheat Yield Estimation by Assimilating GLASS LAI Into a Crop Growth Model With the Proposed Bayesian Posterior-Based Ensemble Kalman Filter,” IEEE Trans. Geosci. Remote Sens., vol. 61, no. Mcmc, 2023, doi: 10.1109/TGRS.2023.3259742.
6. D. Paudel et al., “Machine learning for large-scale crop yield forecasting,” Agric. Syst., vol. 187, no. June 2020, p. 103016, 2021, doi: 10.1016/j.agsy.2020.103016.
7. K. Johansen et al., “Predicting Biomass and Yield in a Tomato Phenotyping Experiment Using UAV Imagery and Random Forest,” Front. Artif. Intell., vol. 3, no. May, 2020, doi: 10.3389/frai.2020.00028.
8. J. Fan, J. Bai, Z. Li, A. Ortiz-Bobea, and C. P. Gomes, “A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction,” Proc. 36th AAAI Conf. Artif. Intell. AAAI 2022, vol. 36, pp. 11873–11881, 2022, doi: 10.1609/aaai.v36i11.21444.
9. E. Cheng et al., “Wheat yield estimation using remote sensing data based on machine learning approaches,” Front. Plant Sci., vol. 13, no. December, pp. 1–16, 2022, doi: 10.3389/fpls.2022.1090970.
10. Ma, Y. and Zhang, Z., 2022. A Bayesian domain adversarial neural network for corn yield prediction. IEEE Geoscience and Remote Sensing Letters, 19, pp.1-5.
11. P. Abbaszadeh, K. Gavahi, A. Alipour, P. Deb, and H. Moradkhani, “Bayesian Multi-modeling of Deep Neural Nets for Probabilistic Crop Yield Prediction,” Agric. For. Meteorol., vol. 314, no. December 2021, p. 108773, 2022, doi: 10.1016/j.agrformet.2021.108773.
12. R. Tripathy, K. N. Chaudhari, G. D. Bairagi, O. Pal, R. Das, and B. K. Bhattacharya, “Towards Fine-Scale Yield Prediction of Three Major Crops of India Using Data from Multiple Satellite,” J. Indian Soc. Remote Sens., vol. 50, no. 2, pp. 271–284, 2022, doi: 10.1007/s12524-021-01361-2.
13. W. Zhuo et al., “Crop yield prediction using MODIS LAI, TIGGE weather forecasts and WOFOST model: A case study for winter wheat in Hebei, China during 2009–2013,” Int. J. Appl. Earth Obs. Geoinf., vol. 106, p. 102668, 2022, doi: 10.1016/j.jag.2021.102668.
14. K. Gavahi, P. Abbaszadeh, and H. Moradkhani, “DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting,” Expert Syst. Appl., vol. 184, no. July, p. 115511, 2021, doi: 10.1016/j.eswa.2021.115511.
15. J. Sun, Z. Lai, L. Di, Z. Sun, J. Tao, and Y. Shen, “Multilevel Deep Learning Network for County-Level Corn Yield Estimation in the U.S. Corn Belt,” IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 13, pp. 5048–5060, 2020, doi: 10.1109/JSTARS.2020.3019046.
16. R. J. Donohue, R. A. Lawes, G. Mata, D. Gobbett, and J. Ouzman, “Towards a national, remote-sensing-based model for predicting field-scale crop yield,” F. Crop. Res., vol. 227, no. February, pp. 79–90, 2018, doi: 10.1016/j.fcr.2018.08.005.
17. A. Elders et al., “Estimating crop type and yield of small holder fields in Burkina Faso using multi-day Sentinel-2,” Remote Sens. Appl. Soc. Environ., vol. 27, no. May, p. 100820, 2022, doi: 10.1016/j.rsase.2022.100820.

18. Y. Liu et al., “Rice Yield Prediction and Model Interpretation Based on Satellite and Climatic Indicators Using a Transformer Method,” Remote Sens., vol. 14, no. 19, 2022, doi: 10.3390/rs14195045.
19. S. Khaki, H. Pham, and L. Wang, “Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning,” Sci. Rep., vol. 11, no. 1, pp. 1–14, 2021, doi: 10.1038/s41598-021-89779-z.
20. L. Martinez-Ferrer, M. Piles, and G. Camps-Valls, “Crop Yield Estimation and Interpretability with Gaussian Processes,” IEEE Geosci. Remote Sens. Lett., vol. 18, no. 12, pp. 2043–2047, 2021, doi: 10.1109/LGRS.2020.3016140.
21. S. Yang et al., “Integration of Crop Growth Model and Random Forest for Winter Wheat Yield Estimation from UAV Hyperspectral Imagery,” IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 14, pp. 6253–6269, 2021, doi: 10.1109/JSTARS.2021.3089203.
22. M. Ayub, N. A. Khan, and R. Z. Haider, “Wheat Crop Field and Yield Prediction using Remote Sensing and Machine Learning,” 2nd IEEE Int. Conf. Artif. Intell. ICAI 2022, no. March 2021, pp. 158–164, 2022, doi: 10.1109/ICAI55435.2022.9773663.
23. M. G. Ziliani et al., “Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop model,” Agric. For. Meteorol., vol. 313, no. July 2021, p. 108736, 2022, doi: 10.1016/j.agrformet.2021.108736.
24. A. V. M. Ines, N. N. Das, J. W. Hansen, and E. G. Njoku, “Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction,” Remote Sens. Environ., vol. 138, pp. 149–164, 2013, doi: 10.1016/j.rse.2013.07.018.
25. Q. M. Ilyas, M. Ahmad, and A. Mehmood, “Automated Estimation of Crop Yield Using Artificial Intelligence and Remote Sensing Technologies,” Bioengineering, vol. 10, no. 2, 2023, doi: 10.3390/bioengineering10020125.
26. J. Wang, P. Wang, H. Tian, K. Tansey, J. Liu, and W. Quan, “A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables,” Comput. Electron. Agric., vol. 206, no. 17, p. 107705, 2023, doi: 10.1016/j.compag.2023.107705.
27. Y. Ma, Z. Yang, and Z. Zhang, “Multisource Maximum Predictor Discrepancy for Unsupervised Domain Adaptation on Corn Yield Prediction,” IEEE Trans. Geosci. Remote Sens., vol. 61, pp. 1–15, 2023, doi: 10.1109/TGRS.2023.3247343.
28. C. Liao, J. Wang, B. Shan, Y. Song, Y. He, and T. Dong, “Near real-time yield forecasting of winter wheat using Sentinel-2 imagery at the early stages,” Precis. Agric., vol. 24, no. 3, pp. 807–829, 2023, doi: 10.1007/s11119-022-09975-3.
29. J. Campoy, I. Campos, J. Villodre, V. Bodas, A. Osann, and A. Calera, “Remote Sensing-based crop yield model at field and within-field scales in wheat and barley crops,” Eur. J. Agron., vol. 143, no. December 2022, p. 126720, 2023, doi: 10.1016/j.eja.2022.126720.