International Journal of Multidisciplinary Research and Growth Evaluation.

Advanced Machine Learning Technique for Monitoring Cotton Production Area in Pakistan (Punjab) Using Sentinel-2 Remote Sensing Data

Zakria Zaheen [1], Muhammad Awais Khan [2], Shawkat Ali [3], Hidayat Ullah [4], Abdul Basit [5], Jiahua Zhang [6*]
[1-6] College of Computer Science and Technology, Qingdao University, Qingdao, Shandong, China

* Corresponding Author: **Jiahua Zhang**

Article Info

ISSN (online): 2582-7138
Volume: 05
Issue: 06
November-December 2024
Received: 06-10-2024
Accepted: 09-11-2024
Page No: 817-828

Abstract
In recent years, remote sensing has developed agriculture by leveraging satellite imagery to achieve various goals, including land classification, crop area identification, growth monitoring, pattern recognition, and agricultural surveys. This technology has been extensively used for major crops like wheat, cotton, and rice. Cotton plays a pivotal role in the global agricultural sector and economy, providing fiber and substantial income while yielding valuable by-products like seeds and banola. Researchers worldwide have developed various remote sensing methods to monitor crop production, yet there remains a research gap in Pakistan concerning the utilization of Sentinel images for monitoring cotton growth. Therefore, it is crucial to employ remote sensing technology for more precise identification of cotton growth areas in Pakistan to optimize production and resource management. In our study, we propose a remote sensing-based system using efficient machine learning models to predict cotton crop areas from satellite images. This system utilizes Sentinel-2 MSI surface reflectance images from the years 2020-2021, filtered to include less than 20% cloud cover over Punjab. It focuses on the red (B4), green (B3), and blue (B2) bands for true-colour visualization, with reflectance values ranging from 0 to 2500 and a gamma correction of 1.1, operating at a 10-meter resolution and exporting a maximum of 7,699,027 pixels. Our system feeds satellite images into a machine learning model to accurately identify and calculate the total cotton cultivation area in Punjab. It has demonstrated an accuracy rate of 90-95% in predicting cotton regions, validated against statistical data from the authoritative publication "Crop Area and Production" by the Ministry of National Food Security and Research, Government of Pakistan. By providing detailed insights into growth patterns, this technology empowers farmers to make informed decisions on optimal cotton cultivation practices. This information is essential for enhancing agricultural productivity, sustainability, and economic outcomes.

DOI: https://doi.org/10.54660/.IJMRGE.2024.5.6.817-828

Keywords: Remote Sensing, Cotton Area, Random Forest, Satellite Images, Sentinel-2, Machine Learning

1. Introduction

Over the past decade, remote sensing-based crop growth identification and monitoring have gained significant popularity. Remote sensors provide high-resolution, multispectral, and high-quality satellite images that permit researchers to identify diverse land characteristics and conduct numerous analyses for various research objectives. The satellite imagery collected offers excellent resolutions in spectral, spatial, radiometric, and temporal domains, allowing for comprehensive crop monitoring over large areas [1].

The Sentinel-2 (S2) satellite, in particular, has facilitated the collection of images with high frequency and medium spectral resolution. Sentinel-2 images offer 10–60-meter spatial resolution across 13 spectral bands, enabling frequent revisits for precise monitoring of vegetation, land use, and environmental changes globally [2, 3].

This capability enables precise and continuous monitoring of crop growth, health, and development, contributing significantly to agricultural research and decision-making processes. These satellite images facilitate in agricultural researches from different perspectives for areas such as crop growth monitoring, disease detection, yield forecasting, crop land estimation weed identification, and water use in agriculture [4].

Cotton (*Gossypium hirsutum* L.) is a crucial commercial crop in Pakistan, with the Multan region in Punjab being the largest producer. In 2018, Multan's cotton crop area was around 312,760 hectares, producing approximately 626,316 tons, significantly contributing to the local GDP. Due to rising demand, Punjab's cotton area expanded to 2.5 million hectares during 2019-2020, accounting for 78% of national production. Climate changes, including variations in humidity, precipitation, temperature, and sunshine, have impacted productivity. Traditionally, annual statistics are released at the fiscal year's end, but timely and accurate data throughout the growing season are essential for better prediction and forecasting [5, 6].

Cotton crop monitoring and yield identification are crucial in Pakistan due to its significant economic impact, contributing to GDP, exports, and rural employment. Monitoring ensures timely interventions like pest control and irrigation management, crucial for yield and quality. Accurate yield estimation aids in market planning and policy decisions, ensuring sustainable production amidst climate challenges. This enhances agricultural resilience and food security, supporting overall economic development and sustainability in Pakistan [7].

Traditional methods to monitor cotton crop production in Pakistan often include field surveys conducted by agricultural extension workers or local agronomists. These surveys involve visiting cotton fields to assess crop health, growth stages, pest infestations, and estimating yield based on visual observations and local knowledge. Additionally, farmers may provide information through interviews or participatory methods to gauge crop conditions and potential challenges such as water availability and pest outbreaks [8, 9]. Theses Traditional methods of cotton crop monitoring and yield estimation in Pakistan face several challenges [10].

Field surveys are a costly and labor-intensive method for crop management, requiring significant manpower and time. They may not cover remote or geographically dispersed fields, affecting data representativeness. Traditional methods may not be available promptly, delaying critical decisions. Human observation alone may overlook subtle changes in crop health, impacting yield forecasts and management strategies. Additionally, weather conditions can affect the ability to conduct field surveys, leading to inconsistent data collection and analysis. Therefore, these methods are not reliable and can be subject to bias Addressing these issues requires integrating modern technologies like satellite imagery (e.g., Sentinel-2) and remote sensing techniques, which offer more objective, comprehensive, and timely insights into cotton crop conditions and yield estimation, thereby enhancing agricultural efficiency and sustainability in Pakistan [11].

The discussion above highlights the motivation for research in this area. Research in cotton crop growth is crucial for sustaining agricultural productivity and GDP. Traditional methods are labor-intensive and susceptible to human error. Remote sensing technology offers reliable satellite imagery for accurate agricultural analysis. Adoption of remote sensing techniques is essential in Pakistan to improve crop monitoring and yield estimation accuracy, enhancing agricultural efficiency.

In the current study, we propose a remote sensing-based cotton crop monitoring system using an efficient machine learning model. Sentinel-2 images from the Punjab region in Pakistan from 2020-2021 are extracted for this purpose. These images are utilized to train the machine learning model using pre-labelled data to identify cotton regions and predict total cotton area. The model's outcomes are then validated against statistical data obtained from official surveys conducted by the Government of Pakistan.

2. Related Work

Numerous techniques have been proposed by various researchers utilizing remote sensing to achieve agricultural objectives such as crop prediction, crop monitoring, disease identification, and crop water requirement prediction. In this section, we will elaborate on research work similar to our proposed system to highlight their significance and limitations.

Rodriguez-Sanchez, Li, and Paterson 2022 [12] suggested a cost-effective method for calculating cotton yield output based on RGB camera pictures. They employed an SVM for supervised machine learning classification, which was initially trained with only one plot picture. Their suggested solution has less complexity because it only requires annotation of one RGB picture, as well as a shorter deployment time for the SVM models. Their suggested approach can detect shifting situations and discern cotton pixels more effectively in single plot photos with varying numbers of cotton bolls. They demonstrated reliable cotton boll counting. Furthermore, their proposed approach was helpful in finding yield differences across several commercial cultivars and breeding lines. And increase the efficiency of decision making for cultivation. And increase the efficiency of decision-making for plant breeding as well as improve the use of resources by speeding up the analysis of all field experiments.

In year 2020 Rao, Nisarga, Srikanth, and Prabhuraj estimated cotton crop production using Sentinel 2A MSI data. They carried out their research in Raichur *taluk*, Raichur district, in the Indian state of Karnataka during the cotton growing season 2018-19. They used sentinel2 images in which they used spectral signatures and temporal profiles of vegetation indices for identification of cotton crop grown under irrigated and rainfed conditions. They used supervised digital image classification of cotton crop and showed classification accuracy of 88.28 % by using MSI band combination of blue (490 nm), green (560 nm), red (665 nm) and near-infrared (842 nm), having 10-meter spatial resolution.

Detected cotton fields from remote sensing images using several deep learning networks, which were previously used to identify water. Their suggested approach featured feature fusion, which involved image down sampling followed by trans-convolution for image up sampling. Using this strategy, they combined characteristics from various scales in the down-sampling phase into the up-sampling process. They confirmed the proposed DenseNet findings with ground truth and compared them to four common CNNs: ResNets, VGG, SegNet, and DeepLab v3+. Their testing findings indicated that the suggested DenseNet model is superior to several popular CNNs utilizing the same datasets, as it shown clear capabilities in discriminating cotton fields from mountain shadows, water bodies, villages, bare terrain, Clouds etc. As a result, they found that a deep neural network architecture based on DenseNet is a trustworthy alternative for extensively used multi-spectral classification applications.

Investigated the use of Sentinel-2 images for continuous monitoring of cotton stem water potential. A reliable measure for assessing the water condition of cotton is stem water potential (SWP). They tracked cotton growth on a vast scale using satellite remote sensing. Their main objective is to measure cotton water stress on a big scale and at a high temporal frequency. To assess cotton water stress throughout the growth season they developed linear regression and machine learning models using the midday SWP samples they measured in their proposed method based on the dates of Sentinel-2 image capture. They used vegetative indicators and several Sentinel-2 spectral bands to estimate SWP to train Linear Regression. The best possible accuracy was attained by using the original spectral bands. Additionally, they discovered that the red edge and SWIR bands were the most significant spectral bands and that the vegetation indices based on these bands were especially useful. Their findings indicate that the machine-learning-based approach is the favored one and that the suggested system has the potential for continuous SWP monitoring at large scales.

Looked into the identification of cotton root rot using random forest and multifeatured selection from Sentinel-2 photos. In their study, they demonstrated that it is possible to detect cotton root rot and produce prescription maps using the spectral model, texture model, and spectral-texture model based on Sentinel-2A data. With an accuracy of 92.95%, they discovered that the spectral model is the most effective in detecting root rot, compared to 84.81% and 91.87% for the texture and spectral-texture models, respectively. Additionally, their method indicated that the spectral-texture model is better suited for cotton fields with minor and dispersed infestations, whereas the spectral model is better suited for cotton field identification. These findings have validated the viability of detecting cotton root rot generate guidelines maps with Sentinel-2 imagery's many properties, and add texture information to enhance the detection of root rot.

Wang, Zhai, and Zhang (2021) [16] used a time series of Sentinel-2 images to do autonomous cotton mapping. They used spectral characteristics during the bolls opening stage to construct the white bolls index in their study. Other crops are unable to see the white bolls, which are unique to cotton. Based on a Sentinel-2 time series, they developed the White Bolls index using linear discriminant analysis. They used WBI to quickly map cotton without the need of training samples. Using a WBI time series, they were able to determine when the bolls opened. They evaluated the suggested system's performance and contrasted it with other classifiers like SVM and 1DCNN.They demonstrated that WBI functioned effectively in the absence of training data. Given the promising findings of their proposed study, WBI need to be applied going forward to evaluate cotton varieties with subordinate white bolls.

According to Hu *et al* mapping cotton fields in Northern Xinjiang, China, was beneficial when Sentinel-1/2 data and machine learning algorithms were used. In order to map cotton fields, they suggested a novel framework that included the utilization of optical and SAR data from various phenological periods, random forest classifiers, and the multi-scale image segmentation approach for cotton field extraction. Their suggested technique demonstrated a decrease in cloud cover interference and noise throughout the extraction process, demonstrating its effectiveness for high-precision and high-resolution cotton field mapping. With a spatial resolution of 10 m, they generated a cotton field map for northern Xinjiang in 2019 using the suggested framework. S2's red edge band and S1 were determined to be particularly crucial for mapping cotton crops. They also came to the conclusion that crop mapping might be much enhanced by combining optical and microwave images. There is a lot of possibility for using their suggested framework for mapping and tracking different crops.

Table 1: State of the Art Work Limitation and Strengths

Reference	Methodology	Limitations with Perspective of Cotton Crop Prediction	Strengths
Rodriguez-Sanchez, Li, and Paterson (2022) [12]	RGB camera images, SVM for supervised classification, trained with a single plot image	Limited to single plot images; may not scale well to larger, diverse cotton fields or different environments	Cost-effective, reliable cotton boll counting, useful for breeding programs
Rao, Nisarga, Srikanth, and Prabhuraj (2020)	Sentinel-2A MSI data, supervised image classification, spectral signatures, temporal profiles	Specific to Raichur district; requires multiple spectral bands; may not generalize to other regions or conditions	High classification accuracy (88.28%), effective for different growing conditions
Li *et al*. (2021)	Deep learning networks, feature fusion, DenseNet, image down-sampling and up-sampling	Computationally intensive; may require significant processing power; needs high-quality labeled data for training	Superior performance compared to other CNNs, effective feature aggregation
Lin *et al*. (2020)	Sentinel-2 imagery, linear regression and machine learning models for SWP estimation	Focused on water potential estimation; may not directly predict cotton yield or area accurately	High accuracy with original spectral bands, useful for large-scale monitoring
Li *et al*. (2020)	Multi-feature selection, random forest classification, spectral, texture, and spectral-texture models	May not generalize well to all cotton fields; especially fields with different conditions or small-scale infestations	High accuracy for root rot identification, effective for creating prescription maps
Wang, Zhai, and Zhang (2021) [16]	Time series of Sentinel-2 images, White Bolls Index (WBI), linear discriminant analysis	WBI may not be significant or applicable in all cotton types or regions; may miss early or late-season crops	No training samples needed, rapid cotton mapping, effective WBI usage
Hu *et al*. (2021)	Combination of Sentinel-1/2 data, machine learning techniques, random forest classifiers, multi-scale image segmentation	Requires multi-spectral data; may not apply to non-cotton crops; needs extensive data processing	High-resolution and high-precision mapping, effective cloud cover and noise reduction

3. Proposed Methodology

The proposed system aids in identifying cotton areas in the Punjab province of Pakistan. It accomplishes this by acquiring satellite images through remote sensing from 2020 to 2021. These images are then pre-processed using machine learning techniques and analyzed by a random forest algorithm to identify cotton fields, subsequently computing the total cotton area. The entire process is illustrated in Figure 1.

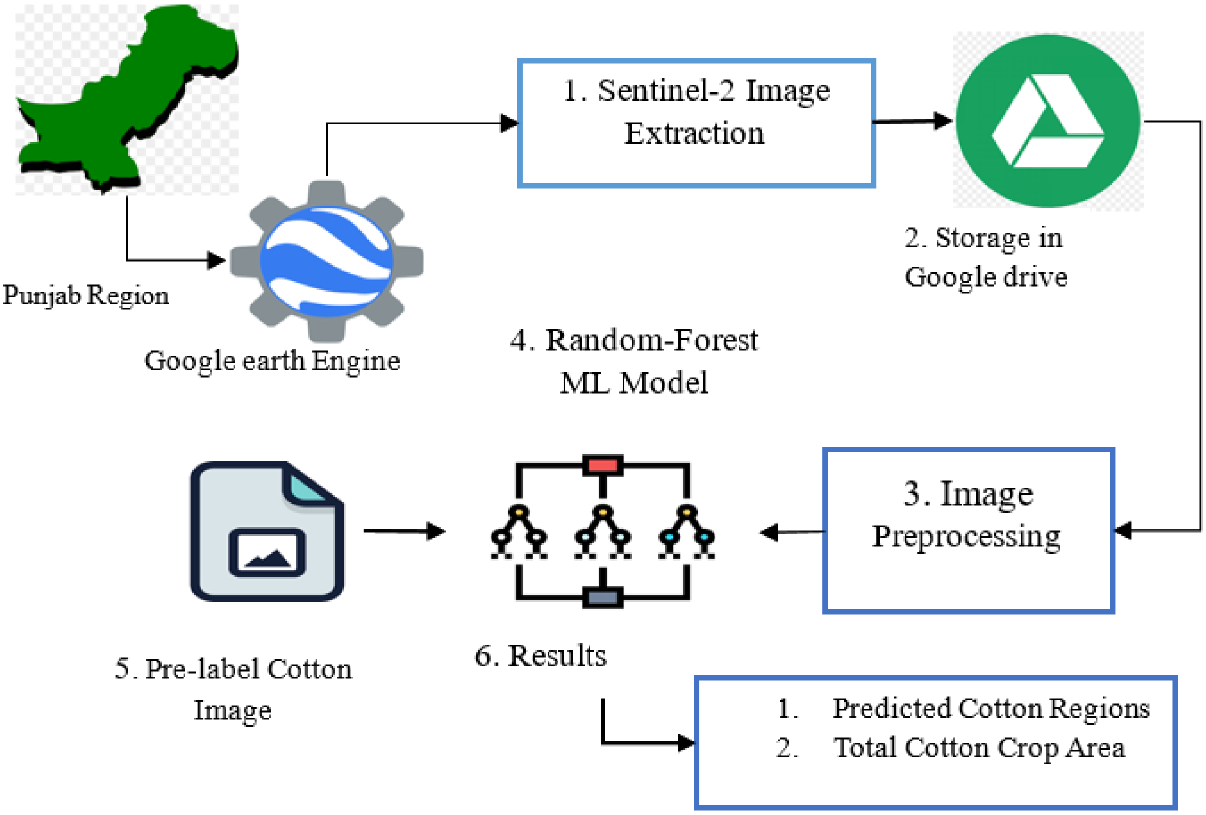

Fig 1: Cotton Area Identification & Prediction Model Using Remote Sensing Sentinel -2 Images

3.1. Study Area

The study area for the proposed system is the Punjab province, which is considered a major cultivation area in Pakistan. According to the Government of Pakistan's statistics, the total cotton produced in Pakistan during the year 2020-2021 was 5.64 million bales, with a significant portion—approximately 4.5 million bales—cultivated in Punjab. Given its importance as a key region for cotton production, we selected the Punjab province for the proposed system [19]. The major districts and divisions are listed in Table 2. The study area location in the Pakistan map along with a description of major divisions in Punjab province also shown in Figure 2.

Table 2: List of Punjab Division and Districts counts.

Division	Districts	Division	Districts
Bahawalpur	3	Lahore	4
Multan	4	Rawalpindi	4
Sahiwal	3	Gujranwala	6
Sargodha	4	Dera Ghazi Khan	4
Faisalabad	4	Gujrat	4

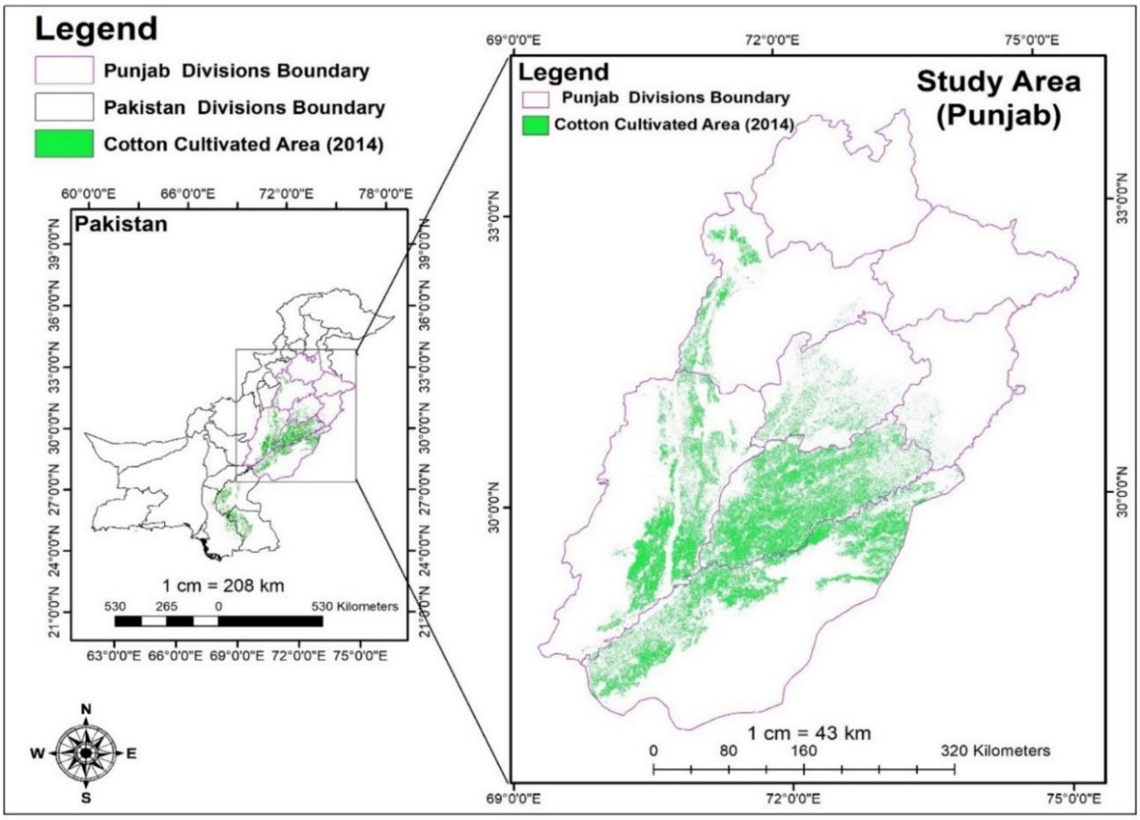

Fig 2. Highlighted Punjab Province in Map of Pakistan

3.2. Sentinel-2 Image Extraction

The European Space Agency (ESA) provided Sentinel-2 imagery; due to its advanced features it is considered as valuable source in precision agriculture. Some important and prominent features of sentinel2 imagery are high spatial resolution (10 meters for visible and near-infrared bands), frequent revisit times (5 days at the equator with two satellites), and a broad spectral range (13 bands from visible to shortwave infrared). These prominent features allow researchers in detailed monitoring and analysis of crop health, soil conditions, and water management as well, making Sentinel-2 an influential means for enhancing agricultural yield and sustainability (Segarra *et al*., 2020) [19]. In the proposed system, Sentinel-2 images were utilized, and we extracted them from the COPERNICUS/S2_SR collection using Google Earth Engine. The extraction process is streamlined by choosing appropriate parameters setting as listed in Table 3, and followed the algorithm 1 described below.

Table 3: Sentinel-2 Image Extraction Parameters

Parameter	Description	Used Values
City	City for image extraction.	Punjab City
Selected Year	Year of image extraction.	2020-2021
Selected Month	Month of image extraction.	4-9(April-September)
Image Collection	Sentinel-2 Surface Reflectance images.	COPERNICUS/S2_SR
Cloud Cover Filter	Filters images with less than 30% cloud cover.	CLOUDY_PIXEL_PERCENTAGE < 30
Composite Generation	Generates monthly composites.	12 monthly composites
Export Parameters	Settings for saving image to Google Drive.	GeoTIFF, Scale: 10m, CRS: EPSG:4326
Bands	Specific bands used.	B2, B3, B4, B5, B6, B7, B8, B8A

The text provides a detailed process for defining an Area of Interest (AOI) for Pakistan, setting the city and time range, filtering the AOI based on the city name, normalizing the image, masking cloud and cirrus pixels using the QA60 band, filtering the Sentinel-2 Image Collection, generating monthly composite images, stacking the image collection, replacing masked pixel values with mean values from neighboring months, visualizing the composite image, and exporting the final processed image to Google Drive. The process involves loading administrative boundaries for Pakistan, filtering the region of interest, setting the city and time range, normalizing the image, filtering the Sentinel-2 Surface Reflectance image collection, generating monthly composite images, stacking the image collection, replacing masked pixel values with mean values from neighboring months, and exporting the final processed image to Google Drive with specified settings.

We collected Sentinel-2 images for Punjab province, which is organized into 10 districts, with a total of 41 cities. For the cotton cultivation period from April to September in the years 2020-2021, we extracted 24 images from different cities. Details of the dataset collection are provided in Table 4.

Table 4: Dataset Description

Region	Image Description	Region	Image Description
Mianwali	April 2020	Bahawalpur	April 2021
Kasur	May 2020	Multan	May 2021
Sialkot	June 2020	Rahim Yar Khan	June 2021
Gujranwala	July 2020	Lodhran	July 2021
Sahiwal	August 2020	Bahawalnagar	August 2021
Okara	September 2020	Vehari	September 2021
Rawalpindi	April 2020	Dera Ghazi khan	April 2021
Jhelum	May 2020	Rajanpur	May 2021
Attock	June 2020	Muzaffar Garh	June 2021
Chakwal	July 2020	Sargodha	July 2021
Faisalabad	August 2020	Khushab	August 2021
Jhang	September 2020	Bhakkar	September 2021

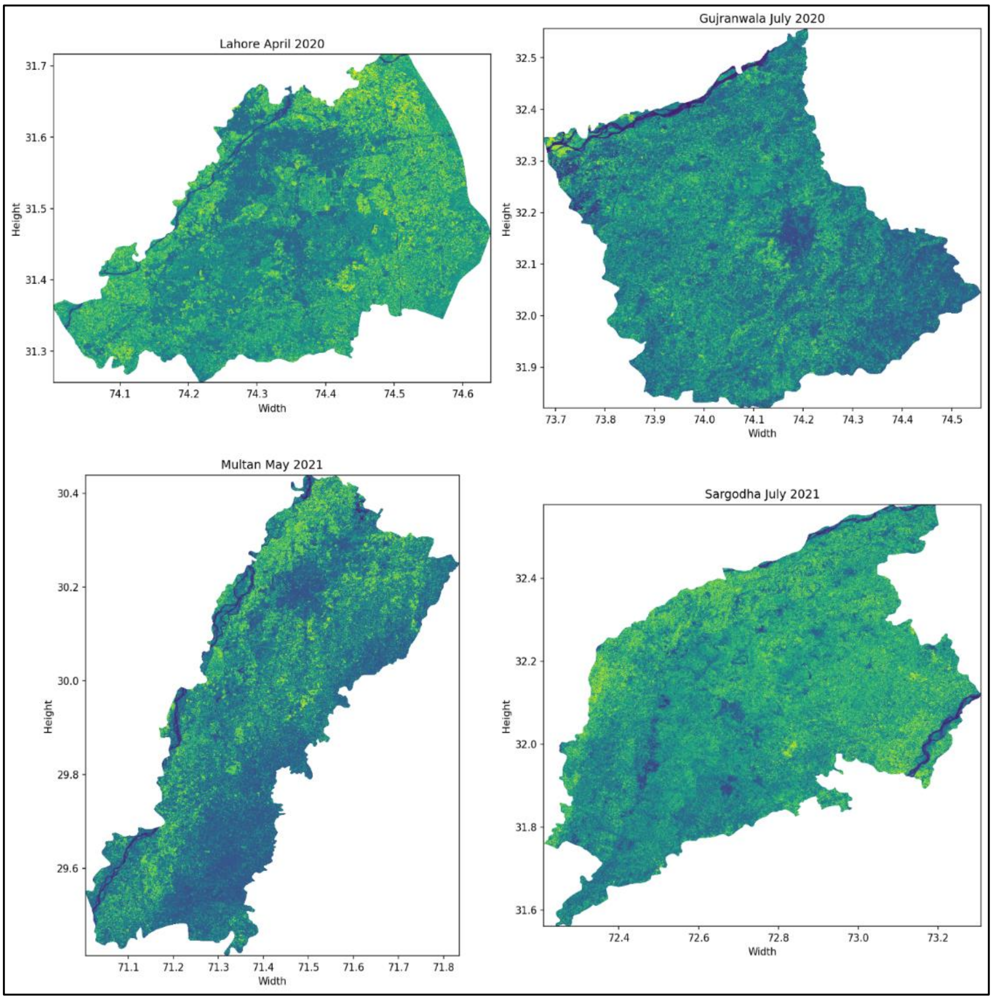

Fig 3: NDVI Images from Sentine-2 Remote Sensing Collection of Punjab

3.4. Configuring the Random Forest Model

The proposed system utilizes an efficient machine learning algorithm, Random Forest, for the identification of cotton areas and computation of total cotton area within Sentinel-2 images. Random Forest is a well-known and efficient ensemble learning model, first introduced by Breiman. It consists of multiple classification trees, specifically K decision trees, combined with P random choices. In the classification process for a particular problem, the variable input is fed into the predictors of K-trees to obtain K predictions. These predictions are then aggregated, and the most popular prediction is selected as the final result. Each tree in the Random Forest is trained using a selected subset of random input variables. Random Forest is considered efficient because it can select features with high importance and rich information content. In random forest prediction of one decision tree based on the input features. Many researchers have employed Random Forest for computationally complex research problems. For instance, Lawrence and Moran (2015) [21-22] conducted a comparison of different classifiers using consistent procedures and 30 different datasets. They found that the average classification accuracy of Random Forest was the highest, significantly outperforming other machine learning methods such as Support Vector Machines (SVM). Consequently, it is believed that the Random Forest model offers superior performance for cotton field extraction. Additionally, the computational intensity of Random Forest classifiers is lower than that of other decision tree ensemble methods. They effectively integrate variables and require minimal adjustment and supervision [18]. The computational framework of Random Forest is supported by three major steps, as highlighted by numerous researchers [23, 24]. A single decision tree represented by DT_K the prediction for an input sample x can be expressed in Equation 1:

$$y_k = DT_k X \qquad \text{Eq1}$$

As, there are many decisions tree inside random forest so to achieve final prediction random forest first compute aggregation of all predictions obtained from each individual decision tree. It can be express as shown in Equation 2,

$$y = mode(y_1, y_2, y_3 \ldots\ldots y_k \qquad \text{Eq2}$$

Then, all aggregated predictions are averaged to compute final outcome of random forest. This computation is expressed in equation 3.

$$y_k = \frac{1}{k}\sum_{k=1}^{k} y_k \qquad \text{Eq3}$$

The proposed random forest model takes Sentinel-2 images as input and performs necessary preprocessing to enhance computational efficiency for cotton identification. After preprocessing, suitable feature mapping between the labeled data and input data is conducted. We used label data of cotton masked for the Punjab and Sindh regions. The dataset 'Crops Area and Production (Districts Wise)' published by the Food and Agriculture Organization of the United Nations (FAO) provides comprehensive statistics on crop production. This data is available in FAO Data Catalog. The input data is then appropriately divided into training and testing sets. The model is trained on the training portion of the data and predictions are made on the testing portion. The entire process is described below and depicted in the given flowchart. The major steps involved to predict cotton regions in an NDVI (Normalized Difference Vegetation Index) image using a Random Forest classifier. Reading the NDVI and labeled data. The major steps involved in predicting cotton regions in an NDVI (Normalized Difference Vegetation Index) image using a Random Forest classifier. Reading the NDVI and labeled data. The process involves preprocessing input data, sampling it for training, splitting it into training and testing sets, training a Random Forest classifier, evaluating its performance, predicting cotton regions in NDVI images, and visualizing the results.

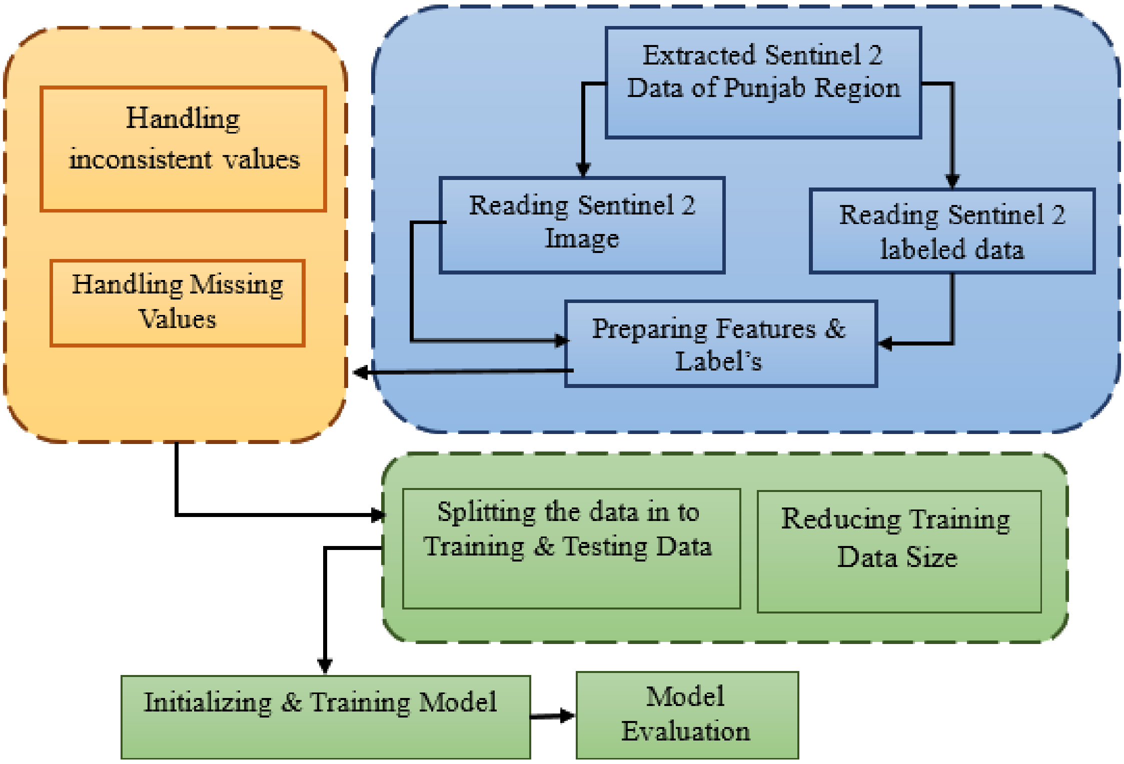

Fig 4: Working Flowchart of Random Forest

4. Experiments and Results

The proposed system, once trained with Sentinel-2 images collected from the Punjab region during the years 2020-2021, utilized a total of 24 images as detailed in Table 4. These images were then used to train the Random Forest model following the steps outlined in the flowchart (see Figure 4). The entire process of training and testing the Random Forest model on Sentinel-2 images from the Punjab region is described here.

We collected a total of 24 images from the Punjab region for the years 2020-2021, specifically during the peak cotton cultivation months of April to September. From these 24 images, we selected 10 images in total, with 5 from each year. These images were then used in the Random Forest model to identify cotton regions and calculate the total cotton cultivation area. The details of the selected images, along with their regions, durations predicted cotton area and accuracy is described in Table 5.

Table 5: Experimental Data

City	Month/Year	Predicted Region (Hectare)	Accuracy
Faisalabad	May, 2020	267.811	99.57%
Kasur	May, 2020	173	99.77%
Vehari	June,2020	116	99.80%
Bahawalpur	July, 2020	235.493	99.63%
Khanewal	August, 2020	162.052	99%
Rajanpur	May, 2021	182..831	99.57%
Multan	May, 2021	392.337	99.67%
Mianwali	June, 2021	220.886	99.50%
Bhakkar	July, 2021	192.98	99.53%
Bahawalnagar	August, 2021	175.581	99.43%
Lodhran	July, 2020	373.299	99.63%

The predicted cotton areas are contoured using blue and red colors. Figure 5-9 shows the prediction results for Vehari, Lodhran, Khanewal, Rajanpur, and Multan respectively. The blue and red highlights indicate the cotton areas in each respective region. The predicted cotton areas are contoured using blue and red colors.

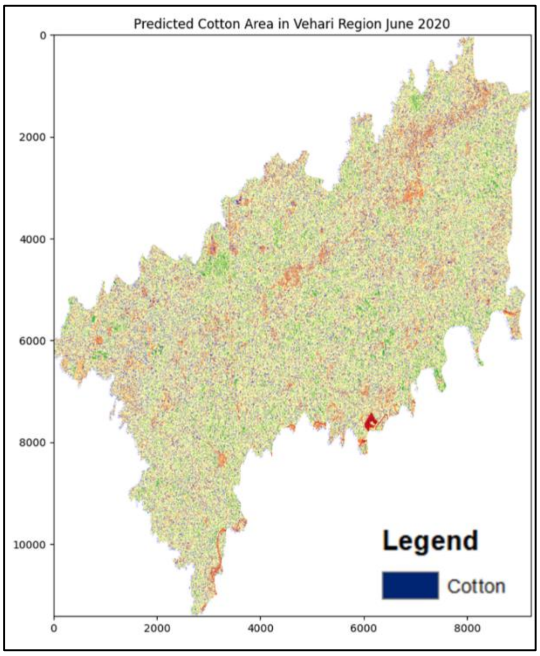

Fig 5: Predicted Cotton Area in Vehari for June 2020 using Random Forest

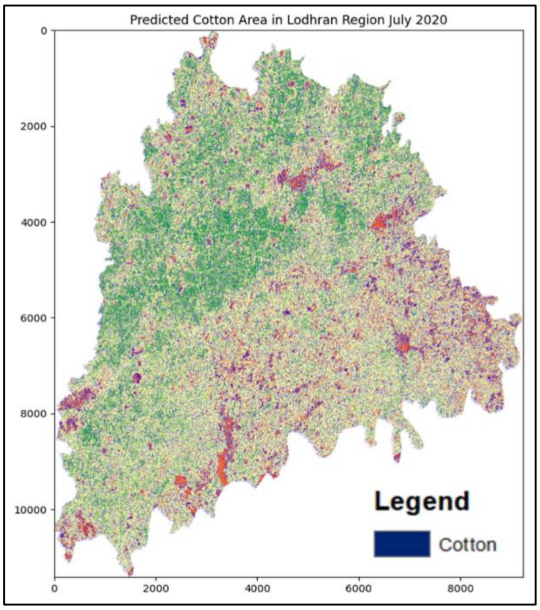

Fig 6: Predicted Cotton Area in Lodhran for July 2020 using Random Forest

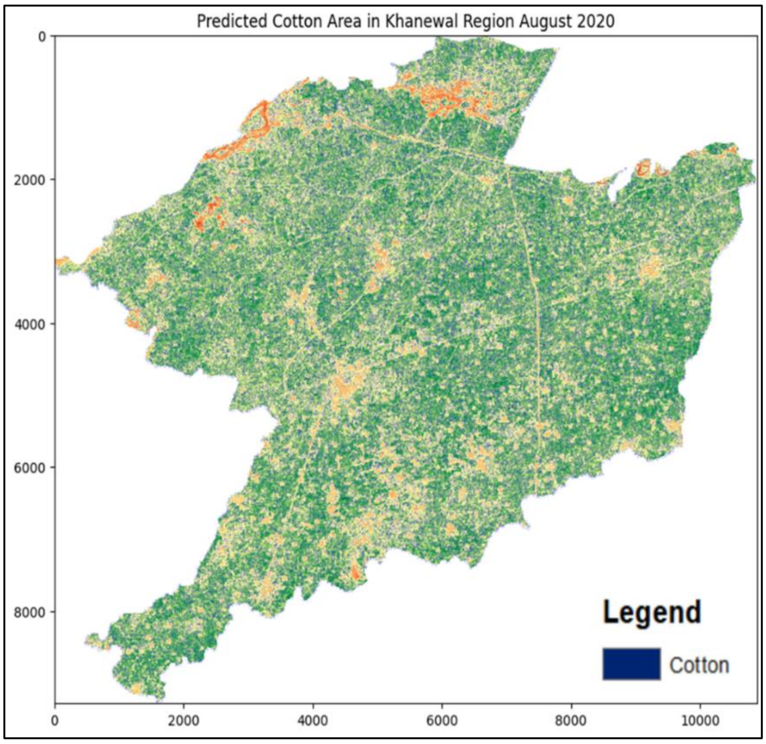

Fig 7: Predicted Cotton Area in Khanewal for August 2020 using Random Forest

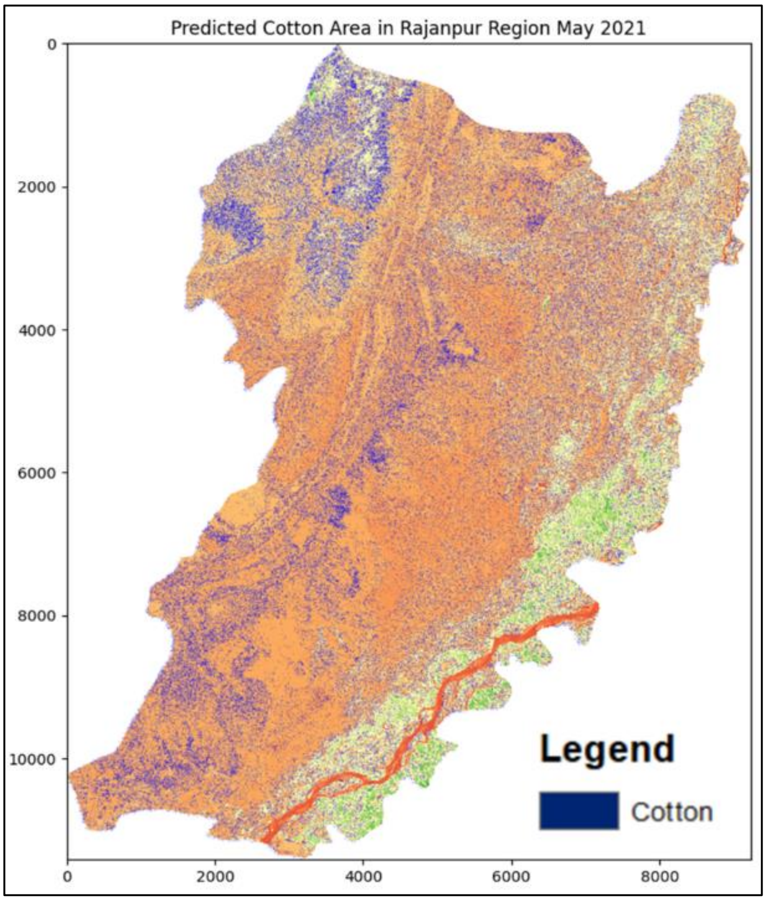

Fig 8: Predicted Cotton Area in Rajanpur for May 2021 using Random Forest

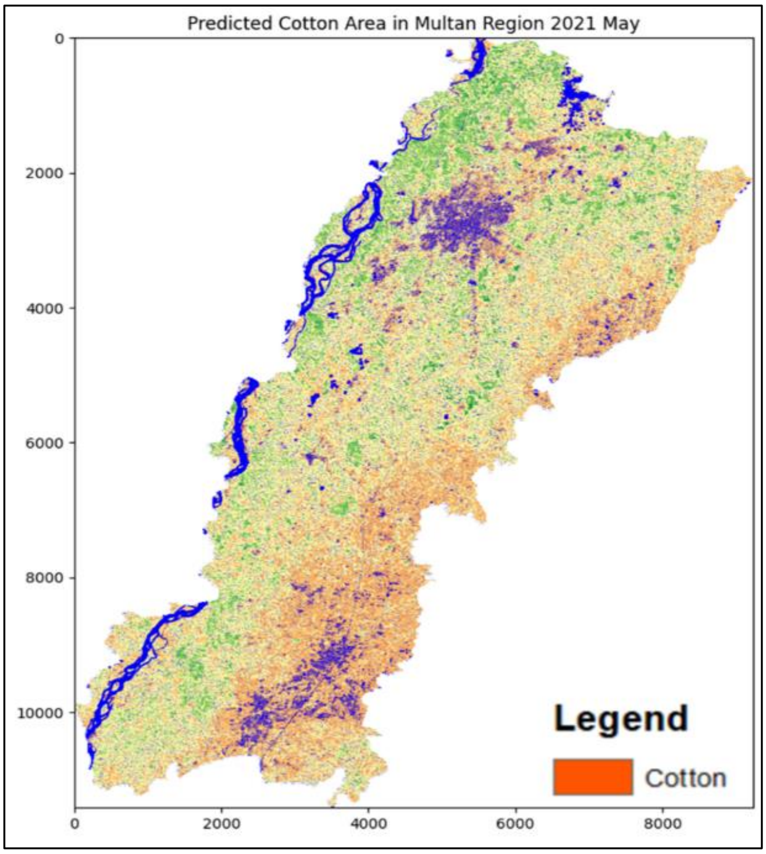

Fig 9: Predicted Cotton Area in Multan for May 2021 using Random Forest

The proposed system generated results with a high accuracy of approximately 90-99%. To ensure the authenticity of the proposed system, we compared the prediction results with the statistics mentioned in "Crops Area and Production (Districts Wise)" published by the Government of Pakistan, Ministry of National Food Security & Research. This publication comprises the latest district-wise data of major and minor crops statistics for the years 2020-21 and 2021-22.We described the results of 5 cities in Table 6. The individual month values have to be compared with the statistics value, which is the average of the whole year. Considering the fact that cotton is not cultivated year-round, a 4-month window (May-August) is more viable for cotton crop prediction. We computed the appropriate average value from the total cotton area achieved by our proposed system using the following formula given inn Equation 4:

$$\text{Cotton Crop Area in Hectacres} * \frac{4(\text{May} - \text{August})}{12(\text{total month in a year})} \qquad \text{Eq (4)}$$

Table 6: Experimental Data Comparison with Statistics

City	Avg value for 20-21 (May-August)	Statistics Data Value
Vehari	38	46
Lodhran	124	134
Khanewal	54	88
Rajanpur	60	77
Multan	130	105

We considered five cases to evaluate the results of our proposed system. For Vehari, the system estimated the cotton cultivation area at 38 hectares, which is slightly lower than the statistical data value of 46 hectares, indicating a minor discrepancy. In Lodhran, the predicted cotton crop area was 373 hectares, but after averaging, the value came to 124 hectares, which is slightly smaller than the actual statistics of 134 hectares. For Khanewal, the system predicted a cotton region of 162 hectares, with an average value of 54 hectares, showing a relatively close match to the actual statistics of 88 hectares. In Rajanpur, the predicted cotton cultivation area was 182.831 hectares, while the average value of 60 hectares is somewhat below the accurate statistics of 77 hectares. Finally, in Multan City, the system estimated the cotton crop area for the year 2020-2021 to be 130 hectares. Although this value is slightly higher than the actual figure of 105 hectares, the overall accuracy remains high with a relative error of approximately 22%. This discrepancy can be attributed to factors such as variations in satellite imagery resolution, atmospheric conditions affecting the NDVI values, and potential temporal changes in crop cover during the assessment period.

Factors contributing to this difference may include the heterogeneity of the landscape, misclassification of land cover types, and limitations in the temporal resolution of the satellite data used [25]. The observed difference could be due to underestimation of NDVI values in regions with dense vegetation cover or overlapping crop cycles, which might not have been adequately captured in the satellite images. The comparison also depicted in graph as shown in Figure 10.

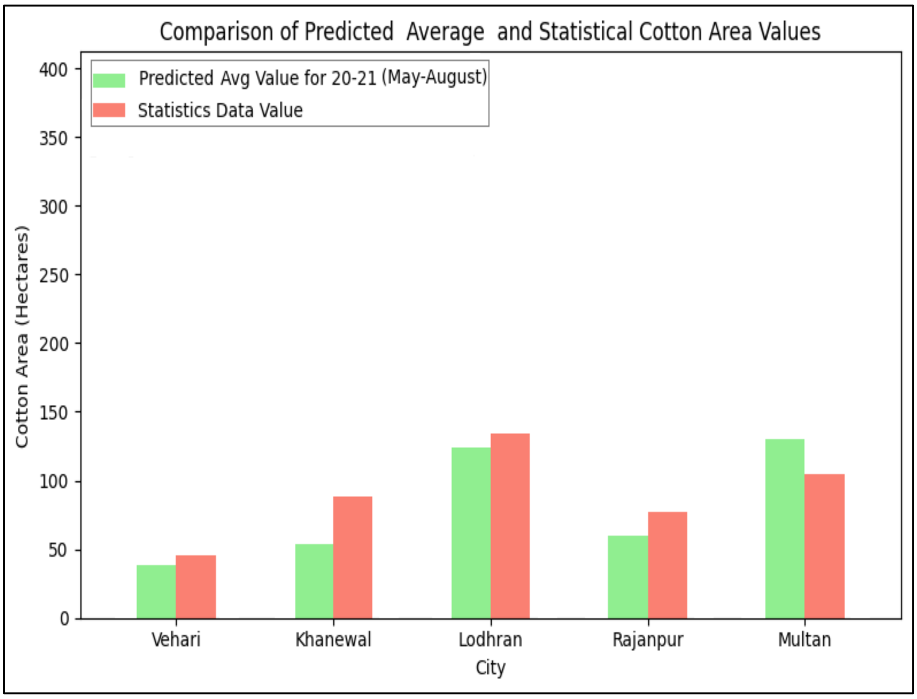

Fig 10: Comparison Graph of Prediction and Statistics

These comparisons demonstrate the effectiveness of our system in estimating cotton crop areas with high accuracy, despite minor variations from the actual statistics. The observed differences highlight the inherent challenges in remote sensing-based agricultural monitoring, such as spatial resolution limitations, atmospheric effects, and temporal variability. Future improvements in sensor technology and data processing algorithms can further enhance the precision of crop area estimations [26].

5. Conclusion

This study proposed a remote sensing-based cotton crop monitoring system using Sentinel-2 satellite imagery and a Random Forest machine learning model. The system aimed to address the existing research gap in cotton growth monitoring in Pakistan, specifically within the Punjab region. The results demonstrate that the system achieved a high accuracy rate of 90-99% in predicting cotton cultivation areas, validating the effectiveness of utilizing Sentinel-2 images for precise agricultural monitoring.

The Random Forest model, trained on Sentinel-2 images from 2020-2021, effectively identified and computed cotton regions, offering reliable predictions that were compared against official statistical data from the Government of Pakistan. Despite minor discrepancies, the proposed system showcases significant promise for enhancing cotton crop monitoring and management, providing valuable insights for optimizing production and resource management in Punjab.

6. Future Work

Future research could focus on broadening the scope of the monitoring system to include other crop types and regions, enhancing its applicability and utility in diverse agricultural contexts. Integrating temporal analysis to monitor crop dynamics across multiple growing seasons could provide deeper insights and improve adaptive management strategies. Additionally, exploring hybrid machine learning models and advanced algorithms may further refine the system's robustness. Developing a real-time monitoring framework with regular updates would enable timely, data-driven decision-making for farmers and stakeholders, ensuring optimal resource utilization and sustainability.

7. Limitations

The study faced several limitations that could be addressed to improve its robustness and scalability. While Sentinel-2 provides high-resolution imagery, which is beneficial for accurate cotton area estimation, it also requires significant computational resources and large memory capacity to process and analyze the data efficiently. Additionally, the model's high accuracy (95–99%) is largely due to being trained on standardized single-class cotton label data, which limits its generalization to other crop types. The spatial resolution of Sentinel-2 images may still struggle to capture small-scale variations in cotton fields, especially in heterogeneous landscapes. Atmospheric variations can also affect NDVI values, introducing potential discrepancies in cotton area estimation. The temporal focus on specific months may not fully capture seasonal growth dynamics, which could affect prediction accuracy. Furthermore, limited availability of ground-truth data for validation in certain regions may impact the reliability of model predictions. Addressing these limitations, along with optimizing computational resources and expanding model capabilities to handle multiple crop classes, can further enhance the system's performance and applicability for agricultural monitoring.

8. References

1. Carneiro FM, de Brito Filho AL, Ferreira FM, Junior GDFS, Brandao ZN, da Silva RP, Shiratsuchi LS. Soil and satellite remote sensing variables importance using machine learning to predict cotton yield. Smart Agricultural Technology. 2023;5:100292.
2. Yang B, Zhu W, Rezaei EE, Li J, Sun Z, Zhang J. The optimal phenological phase of maize for yield prediction with high-frequency UAV remote sensing. Remote Sensing. 2022;14:3-18. doi:10.3390/rs14071559.

3. Gao M, Yang F, Wei H, Liu X. Individual maize location and height estimation in field from UAV-borne LiDAR and RGB images. Remote Sensing. 2022;14(10):2292. doi:10.3390/rs14102292.
4. Li H, Wang G, Dong Z, Wei X, Wu M, Song H, *et al*. Identifying cotton fields from remote sensing images using multiple deep learning networks. Agronomy. 2021;11(1):174.
5. Pakistan Bureau of Statistics. Agricultural statistics of Pakistan. Islamabad: Government of Pakistan; 2020. Available from: https://www.pbs.gov.pk.
6. Abbas S. Climate change and cotton production: an empirical investigation of Pakistan. Environmental Science and Pollution Research. 2020;27(23):29580-8.
7. Mahmood Q, Azam M, Khan Z. Importance of cotton crop monitoring and yield estimation for sustainable agriculture in Pakistan. International Journal of Agriculture Innovations and Research. 2019;7(2):289-98.
8. Beshir B, Dabo A, Idris YA. Traditional methods of crop production and its implications for sustainable agriculture in Ethiopia. International Journal of Agricultural Science, Research and Technology. 2015;5(1):37-46.
9. Prasad R. Traditional farming systems for sustainable agriculture in developing countries. Journal of Sustainable Agriculture. 2007;31(1):43-67.
10. Rasul G, Shahzad MI, Mahmood R. Issues in agriculture and development: Traditional versus modern practices in Pakistan. International Journal of Development Research. 2014;4(7):1452-8.
11. Ali H, Shah SMM, Ashfaq M. Role of remote sensing and GIS techniques for sustainable cotton crop management: A review. Journal of Cotton Research. 2018;1(1):1-17.
12. Rodriguez-Sanchez J, Li C, Paterson AH. Cotton yield estimation from aerial imagery using machine learning approaches. Frontiers in Plant Science. 2022;13:870181.
13. Rao PP, Nisarga I, Srikanth P, Prabhuraj D. Cotton crop production estimation using Sentinel 2A MSI data. 2020. doi:10.13140/RG.2.2.25533.44000.
14. Lin Y, Zhu Z, Guo W, Sun Y, Yang X, Kovalskyy V. Continuous monitoring of cotton stem water potential using Sentinel-2 imagery. Remote Sensing. 2020;12(7):1176.
15. Li X, Yang C, Huang W, Tang J, Tian Y, Zhang Q. Identification of cotton root rot by multifeature selection from Sentinel-2 images using random forest. Remote Sensing. 2020;12(21):3504.
16. Wang N, Zhai Y, Zhang L. Automatic cotton mapping using time series of Sentinel-2 images. Remote Sensing. 2021;13(7):1355.
17. Hu T, Hu Y, Dong J, Qiu S, Peng J. Integrating Sentinel-1/2 data and machine learning to map cotton fields in Northern Xinjiang, China. Remote Sensing. 2021;13(23):4819.
18. Pakistan Bureau of Statistics. Agriculture statistics. 2023. Available from: http://www.pbs.gov.pk/content/agriculture-statistics.
19. Segarra J, Buchaillot ML, Araus JL, Kefauver SC. Remote sensing for precision agriculture: Sentinel-2 improved features and applications. Agronomy. 2020;10(5):641. doi:10.3390/agronomy10050641.
20. Lawrence RL, Moran CJ. Comparison of classification algorithms for forest type mapping using LiDAR and multispectral data. Remote Sensing of Environment. 2015;158:281-94. doi:10.1016/j.rse.2014.11.015.
21. Gislason PO, Benediktsson JA, Sveinsson JR. Random forests for land cover classification. Pattern Recognition Letters. 2006;27(4):294-300. doi:10.1016/j.patrec.2005.08.011.
22. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer; c2009.
23. Cutler DR, Edwards Jr TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. Random forests for classification in ecology. Ecology. 2007;88(11):2783-92. doi:10.1890/07-0539.1.
24. Huang J, Chen J, Zheng W. Assessment of land cover classification accuracy using remote sensing data: The impact of landscape heterogeneity, misclassification, and temporal resolution. International Journal of Remote Sensing. 2018;39(24):8634-47. doi:10.1080/01431161.2018.1526470.
25. Klein JA, Myers AM. Advances in remote sensing for agricultural monitoring: Addressing challenges and future directions. Remote Sensing of Environment. 2020;249:112030. doi:10.1016/j.rse.2020.112030.