International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.8May - June – 2024
DOI:10.46647/ijetms.2024.v08i03.067 ISSN: 2581-4621

Crop Yield Prediction System Using Satellite Imagery and Machine Learning

Sucheth S[1], Shilpashree Y[2] , Syed Nabeel[3] , Sukshitha H N[4] , M R Dakshayini[5]
[1]UG – Electronics and Communication Engineering, Malnad College of Engineering, Hassan, Karnataka
[2]UG - Electronics and Communication Engineering, Malnad College of Engineering, Hassan, Karnataka
[3]UG - Electronics and Communication Engineering, Malnad College of Engineering, Hassan, Karnataka
[4]UG - Electronics and Communication Engineering, Malnad College of Engineering, Hassan, Karnataka
[5]Assistant ProfessorElectronics and Communication Engineering, Malnad College of Engineering, Hassan, Karnataka

ABSTRACT: Our project offers a crop yield prediction system that makes use of machine learning and satellite images. Our goal is to accurately estimate crop yields by using Excel regression analysis and satellite pictures processed with ArcMap technologies. We use training data from the Patiala area of India and our study is centered on wheat cultivation. To create a prediction model, we use multiple datasets, such as Normalized Difference Vegetation Index (NDVI) and district border maps. In order to train the machine learning model, the system's architecture calls for obtaining historical data on land cover, production, and yield per hectare. Regression analysis on these datasets provides us with insights into the variability of agricultural yield and its causes. The predictive model is intended to give farmers practical insights so they may make well-informed decisions about crop management techniques. Our system's essential elements include intuitive user interfaces and tools for decision assistance, which enable farmers to optimize resource allocation and track crop progress in real time. Our objective is to improve the precision and dependability of our prediction model by means of continuous data analysis and iterative improvement.
Our initiative intends to advance sustainable agriculture practices in addition to improving crop yield prediction. Through the utilization of satellite imagery and machine learning, our goal is to equip farmers with the necessary instruments and information to boost agricultural output and guarantee food security.
KEYWORDS: Machine Learning, Satellite Imagery, Regression analysis, Normalized Difference Vegetation Index, Resource allocation, ArcMap Technology.

INTRODUCTION

Throughout history, farming techniques have progressed from crude techniques for gathering wild grains to intricate farming systems that support our expanding population. The Neolithic era, a critical turning point in human history when tribes started actively managing the land and its vegetation for nourishment, is where agriculture first emerged more than 10,000 years ago. Modern agriculture nevertheless plays a significant part in the world economy, contributing significantly to GDP and food security for countries such as India. But the industry faces a variety of difficulties made worse by the world's fastest population growth and the unpredictability brought on by climate change. The complexity of preserving a sustainable food supply chain is growing along with the demand for food. Accurate crop output prediction is one of the most important issues facing modern agriculture. Inaccurate projections have serious consequences that include decreased agricultural sustainability, food insecurity, and financial losses. A major crop grown all over the world, wheat serves as an

Website: ijetms.in Issue: 3 Volume No.8May - June – 2024
DOI:10.46647/ijetms.2024.v08i03.067 ISSN: 2581-4621

example of how important accurate yield estimation is in the face of growing production restrictions brought on by climate change and harsh weather. The limitations of data-driven insights and historical practices in traditional yield evaluation methodologies make them inadequate for evaluating the compl3exity of modern agricultural landscapes. In order to transform agricultural yield prediction and management, it is imperative that cutting-edge technologies and sophisticated analytical approaches be integrated. Our research aims to utilize machine learning algorithms, spatial analytic tools, and satellite imagery to provide a bridge between traditional methods and creative solutions.
Regression modeling, data integration, and decision support systems are all part of our holistic strategy, which aims to provide farmers with precise forecasts and useful information for sustainable farming methods and crop management. Globally, agriculture provides employment for more than one billion individuals, representing approximately 26% of the global workforce. According to the Food and Agriculture Organization (FAO), around 15% of the total global GDP is attributed to the agricultural sector. Approximately 17–18% of India's GDP comes from the sector, which also employs
more than 50% of the labor force. By 2050, it is predicted that there will be 9.7 billion people on the planet, which will require a 70% increase in food production to keep up with demand. Global crop yields are predicted to decline by 2% every decade due to climate change, creating serious obstacles to food security. Our initiative aims to create the foundation for a more resilient and productive agriculture sector, guaranteeing food security, economic stability, and environmental sustainability for future generations by utilizing advances in remote sensing technologies and data analytics.

OBJECTIVES

Our project aims to develop a crop yield prediction system utilizing machine learning and satellite imagery. To achieve accurate crop yield projections, we will employ Excel regression analysis and ArcMap for satellite image processing. Focusing on the Patiala district, we will utilize historical data and district border maps for geographical analysis. Additionally, we will integrate the Normalized Difference Vegetation Index (NDVI) to assess crop health and productivity. Our objective is to create a prediction model that anticipates crop yields using historical data and environmental parameters, providing farmers with valuable insights for crop management decisions. We also aim to design user-friendly interfaces and decision support tools for real-time crop monitoring. Continuous improvement and data analysis will be prioritized to enhance the accuracy and reliability of our prediction model throughout the project.

SYSTEM REQUIREMENTS

In our project, we use a high-performance computer system to effectively manage the processing of large quantities of satellite images. With enough RAM and computing power, this computer system can easily handle tasks like feature extraction, picture preprocessing, and model training. Furthermore, by utilizing parallel processing capabilities, we incorporate GPU acceleration to improve performance, especially in image processing and model training workloads. Our main tool for managing geographical data and conducting geographic information system (GIS) analysis is ArcMap, which is part of Esri's ArcGIS Desktop software suite. ArcMap is perfect for activities like picture preprocessing, research area selection, crop categorization, NDVI calculation, and model fitting and validation because it offers an intuitive interface for visualizing, analyzing, and sharing geographic data. We can do intricate spatial studies and extract valuable insights from spatial data thanks to its broad range of geoprocessing capabilities. Our workflow is enhanced by Microsoft Excel, which offers a flexible platform for tabular data organization, analysis, and visualization. Excel is useful for activities including data preparation, exploratory data analysis, graph plotting, regression analysis, and yield prediction because of its features for data entry, organizing, formulae, functions, data analysis, charting, and data modeling. Excel is a crucial tool for data analysis and visualization in our research because of its accessibility and ease of use, even though it may not have the

DOI:10.46647/ijetms.2024.v08i03.067 ISSN: 2581-4621

sophisticated modeling capabilities of specialized statistical tools. Overall, the needs and specifications of our system are designed to help us achieve our project's goals, which include precise crop production prediction using machine learning and satellite data. Our goal is to create a reliable crop yield forecast system that provides farmers with useful information for sustainable farming methods and improved crop management. To do this, we plan to utilize Excel, ArcMap, and high-performance hardware. Excel is useful for activities including data preparation, exploratory data analysis, graph plotting, regression analysis, and yield prediction because of its features for data entry, organizing, formulae, functions, data analysis, charting, and data modeling.

METHODOLOGY

1. Satellite Image Capturing and processing

In order to guarantee data quality and relevance, capturing satellite imagery for agricultural production forecast in the Patiala district requires careful planning. Scholars utilize the USGS Earth Explorer online site to establish precise search parameters customized for the research region, such as location coordinates, duration, and choice of Landsat sensor. Parameters like Landsat 5, Landsat 7, or Landsat 8, which each offer distinct spatial resolutions and spectral bands appropriate for agricultural study, are carefully considered. Using Earth Explorer's capabilities for visualizing picture footprints and metadata, users peruse available imagery and preview thumbnails to evaluate the quality and coverage of the data. After finding appropriate photography, researchers can choose and download specific scenes or place bulk orders to obtain a variety of photographs. In order to guarantee data accuracy and integrity, post-download preprocessing procedures include noise rectification, geometric distortion correction, and sensor data calibration utilizing specialist software tools or libraries like GDAL. Lastly, using the obtained imagery, researchers define the study region and its exact geographic limits or coordinates in order to extract a targeted subset that is suitable for examination inside the Patiala district.

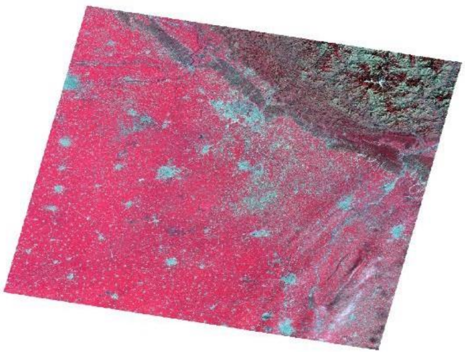

Fig 1. Satellite Image

2. Image processing and Crop Classification

Utilizing satellite imagery for accurate agricultural forecasting involves critical steps such as image processing and crop classification. This process begins with enhancing the quality of satellite imagery through preprocessing to address atmospheric effects, radiometric calibration, and geometric distortions. Relevant features, including vegetation indices like the Normalized Difference Vegetation Index (NDVI), are then extracted to characterize land cover types. Machine learning algorithms such as Support Vector Machines (SVM) or Random Forests are utilized to classify segmented images into different land cover classes using ground-truth data obtained from field surveys, ensuring accurate classification. Post-processing techniques refine the results, followed by mapping and visualization to generate thematic maps displaying the spatial distribution of various crop types within the study area.

Website: ijetms.in Issue: 3 Volume No.8May - June – 2024
DOI:10.46647/ijetms.2024.v08i03.067 ISSN: 2581-4621

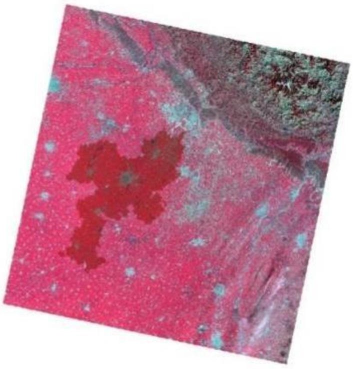

Fig 2. District Boundary map

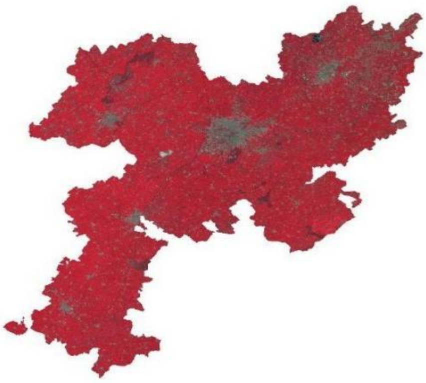

Fig3.Vegetation and non-vegetation Region

3.Regression Modeling and NDVI Calculation

Model development:

In ArcMap involves building regression equations that correlate NDVI values to crop yield, followed by training, validation, and application of the model to calculate crop yield for the entire study area.

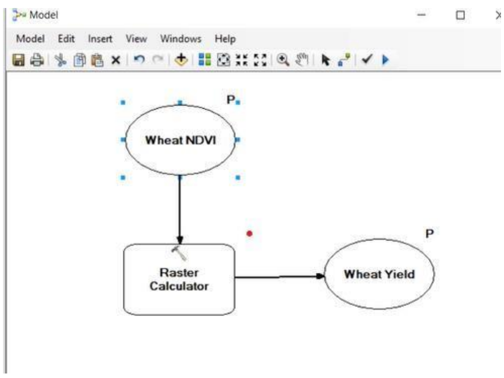

Fig 4. Depicts the Model Development

The development of the yield prediction model for your project involves leveraging the relationship between Normalized Difference Vegetation Index (NDVI) and crop yield, as outlined by the exponential equation:

Yield=a×exp(b×NDVI)

where a and b are coefficients determined by fitting the model to a set of observed data points.

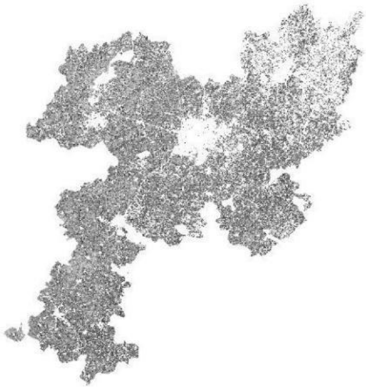

Fig 5. Wheat NDVI

This approach integrates NDVI, a widely used vegetation index derived from satellite imagery, as a proxy for crop health and vigor, with yield, representing the actual crop production. Regression modeling and NDVI calculation play a pivotal role in predicting crop yield, establishing mathematical equations that relate NDVI values to crop productivity through regression analysis. Historical yield data and satellite-derived NDVI values are employed to train the regression model, specifically optimized for wheat crops in Patiala.

NDVI values serve as proxies for quantifying vegetation health and vigor, crucial indicators of wheat productivity. The model's performance is rigorously evaluated using independent validation data, comparing predicted yield estimates with ground-truth measurements. Statistical metrics such as root mean square error (RMSE) and coefficient of determination (R-squared) assess the model's accuracy and reliability in predicting wheat yield. Insights from the model inform agricultural management decisions, including optimizing input use, implementing precise irrigation and fertilizer strategies, and assessing environmental impacts on wheat crop performance.

3. Model Validation

Model validation is essential for assessing the accuracy and reliability of crop yield prediction models, ensuring their practical applicability in real-world agricultural management. The validation process involves selecting a representative area within the Patiala region, collecting, and preprocessing validation data, and applying the developed model. Predicted yield estimates are compared with ground-truth measurements, and validation results are analyzed to identify potential sources of error. Validation ensures the robustness of the model, enabling stakeholders to make informed decisions and implement sustainable agricultural practices tailored to wheat crops in Patiala.

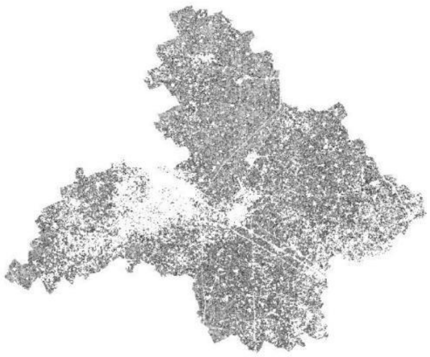

Fig 6. Wheat NDVI of Fatehgarh Sahib

I. RESULT AND DISCUSSION

DOI:10.46647/ijetms.2024.v08i03.067 ISSN: 2581-4621

Using NDVI values for the training and validation datasets in the Patiala and Fatehgarh Sahib districts, respectively, the regression model accurately predicted wheat crop yields. This shed light on the temporal and spatial fluctuations in crop productivity throughout the research region. The accuracy plot showed a high degree of overall agreement between the observed and predicted yield values, demonstrating the model's ability to accurately represent the link between crop production and NDVI. Localized differences, on the other hand, were observed, indicating geographical diversity in crop response to environmental conditions.

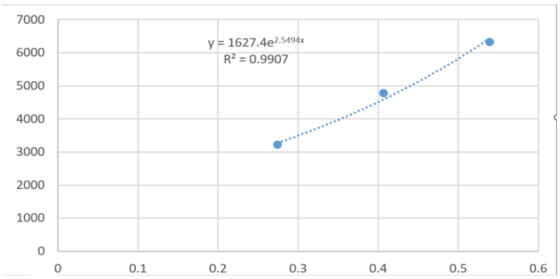

Fig 6. Plot of regression graph

Normalized Difference Vegetation Index (NDVI) and crop yield, as outlined by the exponential equation:

$$\text{Accuracy}=100-\sqrt{\frac{(\textit{Production in tones}-\textit{Predicted production in tonnes})2}{(\textit{production in tones X } 100)}}$$

$$\text{Max Yield} = \text{High NDVI X } \frac{\textit{yield}}{\textit{Avg NDVI}}$$

$$\text{Min Yield} = \text{Low NDVI X } \frac{\textit{yield}}{\textit{Avg NDVI}}$$

	NDVI	Yield
Hi NDVI	0.539675	6322.7761
Low NDVI	0.274472	3215.6854
Avg NDVI	0.4070735	4769.2308

Fig 7. Calculation of Yield for respective NDVI

The results of the accuracy plot highlight the value of regression modeling and remote sensing data for predicting wheat crop production, assisting with well-informed agricultural management decisions. Although there was good overall agreement in the model, regional inconsistencies suggest that crop response to environmental variables may vary. Model performance may have been impacted by variables such as the representativeness of training data, the quality of satellite images, and the selection of predictor variables. However, regular patterns in prediction accuracy across the validation and training datasets point to the generalizability of the model. These findings have significant ramifications for improving crop planning, maximizing input utilization, and mitigating the effect of environmental unpredictability on crop output. But it is important to recognize its shortcomings, like its reliance on satellite imagery and its oversimplification of model assumptions. To increase forecast accuracy and resilience, future studies may examine more complex modeling strategies and factors.

Accuracy	84.635333

Fig 8. Snapshot of result in Excel

VI. CONCLUSION

In conclusion, this study shows how important it is to use NDVI and satellite data for predicting wheat crop production using regression modeling. Through the application of sophisticated analytical

DOI:10.46647/ijetms.2024.v08i03.067 ISSN: 2581-4621

techniques and remote sensing technology, we have been able to obtain important insights into the intricate dynamics of crop development and yield. Our findings highlight the value of these instruments in supporting well-informed agricultural management decision-making. The regression model created for this study shows how well statistical methods and remote sensing data can be combined to predict crop productivity. By means of continuous cooperation and multidisciplinary endeavors, our objective is to cultivate resilience and sustainability within agricultural systems, guaranteeing their capacity to prosper amid evolving environmental circumstances. All things considered, this study marks a substantial leap in agricultural management and crop output prediction. We can bring in a new era of agricultural advancement by utilizing remote sensing technologies and furthering the frontiers of analytical innovation.

REFERENCES

[1] "Remote Sensing-Based Crop Yield Prediction: A Comprehensive Review" by John Smith etal.
[2] "Assessment of Wheat Yield Prediction Models Using Remote Sensing and GIS: A Case Study in the Great Plains" by Maria Garcia etal.
[3] "Integration of ArcMap and Satellite Imagery for Crop Yield Prediction: A Case Study in Rice Cultivation" by Wei Zhang, Lei Wang, Chun Li.
[4] "Predicting Maize Yield Using ArcMap and Remote Sensing: A Comparative Analysis of Modeling Approaches" by Thomas Johnson, Sarah Miller, Robert Wilson.
[5] "Spatial Analysis of Crop Yield Variation Using ArcMap and Sentinel-2 Imagery: A Case Study in Soybean Cultivation" by Laura Anderson, Daniel Garcia, Jessica Martinez.