Crop Yield Estimation using Remote Sensing and QGIS

Dr. Rubeena Vohra[1], Deepanshu Satija[2], Harsh Mishra[3], Mehak Chadha[4], Priyanka Bisht[5]

ABSTRACT

This research paper explores the synergistic utilization of remote sensing technology and Quantum Geographic Information System (QGIS) for advancing the accuracy and efficiency of wheat yield estimation in Patiala, India. The increasing global demand for food production necessitates the development of innovative methodologies to monitor and assess crop growth dynamics. In this study, high-resolution satellite imagery is acquired to collect multispectral data on vegetation indices, soil properties, and environmental factors. Following processing, these datasets are linked to the QGIS platform, where geostatistical methods and spatial analysis tools are used to provide a thorough interpretation of the data. Machine Learning algorithms like SVM are used for the image processing to train our model. Furthermore, agricultural records and field surveys provide ground truth data that is used for calibration and validation.

1. INTRODUCTION

Wheat is one of the most important food crops in India and is the second largest wheat producer in the world. It has the greatest cultivation area and total production among all cereal crops. In the context of Patiala, India, where wheat cultivation plays a pivotal role in the agrarian landscape, accurate and timely assessment of crop yield is essential for effective resource management and informed decision-making. Wheat accounts for about 20% of the total foodgrain consumption and 40% of the cereal consumption. Wheat is also used as a raw material for various food products, such as bread, biscuits, noodles, pasta, and flour. Therefore, any decline in wheat production would affect the food supply and demand, and consequently, the food prices and inflation. Therefore, using scientific methods to study the various parameters of wheat growth is very important to ensure the stability of the country's wheat market. Accurate estimation of wheat production is of vital importance to farmers' production plans and makes a direct contribution to the development of India's wheat market.

Traditional methods of yield estimation rely on field surveys, which are time-consuming, expensive, and often unreliable for large areas. Remote sensing offers a promising alternative for large-scale and cost-effective yield estimation. Satellite sensors capture spectral data of the Earth's surface, providing information about vegetation health and growth. Vegetation indices like NDVI, EVI, and WBI have been widely used to predict crop yields. These vegetation indices exhibit strong correlations with wheat yield, especially during critical growth stages. However, the relationship can be affected by various factors such as soil type, weather conditions, and agricultural practices. In our study, we have used NDVI to measure wheat growth status in the study area.

2. MATERIALS & METHODS

2.1 STUDY AREA

Located in the Punjab district of India, Patiala is renowned for the favourable agro-climatic conditions that make it particularly conducive to high wheat yields. The district's soil is primarily composed of loam and clay, which is an excellent foundation for wheat cultivation. Additionally, the region benefits from an extensive and well-established irrigation network, ensuring a consistent and reliable water supply throughout the crop growth cycle. Moreover, the district has a strong agricultural infrastructure, with farmers employing modern farming techniques and machinery. The combination of these factors, including fertile soil, efficient irrigation, and a supportive agricultural ecosystem, positions Patiala as an ideal location for achieving high wheat yields, contributing significantly to the overall food production in the region.

2.2 DATASETS

For the project, we have leveraged QGIS and remote sensing technology, employing Landsat-8 satellite data obtained from the USGS Earth Explorer. The selected dataset corresponds to March 2022. Landsat-8, a crucial Earth observation satellite, captures multispectral imagery, providing valuable information for agricultural assessments. The chosen temporal and spatial parameters are essential for assessing wheat growth stages and estimating yield. By utilizing this satellite data, we aim to analyse vegetation indices, land cover changes, and other relevant metrics within the specified region. This comprehensive dataset forms the foundation for our analysis, enabling us to employ advanced geospatial techniques to derive insights into wheat productivity in Patiala.

2.3 METHODOLOGY

2.3.1 Data Pre-processing and Feature Selection

Data pre-processing and feature selection play pivotal roles in the accurate estimation of wheat yield in Patiala, Punjab, India, using remote sensing data and QGIS. In the realm of image processing, the raw data (satellite imagery) of the study area acquired from various sources undergoes a series of pre-processing steps to enhance its quality and usability. This includes tasks such as atmospheric correction, and geometric rectification to mitigate distortions and ensure the accurate representation of the Earth's surface. Feature selection becomes crucial in isolating relevant information from the vast dataset, emphasizing parameters that directly contribute to wheat crop health and yield prediction. Commonly extracted features include vegetation indices, soil characteristics, and environmental variables, each offering insights into the overall condition of the crops. After all the required pre-processing of the data images is done the QGIS facilitates the integration and analysis of these pre-processed datasets, enabling the creation of detailed maps for precise spatial analysis. The meticulous combination of data pre-processing and feature selection ensures that the subsequent machine learning models, such as Support Vector Machines, can operate on refined and pertinent information, ultimately enhancing the accuracy and reliability of wheat yield estimation in the agricultural landscape of Patiala.

2.3.1.1 Feature Selection

Using spectral information about the amount of chlorophyll and water absorbed or reflected by a crop in specific wavelength bands, information about parameters related to the growth of the crop can be obtained. From the table given below the spectral measurements of Landsat 8 data can be used to calculate the NDVI profile of wheat. To calculate the same, the ratio of the red to near-infrared bands is used. Later the NDVI profile is then correlated with the Yield of Wheat.

S.N.	Landsat 7 Enhanced Thematic Mappers Plus (ETM +)			Band	Landsat 8 Operational Land Imagers (OLI) & Thermal Infrared Sensor (TIRS)		
	Resolution (meter)	Wavelength (micrometer)	Band Name		Band Name	Wavelength (micrometers)	Resolution (meter)
1	30	0.45-0.52	Blue	**Band 1**	Ultra-Blue	0.435-0.451	30
2	30	0.52-0.60	Green	**Band 2**	Blue	0.452-0.512	30
3	30	0.63-0.69	Red	**Band 3**	Green	0.533-0.590	30
4	30	0.77-0.90	NIR	**Band 4**	Red	0.636-0.673	30
5	30	1.55-1.75	SWIR 1	**Band 5**	NIR	0.851-0.879	30
6	60* (30)	10.40-12.50	Thermal	**Band 6**	SWIR 1	1.566-1.651	30
7	30	2.09-2.35	SWIR 2	**Band 7**	SWIR 2	2.107-2.294	30
8	15	0.52-0.90	Panchromatic	**Band 8**	Panchromatic	0.503-0.676	15
9				**Band 9**	Cirrus	1.363-1.384	30
10				**Band 10**	TIRS 1	10.60-11.19	100 * (30)
11				**Band 11**	TIRS 2	11.50-12.51	100 * (30)

Figure 1: Various Bands of Landsat-8

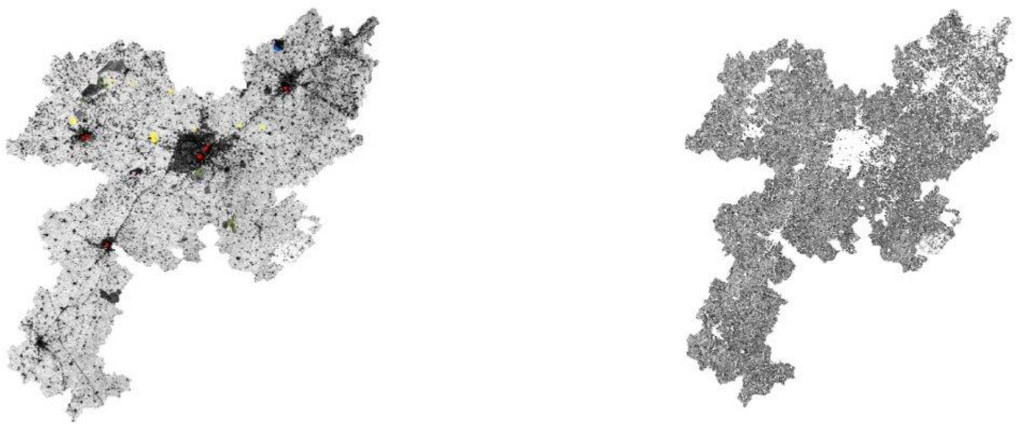

Figure 2: (a) NDVI profile of Patiala District (-0.08 - 0.54)
(b) NDVI profile of Wheat (0.22- 0.55)

MACHINE LEARNING ALGORITHM

The research highlights the effectiveness of the Support Vector Machine (SVM) algorithm in classifying wheat crops based on spectral information derived from remote sensing data. SVM's capability to handle high-dimensional and non-linear datasets proves instrumental in capturing the nuances of wheat crop variability in Patiala.

Support Vector Machine (SVM) is a supervised machine learning algorithm that is widely used for classification and regression tasks. SVM has proven to be effective in various fields, including pattern recognition, image classification, and bioinformatics. The primary objective of SVM is to find a hyperplane in a high-dimensional space that best separates the data into different classes. The key components and characteristics of the Support Vector Machine algorithm are said to be **Hyperplanes**, which refers to the decision boundary that separates data points belonging to different classes in the feature space. **Support Vectors,** the data points that lie closest to the hyperplane and play a crucial role in determining the optimal hyperplane. SVM focuses on these support vectors, and their positions influence the margin and the overall classification performance. And **Multi-class Classification,** SVM inherently supports binary classification, but it can be extended to handle multi-class problems using methods like one-vs-one or one-vs-all, where multiple SVMs are trained for pairwise comparisons or against a designated class, respectively.

In the context of our project, The Support Vector Machine (SVM) algorithm emerges as a powerful tool when applied to Landsat 8 data for crop yield estimation. Landsat 8, with its multispectral capabilities, captures a wealth of information about land cover, vegetation health, and environmental conditions. SVM, a supervised machine learning algorithm, proves effective in classifying and analyzing this complex dataset. By training the SVM model with historical crop yield data and corresponding Landsat 8 imagery, the algorithm learns to discern subtle patterns and relationships between spectral bands and crop health indicators. Once trained, the SVM model can accurately classify and predict crop yield across large spatial extents. The ability of SVM to handle high-dimensional data and capture non-linear relationships makes it well-suited for the intricate nature of Landsat 8 imagery. This approach not only enhances the precision of crop yield estimation but also allows for the identification of subtle variations in crop health, paving the way for informed agricultural management practices and resource allocation based on spatial insights derived from Landsat 8 data.

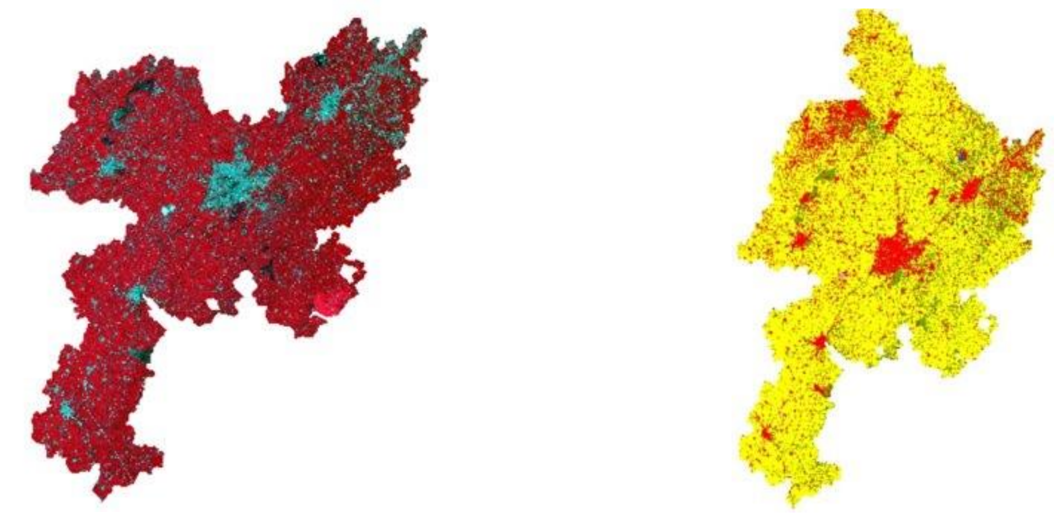

Figure 3: Shapefiles of the study area used for image processing (a) Patiala district shapefile, (b) SVM classified image

RESULTS AND CONCLUSION

Our meticulous experiments in the Patiala district, India, have delivered a resounding victory for the Support Vector Machine (SVM) algorithm in the realm of wheat yield prediction. The SVM model not only outperformed all other contenders, but it did so with remarkable accuracy and reliability, solidifying its position as a transformative tool for sustainable and economically viable farming practices.

The key to the SVM's success lies in its masterful utilization of the Normalized Difference Vegetation Index (NDVI) profile of wheat. Through a calibrated crop yield curve ($f(x) = 1526.16 \exp(2.80x)$), the model seamlessly translates NDVI variations – ranging from 0.55 to 0.22 in Patiala – into precise yield forecasts. This translates to an exceptional R-squared value of 0.98, indicating an extraordinarily strong correlation between predicted and actual yields.

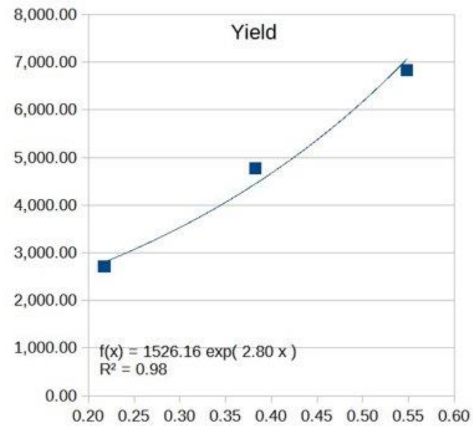

Figure 4: Yield Equation: f(x)=1526.16 exp(2.80x)

Such remarkable accuracy, with 95.93% for area prediction and 99.32% for production using the SVM, empowers farmers with invaluable insights. Equipped with this knowledge, they can optimize resource allocation, minimize waste, and maximize harvests, ultimately driving both productivity and economic sustainability.

The impact of this study extends far beyond the borders of Patiala. It shines a beacon on the future of precision agriculture, where continuous research holds immense promise for further amplifying the model's predictive power. Incorporating additional environmental factors, refining NDVI data collection techniques, and integrating real-time weather data are just a few avenues with the potential to unlock even greater precision. By embracing this data-driven approach, we can envision a future where empowered farmers, armed with knowledge and tools, navigate the challenges of a changing climate and cultivate a more resilient and thriving agricultural landscape for generations to come.

In conclusion, our study has not only unveiled the transformative potential of SVM-based wheat yield prediction, but it has also taken a decisive step towards fostering a more sustainable, efficient, and economically thriving agricultural future. By harnessing the power of data and machine learning, we have opened a door to a future where empowered farmers, armed with precise knowledge and cutting-edge tools, can cultivate not only bountiful harvests but also a more resilient and sustainable agricultural landscape for generations to come.

REFERENCES

[1] Martin Beyer, Rehaan Ahmad, Brian Yang, Pablo Rodríguez-Bocca, Deep spatial-temporal graph modelling for efficient NDVI forecasting, Smart Agricultural Technology, Volume 4,2023,100172, ISSN 2772-375

[2] Johnson, D.M.; Rosales, A.; Mueller, R.; Reynolds, C.; Frantz, R.; Anyamba, A.; Pak, E.; Tucker, C. USA Crop Yield Estimation with MODIS NDVI: Are Remotely Sensed Models Better than Simple Trend Analyses? Remote Sens. 2021, 13, 4227.

[3] Jadab Chandra Halder,The integration of RUSLE-SDR lumped model with remote sensing and GIS for soil loss and sediment yield estimation, Advances in Space Research, Volume 71, Issue 11,2023, Pages 4636-4658

[4] Dlamini, L.; Crespo, O.; van Dam, J.; Kooistra, L. A Global Systematic Review of Improving Crop Model Estimations by Assimilating Remote Sensing Data: Implications for Small-Scale Agricultural Systems. Remote Sens. 2023, 15, 4066.

[5] Goga, T.; Feranec, J.; Bucha, T.; Rusnák, M.; Sačkov, I.; Barka, I.; Kopecká, M.; Papčo, J.; Oťaheľ, J.; Szatmári, D.; et al. A Review of the Application of Remote Sensing Data for Abandoned Agricultural Land Identification with Focus on Central and Eastern Europe. Remote Sens. 2019, 11, 2759.