Proceedings of 13th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2023 Month: April Volume: 13

Paddy Crop Yield Modeling and Estimation: A Case Study of Bandipur Rural Municipality, Nepal

Milan K.C [a], Krishna Prasad Bhandari [b], Bikash Sherchan [c]

[a, b, c] *Department of Geomatics Engineering, Pashchimanchal Campus, Institute of Engineering, Tribhuvan University, Nepal*
[a] kcmilan28@gmail.com, [b] bhandarikrishna@wrc.edu.np, [c] bikash.sherchan@wrc.edu.np

Abstract

Being Nepal's primary crop, rice is crucial for both food security and nutrition. For this reason, monitoring, forecasting, and prediction are crucial for organizations working on food security, both governmental and non-governmental. Using remote sensing images and NDVI data, it is possible to evaluate the development and health of crops and develop models for predicting agricultural yields. This study focuses on monitoring paddy fields in Nepal's Seratar and Nahala land pooling regions using Sentinel 2 photos and NDVI series derived every five days. In the study, the link between the NDVI values and a number of land management factors—including water table, soil type, fertilizers, transplanting date, and sowing date—is investigated. NDVI values are collected from many sentinel pimages using Google Earth Engine to expedite data processing and analysis. Using time-series NDVI data, a regression model is developed to forecast rice production, finding the factors that have a high link to NDVI values. A socioeconomic analysis of the crop yield model in connection to Nepal and the rest of the world is also included in the research, underlining the necessity of planning and policies to maintain the distribution of the demand-supply chain for agricultural production.

Keywords

NDVI, paddy, regression, time series, yield

1. Introduction

Rice, sometimes referred to as paddy, is the second most widely produced grain crop in the world. Almost half of the world's population eats rice as their main staple food. Almost 70% of the population in Nepal is directly dependent on agriculture, which generates 21% of the country's gross domestic product (GDP). Up to the year 1985, Nepal was a net exporter of rice, and in the 1960s, it was shipping up to $45 million worth of rice to India every year [1]. The statistics dictate that Nepal would have to import 531,000 tons of rice from India in 2015 at a cost of almost US$210 million. According to the MoALD 2018/19 report, rice accounts for almost 70% of all arable land, which is just about 10% of Nepal's total land area. In Nepal, an average of 3.506 tons are generated per hectare. Despite the fact that educated youth and children of farmers are either fleeing or abandoning the countryside, agriculture remains the country's main economic sector in Nepal [2]. Young people, educated people, and farmers will continue to leave the agricultural field as a result of the lack of methodologies and mechanization, which will have a direct effect on agricultural production. Systems and mechanization hence necessitate revolution.

Accuracy and timing of rice monitoring and mapping is crucial on a worldwide scale [3]. The evaluation of yield, net production, and irrigation water supply requirements is made easier by mapping paddy crops. because local governments have historically used statistical data to decide the subsidies. Yet, there are often noticeable discrepancies between different datasets from different government ministries or organizations. Thus, it is vital to develop reliable techniques for obtaining accurate and reliable estimations of the area under rice cultivation in order to support governmental agricultural policy [4].

Land consolidation has the potential to be a key factor in altering the agricultural system by enabling the creation of more suitable and efficient field forms and sizes. Innovative technologies and approaches that can reduce the demand for human labor and raise industrial productivity can then be more easily included as a result. Also, the majority of men have relocated to urban areas or other countries in pursuit of employment, which has led to an increasing feminization of the farming industry. Hence, although being significant, the increase in rice production is very disappointing.

Monitoring vegetation is essential for the wellbeing of the environment as a whole because of the carbon that plants store in the ecosystems around them. Monitoring of rice crops enables assessments of a country's food security, planning, and sustainable resource management [5]. That is only possible with the use of remote sensing technologies. The world's rice fields have been mapped using remote sensing technology. The characteristics of RS data and the toolset provided by this technology, both of which have been present for a long, motivated a search for appealing non-destructive approaches to acquire crucial data for agriculture.

Using indices based on spectral features of vegetation within the optical area, the relative vegetation cover at different scales has been mapped and studied in several remote sensing investigations. Despite the fact that optical remote sensing is a typically efficient mapping technique, there are a number of limitations to its usage for vegetation distinction due to cloud interference, air attenuation, and other issues. The normalized differential vegetation index (NDVI) is the most widely used and is helpful for evaluating historical changes; nevertheless, it is sensitive to changes in the canopy background and saturates in areas with comparably high amounts of vegetation [6].

Nonetheless, the operational applicability of such technologies over S2 data has not yet been addressed in the context of developing countries which is precisely the gap that motivates this work. This paper proposes a methodology for paddy monitoring by extracting NDVI data from Sentinel 2 pictures acquired every five days. For understanding and analyzing the NDVI value of various paddy phenological stages with various land management factors, such as water level, degree of damage, type of soil, fertilizers used, date of transplantation, quantity of seed, sowing date, etc., the land pooling area of Seratar and Nahala in the Bandipur Rural Municipality of Nepal has been chosen as a study area.

2. Objectives and Study Area

The objective is set and categorized into two types.

General Objectives

- To develop a yield prediction model and cost-benefit analysis of the crop yield production from sentinel 2 imagery

Specific Objectives

- To compare the yield predictions by using land management factors and time series NDVI.

- To assess the gap between predicted yield and actual yield.

3. Methodology and Data Analysis

3.1 Study Area

Regarding the study area, it was selected Seratar of Bandipur Rural Municipality of Tanahun district. It is about 80km East from Pokhara and 12km South from the Dumre bazaar and extends within the latitude of $27^{0}55^{\prime}57.9^{\prime\prime}$ N latitude and $84^{0}22^{\prime}13.1$" E longitude in Gandaki province of Nepal.

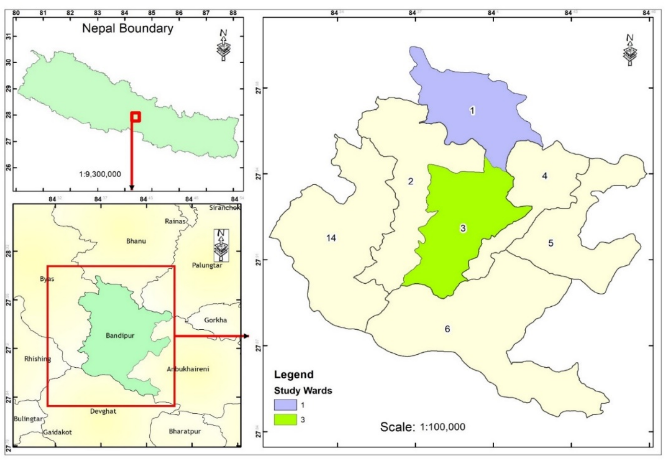

Figure 1: Study Area

3.2 Data and software used

Sentinel-2 imagery with a spatial resolution of 10 m was gathered between the dates of July 14 and November 28, 2021, and used to derive the NDVI time series for this investigation. Information on rice crop management and yield from the prior year (2021) was acquired from the farmers using interviews and questionnaires. A Garmin GPS map 62s device was used to trace the polygon of each farmer's parcel. A portable GPS was used to digitize the field with a 4m accuracy. Excel and the Power BI Desktop were used to analyze and visualize the data.

Records of yield for each parcel were compiled using data on land usage and the weight of the harvested rice crop in the field. Farmers provided 40 production reports total for the paddy fields in 2021. The remaining ten, designated as "Test Parcels," were used to evaluate the effectiveness of the models developed using the root mean squared errors (RMSE) values. Of these, thirty of them were employed in regression analysis as dependent variables to get yield models.

3.3 Image Processing and Classification

The NDVI time series for the research region is derived by Google Earth Engine using Sentinel-2 satellite photos as part of its machine learning technique. Theoretically, using remote sensing data or goods might offer a better knowledge of crop growth at different stages to create more precise production estimations and estimation. The NDVI values of the various rice stages in 2021 were determined using Sentinel 2 pictures of the research region. The start of the season (SoS), the peak of the season (PoS), and the end of the season (EoS), respectively, of the rice crop cycle are identified for the study's focal region as mid-July, mid-September, and mid-November. S2 sensors offer a total of 13 spectral bands with a spatial resolution ranging from 10 m to 60 m, as is well known. The traditional RGB and near-infrared (NIR) bands with a spatial resolution of 10 m are intended primarily for usage on land among these spectral bands.

The map of land use and cover was produced using the Landsat-8. The area that has been set aside for agriculture is separated and formed into a polygon. This agricultural polygon obscures the Landsat 8 mosaic image. The masked agricultural mosaic image was split into several rice kinds using a field sample. The approach used for classification is supervised classification. This map's objective is to look at how rice varieties are grown in the study region.

3.4 Data Preparation and Statistical Analysis

The information gathered from the farmers and taken from the Google Earth Engine platform was entered into an excel spreadsheet. The interview's data was standardized and coded. The recovered and processed hand-held GPS field polygons were used to build a database for the statistical analysis.

Using a simple linear regression (SLR) model with the formula y=a+bx, the correlations between data regarding the specific crop management and each NDVI and yield at various growth stages were originally explored. In the above equation, y is the expected value of the given x variable, b is the slope, and c is the constant.The value of y is predicted by the least square regression line.

When fitting regression models using stepwise regression, the selection of the predictor variables is done automatically. Based on a predetermined criterion, a variable is taken into consideration for addition to or exclusion from the collection of explanatory variables in each step. Stepwise backward regression was applied to create the model for this study. Only factors that firmly and significantly characterize the yield variability are chosen using this strategy. The exploration of unexpected coefficient signs and the identification of relationships between parameters involved several ”trial and error” attempts. From the yield data, a few samples were randomly chosen and set aside for model testing. The final set of data was then subjected to multiple regression. The paddy data set in this instance has 70 entries altogether. 35 records from the year 2021 were used as a training set, while the remaining 35 records from the year 2022 were used for validation.

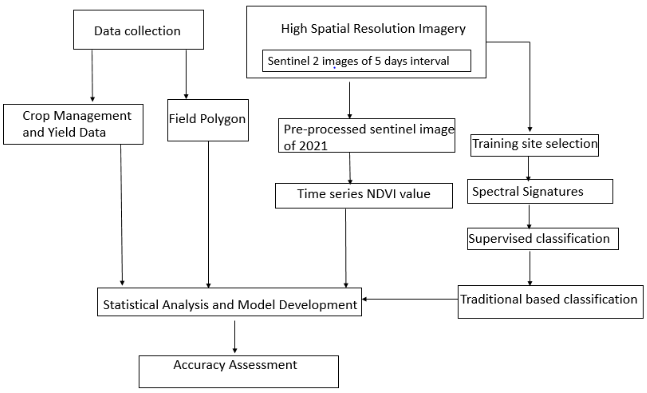

Figure 2: Methodology of the Study

4. Results

4.1 Land Use Land Cover and Agricultural Area Map

The majority of the study region is dominated by the agricultural sector. It covers an area of 43.96% of the entire land, while thin woodland, covering an area of 29.62% of the whole area, is in second place. Corresponding to this, residential and thick forest land have lower percentages of coverage than other types of land (23.52% and 2.90%, respectively).

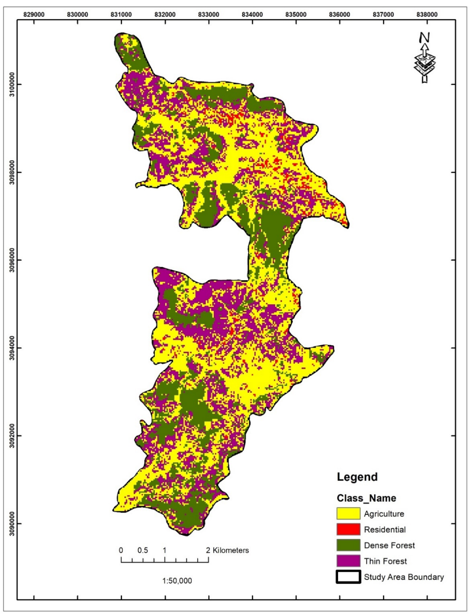

Figure 3: LULC Map of Study Area

Table 1: Land Use Land Cover

Class	Percentage (%)	Area (in Hectare)
Agriculture	43.96	1261.44
Residential	23.52	67.50
Dense Forest	2.90	8.30
Thin Forest	29.62	85.00

4.2 NDVI and Field Level Yield Data

The Normalized Difference Vegetation Indices have been produced at regular varied dates of picture of sentinel 2 from July to December 2021 in order to define and associate the yield data with the indices. The density of green on the plot of land may be calculated using NDVI. The computation made use of the red and near-infrared wavelengths. The respectable real field level yield and the time series NDVI have been connected. For each particular plot, the NDVI categories have been divided into maximum, median, and minimum NDVI. The term ”maximum NDVI” describes the highest NDVI value for any given pixel inside the plot. The median and minimum NDVI also refer to the median and minimum values in the plot’s pixel.

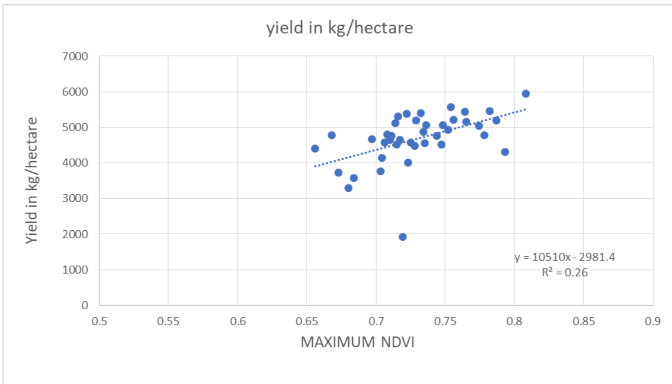

Figure 4: Maximum NDVI Vs Yield in kg/hectare

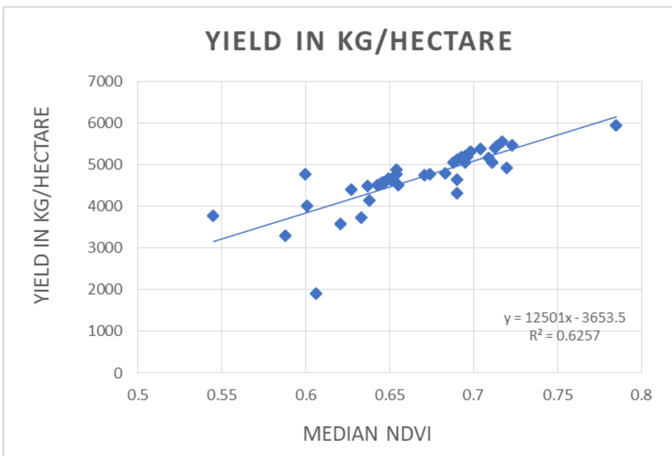

Figure 5: Median NDVI vs Yield in kg/hectare

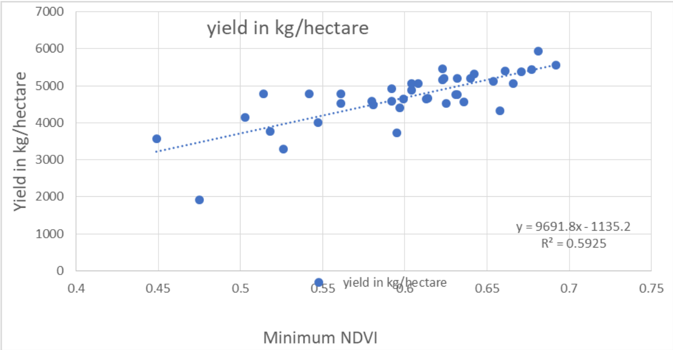

Figure 6: Minimum NDVI vs Yield in kg/hectare

4.3 NDVI and Growth Stages of Paddy

Vegetative, reproductive, and ripening stages make up the entire paddy development period. The paddy's life cycle, which lasts between 55 and 85 days, has a vegetative phase that lasts the longest. Around 30 days are spent in the reproductive phase, while 15 to 40 days are spent in the ripening phase.

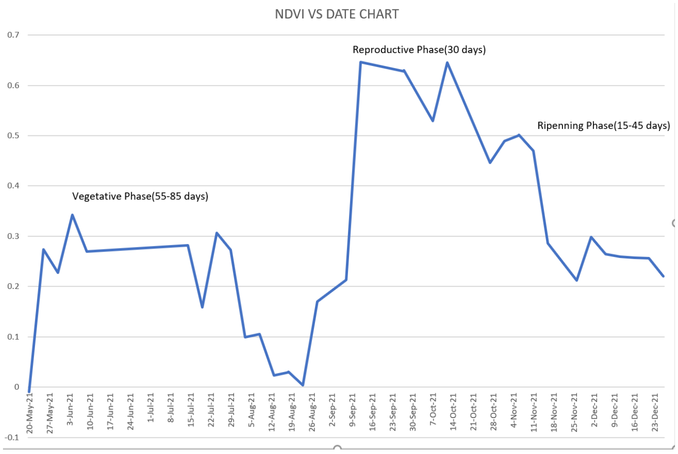

Figure 7: NDVI and different Phases of paddy

4.4 Impact of Land Management Factors on Yield and NDVI

4.4.1 Varieties of Paddy

Many regions of the nation cultivate different types of paddy. New hybrid kinds and very productive paddy seeds from accredited sources are also used nowadays. Yet the following various varieties sample were discovered in the study region.

Table 2: Variety of Paddy Grown

Varieties	Count	Yield(Kg/Hectare)	Min NDVI	Median NDVI	Diff. NDVI
Sabitri	21	4599.12	0.47	0.588	0.118
Makwanpure	6	4692.56	0.547	0.633	0.086
Ramdhan	13	4843.77	0.459	0.545	0.086

The table made it evident that the Sabitri variety of rice, with an average yield of 4599.12 kg/hectare, is a form of paddy that is extensively produced in the region of this research. In Nepal, a decent yield was thought to be between 3.5 and 4 tons/hectare on average. While the Makwanpure variety has a high NDVI rating, its average yield of 4692.56 kg/hectare was almost identical to that of Sabitri. Ramdhan is the second most often used variety in Seratar, Bandipur, and has the greatest average yield at 4843.77 kg/hectare. Comparatively low to the NDVI of the other two varieties is the Ramdhan.

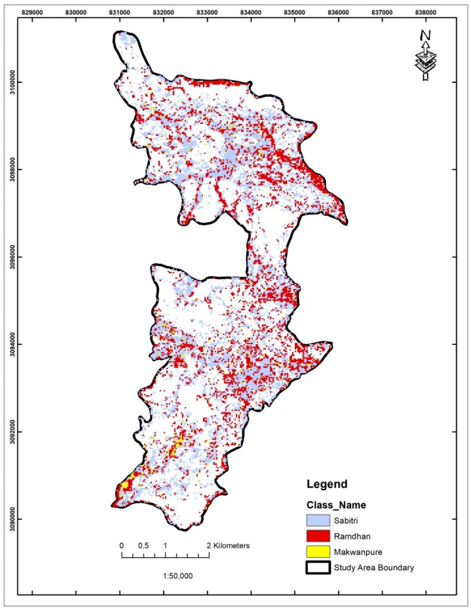

Figure 8: Rice Distribution Map

4.4.2 Damages due to Pest and Diseases

If we do not find a timely solution to pest and disease problems, they immediately impact agricultural productivity. Thus, the crop should be regularly observed. There are several sorts of pests that are prevalent in Nepal throughout the growing of the rice crop. These include the gall Mudge (also known as dhungrekira), leaf folder (also known as pat berne), brown plant hopper (also known as khairafadkekira), stem borer (also known as gawarokira), and others. And by consulting with professionals, these pests may be readily managed.

Table 3: Table Showing Damage's Impact on Yield and NDVI

Damages	Count	Yield (kg/Hectare)	Min NDVI	Median NDVI
High Damage	7	3981.33	0.563	0.549
Medium Damage	8	4543.83	0.561	0.631
No Damage	15	4901.37	0.642	0.674

4.4.3 Date of Transplation

Nepalese farmers typically transplant paddy rice when the seedlings are about 30 days old. The majority of farmers transplanted between the 80th and 110th day of the year, as shown in the figure, with sowing dates ranging from as early as the 20th jestha to the 15th Asad in 2021.Each field's transplantation time depends on a variety of variables, such as manpower availability, water level, etc. In our nation, the majority of transplants were performed manually. Also, we could not discover any statistically significant impact on yield due to different transplanting dates in the regression test.

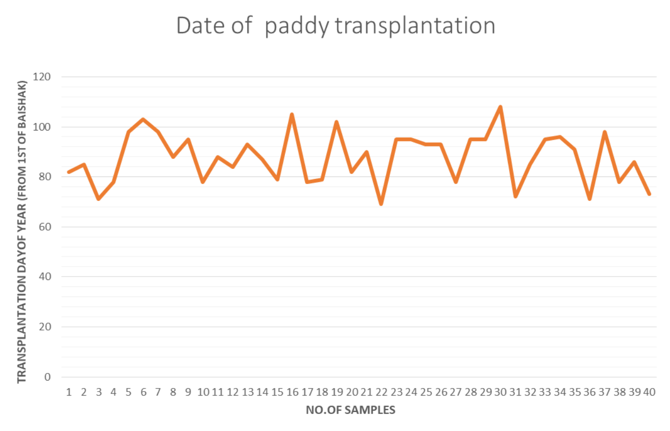

Figure 9: An illustration of the timing of paddy transplation

4.4.4 Date of Harvesting

According to the data, most farmers harvested paddy between 218 and 234 days starting on the first of Bhaishak, with an average yield of 4922.31 kg/hectare and an NDVI of 0.68, which is the highest value in comparison to the other harvesting days. This demonstrates unequivocally that the 218–234 days of the year are ideal for rice harvesting in the studied region. We calculate the correlation coefficient as r=0.653, which indicates a moderate association between the date of harvesting and production. It follows that moving up the harvest date can result in a smaller increase in crop output.

Table 4: Table showing the date of harvesting

DOY	Count	Average Yield	Average NDVI
192-200	6	4185.85	0.62
202-210	13	4864.26	0.67
211-217	12	4580.10	0.67
218-234	9	4922.31	0.62

4.4.5 Effect of Organic Manure in Yield

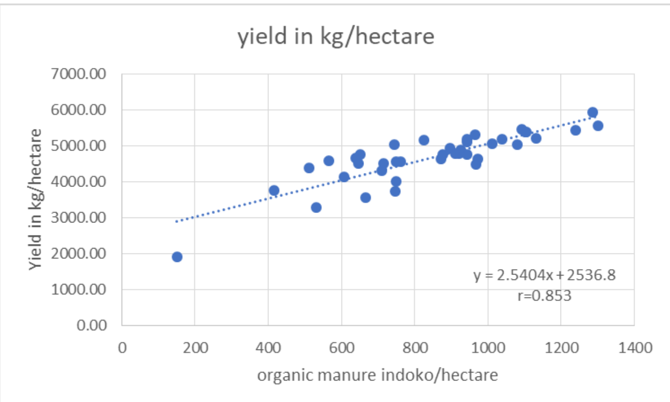

Figure 10: Organic manure vs yield in kg/hectare

It is evident from this graph that the amount of organic manure applied and the yield in kg/hectare are directly related. And the correlation coefficient (r=0.853) shows that there is a significant association between the amount of organic manure applied (in doko/hectare) and the yield (in kg/hectare).

4.5 Model Development

In the case of rice crop yield estimation, taking into account remote sensing data may logically give a better understanding and knowledge of seasonal crop growth in order to provide more accurate yield projections. Multiple regression analysis is used to determine whether there is a statistically significant association between sets of data and to spot patterns in those sets of variables. The relationship is demonstrated here in a way that demonstrates how crop yield depends on at least two crop parameters, and maybe more, which were presented here as mean NDVI.

In order to choose the optimal subset that best represents the field level NDVI variability, all relevant parameters were included into the stepwise multiple regression. Two factors were chosen as predictors explaining 96.30% (R2adj) of yield variability after much trial and error. The major goal of this was to determine which variables may account for the yield variability in order to prevent autocorrelation from occurring while the final model was being built.The regression equation can be written as:

Yield=1.004 X (OM)-395.979 X (Damage high) + 3147.7 X (Median NDVI) + 3073.46 X (Min NDVI)

where,
Yield= predicted yield in kg/hectare,
OM= Organic Manaure in doko/hectare,
Damage high = Damage caused by the various factors in high level , Value is 1 if field have high damage else value is zero,
Median NDVI = Aggregated medium NDVI value,
Min NDVI= Aggregated minimum NDVI value.

This model suggested that the high damage to paddy caused by illnesses has a detrimental impact on yield. And only when farmers in the study region utilized organic manure at a higher rate than other fertilizers would paddy be harvested.

Table 5: Model Development Parameters

Predictor	Coefficients	Standard Error	t -Stat	P-value
Intercept	0			
Organic manure	1.004	0.3022	3.324	0.0023
Damage high	-395.979	143.08	-2.7674	0.0094
Median NDVI	3147.696	1484.9	2.119	0.042
Min NDVI	3073.463	1499.9	2.049	0.049

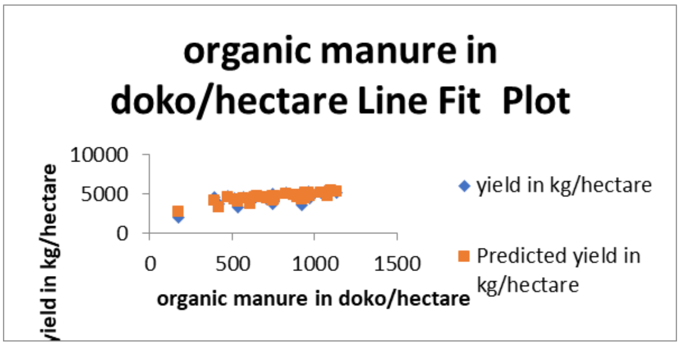

Figure 11: Organic Manure Fit Plot

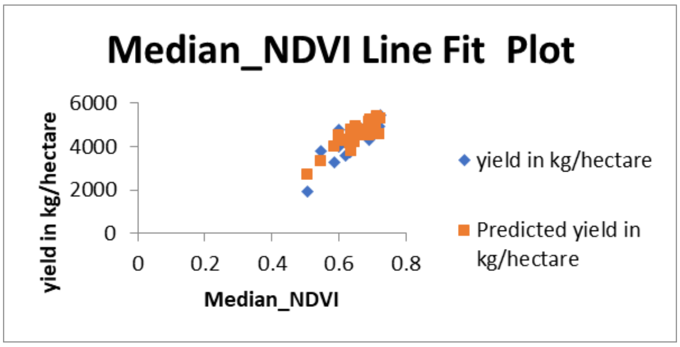

Figure 12: Median NDVI Fit Plot

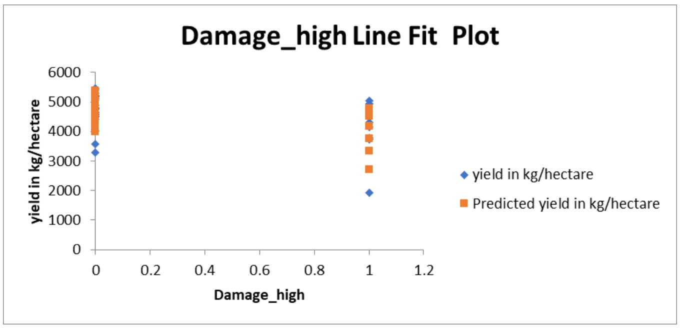

Figure 13: Damage High Fit Plot

4.6 Model Validation and Gap Assessment

Out of 70 sample data, 35 sample data from the year 2021 were utilized to create the yield prediction model in the current study. We may infer from this study that the model has an accuracy of 91.22%. For testing and validation purposes, we obtain 35 samples from the same field and location in 2022. We were able to attain the amazing degree of accuracy described above by including land management elements used by farmers in 2022 that have a link with the model. So, we may use the model to estimate and anticipate yields, which rely on many aspects of land management.

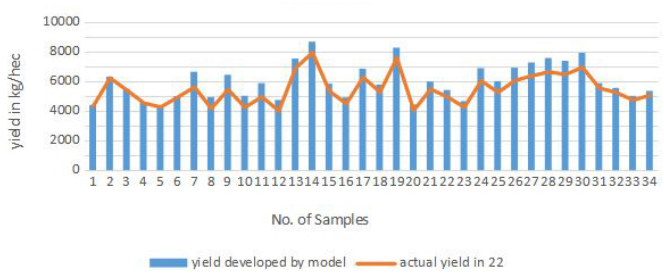

Figure 14: Graph showing the Actual Yield and Predicted Yield of 2022

4.6.1 Rice Prediction Map of Study Area

If we address or manage the land management factors in the research, we may achieve the maximum output of 5 to 6 tons per hectare. This rice production forecast map shows that most places will produce more than 5 tons of rice per hectare, while other areas will produce less than 3 tons.

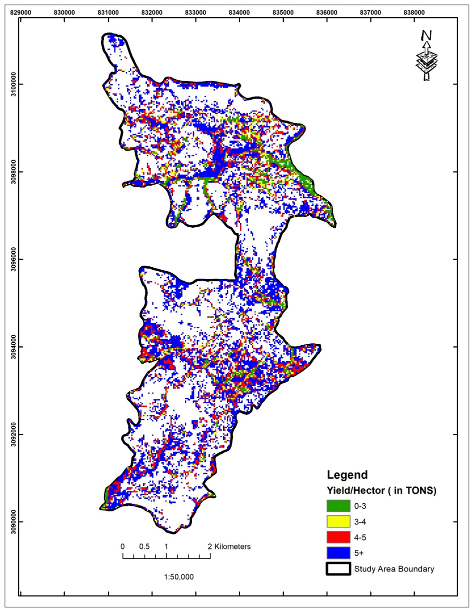

Figure 15: Rice Prediction map of study area

5. Discussion and Conclusion

With the integration of remote sensing data, land management variables, and management features, a field-level yield forecast model has been created. In comparison to using NDVI alone, the study shows that integrating NDVI with land management characteristics can improve field-level yield prediction. There is a good association between the NDVI and field level yield (r=0.79), and the damage from pests and diseases, fertilizers, and the NDVI are all determined to be major contributors determining yield variability.

In conclusion, Crop yield modeling is a crucial agricultural tool that aids farmers in estimating how much crop they may anticipate to harvest. It entails utilizing mathematical models and statistical techniques to forecast a crop's production based on a variety of variables, including weather patterns, soil characteristics, and crop management techniques. Data collection on the numerous factors that influence crop growth and production is the first step in the modeling process. The link between these variables and crop yield is then shown mathematically using this data. To assure the model's accuracy and dependability, it is then calibrated and verified using historical data.The model may then be used to forecast crop production in a variety of settings, such as shifting weather patterns, altered soil properties, or new crop management techniques. Farmers need to know this information in order to make wise choices about crop management methods including when to plant, when to apply fertilizer, and when to irrigate their fields.

References

[1] Editorial. Nepal's rice economy. *Nepali Times*, 2019.

[2] Rakesh Kumar. The price of rice. *Nepali Times*, March 1, 2019.

[3] Xiangtao Fan, Hao Liu, Jianhua Xiao, S Ross, B. Brisco, R Brown, and Gordon Staples. Rice monitoring and production estimation using multitemporal radarsat. *Remote Sensing of Environment*, 76:310–325, 06 2001.

[4] Qiangzi Li, Huanxue Zhang, Xin Du, Ning Wen, and Qingshan Tao. County-level rice area estimation in southern china using remote sensing data. *Journal of Applied Remote Sensing*, 8:083657, 03 2014.

[5] Murali Krishna Gumma, Andrew Nelson, Prasad S. Thenkabail, and Amrendra N. Singh. Mapping rice areas of South Asia using MODIS multitemporal data. *Journal of Applied Remote Sensing*, 5(1):053547–053547, January 2011.

[6] Yingxin Gu, Bruce Wylie, Daniel Howard, Khem Phuyal, and Lei Ji. Ndvi saturation adjustment: A new approach for improving cropland performance estimates in the greater platte river basin, usa. *Ecological Indicators*, 30:1–6, 07 2013.

[7] Guangming Zhao, Yuxin Miao, Hongye Wang, su Minmin, Mingsheng Fan, Fusuo Zhang, Rong-feng Jiang, Zujian Zhang, Cheng Liu, Penghuan Liu, and Dequan Ma. A preliminary precision rice management system for increasing both grain yield and nitrogen use efficiency. *Field Crops Research*, 154:23–30, 12 2013.

[8] I Wayan Nuarsa, Fumihiko Nishio, and Chiharu Hongo. Spectral characteristics and mapping of rice plants using multi-temporal landsat data. *Journal of Agricultural Science*, 3:54–67, 03 2011.

[9] I Wayan Nuarsa, Fumihiko Nishio, and Chiharu Hongo. Rice yield estimation using landsat etm+ data and field observation. *Journal of Agricultural Science*, 4:45–56, 12 2011.

[10] Yan Zhao, Andries Potgieter, Miao Zhang, Bingfang Wu, and G. Hammer. remote sensing predicting wheat yield at the field scale by combining high-resolution sentinel-2 satellite imagery and crop modelling. *Remote Sensing*, 12, 03 2020.

[11] Ahmed Kayad, Marco Sozzi, Simone Gatto, Francesco Marinello, and Francesco Pirotti. Monitoring within-field variability of corn yield using sentinel-2 and machine learning techniques. *Remote Sensing*, 11(23), 2019.

[12] Diego Gómez, Pablo Salvador, Julia Sanz, and Jose Luis Casanova. Potato yield prediction using machine learning techniques and sentinel 2 data. *Remote Sensing*, 11:1745, 07 2019.

[13] Jiaxuan You, Xiaocheng Li, Melvin Low, D. Lobell, and Stefano Ermon. Deep gaussian process for crop yield prediction based on remote sensing data. In *AAAI Conference on Artificial Intelligence*, 2017.

[14] Alex O. Onojeghuo, George A. Blackburn, Qunming Wang, Peter M. Atkinson, Daniel Kindred, and Yuxin Miao. Mapping paddy rice fields by applying machine learning algorithms to multi-temporal sentinel-1a and landsat data. *International Journal of Remote Sensing*, 39(4):1042–1067, 2018.