

Integrating Remote Sensing And Weather Data For Accurate Rice Yield Prediction In Punjab, India

Parmod Kumar¹, Rajesh Jolly¹, Ripudaman Singh²

^{1.} Lovely Professional University Phagwara, Punjab

^{2.} Ameti University, Noida, Uttar Pradesh

Corresponding Author: parmod.nehra25@gmail.com

Abstract-

Rice cultivation, crucial for global food security, faces challenges due to limited land availability, climate change, and unpredictable weather patterns. Accurate prediction of rice yield before harvest is essential for informed decision-making at international, national, and regional levels. Traditional methods like Crop Cutting Experiments (CCE) are labor-intensive and time-consuming. This study explores the use of Remote Sensing (RS) and Geographic Information System (GIS) technologies along with weather parameters to predict rice yield in Bathinda, Ludhiana, and Gurdaspur districts of Punjab, India, for the years 2019 and 2020. Spectral indices (NDVI, LSWI) from Sentinel-2 MSI satellite data, along with weather parameters (temperature, rainfall) and crop length period, were utilized as independent variables, while CCE yield served as the dependent variable. Stepwise regression analysis was employed for yield prediction.

Results indicate varying influences of independent variables on rice yield across districts and years. Bathinda consistently exhibited higher model accuracy, followed by Ludhiana and Gurdaspur. The models showed improved performance in predicting crop yield in 2020 compared to 2019. Statistical parameters such as RMSE, NRMSE, MAE, and MAPE were used to assess model accuracy, with Bathinda consistently demonstrating superior performance. Scatter plots of predicted versus observed yields illustrate the model's capability to approximate actual yields, with around 55% and 49% of data points aligning closely for Bathinda in 2019 and 2020, respectively. Similarly, Gurdaspur and Ludhiana districts

showed trends aligning with observed values, albeit with slightly lower percentages.

In summary, the study demonstrates the potential of RS and GIS technologies, coupled with weather parameters, in accurately predicting rice yield, aiding stakeholders in decision-making processes related to agriculture, market dynamics, and resource allocation.

Keywords: Yield, Stepwise regression analysis, Crop Cutting Experiments, Spectral indices (NDVI, LSWI)

Introduction

Rice is one of major kharif crop in the world and primary food of the half of the world and 90% of rice produced and consumed in Asia. India is second largest rice producing country in the world (Food and agriculture organization for the United States) and account its cultivation is also a key source of the farmers. As the population is increasing day by day, several challenging are arising like limited land, global warming and unpredictable weather condition therefore to maintain & increase the rice production is becoming challenging task.

Rice crop growth and development are heavily influenced by a variety of environmental factors, resulting in large differences in crop yields from year to year, therefore to predict the crop yield with higher accuracy before harvesting at international, National and regional scale can become the boon for farmers, planner as well as the government agencies also (Basso et al., 2013) . Monitoring prospective yields precisely and timely is critical for decision making since it affects markets, export-import decisions, and farm income budgeting (Zhao et al., 2020). The traditional method of yield estimation using crop cutting experiment (CCE) is labour intensive and time consuming process which is difficult also to apply on a large scale area((Reynolds et al., 2000).

Remote Sensing and GIS technology has extensively applied for different sectors like agriculture, forest as well as for urban sector also (Bouvet et al. 2009, Pan et al. 2010, Niel et al. 2003; Nuarsa et al. 2011). Several developments took place in recent years especially for the improvement of spatial, spectral and temporal

resolution of the sensor made it more vital for accurate, timely and temporal assessment of crop acreage, crop health, crop yield and production before harvesting crops. Remote sensing is now capable to monitor crop growth parameters like crop vigor, biomass (Fu et al., 2014), leaf area index (Haboudane et al., 2004; Verger et al., 2014; Liang et al., 2015), and chlorophyll content (Haboudane et al., 2002), which can be estimated accurately with the vegetation indices (VIs) using combination of different spectral band of satellite imagery. The use of satellite spectrum data to estimate agricultural yields is an appealing possibility since yield is connected to crop vigour, which is related to crop spectral response, which is related to crop spectral response recorded by satellite sensors (Barnett & Thompson, 1982). Most of researchers have used only satellite based indices like Normalized difference vegetation index (NDVI), Land surface wetness index (LSWI) but there are very few studies in which both satellite based and weather parameters have used to predict the rice yield.

In this study satellite based indices NDVI, LSWI along with different weather parameters minimum & maximum temperature and rainfall, and crop growing days (crop length period) were used as independent variable. Crop cutting experiment (CCE) derived predicted yield results were also used as dependent variable.

Study area - Bathinda, Ludhiana and Gurdaspur (Figure.1) are major paddy growing districts in Punjab and these three districts are covering four different agro-climatic regions namely western plain area, northern plain semiarid, Northern plain dry sub humid and Western Himalayas (Siwalik and Kandi area sub humid regions. Bathinda and Ludhiana districts are in south of Sutlej river which is known as malwa and Gurdaspur district is between Ravi and Beas river which is known as majha region of Punjab. Gurdaspur and Ludhiana is mainly paddy dominant district while cotton and paddy crops are grown in Bathinda district due to water scarcity in southern and western part of Bathinda district.

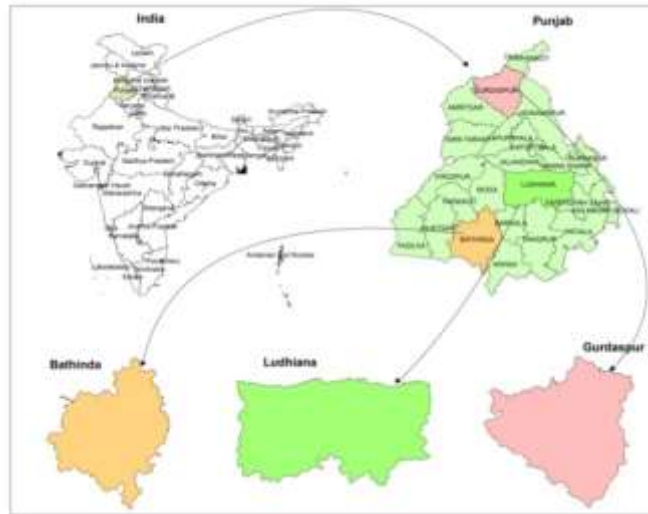


Fig 1. Study area

Methodology- Rice is major crop in all 3 study districts and the analysis was carried out for the 2 year 2018 & 2019. Various crop centric spectral indices like NDVI and LSWI alongwith the weather parameters like min & max temperature, rainfall were used as independent parameters as per kharif rice crop calendar (June to October). Crop length period/ growing degree days was also used as independent parameters.

Generation of Spectral indices- The European space agency provides the Sentinel-2 MSI satellite data at 10 meters spatial resolution and at 5 days (Sentinel-2 A/B) temporal resolution. The data is freely available in 13 spectral band from coastal & aerosol to short wave infrared. As we know the crop vigor and surface wetness are major factor which affect the crop yield therefore the two major spectral indices NDVI and LSWI was considered as the one of independent variables.

Normalize difference vegetation index (NDVI) - basically quantifies vegetation health in terms of greenness and leaf vigour using the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs). The NDVI values range between -1 to 1, negative values likely for the water and values near 0 to 0.2 likely for fallow land or built up areas or bare soil and the value ranging from 0.2 to 1 shows crop health and its density. Higher the NDVI value better is the crop growth. As the crop vigor is considered one major parameter of crop yield and

it is considered that the higher the vigor higher the crop yields. The following equation was used to calculate the NDVI:

$$\text{NDVI formula} = (\text{NIR}-\text{R}) / (\text{NIR}+\text{R})$$

Land Surface Wetness Index (LSWI) - quantifies the crop moisture content using the difference between near infrared (NIR) which (vegetation strongly reflects) and short wave infrared (SWIR) (strongly absorbed by water present in the crop canopy) bands. LSWI shows crop moisture content and if bare field then it shows surface soil moisture condition. AS we know the surface moisture also plays important in plant growth therefore it is also considered as major factor which affects the crop growth and yield. LSWI values also ranges between -1 to +1.

$$\text{LSWI formula} = (\text{NIR}-\text{SWIR}) / (\text{NIR}+\text{SWIR})$$

Crop Length period Assessment- After identification sowing/transplanting dates using Multi-date Sentinel-1 SAR satellite data and harvesting date identification using Sentinel-2 A/B Multidate optical data, the Crop length period was assessed using the paddy date wise sowing/transplanting layer and using harvesting date's pattern. Crop length period (in no. of days) was calculated for each paddy-classified pixel. Crop length period was also taken as independent parameters.

Crop Length period assessment = Pixel wise harvest date- pixel wise sowing date

Minimum & Maximum Temperature- Indian meteorological department captured the daily minimum and maximum temperature data at 2 meters height and provide the data in form of Gridded. The data is freely available at IMD Pune Website (https://www.imdpune.gov.in/cmpg/Griddata/Max_1_Bin.html). As particular crop grows in favorable weather condition therefore changes in minimum and maximum temperature also plays an important role. The daily data (0.5 degree spatial resolution) was downloaded and data was converted into tagged image file format (.Tiff) using Python language and arranged fortnightly average was taken.

Rainfall- Indian meteorological department (IMD) captured the daily precipitation and provides the daily rainfall in gridded format.

The IMD provides the rainfall data at 0.25 degree spatial resolution at free of cost

(https://www.imdpune.gov.in/cmpg/Griddata/Max_1_Bin.html).

The daily data was downloaded in gridded format and arranged fortnightly (sum of 15 days) and converted into tiff format using python language. Rainfall is one of major parameters which mostly affect the crop growth as well as it yield therefore the rainfall was also taken as an independent variable while computing yield.

Crop Cutting Experiment Yield- The Crop Cutting Experiment (CCE) is a technique that uses small subplots inside farmed fields to provide an accurate, cost-effective estimate of crop yields and other information. Crop cutting experiments are conducted by agriculture department every year to estimate the crop production in the country as per the guidelines of department of agriculture. Crop cutting experiments data was taken from Punjab Agriculture University and Punjab agriculture department for all the 3 districts for 2019 & 2020. These crop cutting experiments were used to compare the model predicted result and statistical analysis were computed.

All the independent variables were arranged fortnightly. NDVI, LSWI, minimum and maximum temperature, rainfall, crop length period were used as independent and CCE yield was used as dependent variable. Step wise regression was used to predict the yield using independent and dependent variable using Statistical package for Social Science (SPSS). The detailed methodology is depicted in below given in the Figure 2.

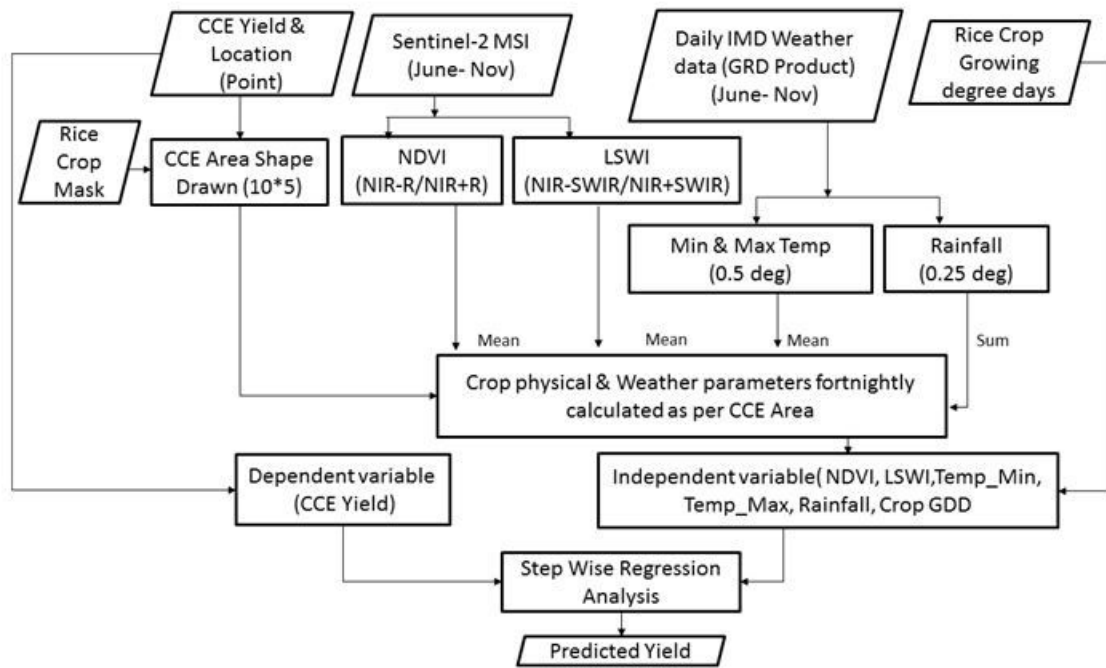


Fig2. Flow diagram of Yield estimation

Results & Discussion

Rice yield was estimated against the Crop cutting experiment (observed yield) using independent (NDVI, LSWI, minimum and maximum temperature, rainfall, crop length period) and dependent variables (CCE yield). Stepwise regression was used to predict the yield results for the year 2019 & 2020.

The model shows that in 2019, Bathinda's crop yield was significantly influenced by Oct_F1_NDVI, sep19_F1_rf, aug_F2_tmx, and july19_F1_rf, with a relatively high R Square value of 0.601. However, in 2020, the model's R Square value decreased to 0.425, with Oct_F1_Ndvi, July_F1_LS, and crop length being significant predictors. In comparison, Ludhiana's crop yield in 2019 was primarily affected by Oct_F2_LSWI, with an R Square value of 0.189, while in 2020, June20_F2_rf and July_F1_LSWI were significant predictors, resulting in an R Square value of 0.27. Gurdaspur's crop yield in 2019 was mainly influenced by Oct19_F2_tmx and Aug19_F1_rf, with an R Square value of 0.388, whereas in 2020, Aug20_F2_tmin, July20_F2_tmin, and Sep_F2_LSWI were significant predictors, resulting in an R Square

value of 0.308. Overall, Bathinda had the highest R Square values in both years, indicating that the models were more accurate in predicting crop yield for Bathinda compared to Ludhiana and Gurdaspur. The result are depicted in Table:1

Table. 1 Statistical parameters of Empirical model

District	Year	Model	R Square	Adjusted R Square	Std. Error of the Estimate	F	Sig.
Bathinda	2019	$Y=13238.45+(2732.118*Oct_F1_NDVI+27.81*sep19_F1_rf+-295.6*aug_F2_tmx+(12.534*july19_F1_rf))$	0.601	0.555	213.629	13.17	.000f
	2020	$Y=2493.679+(6645.464*Oct_F1_Ndvi+-3090.043*July_F1_LS+(-18364*crop_length))$	0.425	0.379	259.237	9.13	.000c
Ludhiana	2019	$Y=3287.68+(3170.95*Oct_F2_LSWI)$	0.189	0.175	500.0286	14.181	.000a
	2020	$Y=3382.808+(23.106*June20_F2_rf)+3478.97*July_F1_LSWI$	0.27	0.246	495.4044	11.455	.000b
Gurdaspur	2019	$Y=16281.9+(637.95*Oct19_F2_tmx+5.158*Aug19_F1_rf)$	0.388	0.355	606.203	11.740	.000b
	2020	$Y=1340.113+(2483.368*Aug20_F2_tmin+(-2220.366*July20_F2_tmin)+(-3696.95*Sep_F2_LSWI))$	0.308	0.246	387.732	4.906	0.01

After predicted the result using independent and dependent variable and various statistical parameters (Table 2) were used to model accuracy among different districts. Bathinda consistently had the lowest error metrics across both years, indicating superior model performance in predicting crop yield. In 2019, Bathinda and Ludhiana had comparable RMSE values of 0.50 and 0.49, respectively, while Gurdaspur had a slightly higher RMSE of 0.58. In 2020, Bathinda's RMSE reduced significantly to 0.25, showcasing a notable improvement in model accuracy. Ludhiana and Gurdaspur also saw a decrease in RMSE to 0.48 and 0.37,

respectively, suggesting enhanced predictive capabilities in these districts as well.

Regarding the Normalized RMSE (NRMSE) (%), in 2019, Bathinda and Ludhiana had similar values of 12.72% and 12.18%, respectively, while Gurdaspur had a higher NRMSE of 17.27%. In 2020, the NRMSE values for all districts decreased, with Bathinda showing the lowest at 5.64%, followed by Ludhiana at 11.08%, and Gurdaspur at 10.11%. These results highlight Bathinda's consistent improvement in model performance.

The Mean Absolute Error (MAE) for Bathinda was the highest among the three districts in 2019, standing at 0.45, while Gurdaspur and Ludhiana had MAE values of 0.49 and 0.37, respectively. However, in 2020, Bathinda maintained its lead with the lowest MAE of 0.18, followed by Ludhiana at 0.34, and Gurdaspur at 0.31. This indicates that the model's predictions for Bathinda were the closest to the actual crop yield in both years.

Looking at the Mean Absolute Percentage Error (MAPE), in 2019, Ludhiana had the lowest MAPE at 1.84%, followed by Gurdaspur at 3.38%, and Bathinda at 11.32%. In 2020, Bathinda again had the lowest MAPE at 0.36%, followed by Gurdaspur at 1.15%, and Ludhiana at 6.32%. This further supports Bathinda's superior performance in predicting crop yield.

In summary, Bathinda consistently demonstrated the most accurate predictions of crop yield across both years, with Ludhiana showing improvement from 2019 to 2020, and Gurdaspur exhibiting varying results between the two years.

Table. 2 Statistical analysis for model training and validation accuracy

	Bathinda		Ludhiana		Gurdaspur	
	2019					
Statistical Criteria	Training	Validation	Training	Validation	Training	Validation
RMSE	0.50	0.59	0.49	0.38	0.58	0.50
NRMSE (%)	12.72	15.16	12.18	9.10	17.27	4.64

MAE	0.45	0.52	0.37	0.29	0.49	0.38
MAPE	11.32	14.18	1.84	-3.17	3.38	10.85
2020						
RMSE	0.25	0.41	0.48	0.50	0.37	0.73
NRMSE (%)	5.64	9.84	11.08	11.85	10.11	21.36
MAE	0.18	0.30	0.34	0.38	0.31	0.59
MAPE	0.36	6.32	-1.70	-5.43	1.15	9.76

The predicted and Crop cutting experiment (observed values) were plotted for 2019 using scatter plot and observed (Figure 3) that around 55% of data is expressing the similar results as compared to observed yield.

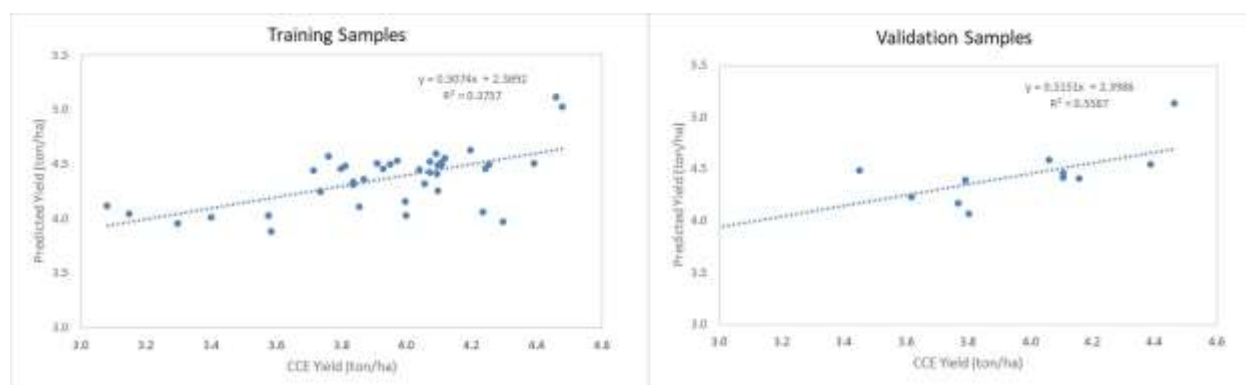


Fig.3 Scatter plot of CCE Yield and Predicted yield for training (Left) & validation (Right) of Bathinda district for the year 2019

The predicted and Crop cutting experiment (observed values) were also plotted for 2020 using scatter plot and observed (Figure 4) that around 49% of data is expressing the similar results as compared to observed yield.

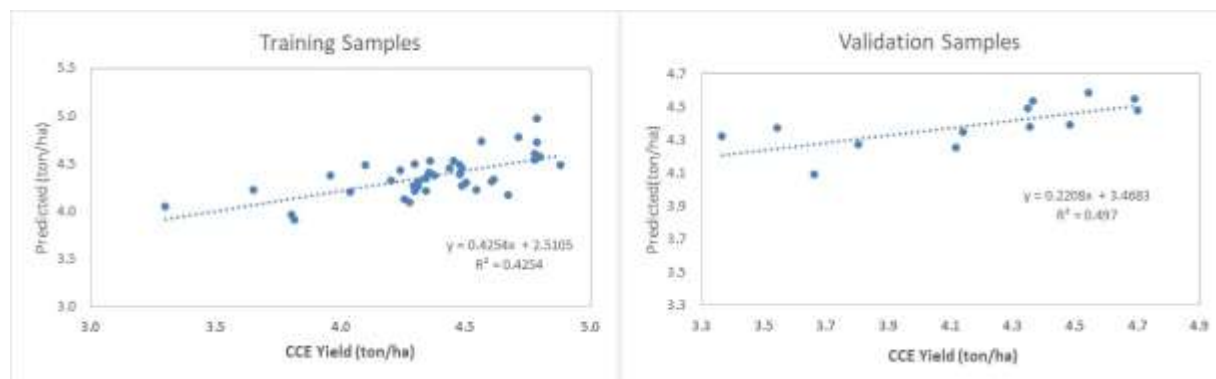


Fig.4 Scatter plot of CCE Yield and Predicted yield for training (Left) & validation (Right) samples of Bathinda district for the year 2020

The predicted and Crop cutting experiment (observed values) were plotted for 2019 using scatter plot and observed (Figure 5) that around 39% of data is expressing the similar results as compared to observed yield.

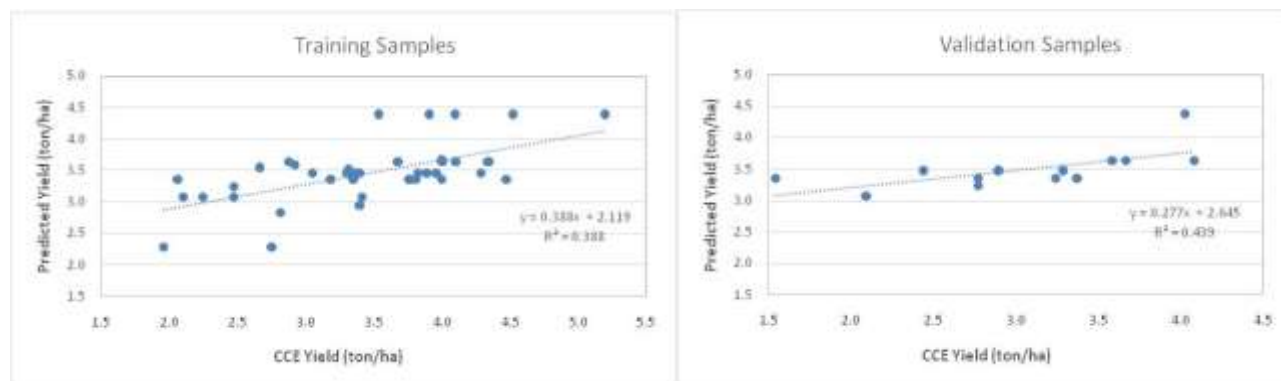


Fig.5 Scatter plot of CCE Yield and Predicted yield for training (Left) & validation (Right) of Gurdaspur district for the year 2019

The predicted and Crop cutting experiment (observed values) were also plotted for 2020 using scatter plot and observed (Figure 6) that around 31% of data is expressing the similar results as compared to observed yield.

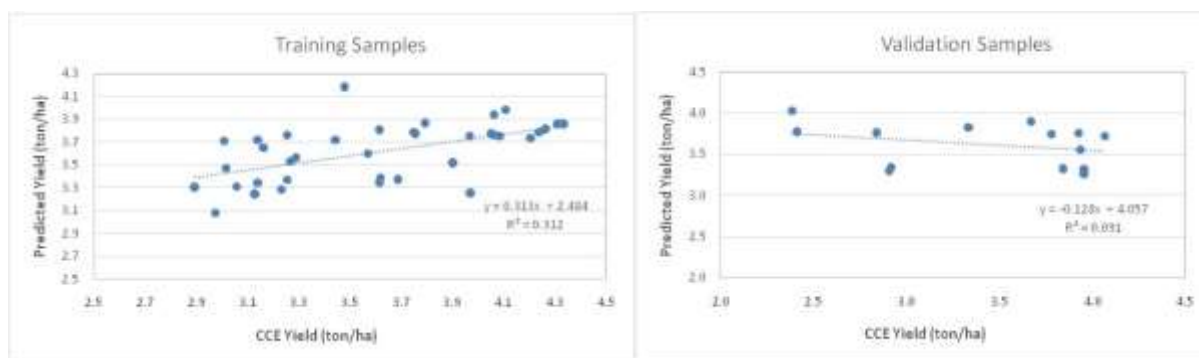


Fig.6 Scatter plot of CCE Yield and Predicted yield for training (Left) & validation (Right) of Gurdaspur district for the year 2020

The predicted and Crop cutting experiment (observed values) were also plotted for 2019 using scatter plot and observed (Figure 7) that values are around trend line.

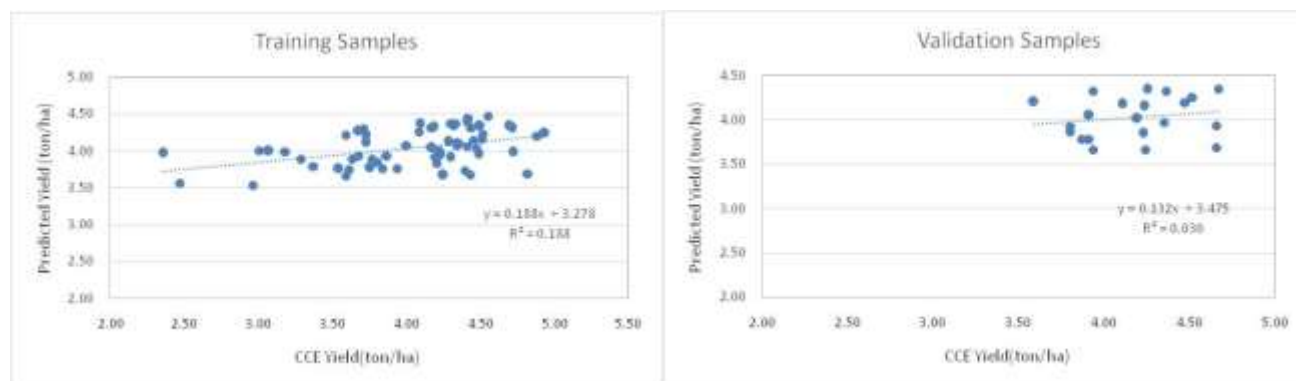


Fig.7 Scatter plot of CCE Yield and Predicted yield for training (Left) & validation (Right) of Ludhiana district for the year 2019

The predicted and Crop cutting experiment (observed values) were also plotted for 2020 using scatter plot and observed (Figure 8) that values are around trend line.

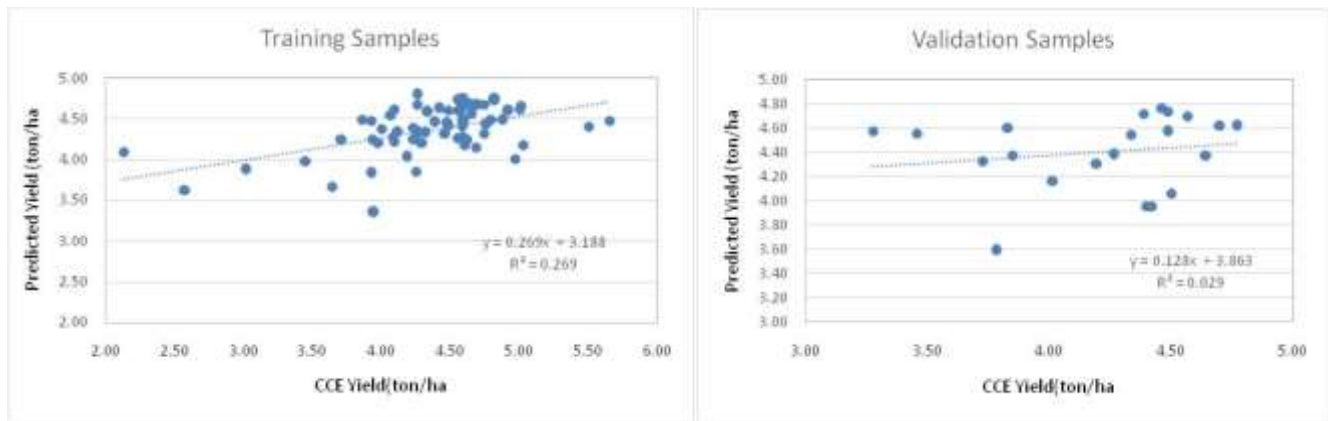


Fig.8 Scatter plot of CCE Yield and Predicted yield for training (Left) & validation (Right) of Ludhiana district for the year 2020

Conclusion

The study demonstrates the potential of utilizing Remote Sensing (RS) and Geographic Information System (GIS) technologies, along with weather parameters, for accurately predicting rice yield in Bathinda, Ludhiana, and Gurdaspur districts of Punjab, India, for the years 2019 and 2020. By employing spectral indices such as NDVI and LSWI from Sentinel-2 MSI satellite data, along with weather parameters like temperature, rainfall, and crop length period, the study aimed to develop models for predicting rice yield, thereby aiding stakeholders in decision-making processes related to agriculture, market dynamics, and resource allocation.

The results of the study indicate varying influences of independent variables on rice yield across districts and years. Bathinda consistently exhibited higher model accuracy compared to Ludhiana and Gurdaspur, with the models showing improved performance in predicting crop yield in 2020 compared to 2019. Statistical parameters such as RMSE, NRMSE, MAE, and MAPE were used to assess model accuracy, with Bathinda consistently demonstrating superior performance.

In summary, the study highlights the effectiveness of RS and GIS technologies, coupled with weather parameters, in accurately predicting rice yield. Bathinda consistently emerged as the district with the most accurate predictions, followed by Ludhiana and Gurdaspur. These findings underscore the potential of these technologies in facilitating informed decision-making processes for agriculture, thereby contributing to global food security efforts.

References

1. Basso, B., Cammarano, D and Carfagna, E. 2013. Review of crop yield forecasting methods and early warning systems. In Proceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics, FAO Headquarters, Rome, Italy .18:19.
2. Bouvet, A., Le Toan, T. & Lam-Dao, N. (2009). Monitoring of the Rice Cropping System in the Mekong Delta Using ENVISAT/ASAR Dual Polarization Data. IEEE Trans. Geoscience and Remote Sensing, 47, pp. 517 –526.
3. Fu, Y., Yang, G., Wang, J., Song, X., Feng, H., 2014. Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyper spectral measurements. Comput. Electron. Agric. 100,51–59.
4. Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sens. Environ. 81, 416–426.
5. Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. Remote Sens. Environ. 90, 337–352.
6. Pan, X.Z., Uchida, S., Liang, Y., Hirano, A. & Sun, B. (2010). Discriminating different land use types by using multitemporal NDXI in a rice planting area. International Journal of Remote Sensing, 31, 585–596.
7. Reynolds, C.A., Yitayew, M., Slack, D.C., Hutchinson, C.F., Huete, A., Petersen, M.S., 2000. Estimating crop yields and production by integrating the FAO Cropspecific Water Balance model with real-time satellite data and ground-based ancillary data. Int. J. Remote Sens. 21, 3487–3508.
8. Niel, T.G.V. & McVicar, T.R. (2003). A Simple method to improve field-level rice identification: toward operational monitoring with satellite remote sensing. Australian Journal of Experimental Agriculture, 43, 379–387.
9. Liang, L., Di, L., Zhang, L., Deng, M., Qin, Z., Zhao, S., Lin, H., 2015. Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method. Remote Sens. Environ. 165, 123–134.
10. Nuarsa, I.W., Nishio, F. & Hongo, C. (2011). Spectral Characteristics and Mapping of Rice Plants Using Multi-Temporal Landsat Data. Journal of Agriculture Science, 3, 54-67.

11. Verger, A., Vigneau, N., Chéron, C., Gilliot, J., Comar, A., Baret, F., 2014. Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sens. Environ.* 152, 654–664.
12. Zhao, Y., Potgieter, A. B., Zhang, M., Wu, B., & Hammer, G. L. (2020). Predicting wheat yield at the field scale by combining high-resolution Sentinel-2 satellite imagery and crop modelling. *Remote Sensing*, 12, 1024.