Wheat Crop Field and Yield Prediction using Remote Sensing and Machine Learning

Maheen Ayub
Department of Computer Science and Information Technology
NED University of Engineering and Technology
Karachi, Pakistan
ayub.pg3200204@cloud.neduet.edu.pk

Najeed Ahmed Khan
Co-PI NCAI Smart City lab
Department of Computer Science and Information Technology
NED University of Engineering and Technology
Karachi, Pakistan
najeed@neduet.edu.pk

Rana Zeeshan Haider
Baqai Institute of Hematology
Baqai Medical University
Karachi 75340, Pakistan
Karachi, Pakistan
zeeshan3335@yahoo.com

Abstract—**Agriculture plays an important role in the growth of a country's economy. Crop area and yield predictions using machine learning are important investigation domains in current research fields. Wheat is the most important food crop in Pakistan which is cultivated in the Rabi season. Weather conditions, Remote Sensing (RS) data, and Machine learning (ML) technologies can be used to forecast wheat yield before actual harvesting to assist the management of wheat production, trade, and storage. In this paper, a supervised ML based framework is proposed that extracts features/Vegetation Indices (VIs) including Enhanced Vegetation Index (EVI), Normalized Difference Vegetation Index (NDVI), Red Edge Normalized Difference Vegetation Index (RENDVI), and Normalized Difference Moisture Index (NDMI) from Sentinel-2 Satellite images and contributes for: estimation of wheat area, and identification of most effective VIs in wheat area estimation, prediction of wheat yield, and identification of most effective VIs and meteorological parameters in wheat yield prediction. In the initial experimental setup, good performance output obtained using the Random Forest (RF) machine learning algorithm therefore in this framework RF machine learning algorithm is focused on wheat area estimation and generation of Land Use Land Cover (LULC) maps which is capable of estimating area with an accuracy of 84%, consumer's accuracy of 81%, producer's accuracy of 83% and kappa statistics of 0.80. LULC maps are used for wheat yield prediction. Multivariate regression forward stepwise technique is applied for yield prediction and selection of effective VIs and meteorological parameters. The adjusted coefficient of determination (R2) between reported and predicted yield found 0.84 with an error of 46.14 Kg/ha for yield prediction.**

Keywords—**Wheat, Crop Yield, Yield Prediction, Machine Learning, Remote Sensing, Feature Extraction, Yield Forecasting**

I. Introduction and Literature Review

Agricultural monitoring plays an important role in food production management. Crop yield prediction helps in planning trade, national food policy, and security. Government can make timely import and export decisions based on accurate yield predictions to strengthen national food security. The significance of these estimates is even greater in countries including Pakistan, which are vulnerable to population growth and climate change [1]. Agriculture is the most important field of our economy. The majority of our population depend on this industry, either indirectly or directly. It accounts for half of the employed labor force [2]. It is the main source of foreign exchange earnings, contributing about 24% of GDP. It serves the needs of food of the entire urban and rural population. Planners and policymakers are becoming more aware of its significance. Policymakers and planners need to obtain timely and accurate area and production statistics for crops. For important crops such as wheat, rice, maize, sugarcane, cotton, etc., policymakers need reliable and timely statistics [3].

There are two cropping seasons in Pakistan. The first is known as "Kharif" and the second is known as "Rabi". The first sowing season begins from April to June and is harvested from October to December. "Kharif" crops include rice, sugarcane, cotton, maize, moong, mash, bajra, and jowar. The second sowing season starts from October to December and is harvested from April to May. Wheat, gram, lentil (masoor), tobacco, rapeseed, barley, and mustard are all "Rabi" crops. The most significant crop in "Rabi" is wheat. Wheat contributes 8.7% to value addition in agriculture and 1.7% to GDP. Wheat crop production increased % to 24.946 million tons over last year's production of 24.349 million tons. Over the previous year's area (8,678 thousand hectares), the area under cultivation expanded by 1.7% to 8,825 thousand hectares. Increased cultivated area, healthy grain formation, and higher crop yields all contributed to increased production [4].

Crop yield estimation has traditionally relied on ground-based data collection methods, which are time-consuming, tedious, and fail to capture the spatial variability of crops [5]. Using emerging IT techniques e.g. Image processing on satellite images has been used in agricultural monitoring and yield forecasting [6]. It reduces the effort needed by the traditional methods. Satellite Remote Sensing (RS) outperforms other monitoring techniques by offering a timely, comprehensive, and up-to-date image of large-scale crop monitoring at various stages [7, 8]. Multiple Machine Learning (ML) models have been used along with satellite and climate data to predict wheat yield [9]. Neural networks models have also been used to predict wheat prices [10].

The proposed framework can significantly tackle the challenges faced by the government to improve wheat

978-1-6654-6896- 1/22/$31.00 ©2022 IEEE

productivity and thus help them make appropriate decisions to maximize its yield rate. This paper addresses the following research questions (RQ).

RQ1: Which Vegetation Indices (VIs) (e.g. Enhanced Vegetation Index (EVI), Normalized Difference Vegetation Index (NDVI), Red Edge Normalized Difference Vegetation Index (RENDVI), and Normalized Difference Moisture Index (NDMI)) mainly contribute towards the wheat area and yield estimation?

RQ2: Which meteorological parameters (e.g. Sunshine duration, humidity, etc.) have a major impact on wheat yield prediction?

NDVI product of MODIS data used for identification of wheat area and prediction of wheat yield for two Punjab districts Faisalabad and Chakwal using an unsupervised clustering technique [11]. The NDVI profiles were generated for the study area and then through mapping temperature and rainfall with NDVI profiles, the comprehensive effect of two metrological parameters on wheat production observed. It is recorded that the meteorological variables affect wheat crop development at various stages, including germination, vegetation, and maturity, and can have positive impact if they occur in sufficient quantities and at critical growth stages. Wheat yield first estimated using linear regression and then with non-linear regression to improve yield estimates. The Coefficient of determination (R2) between the predicted and reported yield found 0.86 for Chakwal and 0.69 for the Faisalabad district.

A model to predict wheat yield before harvest for each of eight individual districts as well as the entire Punjab province is developed [2]. The model uses weather and NDVI data from MODIS to generate Random Forest (RF) statistical models with 15 independent variables from 2001 to 2014 (14 years). Temperature, rainfall, sunshine hours, growing degree days, and MODIS-derived NDVI used to predict yield for each of Punjab province's eight districts in 2014. The Random Forest algorithm applied with 4 approaches. The first method employed separate RFs for each district. The second method employed generic RFs for each district. All eight districts' input data sets are combined, and a generic RF is generated to forecast yield per district for 2014. The RF in the third approach is calculated by taking the average of each independent variable across the eight districts. In the fourth approach, the generic RF, which is used for each district, used to yield forecasting for the year 2014 using the averages of all variables for Punjab province. Using average and generic RFs, the root means square error (RMSE) of the prediction results for the entire Punjab province observed 147.7 kg ha^{-1} and 148.7 kg ha^{-1}, respectively, with a mean error of 4.51 % and 3.8 % and R2 of 0.27 and 0.45. Individual district forecasts had an R2 of 0.95, an RMSE of 175.6 kg ha^{-1}, and a 5.86 % mean error.

A method proposed in [12] to determine the area under wheat and forecast its production in the command area of three distributaries of the Lower Chenab Canal System (LCC), namely Khurrian Wala, Killian Wala, and Mungi distributary. The research carried out throughout the Rabi season in 2011-2012. Field data is collected twice for 36 sample points throughout the study period. Firstly, in peak growing season for collecting training data for wheat area estimation. Supervised classification technique i.e. maximum likelihood classifier (MLC) applied on Landsat 7 satellite images for land cover mapping as wheat and non-wheat classes. The area under the wheat computed accordingly by using the spatial resolution of Landsat 7 i.e., 30 m, and the number of pixels covered by the wheat class. The accuracy of LULC is not reported in the paper. The second ground truth survey conducted near harvesting season for yield estimation. The NDVI computed for wheat class. Predicting the total yield for each distributary, the correlation between max NDVI and reported yield in the ground survey is calculated. The correlation values reported as 0.45 for Khurrian Wala, 0.36 for Killian Wala, and 0.39 for Mungi distributary. The results up to 85% accuracy are obtained using 30 m spatial resolution and 16 days temporal resolution of Landsat 7 data. The author recommended using high-resolution spatial-temporal imagery like 10m or 5m data for more accurate yield estimation results hence Sentinel-II satellite data can be used for better results.

The wheat yield estimated for the U.S. state of Kansas [13] by linearly regressing the NDVI data of the MODIS surface reflectance dataset against the reported yield. The NDVI values that used as input to the Kansas yield regression model selected using a crop type map of winter wheat. The Cropland Data Layer (CDL) supplied by NASS included a wheat crop map for Kansas. The CDL is a rasterized land cover map created with field level training data from extensive ground surveys, farmer reports submitted to the United States Farm Service Agency, and remotely sensed data from Landsat Thematic Mapper (TM), Advanced Wide Field Sensor (AWiFS), and Landsat Enhanced Thematic Mapper (ETM+). These data are fed into a decision tree classifier, which generates a land cover classification that distinguishes between various crop kinds and generates a crop type map. The RMSE between the actual and predicted yield of Kansas found equivalent to a 7% error, the regression slope 1.0334 with the intercept set to 0 and the regression coefficient 0.88. The model generated for Kansas directly applied to the Ukraine. The RMSE between the actual and predicted yield of Kansas found equivalent to 15% error, with the intercept adjusted to 0, the regression slope 0.9433 and the regression coefficient recorded 0.74. According to [2] forecasting models based solely on RS indices are unable to account for yield variability caused by climatic conditions.

A model for forecasting wheat production six weeks before harvest in Pakistan's Punjab province using MODIS and LANDSAT images established in [14]. Multiple VIs including wide dynamic range vegetation index and saturation-adjusted normalized difference vegetation index used to forecast crop yield. The VIs values of wheat used as input to the yield regression model they selected using a wheat/non-wheat map generated by applying a bagged decision tree classifier on Landsat imagery. Visual interpretation of the Landsat time series they used to select training data for the classification of wheat/non-wheat areas based on the phenology of wheat which the analyst had familiar with from their earlier work of other wheat-growing locations in Kansas and Ukraine [13]. The wheat yield they estimated by linearly regressing the peak season values of all 4 VIs against the reported yield. On comparing the yield estimates of all 4 VIs they found that WDRVI provided the most accurate forecast results. The RMSE between actual and predicted yield they computed 11.1.

The work on winter wheat yield prediction for the Huanghuaihai plain in North China has been described at a regional scale in [15] which uses crop biomass estimation by employing data from multiple sources (i.e. soil moisture data, meteorological data, and RS data from MOD13Q1). Two VIs (EVI and NDVI) used for predicting yield. Multiple ML models including SVM, RF, NN, DT, and GPR used for estimating yield. The wheat yields predicted more accurately by RF, GPR, and SVM, with RF demonstrating the strongest generalization ability of the three approaches. The results provided an R2 of 0.75 and a yield error of less than 10%.

The impact of combining Sentinel-2 images with crop stress-related indices to predict wheat yield studied in [16] at the field scale in Australia. The indices related to canopy structure, chlorophyll canopy status, and simulated crop stress used to establish yield prediction models. After calibration and validation of linear regression model with different combinations of VIs across a diverse set of environments, found that a predictive model using a VI or more than one VI (structural and/or chlorophyll) to predict final harvested wheat yield is less accurate than a model combining VIs (structural and chlorophyll) with the simulated SI (crop stress around flowering).

The regional wheat yield in Faisalabad estimated using RS soil moisture data i.e. NDWI of Landsat 7 ETM+ and Landsat 8 OLI satellite images and CERES-Wheat model which is an open-source crop model in [17]. Weather parameters including minimum temperature, rainfall, maximum temperature, and sunshine hours and Soil parameters saturation %age, clay %age, bulk density, silt %age, organic carbon, and soil pH have also been used for yield forecasting. Supervised classification technique Decision Tree Algorithm used for identifying the wheat area. The classified wheat area integrated with the CERES-Wheat model used for yield prediction. The model's RMSE and R2 reported as 284.8 kg/ha and 0.71, respectively.

The ML-based model proposed to predict wheat yield in China in [18] using multi-source data including VIs from Satellite images, climate data, and soil properties. The MODIS 16 days the composite product used for deriving NDVI, EVI, and NIR reflectance values used for yield estimation. The climate variables included mean values of wind speed, temperature, precipitation, sunshine hours, and relative humidity. The soil properties included soil bulk density, soil organic carbon content, and cation exchange capacity of clay. The yield predicted using RF and SVM. The RF model with NIR reflectance provided better results than SVM with an R2 value of 0.74 and RMSE of 758 kg/ha.

The ML algorithms counter-propagation artificial neural networks (CP-ANNs, Supervised Kohonen Networks (SKNs) and), XY-fused Networks (XY-Fs) used to predict wheat yield in the UK using soil parameters and NDVI values [19]. The achieved results provided an average overall accuracy of 78.3%, 81.65%, and 80.92% respectively.

This framework used sentinel-2 satellite images to extract features/obtain VIs and applied a supervised ML algorithm on different combinations of bands/VIs to estimate wheat area and create Land Use Land Cover (LULC) maps. The extracted NDVI and EVI profiles of wheat from the LULC maps along with other meteorological parameters used to forecast yield using multivariate regression as indicated in Fig. 1.

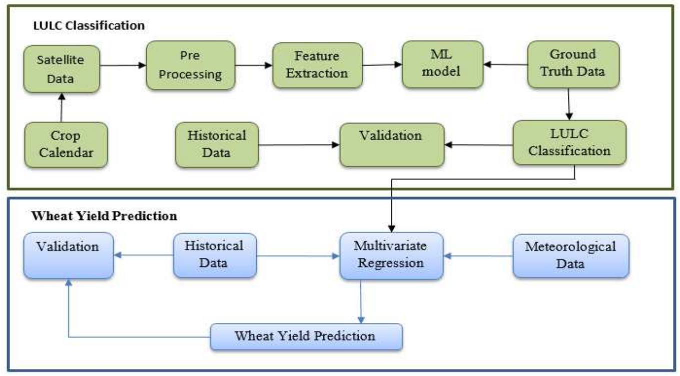

Fig. 1. *Generic Data Flow Diagram*

II. Materials and Methods

A. Study Site

The area of this study is limited to the Faisalabad district of Punjab which is Pakistan's central province and the largest wheat-producing area of the country. Punjab province accounts for more than 70% of the total cropped area in the country [20]. Faisalabad is located in Punjab province, Pakistan at 31.4187° N, 73.0791° E, with an altitude of 184m and a land area of 5,856 km2. The study site is represented in Fig. 2 taken from [11]. It is Punjab's second most populous district, with a population of about 3.5 million people. Extreme weather conditions exist in the district. The average maximum and minimum temperatures in the summer approaching 39 and 27 degrees Celsius, respectively. In the winter, the average maximum and minimum temperatures reach 21 and 6 degrees Celsius, respectively. The winter season ranges from November to March, while the summer season begins in April and lasts till October. Faisalabad's cropping system is based on canal-based irrigation, tube well cultivation, or rainwater harvesting. The agricultural production of the Faisalabad district is incomparable. Millet, maize, sugar cane, and rice are the region's main Kharif crops, while wheat, gram, and barley are the region's winter crops [11].

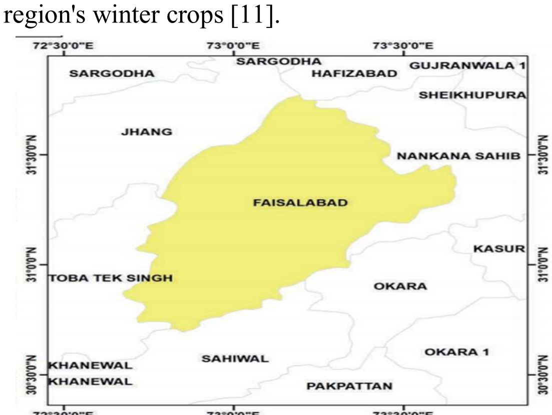

Fig. 2. *Study Area Site*

B. Dataset

This research makes use of optical imagery obtained from the Sentinel-2 sensor. The sensor was chosen primarily because of its availability and open-source nature, as well as the high spatial, spectral, and temporal resolutions it provides. The European Space Agency (ESA) [21] produces and distributes ortho-rectified Sentinel-2 data expressed in reflectance at the 1C level at the top of the atmosphere.

1) Remotely Sensed Data:

a) Data Processing: The Sentinel-2 images processed in Google Earth Engine (GEE) to produce products based on the NDVI, EVI, RENDVI, NDMI, and Normalized Difference Built-up Index (NDBI).

The NDVI first introduced in 1979 [22], and it is currently the most commonly used vegetation index in RS. The normalized difference of the near-infrared (NIR) and red (R) bands of Sentinel-2 images is used to calculate NDVI, which is calculated using Equation (1).

$$NDVI= (NIR-RED)/ (NIR+RED) \quad (1)$$

The EVI is calculated using Equation (2).

$$EVI=2.5*((NIR-RED)/ (NIR+6*RED-7.5*BLUE+1)) \quad (2)$$

The RENDVI is calculated using Equation (3).

$$RENDVI = (NIR-RedEdgeRED)/ (NIR+RedEdgeRED) \quad (3)$$

The NDMI is calculated using Equation (4).

$$NDMI= (NIR - SWIR1)/ (RED + SWIR1) \quad (4)$$

The NDBI is calculated to generate training points for the Built-up/Barren class using Equation (5).

$$NDBI= (SWIR - NIR) / (SWIR + NIR) \quad (5)$$

b) Acquired Scenes: The atmospheric corrected and cloudless Sentinel-2 images of dates 22 March 2018, 12 March 2019, 01 March 2020, and 26 March 2021 are used for this study.

2) Crop Calendar:

The acquisition period is decided based on phenological stages of the wheat crop's growing season which are taken from PMD.

3) Meteorological Data: Meteorological data obtained from *[23]* for the period 2018-2021. Then computed the monthly average for each weather parameter to assess its impact on yield.

4) Historical Data: Historical wheat area and yield data acquired from [24].

5) Ground Truth Field Data: Ground Truth wheat field data collected from SUPARCO. It includes the coordinates of 33 sample wheat points from the Faisalabad District.

III. METHODOLOGY

The detailed workflow of the framework is presented in Fig. 3. The workflow involves two major components: one for area estimation/LULC classification and the other for yield prediction.

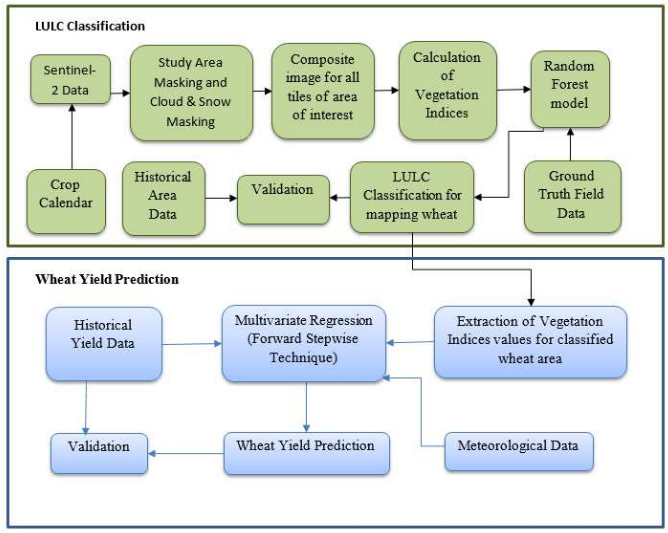

Fig. 3. *Two stage flow of the system – Stage 1 LULC classification, Stage 2 Wheat Yield Prediction*

A. LULC classification

Sentinel-2 satellite images acquired based on the crop calendar to identify the peak growing season of wheat. The image tiles stacked together and clipped to mask the study area for each year from 2018 to 2021. Calculated the VIs including NDVI, EVI, RENDVI, NDMI, and NDBI for all images. LULC Classification for 4 major land covers as wheat, water, built-up, and other vegetation performed using supervised ML algorithm RF to map wheat crop.

1) Training Data: Training data comprises 4 classes. Training data set for all classes each having 33 sample training points generated to apply supervised ML Algorithms. As the acquired ground truth of the wheat class contain 33 sample points, therefore same numbers of sample training points generated for the rest of the classes.

a) Wheat training points: The ground truth wheat field data including coordinates of 33 wheat fields used as sample training points for the wheat class.

b) Water training points: To generate water training points, selected the smaller region from the area of interest to reduce computation as water percentage in the Faisalabad district is very less hence extracting values from the complete region would have been computation extensive. Then discretized the selected area into several points and extracted

NDVI values for each point. The points meeting the threshold value i.e. (-ve) selected as sample points for water class.

c) Built-up/Bare Land training points: To generate training points for this class, discretized the area of interest into several points and extracted NDBI values for each point. The points meeting the threshold value found (>0.05) in the case of Faisalabad district after extracting values of different built-up areas selected as sample points for the Built-up/Bare Land class.

d) Other Vegetation training points: 33 points other than wheat marked by visually analyzing area of interest in Google Earth and used as training points for other vegetation classes.

Multiple Bands and VIs combinations used as input features for the ML algorithm listed in TABLE 1.

TABLE 1. BANDS AND VIS COMBINATIONS

S. No	Bands/VIs Combination	Resolution	Description
1	B4,B3,B2,B8,nd	10 m	Red, Green, Blue & NIR bands and NDVI
2	nd, NIR	10 m	NDVI and NIR bands
3	B5,B6,B7,SWIR 1	20 m	Vegetation Red Edge Bands and SWIR Band
4	B5,B6,B7	20 m	Vegetation Red Edge Bands
5	nd	10 m	NDVI
6	B5,B6,B7,RENDVI,NDMI	20 m	Vegetation Red Edge Bands and RENDVI & NDMI
7	nd, EVI	10 m	NDVI & EVI

After comparing the overall test accuracy, consumer's accuracy, producer's accuracy, and kappa statistics of all Bands/VIs Combination found that RF Algorithm using Bands/VIs Combination of NDVI and EVI provides the most accurate results for LULC. Hence selected this model for generating wheat crop maps and calculation of wheat area. The spatial resolution of Sentinel-2 bands used to calculate NDVI and EVI is 10 meters which means one pixel contains an area of 10*10=100 m2. To compute the overall wheat area of the Faisalabad district, added the area of all classified wheat pixels. The calculated area results were compared with the actual area taken from [24] for validation. The differences between the calculated and actual area computed for each year.

B. Wheat Yield Prediction

The NDVI and EVI values of wheat extracted from wheat crop map along with other meteorological parameters including Humidity, Temperature, Wind Speed, Wind Direction, Mean Sea Level Pressure, Soil Temperature, Soil Moisture, and Sunshine Duration used for yield prediction. Multivariate regression technique applied on different combinations of meteorological parameters with NDVI and EVI to predict wheat yield. The forward Stepwise technique used to select the best combination of variables. It involves first developing models with one variable and then selecting the best model among them, then adding one more variable to it and finding the optimal model using 2 variables. Repeat the process until all variables are tested. Then find the simplest model among all selected optimal models which provide the best results. Regression was applied with variables 1 till 10. The simplest model that gives the best results found the one with 3 variables (NDVI, EVI, and Sunshine Duration). Hence selected this model for yield prediction. The yield results were validated with the historical yield data.

IV. RESULTS

A. LULC Classification

The RF model with Band/VIs combination of NDVI and EVI can estimate wheat area with an average accuracy of 84%. The overall Test Accuracies, Consumer's Accuracies, Producer's Accuracies, Kappa Statistics, Calculated and Actual Wheat Area, and Difference of Calculated Area from Actual Area for years 2018-2021 are listed in TABLE 2. Computed and Actual area using the best-selected model for the year 2018 till 2021 is represented by Fig. 4. Fig. 5 shows the Overall Test Accuracy, Consumer's Accuracy, and Producer's Accuracy using RF model with Band/VIs combination of NDVI and EVI for the year 2018-2021

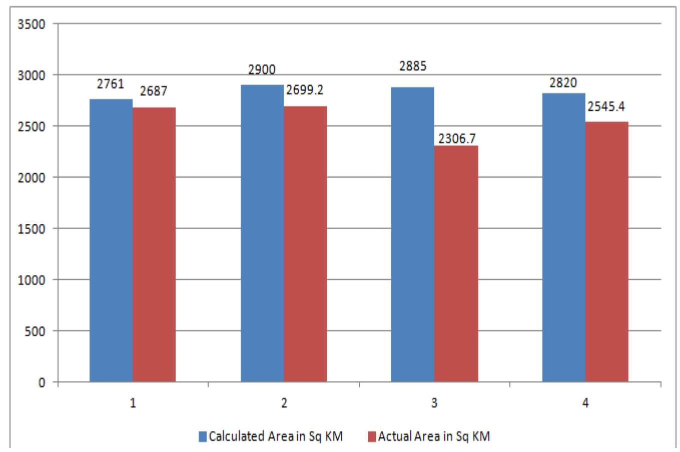

Fig. 4. *Actual vs Calculated Wheat area in Sq KM using RF model with Bands/VIs combination of NDVI and EVI for the year 2018-2021*

Fig. 5. *Overall Test Accuracy, Consumer's Accuracy, and Producer's Accuracy using RF model with Band/VIs combination of NDVI and EVI for the year 2018-2021*

TABLE 2. Overall Test Accuracies, Consumer's Accuracies, Producer's Accuracies, Kappa Statistics, Calculated and Actual Wheat Area and Difference of calculated area from Actual Area for years 2018-2021 using best-selected model

Year	Test Accuracy	Consumer's Accuracy	Producer's Accuracy	Kappa Statistics	Calculated Area in Sq KM	Actual Area in Sq KM	Difference
2018	77.78	76.2	75.9	0.766	2761	2687	74
2019	86.11	86.4	85.9	0.812	2900	2699.2	200.8
2020	87.1	83.5	82.3	0.823	2885	2306.7	578.3
2021	85.29	77.5	86.6	0.805	2820	2545.4	274.6

Wheat crop maps were generated using the above results for years 2018, 2019, 2020, and 2021 are shown in Fig. 6.

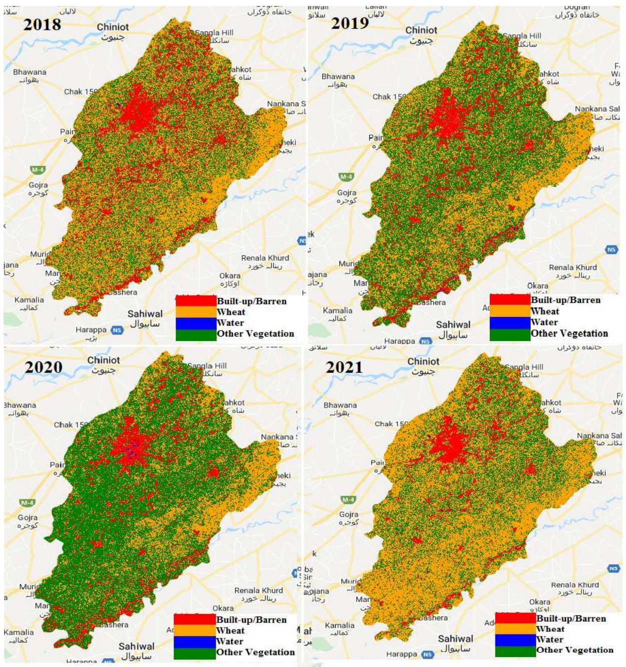

Fig. 6. *LULC Classification as Built-up/Barren Land, Wheat, Water and Other Vegetation for year 2018-2021*

The total area comprising of wheat increased in 2019 as compared to 2018 and it later reduced in 2021which can be seen from LULC maps.

B. *Yield Prediction*

The most accurate results of yield prediction obtained by applying multivariate regression on NDVI, EVI, and Sunshine Duration. In the regression equation obtained (6) Y represents yield, 'a' corresponds to NDVI value, 'b' corresponds to Sunshine duration and 'c' corresponds to EVI.

$$Y=3011.364+11460.46a+12.4211b-11526.9c \qquad (6)$$

The R2 between reported and predicted yield found 0.96. As it is the case of multivariate regression hence adjusted R2 value is considered which found to be 0.84 with a standard error of 46.14 kg/ha for yield prediction.

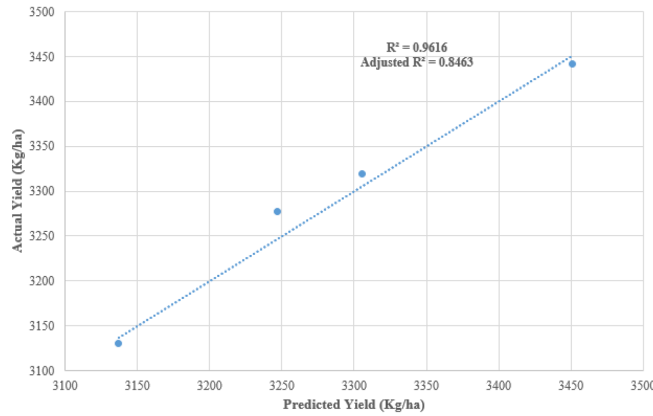

Fig. 7. *Trend line of predicted yield and actual yield for year 2018-2021*

Predicted Yield using Equation (6) and Actual Yield for the year 2018 till 2021 is represented by Fig. 8 and values are listed in TABLE 3.

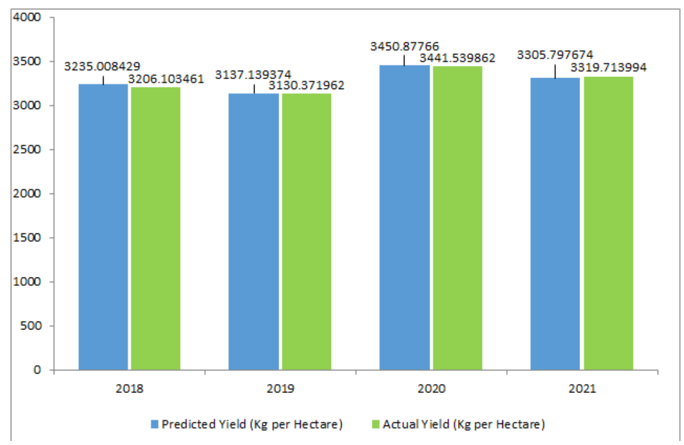

Fig. 8. *Actual and Predicted Wheat Yield using multivariate regression on NDVI, EVI, and Sunshine duration in Kg/ha for the year 2018-2021*

TABLE 3: Predicted and Actual Wheat Yield

Year	Predicted Yield (Kg per Hectare)	Actual Yield (Kg per Hectare)	Difference
2018	3235.008	3206.103	28.905
2019	3137.139	3130.372	6.767
2020	3450.877	3441.54	9.337
2021	3305.797	3319.714	13.916

V. Conclusion

Early forecasting of wheat yield can assist in improving the economic growth of the country by managing consumption, distribution, storage, and shortcoming well in advance. Hence it is needed to have such kinds of algorithms which can predict yield with great accuracy. The high-resolution satellite images of Sentinel-2 can be used for early prediction of the total wheat area before harvesting during Heading, Flowering, and Milk Maturity Stages. This framework compares multiple Bands/VIs combinations for generating LULC maps for 4 land cover classes wheat, water, built-up, and other vegetation and predicting wheat area of Faisalabad district and proposed the best VIs combination i.e. NDVI and EVI using 10m spatial resolution bands which provide accurate LULC map and determine wheat area with an accuracy of 84%, consumer's accuracy of 81%, producer's accuracy of 83%, and kappa statistics of 0.8. NDVI and EVI computed from RS images along with weather parameters are useful to determine yield value in advance before actual harvesting. Meteorological data have a major impact on yield prediction. The impact of 8 meteorological parameters on yield assessed and in the end, only one parameter proposed which provides yield values with greater accuracy. This framework can predict yield value with an error of 46.14 Kg/ha and the adjusted R2 between reported and predicted yield of 0.84 using sunshine duration, EVI, and NDVI. In a study [17], the wheat yield they predicted for Faisalabad using NDWI with an error of 284.8 kg/ha which is greater than this proposed framework as they used Landsat satellite which has low resolution than Sentinel-2. The R2 between the reported and predicted yield found 0.69 by using unsupervised clustering and linear regression technique to identify the wheat area and predict wheat yield for Faisalabad [11]. They did not incorporate meteorological parameters to predict yield results and also the NDVI product of satellite used has low resolution than Sentinel-2. The wheat yield in china predicted in [15] with an R2 of 0.75 and a yield error of less than 10%. Although the soil moisture data and meteorological data incorporated for yield prediction the predicted results reported less accurate than the proposed framework due to the low resolution of satellite images.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Suhaib Bin Farhan, Mr. Syed Shariq Mobeen, Mr. Syed Farhan Ahmed Khalil, and Mr. Syed Faisal Hussain Kazmi, from SUPARCO for their provision in acquiring the ground truth wheat fields' data and meteoblue for their support in accessing meteorological data of Faisalabad district. The authors also acknowledge the NCAI-smart city lab at NED University for their collaboration in this research.

REFERENCES

[1] Saad-ul-Haque, B. Ghauri, and M. R. Khan, "Short term drought monitoring using remote sensing technique: A case study of Potohar region, Pakistan," in *2013 International Conference on Aerospace Science & Engineering (ICASE)*, Islamabad, Aug. 2013, pp. 1–8. doi: 10.1109/ICASE.2013.6785571.

[2] U. Saeed, J. Dempewolf, I. Becker-Reshef, A. Khan, A. Ahmad, and S. A. Wajid, "Forecasting wheat yield from weather data and MODIS NDVI using Random Forests for Punjab province, Pakistan," *Int. J. Remote Sens.*, vol. 38, no. 17, pp. 4831–4854, Sep. 2017, doi: 10.1080/01431161.2017.1323282.

[3] "Agriculture Statistics | Pakistan Bureau of Statistics." https://www.pbs.gov.pk/content/agriculture-statistics (accessed Feb. 08, 2022).

[4] "| Ministry of Finance | Government of Pakistan |." http://www.finance.gov.pk/survey_1920.html (accessed Feb. 08, 2022).

[5] Mjm Cheema, Allah Bakhsh, Talha Mehmood, and U. Liaqat, "Assessment of water allocations using remote sensing and GIS modeling for Indus Basin, Pakistan," 2016, doi: 10.13140/RG.2.1.2068.4566.

[6] I. Ahmad *et al.*, "Yield Forecasting of Spring Maize Using Remote Sensing and Crop Modeling in Faisalabad-Punjab Pakistan," *J. Indian Soc. Remote Sens.*, vol. 46, no. 10, pp. 1701–1711, Oct. 2018, doi: 10.1007/s12524-018-0825-8.

[7] K. Guan *et al.*, "The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields," *Remote Sens. Environ.*, vol. 199, pp. 333–349, Sep. 2017, doi: 10.1016/j.rse.2017.06.043.

[8] Q. Yang, L. Shi, J. Han, Y. Zha, and P. Zhu, "Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images," *Field Crops Res.*, vol. 235, pp. 142–153, Apr. 2019, doi: 10.1016/j.fcr.2019.02.022.

[9] Y. Cai *et al.*, "Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches," *Agric. For. Meteorol.*, vol. 274, pp. 144–159, Aug. 2019, doi: 10.1016/j.agrformet.2019.03.010.

[10] A. Rasheed, M. S. Younis, F. Ahmad, J. Qadir, and M. Kashif, "District Wise Price Forecasting of Wheat in Pakistan using Deep Learning," 2021, doi: 10.48550/ARXIV.2103.04781.

[11] R. Mumtaz, S. Baig, and I. Fatima, "Analysis of meteorological variations on wheat yield and its estimation using remotely sensed data. A case study of selected districts of Punjab Province, Pakistan (2001-14)," *Ital. J. Agron.*, vol. 12, no. 3, Oct. 2017, doi: 10.4081/ija.2017.897.

[12] A. Nazeer, M. M. Waqas, S. Ali, U. K. Awan, M. J. M. Cheema, and A. Baksh, "LAND USE LAND COVER CLASSIFICATION AND WHEAT YIELD PREDICTION IN THE LOWER CHENAB CANAL SYSTEM USING REMOTE SENSING AND GIS," *Big Data Agric.*, vol. 2, no. 2, pp. 47–51, Jan. 2020, doi: 10.26480/bda.02.2020.47.51.

[13] I. Becker-Reshef, E. Vermote, M. Lindeman, and C. Justice, "A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data," *Remote Sens. Environ.*, vol. 114, no. 6, pp. 1312–1323, Jun. 2010, doi: 10.1016/j.rse.2010.01.010.

[14] J. Dempewolf *et al.*, "Wheat Yield Forecasting for Punjab Province from Vegetation Index Time Series and Historic Crop Statistics," *Remote Sens.*, vol. 6, no. 10, pp. 9653–9675, Oct. 2014, doi: 10.3390/rs6109653.

[15] J. Han *et al.*, "Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China," *Remote Sens.*, vol. 12, no. 2, p. 236, Jan. 2020, doi: 10.3390/rs12020236.

[16] Y. Zhao, A. B. Potgieter, M. Zhang, B. Wu, and G. L. Hammer, "Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling," *Remote Sens.*, vol. 12, no. 6, p. 1024, Mar. 2020, doi: 10.3390/rs12061024.

[17] M. Fahad, I. Ahmad, M. Rehman, M. M. Waqas, and F. Gul, "Regional Wheat Yield Estimation by Integration of Remotely Sensed Soil Moisture into a Crop Model," *Can. J. Remote Sens.*, vol. 45, no. 6, pp. 770–781, Nov. 2019, doi: 10.1080/07038992.2019.1692651.

[18] L. Li *et al.*, "Developing machine learning models with multi-source environmental data to predict wheat yield in China," *Comput. Electron. Agric.*, vol. 194, p. 106790, Mar. 2022, doi: 10.1016/j.compag.2022.106790.

[19] X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques," *Comput. Electron. Agric.*, vol. 121, pp. 57–65, Feb. 2016, doi: 10.1016/j.compag.2015.11.018.

[20] muziadmin, "Overview," *Agriculture Department | Government of the Punjab*, Mar. 20, 2014. http://www.agripunjab.gov.pk/overview (accessed Feb. 08, 2022).

[21] "Sentinel Online - ESA - Sentinel Online." https://sentinel.esa.int/web/sentinel/home (accessed Feb. 08, 2022).

[22] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, May 1979, doi: 10.1016/0034-4257(79)90013-0.

[23] "Weather Faisalabad," *meteoblue*. https://www.meteoblue.com/en/weather/week/faisalabad_pakistan_1179400 (accessed Feb. 08, 2022).

[24] "AMIS Agriculture Marketing Wing Punjab." http://amis.pk/ (accessed Feb. 08, 2022).