Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

DEVELOPMENT OF A PREDICTIVE MODEL FOR WHEAT YIELD USING MULTISPECTRAL SATELLITE IMAGERY AND GROUND TRUTH DATA IN FAISALABAD DIVISION, PUNJAB, PAKISTAN

HAFIZ MUHAMMAD BILAWAL AKRAM *
Department of Agronomy, University of Agriculture, Faisalabad, Pakistan.
*Corresponding Author Email: bilawalakram48@gmail.com

SYED AFTAB WAJID
Department of Agronomy, University of Agriculture, Faisalabad, Pakistan.

KHALID HUSSAIN
Department of Agronomy, University of Agriculture, Faisalabad, Pakistan.

SIKANDAR ALI
Department of Irrigation & Drainage, University of Agriculture, Faisalabad, Pakistan.

Abstract

Crop yield estimation has significant importance for policy makers to make timely decisions on import/ export of particular crop. Another method for determining vegetation health and yield is the use of satellite imagery. Although several vegetative indices are being utilized, it is unknown how effective they are at estimating yield. This study compared several satellite-based vegetation indices, including the Enhanced Vegetation Index (EVI), Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), and Modified Soil Adjusted Vegetation Index (MSAVI), to determine which index is most appropriate for the central Punjab, Pakistan cropping area. The research focuses on analyzing the correlation between these vegetation indices and biomass/biological yield and grain yield. Through scatter plots and regression analyses, the study reveals strong positive correlations between these vegetation indices and crop yields, demonstrating their effectiveness as indicators for predicting agricultural productivity. SAVI and MSAVI showed high reliability in semi-arid regions by minimizing soil brightness effects. EVI, with its additional correction for soil and atmospheric influences, proved particularly effective in densely vegetated areas. NDVI also showed a significant correlation with crop yield but was found to be less effective in regions with sparse vegetation due to its sensitivity to soil reflectance. The results revealed that all vegetation indices have a positive correlation with wheat yield, but their predictive power varies. Model-1 of (Wheat Grain Yield), which incorporates all four indices, showed the best performance with an R-squared of 0.91 and a Pearson correlation of 0.95, indicating a strong fit to the observed data. However, its NSE value of 0.89 suggests moderate predictive reliability. Among the vegetation indices, NDVI emerged as the most significant predictor of yield due to its high positive coefficient in the regression models. These findings suggest that the appropriate selection of vegetation indices, considering environmental context, is crucial for accurate yield prediction.

1. INTRODUCTION

The sustainable management of agricultural resources is crucial for ensuring food security, particularly in semi-arid regions where water scarcity and unpredictable climatic conditions pose significant challenges (Mesquita and Milhorance, 2019). Global warming and rising temperatures negatively impact cereal yields, with studies indicating that a 1ºC

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

increase in minimum temperature can reduce cereal production by up to 10% (Rehan et al., 2024). In this context, accurate land use classification and yield prediction are essential for optimizing agricultural practices and enhancing productivity. In the past, the most popular techniques for gathering information on crop nutrition, crop growth, crop yield, and soil nutrition were surveys, field sampling, and laboratory analysis. This work has mostly been completed in the past. Production managers can gather point data, history data, current data, and any other information required in the field of precision agriculture to make decisions on variable-rate activities (Farzand et al., 2023). Remote sensing technologies, such as Sentinel imagery, coupled with field data, have emerged as powerful tools for monitoring crop health, estimating yields, and making informed decisions on resource allocation (Debella-Gilo and Gjertsen, 2021).

Wheat (*Triticum aestivum* L.) is a staple food crop and a significant contributor to the agricultural economy of Pakistan, especially in the Punjab province. Faisalabad, located in this province, is characterized by semi-arid conditions with limited and erratic rainfall patterns (Fahad *et al.,* 2019). These conditions necessitate the efficient management of available resources to maximize wheat yield (SAEED, 2017). However, traditional methods of monitoring crop performance and predicting yields often fall short in terms of accuracy and timeliness. This gap has led to an increasing interest in integrating remote sensing data with ground-based observations to improve the precision of land use classification and yield estimation (Xie and Huang, 2021; Guo *et al.,* 2018).

The use of Sentinel satellite imagery has gained traction in recent years due to its high spatial and temporal resolution, which allows for detailed monitoring of crop phenology and land use patterns (Gao and Zhang, 2021; Holtgrave *et al.,* 2020). Sentinel-2, with its multispectral imaging capabilities, provides critical data for assessing vegetation health and classifying land use based on spectral signatures. Additionally, the integration of Sentinel-1 radar data, which is less affected by cloud cover, enhances the robustness of monitoring in regions prone to frequent cloudiness during key growth stages (De Fioravante *et al.,* 2021; De Luca *et al.,* 2022).

In semi-arid regions like Faisalabad, the challenge lies not only in classifying land use accurately but also in predicting wheat yield under variable environmental conditions. Several studies have demonstrated the effectiveness of combining remote sensing data with machine learning algorithms to model and predict crop yields (Arshad *et al.,* 2023; Kanwal *et al.,* 2021). These models leverage the spectral information from Sentinel imagery along with ground-truth data, such as soil moisture content, nutrient levels, and historical yield records, to produce reliable yield estimates (Ahmad *et al.,* 2018; Ahmad *et al.,* 2020). The predictive models are particularly valuable for making timely decisions regarding irrigation, fertilization, and pest control, which are crucial for optimizing wheat production in resource-limited settings (Qader *et al.,* 2021; Snigdha, 2022).

Despite the advancements in remote sensing and yield prediction techniques, there is a need for localized studies that account for the specific agro-climatic conditions of regions like Faisalabad. The variability in soil types, irrigation practices, and cropping systems

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

across different areas within the same region can significantly impact the accuracy of land use classification and yield prediction models. Therefore, this research aims to address these gaps by utilizing Sentinel imagery and field data to develop a robust framework for land use classification and wheat yield prediction in the semi-arid conditions of Faisalabad.

2. MATERIAL AND METHODS

2.1 Study Area

Study was conducted at Faisalabad division in Central Punjab Pakistan (figure 1) for wheat crop during the growing season of 2021-22. The latitude 31.4504° N, longitude 73.1350° E and the elevation from the sea level 184 m. It has semi-arid climate characteristics. A comprehensive survey was conducted in Faisalabad division to collect the crop management data from the farmers. Stratified random sampling technique was used for the selection of farms. Mobile Agricultural Geo-tagging Information System (MAGIS) was used to collect data.

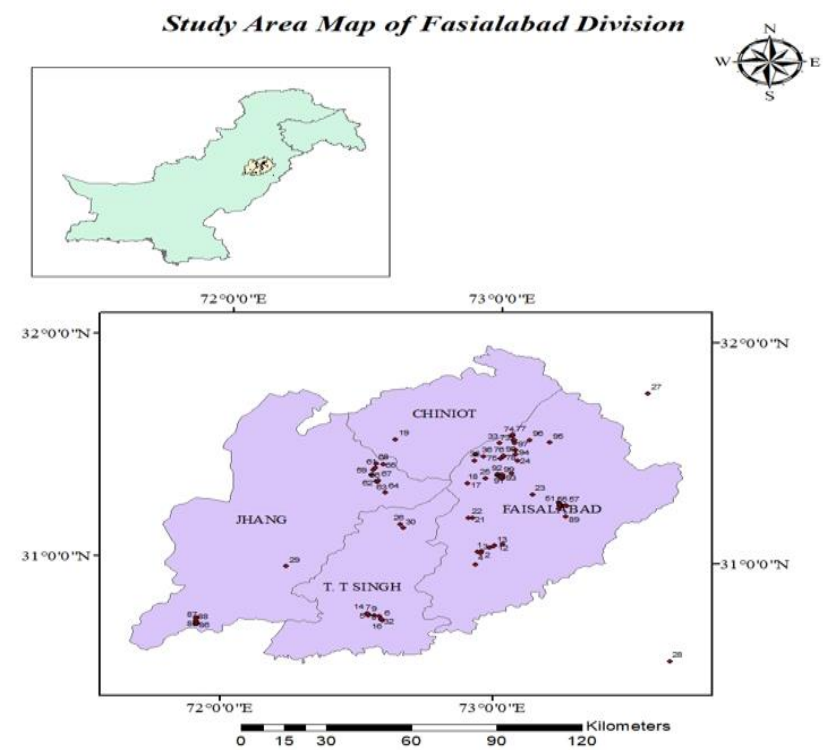

Figure 1: Study Area of Faisalabad Division Punjab Pakistan

There are two dominant cropping seasons, the summer season commonly called ‘‘kharif” that starts from May and ends by October. The winter season called ‘‘rabi” in which sowing starts from November and harvesting is completed in April. Wheat is a major rabi crop while rice and cotton are major kharif crops.

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

Table 1: Physical and Chemical Properties of the Soils

```
*UFAGFA0027  SSOP        SIL      150 Faisalabad,Jaranwala
@SITE        COUNTRY          LAT     LONG SCS FAMILY
 Faisalabad  Pakistan        31.28   73.35 Coarse silty,mixed,hyperthermic,Typic Haplocambids
@ SCOM  SALB  SLU1  SLDR  SLRO  SLNF  SLPF  SMHB  SMPX  SMKE
    BN   .13     6    .6    73     1     1 IB001 IB001 IB001
@  SLB  SLMH  SLLL  SDUL  SSAT  SRGF  SSKS  SBDM  SLOC  SLCL  SLSI  SLCF  SLNI  SLHW  SLHB  SCEC  SADC
    18    AP  .074  .192  .357     1   .68  1.24   -99    14    57    29   -99   8.2   -99   -99   -99
    40   Bw1  .073  .188  .349   .56   .68  1.25   -99    14    56    30   -99   8.2   -99   -99   -99
   110   Bw2  .079  .198  .362  .223   .68  1.24   -99    16    56    28   -99   8.2   -99   -99   -99
   130   BCK  .073  .204  .385  .091   .68  1.19   -99    12    62    26   -99   8.2   -99   -99   -99
   150    CK  .082  .224  .414  .061   .68  1.18   -99    14    65    21   -99   8.3   -99   -99   -99
```

The data in the table.1 provides information about the physical and chemical properties of the soils at Faisalabad. Some key observations can be made:

i. **Soil Series and Texture:** The "SIL" column indicates the soil series, which is a classification based on similar characteristics. The combination of particle size analysis (sand, silt, clay) and texture class provides insights into the soil's physical properties.

ii. **Soil Drainage and Infiltration:** The "SLDR" and "SLINF" columns suggest the soil's drainage and infiltration characteristics, which are important for plant growth and water retention.

iii. **Soil Structure and Bulk Density:** These properties influence soil aeration, water movement, and root penetration.

iv. **Soil pH:** The "8.2" values in the pH column indicate that the soils are slightly alkaline, which may affect nutrient availability and plant growth.

2.2. Field Data

Ground truth data was collected from various wheat fields in the study area showing in figure 1. Different sampling points were selected based on a stratified random sampling method. At each sampling point, the following data were collected: wheat yield, soil type, soil moisture content, and GPS (Global Positioning System) coordinates. Yield data were recorded at the time of harvest, and soil samples were collected for laboratory analysis.

2.3. Sentinel-2 Data

Sentinel-2 data was acquired from (https://earthexplorer.usgs.gov) for the period corresponding to the wheat growing season, from November 2021 to April 2022. The Sentinel-2 data has a spatial resolution of 10 meters and provides multispectral images across 13 bands, which are particularly useful for vegetation analysis and land use classification. A total of 25 tiles were required to cover the Faisalabad Division of Punjab, Pakistan.

2.4. Crop Data

Crop data in Pakistan are gathered and distributed by the Crop Reporting Services (CRS) and the Provincial Agricultural Department. Additionally, these departments use field

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

research, farm visits, and ground truth to create official agriculture statistics on crops. The CRS department uses the following hierarchy to display data.

The source of the crop data was agricultural statistics. At the district and provincial levels, CRS interpreters and researchers working under the temporary agricultural department convened to gather data on crop acreage, average production, and farming challenges faced by farmers (Dempewolf et al., 2014). In this study, we obtained yield data from the Provisional Agriculture and CRS department.

2.5. Vegetation Indices

Spectral vegetation indices (VIs) are mathematical numerical values that represent several spectral bands, primarily in the visible and near-infrared sections of the electromagnetic spectrum. These indices provide a comprehensive assessment of leaf chlorophyll content, leaf area, optical measurements of canopy greenness, and canopy structure. Furthermore, the identification of vegetation patterns and the examination of vegetation health are highly beneficial for the surveillance of crop yield and the management of natural resources (Ye et al., 2008; Funk and Budde, 2009; Subash et al., 2011; Reddersen et al., 2014). In this study indices provided in Table 2 were examined.

NIR, Red, and Blue refer to the spectral brightness measured in the near infrared, red, and blue sections of the electromagnetic spectrum. The value of the adjustment factor L is contingent upon the vegetation cover. Under thick or heavy vegetation conditions, the value is zero, which is equivalent to NDVI. Conversely, for low vegetation, the value is 1.

Table 2: Selected Functionally Various Vegetation Indices (VIs)

Index	Formula
NDVI	**(NIR - RED) / (NIR + RED)**
EVI	**2.5 * (NIR - RED) / (NIR + 6RED - 7.5BLUE + 1)**
SAVI	**(NIR - RED) * (1 + L) / (NIR + RED + L)**
MSAVI	**0.5 * (2NIR + 1 – sqrt ((2NIR + 1)^2 - 8*(NIR - RED)))**

*L: Soil brightness correction factor (typically 0.5)

2.6. Wheat Yield Prediction Model

A regression model was developed to predict wheat yield based on NDVI and other vegetation indices derived from Sentinel-2 imagery. The model was calibrated using the field data collected at the sampling points. The regression analysis was performed using R-studio software, and the model's accuracy was evaluated using the coefficient of determination (R^2) and Pearson correlation (r), and Nash-Sutcliffe Efficiency (NSE).

2.7. Methodology

The terms NIR, Red, and Blue denote the spectral luminosity that is measured within the near infrared, red, and blue regions of the electromagnetic spectrum. The specific value of the adjustment factor L depends on the level of vegetation cover. Under dense or abundant vegetation, the value is zero, which is equal to the Normalized Difference

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

Vegetation Index (NDVI). In contrast, for minimal vegetation, the value is 1. Values of 2.5, 6, and 7.5 are assigned to the correction factors G, C1, and C2, respectively (Bastiaanssen and Ali, 2003). The abbreviations NIR, Red, and Blue refer to the measure of spectral luminosity within the near infrared, red, and blue sections of the electromagnetic spectrum. The precise magnitude of the adjustment factor L is contingent upon the extent of vegetation coverage. In the presence of dense or plentiful vegetation, the value is zero, which corresponds to the Normalized Difference Vegetation Index (NDVI). Conversely, minimum vegetation yields a value of 1. Specifically, the correction factors G, C1, and C2 are assigned values of 2.5, 6, and 7.5, respectively. A regression association between measured yield and vegetation indices (NDVI, EVI, SAVI, and MSAVI) was established at the heading stage, following the technique proposed by Bastiaanssen et al. (1999) given in Eq. (1).

$$\mathbf{Y_{RS} = m\ Wheat_{\ (Vegetation\ Indices)} + c} \quad (1)$$

where YRS denotes the wheat yield based on remote sensing, m and c depicts the corresponding slope and intercepting parameters of above regression equations found by replacing the yield and average Wheat(NDVI) or Wheat(SAVI) or Wheat(MSAVI) or Wheat (EVI) values (From Provisional Agriculture Department and CRS) at the crop heading stage respectively.

3. RESULTS AND DISCUSSIONS

3.1. Correlation between Soil Adjusted Vegetation Index (SAVI) and Biomass/Biological Yield (Y):

The scatter plot in Figure 2 shows a fitted regression line showing the relationship between the Soil Adjusted Vegetation Index (SAVI) and the variable "y" (which is Biomass or Biological Yield.

a) Positive Correlation between SAVI and Biological Yield (Y):

The scatter plot demonstrates a clear positive linear relationship between SAVI and the variable "y" (Biomass or Crop Yield) Figure 2. The points are scattered around the regression line with some variability, indicating that while there is a general trend of increasing yield (y) with increasing SAVI values, there is also some degree of variation that could be attributed to other factors. The confidence interval shaded around the regression line shows the uncertainty of the prediction; narrower intervals indicate more reliable predictions in certain SAVI ranges.

b) Statistical Significance of the Relationship:

In figure 2 the positive slope of the regression line suggests that as SAVI increases, there is a corresponding increase in yield. This observation aligns with the expected behavior since SAVI is designed to minimize the effects of soil background, which is crucial in areas with varying soil conditions. If the p-value of the regression slope is significant

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

(typically less than 0.05), it can be concluded that SAVI is a significant predictor of yield in the study area.

c) Effectiveness of SAVI as A Predictor of Yield:

The positive relationship between SAVI and yield implies that SAVI is an effective indicator for estimating crop biomass or yield in the study area. This finding is consistent with previous research that highlights SAVI's ability to account for soil influences, making it more robust in environments with sparse vegetation or exposed soil surfaces Huete, (1988).

d) Implications for Precision Agriculture:

The ability to use SAVI as a predictor of crop yield has significant implications for precision agriculture. By integrating SAVI-based remote sensing data with other agronomic and environmental variables, farmers and decision-makers can better manage crop inputs, optimize yield, and promote sustainable agricultural practices Mulla, (2013).

e) Comparison With Other Vegetation Indices:

While SAVI has shown a strong correlation with yield, it is important to compare this with other indices like EVI or NDVI to determine the most suitable index for yield prediction in different conditions. SAVI's advantage in areas with significant soil exposure may be complemented by the use of indices like MSAVI, which further adjusts for soil brightness Qi *et al.,*(1994).

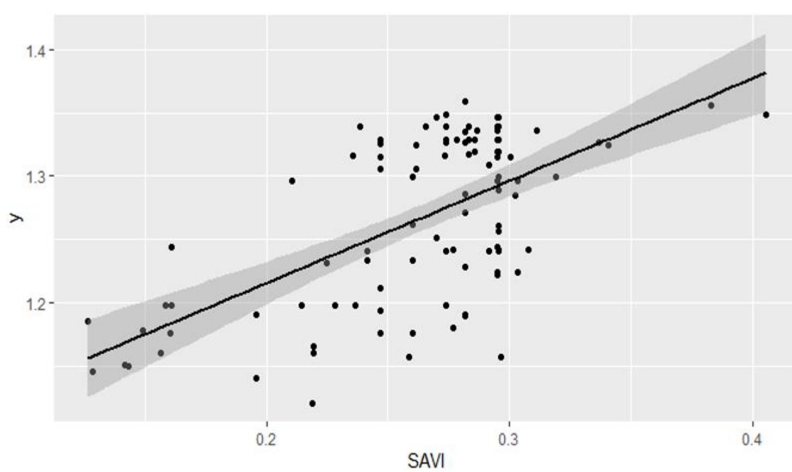

Figure 2: Correlation between SAVI and Biological Yield (Y)

3.2. Correlation between Modified Soil Adjusted Vegetation Index (MSAVI) and Biomass/Biological Yield (Y):

The scatter plot with a regression line showing the relationship between the Modified Soil Adjusted Vegetation Index (MSAVI) and biological yield (Y) figure 3.

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

a) Positive Correlation between MSAVI and Biological Yield (Y):

The scatter plot (figure 3) indicates a positive linear relationship between MSAVI and biological yield (Y). As MSAVI values increase, there is a corresponding increase in biological yield. The points are scattered around the regression line, but the positive slope suggests that higher MSAVI values are associated with higher yields. The shaded region around the regression line represents the confidence interval, indicating the uncertainty of the regression predictions. A relatively narrow interval suggests a reliable predictive relationship between MSAVI and biological yield for most of the observed range.

b) Utility of MSAVI IN Yield Prediction:

The positive association between MSAVI and biological yield indicates that MSAVI, which adjusts for soil brightness and reduces the influence of soil background, is an effective index for monitoring crop health and predicting yield. This is particularly useful in regions like Faisalabad Division, where soil conditions vary and vegetation cover may be sparse or heterogeneous Qi et al. (1994). .The strong relationship observed here between MSAVI and biological yield supports the findings of previous studies, which have shown MSAVI's effectiveness in minimizing soil noise in remote sensing data, making it a valuable tool for precision agriculture Rondeaux et al., (1996).

c) Implications for Agricultural Management:

The use of MSAVI for yield prediction can significantly improve the precision of agricultural management practices, particularly in managing fertilization, irrigation, and other inputs tailored to the specific needs of different field zones. The integration of MSAVI in yield models can help optimize resource use and increase productivity while promoting sustainable practices Thenkabail et al., (2000).

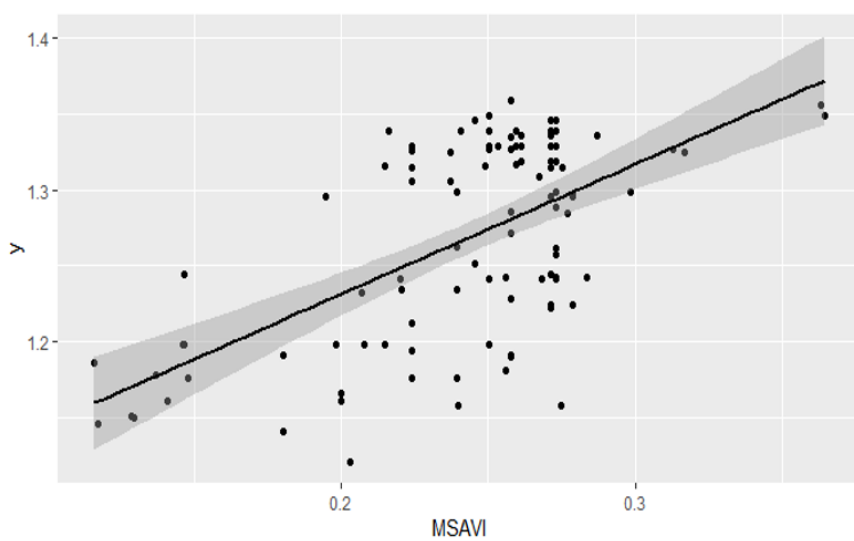

Figure 3: Correlation between MSAVI and Biological Yield (Y)

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

3.3. Correlation between Modified Enhanced Vegetation Index (EVI) and Biomass/Biological Yield (Y)

a) Analysis of Biological Yield and EVI:

The scatterplot figure 4 demonstrates a strong positive linear relationship between Biological Yield (Y) and Enhanced Vegetation Index (EVI). This suggests that as EVI values increase, biological yield tends to increase as well. The regression line and confidence interval provide a clear visual representation of this relationship.

b) Ecological Implications:

EVI is a widely used vegetation index that measures the greenness of vegetation, and it has been shown to be a reliable indicator of plant growth and productivity. The results of this analysis suggest that EVI can be effectively used as a proxy for biological yield Wang, **(2017).**

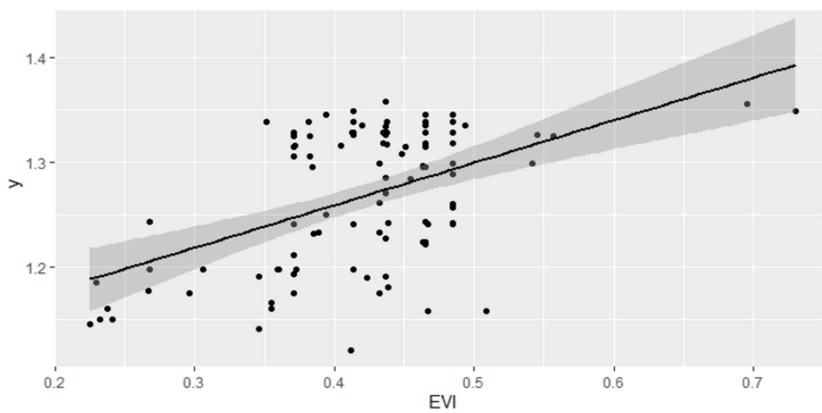

Figure 4: Correlation between EVI and Biological Yield (Y)

3.4. Correlation between Different Vegetation Indices and Biomass/Biological Yield (Y)

The scatterplot matrix in (figure 5) shows a series of pairwise relationships between BLY and vegetation indices such as EVI, SAVI, and MSAVI. There appears to be a positive correlation between BLY and each of the vegetation indices (EVI, SAVI, MSAVI), as indicated by the upward trend in the scatterplots. This suggests that as the values of these vegetation indices increase, the biomass or crop yield (BLY) tends to increase as well. The scatterplot between EVI and BLY, SAVI and BLY, and MSAVI and BLY show relatively tighter clustering, indicating a stronger relationship. The positive correlations between BLY and the vegetation indices (EVI, SAVI, MSAVI) indicate that these indices can be effective predictors of crop yield. In particular, the strong correlation with EVI suggests it might be a reliable indicator for estimating crop biomass or yield during the wheat growing season. This finding aligns with the work of Huete *et al.* (2002), who demonstrated the utility of EVI in accurately reflecting vegetation dynamics due to its reduced sensitivity to atmospheric and soil background conditions.

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

a) Inter-Correlation Among Vegetation Indices (EVI, SAVI, MSAVI, NDVI) with Biological Yield (BLY):

The relationships among EVI, SAVI, NDVI and MSAVI show strong positive correlations, which is expected as these indices are derived from similar spectral bands and are often used for similar purposes (e.g., vegetation monitoring). The scatterplots between these indices form a roughly linear pattern, which supports the idea that they are highly correlated. While EVI and MSAVI both show strong correlations with BLY, MSAVI may have a slight advantage in conditions where soil influence is significant, as suggested by Qi *et al.* (1994). MSAVI incorporates a soil brightness correction factor, making it potentially more robust in arid and semi-arid environments like Faisalabad Division.

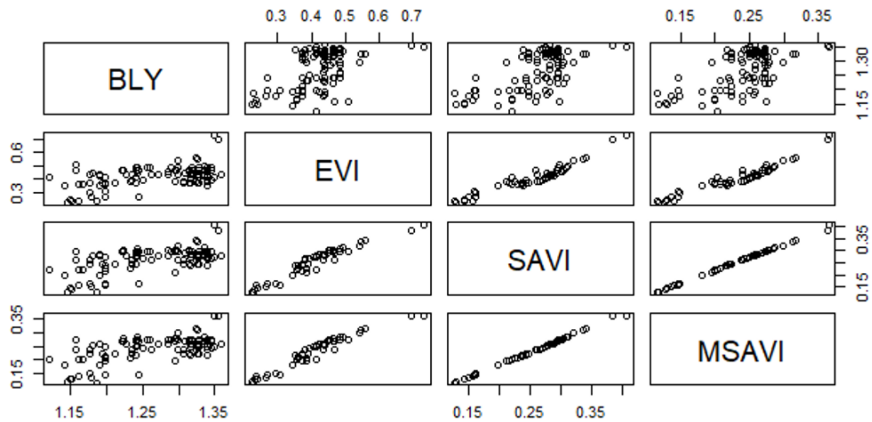

Figure 5: Scatterplot matrix displaying the relationships among different vegetation indices with biological yield (BLY)

3.5. Correlation between Soil Adjusted Vegetation Index (SAVI) and Grain Yield (Y)

The scatter plot presented in figure 6 illustrates the relationships between Soil-Adjusted Vegetation Index (SAVI) and grain yield (Y).

a) Relationship between SAVI and Grain Yield (Y):

The scatter plot shows a positive linear relationship between SAVI and grain yield, indicating that as SAVI increases, grain yield also increases (figure 6). The regression line suggests a relatively strong correlation, with most data points closely following the trend line.

SAVI is specifically designed to minimize the influence of soil brightness, making it more suitable for areas with sparse vegetation where soil reflectance might otherwise distort measurements (Huete, 1988). The positive correlation observed between SAVI and grain yield indicates that healthier vegetation, as signaled by higher SAVI values, is associated with higher grain yields. This suggests that SAVI could be a reliable indicator for predicting crop performance, particularly in regions with varying soil backgrounds where other indices like NDVI might be less effective. SAVI's ability to account for soil reflectance

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

provides more accurate estimates of vegetation health and productivity in semi-arid and arid regions (Huete et al., 2002).

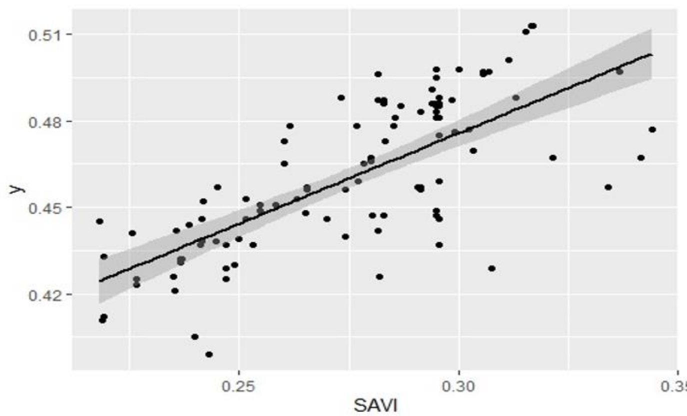

Figure 6: Correlation between SAVI and Grain Yield (Y)

3.6. Relationship between NDVI and Grain Yield:

The plot (figure 7) demonstrates a clear positive correlation between NDVI and grain yield, similar to the SAVI plot. The trend line shows a strong linear relationship, with data points scattered around the line but still maintaining a discernible pattern. NDVI is one of the most widely used vegetation indices for assessing vegetation health by measuring the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs) (Rouse et al., 1974). The strong correlation between NDVI and grain yield supports the notion that NDVI can be effectively used to monitor crop health and predict yields. Higher NDVI values correspond to healthier, more productive crops, which directly translates to increased grain yields. This index is highly valuable in precision agriculture applications where continuous monitoring of crop growth conditions is essential for timely management interventions (Tucker, 1979).

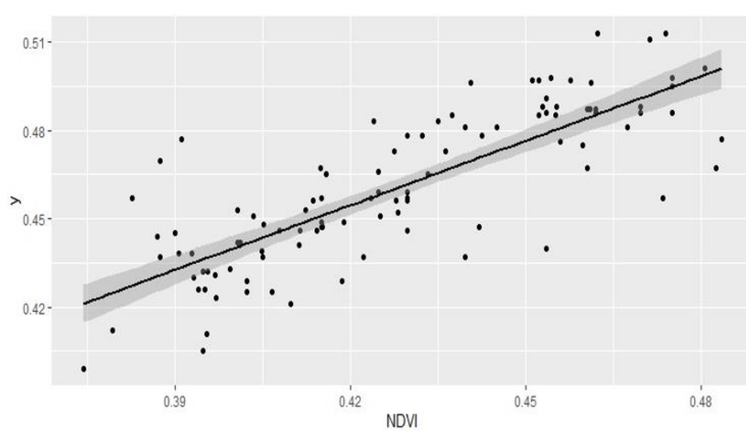

Figure 7: Correlation between NDVI and Grain Yield (Y)

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

3.7. Relationship between EVI and Grain Yield (Y):

The EVI plot also shows a positive linear relationship with grain yield. The data points are well aligned with the regression line, suggesting a strong correlation figure 8.

EVI is designed to enhance the vegetation signal with improved sensitivity in high biomass areas and reduced atmospheric influences (Huete et al., 2002). Unlike NDVI, EVI uses additional blue band reflectance to correct for soil background signals and atmospheric effects, making it particularly useful in areas with dense vegetation where NDVI tends to saturate. The strong correlation observed in the plot suggests that EVI, like NDVI, MSAVI and SAVI, is a reliable predictor of grain yield. This is particularly relevant in regions with diverse vegetation cover, where EVI can outperform NDVI by reducing saturation effects and providing a more accurate representation of vegetation health and productivity (Jiang et al., 2008).

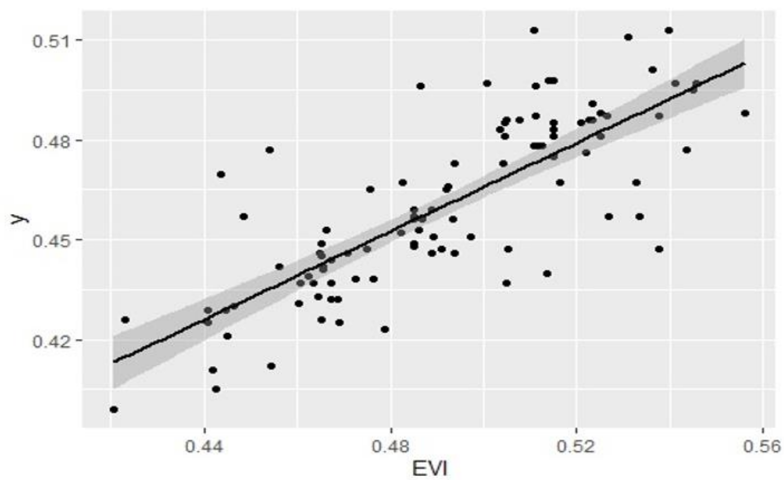

Figure 8: Correlation between EVI and Grain Yield (Y)

3.8. Relationship between MSAVI and Grain Yield (Y):

The scatter plot shows a positive linear relationship between MSAVI and grain yield, indicating that as MSAVI values increase, grain yield also increases in figure 9. The regression line suggests a strong correlation, with most data points closely following the trend line. This alignment of data points around the regression line reflects the reliability of MSAVI in capturing the variations in grain yield.

MSAVI is a modified version of the Soil-Adjusted Vegetation Index (SAVI) and is designed to further reduce the influence of soil brightness, especially in regions where vegetation cover is low or the soil surface is exposed (Qi et al., 1994). The positive correlation between MSAVI and grain yield observed in this study suggests that MSAVI is effective in distinguishing healthy vegetation from less productive areas, which translates to higher grain yields. This makes MSAVI a robust indicator for predicting crop performance,

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

particularly in heterogeneous agricultural landscapes where soil background effects are a concern. The use of MSAVI is advantageous in semi-arid and arid environments, where soil reflectance can significantly impact the accuracy of other vegetation indices like NDVI (Qi et al., 1994; Huete, 1988).

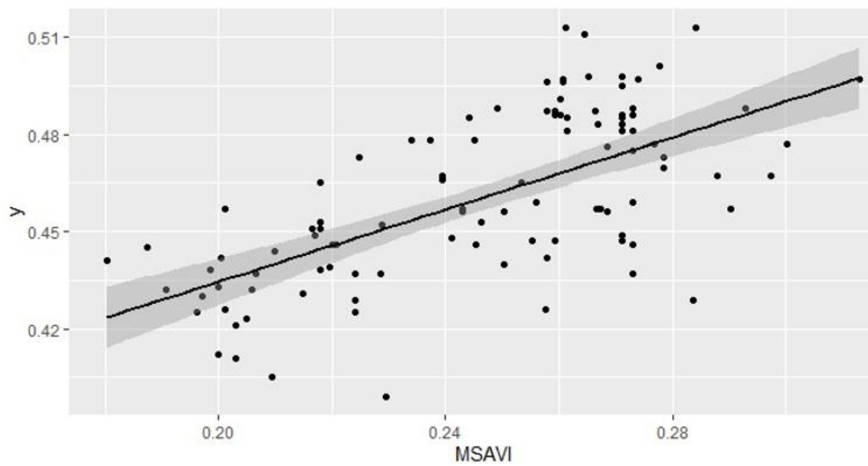

Figure 9: Correlation between MSAVI and Grain Yield (Y)

3.9. Correlation between Vegetation Indices and Grain Yield (GNY):

a) Analyzing the Scatterplots: A Visual Exploration of Different Vegetation Indices:

The figure 10 represents the relationship between the variable grain yield (GNY) and four different vegetation indices: NDVI, EVI, SAVI, and MSAVI. Each scatterplot provides visual insight into how these vegetation indices correlate with GNY values. The interpretation of these scatterplots is as follows:

b) NDVI (Normalized Difference Vegetation Index) VS. GNY:

The scatterplot shows a positive correlation, where higher GNY values are associated with higher NDVI values (figure 10). This suggests that as GNY increases, vegetation health, as indicated by NDVI, improves. NDVI is widely recognized as a measure of vegetation greenness and is sensitive to changes in vegetation cover and health (Tucker, 1979).

b) EVI (Enhanced Vegetation Index) VS. GNY:

The EVI scatterplot in (figure 10) also shows a positive correlation, but the trend might be more pronounced or distinct compared to NDVI. EVI is designed to optimize the vegetation signal with improved sensitivity in high biomass regions and is less affected by atmospheric conditions (Huete *et al.,* 2002).

c) SAVI (Soil-Adjusted Vegetation Index) VS. GNY:

The SAVI scatterplot indicates a positive correlation similar to NDVI but is adjusted to minimize soil brightness influence in (figure 10). This is particularly useful in areas where

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

soil brightness affects vegetation signals (Huete *et al.,* 2002). The trend line appears less steep than NDVI or EVI, suggesting SAVI's specific suitability in semi-arid or soil-exposed regions.

d) MSAVI (Modified Soil-Adjusted Vegetation Index) VS. GNY:

The MSAVI scatterplot shows a positive correlation (figure 10), possibly stronger than that observed for SAVI. MSAVI further minimizes soil noise, making it highly useful for vegetation monitoring in areas with significant soil background influence (Qi *et al.,* 1994).

Overall, all scatterplots show positive correlations between GNY and the vegetation indices, suggesting that GNY is positively related to vegetation health as represented by these indices.

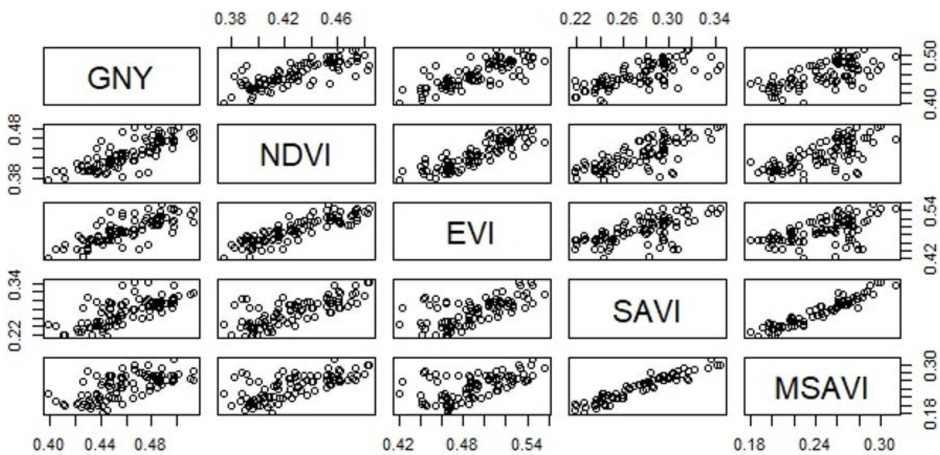

Figure 10: Scatterplot Matrix Displaying the Relationships among Different Vegetation Indices with Grain Yield (GNY)

3.10. Regression Models to Predict Biological Yield of Wheat Based on Various Vegetation Indices:

The figure 11 presents four regression models that predict wheat biological yield based on various vegetation indices: EVI, MSAVI, NDVI, and SAVI. Each model includes a regression equation and corresponding statistical metrics (R-squared, Pearson correlation coefficient, and Nash-Sutcliffe Efficiency).

Multiple Models: Different combinations of vegetation indices were used to develop the models, suggesting that multiple factors influence wheat biological yield.

Varying Performance: The statistical metrics indicate varying model performance. Model-1of Wheat Biological Yield (using all four indices) achieved the highest R-squared (0.58) and Pearson correlation (0.93), suggesting the best fit to the data. However, its NSE value (0.78) was moderate.

Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/
Journal of Xi'an Shiyou University, Natural Sciences Edition
ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

Index Importance: The coefficients in the regression equations reveal the relative importance of each vegetation index in predicting yield. For example, in Model-1of (Wheat Biological Yield), SAVI has the highest positive coefficient, indicating its strong positive relationship with yield.

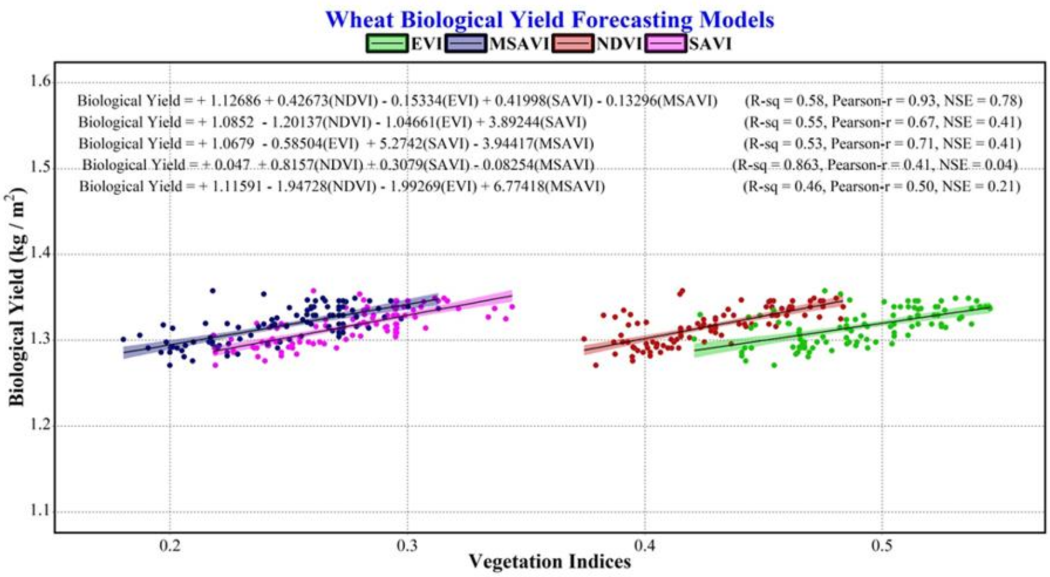

Figure 11: Wheat Biological Yield Forecasting Models

3.11. Regression Models To Predict Grain Yield of Wheat Based on Various Vegetation Indices:

The figure 12 presents four regression models that predict wheat grain yield based on various vegetation indices: EVI, MSAVI, NDVI, and SAVI. Each model includes a regression equation and corresponding statistical metrics (R-squared, Pearson correlation coefficient, and Nash-Sutcliffe Efficiency).

Multiple Models: Different combinations of vegetation indices were used to develop the models, suggesting that multiple factors influence wheat grain yield.

Varying Performance: The statistical metrics indicate varying model performance. Model-1 of (Wheat Grain Yield) using all four indices, achieved the highest R-squared (0.91) and Pearson correlation (0.95), suggesting the best fit to the data. However, its NSE value (0.89) was moderate.

Index Importance: The coefficients in the regression equations reveal the relative importance of each vegetation index in predicting yield. For example, in Model-1of (Wheat Grain Yield), NDVI has the highest positive coefficient, indicating its strong positive relationship with yield.

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

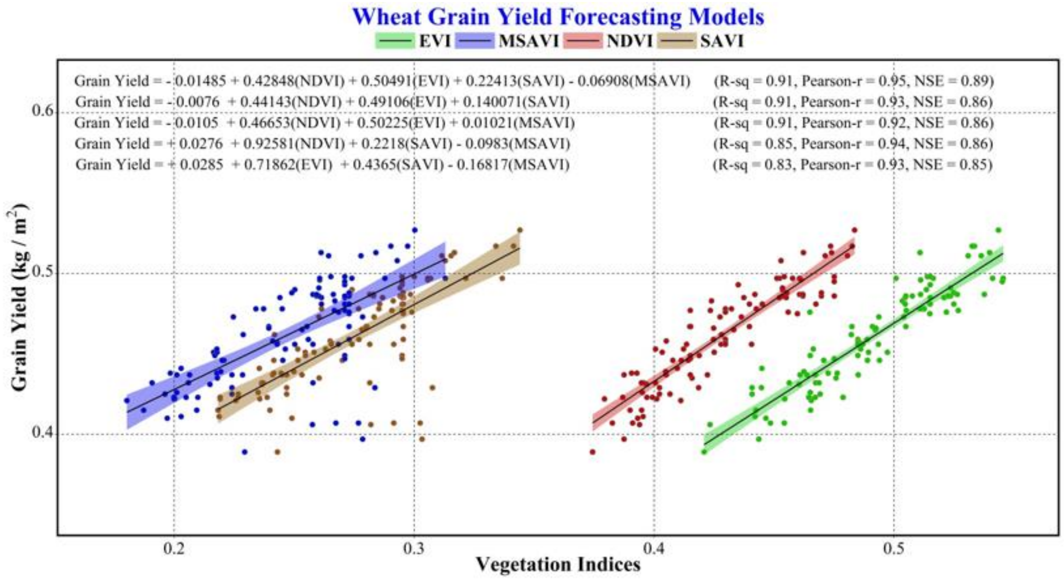

Figure 12: Wheat Grain Yield Forecasting Models

5. CONCLUSION

This research demonstrates the effectiveness of using various vegetation indices—EVI, MSAVI, NDVI, and SAVI—derived from Sentinel-2 multispectral imagery for predicting wheat biological and grain yield in the Faisalabad Division, Punjab, Pakistan. The study developed several regression models incorporating these indices and evaluated their performance using statistical metrics such as R-squared, Pearson correlation coefficient, and Nash-Sutcliffe Efficiency (NSE).

The results indicate that all vegetation indices have a positive correlation with wheat yield, but their predictive power varies. Model-1 of (Wheat Grain Yield), which integrates all four indices, exhibited the best performance with the highest R-squared (0.91) and Pearson correlation (0.95), indicating a strong relationship between the indices and yield. However, the model's NSE value (0.89) was moderate, suggesting some room for improvement in predictive reliability. Among the indices, NDVI was identified as the most significant predictor of yield, as reflected by its high positive coefficient in the regression models.

The findings highlight the potential of using remote sensing-based vegetation indices for accurate and efficient wheat yield prediction, which is crucial for precision agriculture. The study suggests that combining multiple indices can enhance model accuracy and reliability, especially in regions with diverse soil and vegetation conditions. However, the choice of the best model depends on the specific application, balancing between complexity and usability.

To further improve yield predictions, future research should consider integrating additional variables, such as soil properties, climatic factors, and management practices,

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

as well as exploring spatial variability and conducting sensitivity analyses. Such advancements could lead to more robust and operational models for supporting sustainable agricultural practices and optimizing crop management.

Overall, this research provides valuable insights into the application of advanced regression modeling and remote sensing techniques for agricultural yield prediction, offering a pathway for improving decision-making in agricultural management.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

1) Ahmad, I., Saeed, U., Fahad, M., Ullah, A., Habib ur Rahman, M., Ahmad, A. and Judge, J., 2018. Yield forecasting of spring maize using remote sensing and crop modeling in Faisalabad-Punjab Pakistan. J. Indian Soc. Remote Sens. 46: 1701-1711.

2) Ahmad, I., Singh, A., Fahad, M. and Waqas, M.M., 2020. Remote sensing-based framework to predict and assess the interannual variability of maize yields in Pakistan using Landsat imagery. Comput. Electron. Agric. 178: 105732.

3) Arshad, S., Kazmi, J.H., Javed, M.G. and Mohammed, S., 2023. Applicability of machine learning techniques in predicting wheat yield based on remote sensing and climate data in Pakistan, South Asia. Eur. J. Agron., 147: 126837.

4) Bastiaanssen, W.G.M., Ali, S., 2003. A new crop forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. Agric., Ecosyst. Environ. 94, 321–340.

5) Bastiaanssen, W.G.M., Thiruvengadachari, S., Sakthivadivel, R., Molden, D.J., 1999. Satellite remote sensing for estimating productivities of land and water. Int. J. Water Resour. Develop. 15 (1–2), 181–186.

6) De Fioravante, P., Luti, T., Cavalli, A., Giuliani, C., Dichicco, P., Marchetti, M., Chirici, G., Congedo, L. and Munafò, M., 2021. Multispectral sentinel-2 and sar sentinel-1 integration for automatic land cover classification. Land. 10(6): 611.

7) De Luca, G., MN Silva, J., Di Fazio, S. and Modica, G., 2022. Integrated use of Sentinel-1 and Sentinel-2 data and open-source machine learning algorithms for land cover mapping in a Mediterranean region. EuJRS. 55(1): 52-70.

8) Debella-Gilo, M. and Gjertsen, A.K., 2021. Mapping seasonal agricultural land use types using deep learning on Sentinel-2 image time series. Remote Sens. 13(2): 289.

9) Dempewolf, J., Adusei, B., Becker-Reshef, I., Hansen, M., Potapov, P., Khan, A., Barker, B., 2014. Wheat yield forecasting for Punjab Province from vegetation index time series and historic crop statistics. J. Rem. Sens. 6, 9653–9675.

10) Fahad, M., Wajid, S.A., Ahmad, A. and Cheema, M.J.M., 2019. Response of wheat cultivars to deficit irrigation under semiarid conditions of Faisalabad. IJAB. 21: 1004-1012.

11) Farzand, A., Akram, H.M.B., Ahmad, A., Aslam, Z., Bellitürk, K., Büyükfiliz, F. and Sözübek, B., 2023. Plant and Soil Data Management via Intelligent Agricultural Machinery and Field Robots. (Climate Change and Soil-Plant-Environment Interactions. Editors: K. Bellitürk, A. Çelik, M. Kiliç, and F. Büyükfiliz). IKSAD Publishing House. ISBN: 978-625-367-101-3, Chapter 1: 9-36, Print Date: 1 June, 2023, Ankara.

12) Funk, C., Budde, M., 2009. Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. Rem. Sens. Environ. 113, 115–125.

13) Gao, F. and Zhang, X., 2021. Mapping crop phenology in near real-time using satellite remote sensing: Challenges and opportunities. J. Remote Sens.

14) Guo, C., Zhang, L., Zhou, X., Zhu, Y., Cao, W., Qiu, X., Cheng, T. and Tian, Y., 2018. Integrating remote sensing information with crop model to monitor wheat growth and yield based on simulation zone partitioning. Precis. Agric. 19: 55-78.

15) Holtgrave, A.K., Röder, N., Ackermann, A., Erasmi, S. and Kleinschmit, B., 2020. Comparing Sentinel-1 and-2 data and indices for agricultural land use monitoring. Remote Sensing, 12(18): 2919.

16) Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens Environ 83(1–2):195–213.

17) Huete, A.R. (1988). A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25(3), 295-309.

18) Jiang, Z., Huete, A.R., Didan, K. and Miura, T., 2008. Development of a two-band enhanced vegetation index without a blue band. Remote sensing of Environment, 112(10), pp.3833-3845.

19) Kanwal, H.H., Ahmad, I., Ahmad, A. and Li, Y., 2021. Yield forecasting and assessment of interannual wheat yield variability using machine learning approach in semiarid environment. Pak. J. Agric. Sci. 58(2).

20) Mesquita, P. and Milhorance, C., 2019. Facing food security and climate change adaptation in semi-arid regions: lessons from the Brazilian Food Acquisition Program. Sustentabilidade em debate, 10: 30-42.

21) Mulla, D.J. (2013). Twenty-five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosystems Engineering, 114(4), 358-371.

22) Qader, S.H., Dash, J., Alegana, V.A., Khwarahm, N.R., Tatem, A.J. and Atkinson, P.M., 2021. The role of earth observation in achieving sustainable agricultural production in arid and semi-arid regions of the world. Remote Sens. 13(17): 3382.

23) Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). Modified soil adjusted vegetation index (MSAVI). Remote Sensing of Environment, 48(2), 119-126.

24) Reddersen, B., Fricke, T., Wachendorf, M., 2014. A multi-sensor approach for predicting biomass of extensively managed grassland. Comput. Electron. Agric. 109, 247–260.

25) Rehan, M., Wajid, S. A., Hussain, K., Saqib, Z. A., & Hoogenboom, G. (2024). Evaluating agronomic practices and radiation use efficiency in promising rice genotypes under different nitrogen levels. Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/Journal of Xi'an Shiyou University, 67(7), 124-138.

26) Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. Remote Sensing of Environment, 55(2), 95-107.

27) Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA Spec. Publ, 351(1), p.309.

28) SAEED, U., 2017. Yield Forecasting of Wheat (*Triticum aestivum* L.) for different Irrigation and Nitrogen Levels Using Simulations and Satellite Imagery (Doctoral dissertation, UNIVERSITY OF AGRICULTURE FAISALABAD).

ISSN: 1673-064X
E-Publication: Online Open Access
Vol: 67 Issue 09 | 2024
DOI: 10.5281/zenodo.13744102

29) Snigdha, G., 2022. Rice yield estimation using remote sensing and crop simulation model in Nalgonda district, Telangana (Doctoral dissertation, Professor Jayashankar Telangana State Agricultural University).

30) Subash, N., Mohan, H.S.R., Banukumar, B., 2011. Comparing water-vegetative indices for rice (*Oryza sativa* L.)–wheat (*Triticum aestivum* L.) drought assessment. Comput. Electron. Agric. 77, 175–187.

31) Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. Remote Sensing of Environment, 71(2), 158-182.

32) Tucker, C. J. (1979). "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation." Remote Sensing of Environment, 8(2), 127-150.

33) Wang, L., Liu, J., & Li, X. (2017). Estimating crop yield using remote sensing-based vegetation indices: A meta-analysis. Remote sensing of environment, 193, 113-126.

34) Xie, Y. and Huang, J., 2021. Integration of a crop growth model and deep learning methods to improve satellite-based yield estimation of winter wheat in Henan Province, China. Remote Sens. 13(21): 4372.

35) Ye, X., Sakai, K., Okamoto, H.O., Garciano, L., 2008. A ground-based hyperspectral imaging system for characterizing vegetation spectral features. Comput. Electron. Agric. 6 (3), 13–21.