JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 08 Issue 02
2025

Research Article
https://doi.org/10.56979/802/2025

Multispectral Approach for Wheat Yield Estimation Using Deep Learning

Muhammad Rashid[1*], Salman Qadri[1], Abdul Razzaq[1], Sarfraz Hashim[2], Ali Hamza[1], and Muhammad Habib-ur-Rehman[3]

[1]Institute of Computing, Muhammad Nawaz Sharif University of Agriculture Multan, Pakistan.
[2]Department of Agricultural Engineering, Muhammad Nawaz Sharif University of Agriculture Multan, Pakistan.
[3]Institute of Plant Breeding and Biotechnology (IPBB), Muhammad Nawaz Sharif University of Agriculture Multan, Pakistan.
Corresponding author: Muhammad Rashid. Email: 2022-uam-2888@mnsuam.edu.pk

Received: December 22, 2024 **Accepted:** January 26, 2025

Abstract: Automation is becoming increasingly vital across various professions and domains, including agricultural practices. Remote-sensing-based wheat yield estimation has emerged as a superior alternative to traditional yield prediction methods. Historically, wheat yield measurement involved labor-intensive and time-consuming destructive sampling techniques. However, accurate and timely yield forecasts are pivotal for decision-making processes such as crop harvesting plans, milling, marketing, and forward selling strategies, thereby enhancing the efficiency and profitability of the global wheat sector. Presently, producers or productivity officers, often funded by mills, rely on destructive or visual sampling techniques to assess wheat production during the growing season. There's a growing demand for swift and efficient problem-solving methods. Consequently, the adoption of machinery for wheat cultivation has surged, aiming to lower production costs, reduce labor demands on farmers, and enhance harvest efficiency. Although not extensively compared, existing techniques for estimating agricultural output typically employ regression models relying on specific forecasting factors. This study aims to illustrate and compare the effectiveness of utilizing satellite earth observation data for monitoring agriculture, particularly in wheat production. Multiple regression models are compared, utilizing various predictor variables. The study incorporates wheat yield estimation techniques, such as regression models, time series analysis of vegetation indices, remote sensing, phenology measurements, and normalized difference vegetation index (NDVI). Artificial intelligence algorithms, including Random Forest and ordinary least squares, are employed to develop a suggested approach that accurately correlates ground-measured data. This research introduces a novel wheat yield estimation technique, which significantly improves forecasting accuracy and holds promise for enhancing decision-making processes in wheat farming practices.

Keywords: Wheat; Yield Estimation; Multispectral Approach; Time Series Vegetation Index (VI); Normalized Difference Vegetation Index (NDVI)

1. Introduction

The economies of least developed nations (LDNs), such as Pakistan, heavily rely on agriculture, with Pakistan's economy being predominantly agriculture-based [1]. The agriculture sector contributes significantly to Pakistan's GDP, accounting for 21% and exhibiting an annual growth rate of 2.7%. A considerable portion of the labor force, approximately 44%, is engaged in agricultural activities, providing 62% of rural dwellers with their primary source of income. Agriculture stands at the intersection of our way of life and commercial innovation, playing multifaceted roles in the economic system, particularly in developing nations. Wheat, being a staple crop in Pakistan, holds immense economic importance. Three key factors underscore the value of wheat cultivation to the economy: firstly, it serves as a crucial source of food and raw materials for national industries; secondly, it facilitates cash generation from foreign markets; and thirdly, it enables the provision of products and services both domestically and internationally. Throughout history, individuals have cultivated crops on their lands to meet their dietary needs, with the surplus often benefiting animals and birds as well. Globally, agriculture occupies more than 5 billion hectares of land, representing approximately 38% of the total land area [2].

This general introduction highlights the centrality of wheat cultivation in Pakistan's agricultural landscape, emphasizing its pivotal role in the nation's economy, food security, and international trade dynamics.

1.1. Effective factors of Agriculture

Because of a variety of factors, risk control related to agriculture requires particular care. Farming is definitely an industry where the cultivation cycle are tightly linked to natural events and fully depends on the climate, which in turn impact the risk level in various ways based on the specific location (Risk factors in the agriculture sector). Crop cultivation and development are influenced by modify in human and environmental resources [3]. Numerous factors influence Wheat yield, which is the volume of agricultural, produced at a specific location. These variables can be categorize into three main groups: a technical one (agricultural procedures, managing choices), biologically (insects, pests, weeds, diseases), and climatic (climate, soil fertility, topography, water quality), etc.

1.2. Problems in Agriculture

Due to traditional farming practices and a lack of innovative technology application, the nation is having serious problems with the food supply shortfall. The average yield differs greatly from output (Arshad *et al.*, 2023). Additionally, the rapid increase in population seriously puts at risk the nation's nutrition plans. The economic system of Pakistan is based on agriculture, which is plagued by issues including; absence of water supply infrastructure, reduced agricultural yields, unlawful water allocation, lack of basic agricultural technology, insufficient land for farming, salinity and moisture of land, low rate of expansion for related goods, poor yield for each acre and many other factors.

1.3. Agriculture and Remote Sensing

Healthy production of food and resource management is two interrelated topics that are significantly impacted by agriculture and remote sensing. Remote sensing is the practice of gathering data about the surface of the Earth without touching it directly by using spacecraft, aircraft, drones, and other cutting-edge technologies [4]. Remote sensing has a significant impact on agriculture, helping to improve farming methods, maximize crop productivity, and keep an eye on the environment. Crop tracking and administration are two of remote sensing's most important uses in agriculture. Remote sensing may gather in-depth information about the health of crops, indices of vegetation, and growth patterns by using a variety of sensors, such as multidimensional or hyperspectral cameras. The best irrigation and fertilization techniques can be determined using this data, which can also be used to detect illnesses, pests, or nutrient deficits and evaluate the general health of crops. Farmers can prevent production loss and improve resource allocation via early detection of problem regions and prompt action [5].

Precision farming and land use planning are two further uses for remote sensing. It offers useful data for mapping and categorizing agricultural land, figuring out the makeup of the soil and how much moisture is in it, and analyzing topography features. With the aid of this information, farmers are better equipped to choose crops, plan plantings, and apply agricultural inputs like pesticides and herbicides. Remote sensing-enabled precision agriculture techniques enable targeted resource application, minimizing loss and adverse environmental effects while maximizing yield. Additionally, remote sensing helps in the management and monitoring of agricultural natural resources. It aids in determining the availability of water, tracking changes to water bodies, and identifying drought conditions. Farmers and water managers can create effective water management strategies and put into practice accurate irrigation techniques through the integration of remote sensing data with geographical information systems (GIS) [6]. In order to support environmentally friendly land use and conservation efforts, remote sensing also helps in the monitoring of environmental changes such as deforestation, land degradation, and others.

For monitoring agricultural crops, remote sensing has proven to be an invaluable tool. With its extensive coverage and excellent temporal precision, satellite imagery offers useful data on crop growth, vegetation indices, and other important characteristics. It makes it possible to recognize both regional and temporal differences in agricultural conditions, which helps in crop yield forecasting. Crop yield prediction using artificial intelligence approaches, such as machine learning algorithms and deep learning models, has demonstrated encouraging results. These techniques use the correlation between ground truth yield measurements and remote sensing data to train models that can precisely forecast future yields (Ashapure *et al.*, 2020). AI methods that are frequently used include decision trees, random forests, support vector machines (SVM), convolutional neural networks (CNN), and recurrent neural networks (RNN).

In conclusion, remote sensing technology and agriculture have a mutually beneficial interaction that helps to increase productivity in agriculture, efficiency of resources, and environmental sustainability. Farmers may improve their operations, make well-informed decisions, and make a contribution to a more resilient and sustainable food system by utilizing the potential of remote sensing.

1.4. Artificial Intelligence

The creation of technology for computers that can carry out tasks that traditionally require intelligence from humans is referred to as artificial intelligence (AI) [7]. It includes a wide range of methods and strategies designed to help computer programs imitate or emulate mental functions like memory, thought, solving issues, thinking, and judgement. Artificial intelligence (AI) systems are made to analyze and comprehend data, acquire knowledge through trends and situations, and then forecast the future or act on it. They are capable of quickly digest enormous amounts of data, spot trends and patterns, and come up with wise choices or reactions. Within AI, there are various important subfields, including: Robotics, Deep Learning (DL), Natural Language Processing (NLP), Computer Vision (CV), and Machine Learning (ML). Many industries and disciplines, including medical care, banking, logistics, farming, manufacturing, and more, have found uses for AI. It is used for things like identifying fraud, suggestions for driverless vehicles, and automatic upkeep.

Although AI presents many potential and breakthroughs, it also creates a number of moral and social questions, such as those related to confidentiality, prejudice, openness, and the effect on employees. The ethical and advantageous implementation of AI technology is ensured through ongoing research and development in AI, which aims to address these issues. Health is only one of the numerous industries that use artificial intelligence (AI). Following the development of conceptual methods and systems for experts in the 1970s, artificial intelligence in the 1990s, and advanced learning in the 2010s, foundational algorithms developed in the 1950s were followed. Today, AI is helping clinical practice more and more [8].

1.5. Artificial Intelligence in Agriculture

Artificial intelligence (AI) systems are made to analyze and comprehend data, acquire knowledge through trends and situations, and then forecast the future or act on it. They are capable of quickly digest enormous amounts of data, spot trends and patterns, and come up with wise choices or reactions. New robotic methods were consequently developed. The latest techniques supplied food for all on the globe while setting up jobs for billions of people. Artificial intelligence sparked a new age in agriculture. [9]. The agricultural output has been shielded by this technique from a number of circumstances, including population expansion, job issues, and food security concerns.

Although AI presents many potential and breakthroughs, it also creates a number of moral and social questions, such as those related to confidentiality, prejudice, openness, and the effect on employees. The ethical and advantageous implementation of AI technology is ensured through ongoing research and development in AI, which aims to address these issues. Health is only one of the numerous industries that use artificial intelligence (AI). In order to provide a quick summary of how robotics is now being used in agriculture, including cleaning systems using robots and drones, this study analyses the work of numerous scholars. Two automatic plucking approaches are covered together with the various soil moisture sensing methods. In this study, the use of aircraft is addressed as well as the numerous ways they might be utilized for crop tracking and irrigation [10].

Artificial intelligence (AI) has become a game-changing technology in the agricultural sector, revolutionizing a number of farming practices, cultivation techniques, and harvesting methods. Crop Tracking and Management, Precision Agriculture, Crop Forecast and Optimization, Herb and Pest Administration, Farm Automation, Livestock Tracking and Management, Field Management Systems, and Logistics Optimization are a few of the significant uses of AI in farming. AI is making agriculture into a more effective, profitable, and sustainable industry by utilizing information-driven decisions and technology [11].

1.6. Wheat Field

Wheat, a staple crop with substantial significance in international agricultural trade, plays a pivotal role in providing food and feed for global populations. Originating from regions spanning mild temperate to tropical climates, wheat is a vital cereal grain belonging to the Triticum genus of the grass family (Poaceae) [12]. It typically ranges in height from 0.5 to 1.5 meters, characterized by slender, hollow stems bearing grains rich in carbohydrates. Unlike sugarcane, wheat does not cross-pollinate, with cultivars typically exhibiting minimal hybridization. Cultivation of wheat encompasses countries situated across various latitudinal zones, with climates ranging from temperate to subtropical. Wheat occupies approximately 220 million hectares of land globally, contributing to a total production exceeding 700 million tons annually. However, the cultivation area and productivity of wheat exhibit significant variations among nations due to diverse agricultural practices and environmental conditions. Wheat serves as a crucial source of sustenance, fueling various food industries and providing raw materials for essential products.

1.7. Production of Wheat in Pakistan

Wheat, deeply intertwined with Pakistan's agricultural heritage, traces its cultivation roots back to ancient times, fostered by the fertile lands of the great river Indus and its tributaries. The historical region of the Indus Valley Civilization harbored knowledge of wheat production, a tradition that persists to this day, shaping the cultural landscape with staples like "Roti" and "Naan." Wheat cultivation flourishes in regions between latitudes 24 and 34 degrees north, typified by irrigated subtropical zones boasting mild temperatures conducive to its growth.

Out of Pakistan's available 22.0 million hectares for cultivation, wheat claims nearly 16.0 million hectares, constituting a significant portion of the irrigated land at approximately 72.7%. The nation's water resources, including reservoirs, amount to around 135 MAF (million acre-feet), insufficient to meet the crop's estimated requirement of about 10 MAF. Despite these challenges, Pakistan's wheat industry remains robust, operating at about 70% capacity. Annual wheat production in Pakistan hovers between 25 to 30 million tons, a figure influenced by factors such as irrigation availability and rainfall patterns. However, growth in wheat production has been relatively stagnant over the past several decades. The anticipated rebound in wheat production for the 2023–2024 season is forecasted to reach 30 million metric tons, reflecting a 10% increase from the previous year's estimates. Adverse weather conditions, such as floods, can significantly impact harvested areas and yields, highlighting the need for continuous monitoring and forecasting tools to aid decision-makers [13].

Government support, including increased assistance prices for wheat growers, serves as an incentive for farmers to prioritize wheat cultivation over alternative crops. Wheat cultivation spans across three provinces in Pakistan, with Punjab contributing the majority of production at 68%, followed by Sindh at 24%, and Khyber Pakhtunkhwa (KPK) at 8%. Notable wheat-growing regions include the Punjabi division of Bahawalpur and the Sindhi division of Sukkur.

1.8. Ecological Aspects of Wheat Farming:

1.8.1. Productivity of Photosynthesis

Wheat, a vital cereal crop, exhibits moderate efficiency in photosynthesis, converting solar energy and atmospheric carbon dioxide into biomass. While not as efficient as sugarcane, wheat's C3 photosynthetic pathway enables it to utilize approximately 1% of incident sunlight energy for biomass production. With a photosynthesis rate typically ranging between 6–8 μ mol CO2/m2/sec, wheat remains a significant contributor to food and agricultural systems, producing grains essential for human consumption and livestock feed.

Figure 1. Graph showing production of wheat in Pakistan from 1960-2023

1.8.2. Climate

Wheat cultivation thrives in temperate to subtropical climates, requiring a minimum of 400 mm of annual rainfall for optimal growth. In Pakistan, wheat cultivation extends across various ecological regions, including the northwestern, central, and southern parts of the country. Lower Sindh, characterized by its hot and humid climate, offers particularly favorable conditions for wheat farming. Pakistan's climate ranges from temperate arid to semiarid, with average temperatures fluctuating between 4°C in December and January and 38°C in June and July. While the overall climate is conducive to wheat productivity throughout the year, adverse weather conditions, such as droughts or extreme temperatures, pose significant challenges to crop yield. Winter temperatures, especially frost events, can halt or delay wheat development in certain regions. Adequate rainfall is essential for optimal wheat growth, with water scarcity posing a significant constraint to productivity in Pakistan's wheat farming regions.

In conclusion, wheat farming in Pakistan is influenced by ecological factors such as climate and photosynthetic productivity, which impact crop growth and yield. Understanding and addressing these ecological aspects are crucial for optimizing wheat production and ensuring food security in the region.

1.9. Effects of biotic and abiotic variables on the growth of Wheat

Millions of families rely on agriculture for their livelihoods because it expected to be one of the profitable ventures. Agriculture includes both major and minor crops like cotton, sugarcane, rice, maize, and wheat in Pakistan. Contrarily, fruit and vegetables are growing on a commercial basis to generate income and satisfy local dietary needs. Wheat is the sole major product in Pakistan. It clearly plays a significant role in improving the farmers' financial standing as well as supplying the raw materials for the production of sugar and sugar-related goods in the sugar-processing sector. Wheat output in Pakistan is low for several reasons, including poor soil fertility, low seed rates, poor quality of seeds, traditional seeding techniques, and subpar management of farms. Many factors were thought to be contributing to reducing sugarcane output, including the impacts of water, air, and soil pollutants, fertilizer, chemicals, burning during harvest, soil erosion, compactions, competition with food crops, loss of habitat, and influence on biodiversity [14].

Attacks by insects and other pests are one of many factors that contribute significantly to restricting sugarcane productivity. For instance, sugarcane borers decreased the sugarcane's quality, growth, and sucrose level while increasing its fiber content. Trunk bores damaged the sugarcane while feeding on its internal tissues and decreased its production. Islam has reported a significant decline in sugarcane yield in Pakistan due to an invasion of insects and pests [15]. In Pakistan, insect attacks are reducing cane production, but it is thought that Pyrilla, top borers, and Gurdpur borer are the primary culprits for production declines of 15-20%, 10-20%, and 30-35%, respectively. In a few instances, yields of crops have been reported to decrease by as much as 80-85 percent because of insect attack [16].

Pesticides are crucial to the IPM programmer. When necessary, insecticides are helpful. A combination of cultural practices, resistant species, and natural enemy preservation are used. The great ideal ways that combine multiple potential measures of control are those that fall below the IMP threshold level. Early infestation has a significant impact on production throughout a significant time of cane growth, but late infestation, starting in September, has a significant impact on the amount of sucrose in the soil of the cane field. The scientist's focus was diverted by the adverse consequences of the insecticides. The scientific community is working to contain this dangerous scenario. IPM is use as the best strategy for controlling ecological disruption. IPM approaches combine the use of chemical, biological, cultural, mechanical, pheromone sex, and light traps. In Pakistan, insect and pest infestations were reportedly a key factor in the decline in wheat yield.

2. Literature Review

The wheat industry depends heavily on the crop wheat; hence it is essential to anticipate wheat production accurately in order to maximize crop management techniques, ensure effective resource allocation, and make informed decisions about the wheat sector. A potent tool for estimating agricultural yields has evolved that combines remote sensing with artificial intelligence (AI) methods. This study of the literature seeks to investigate and compile the current knowledge on AI-based remote sensing-based wheat yield prediction models. A number of studies have concentrated on creating remote sensing and AI-based models for predicting wheat yield. These models extract pertinent elements for yield prediction using a variety of remote sensing data, such as multispectral or hyperspectral images. Numerous vegetation indicators, including the Green Chlorophyll Index (GCI), Enhanced Vegetation Index (EVI), and Normalized Difference Vegetation Index (NDVI), have been used to gauge the health of the vegetation and forecast wheat yields [17].

This crop's growing significance is due to its use as a source of biofuels (ethanol) supply as well as food and feed. The world's largest producer is Brazil. According to the IGBE (International Gaming Business Exposition), Brazil produces 672 million tonnes annually, or close to one-third of the entire amount produced globally. In such a situation, it is obvious that managers and decision-makers would benefit from instruments that could track the Brazilian wheat's vegetative vigor continually and deliver immediate information about any potential short-term effects of weather conditions on yield forecasts. More specifically, sensors are useful tools for determining the status of the plants and the buildup of green biomass. It is common practice to use satellite Earth observation equipment that provide regular (daily or nearly-daily) information at a coarse spatial resolution for local monitoring of crops, yield calculation, and prediction. Historically, uses for wheat satellite imagery monitoring have included sorting various wheat varietals using spectrum indicators.

A crucial component of agricultural remote sensing (RS) is the timely and non-destructive evaluation of crop output [18]. A new method for RS has been made available by the development of unmanned aerial vehicles (UAVs), which enable the acquisition of high spatial-temporal precision photos on the scale of the region. One of the most crucial factors affecting national food security and individual standard of living is grain yield from crops. One of the most crucial issues pertaining to the wellbeing of the nation and the life of the populace is the cultivation of grains. For the creation of a national food policy, price regulation, and international grain trading, harmless, actual time, and precise crop production forecast over a vast area is essential.

The creation of a time series of high spatial resolution multispectral imagery has been the subject of different researches that combine a high spatial resolution panchromatic image and a coarse spatial resolution multispectral image from one or more devices. When a landscape becomes disperse and most of the pixels are made up of a variety of surfaces, the potential of multitemporal coarse spatial resolution remotely sensed pictures to track vegetation is decreased.

The spectral unmixing model was used to combine TM data with multi-temporal MODIS-NDVI data in order to get time series NDVI data at an accurate Thematic Mapper (TM) scale [19]. Utilizing data on within-pixel fractional cover obtained from a high resolution land-use map, sub-pixel NDVIs for the various land-cover classes are computed by solving a weighted linear system of equations for each pixel of a coarse resolution image. The amount of weight allocated to the various image pixels are calculated by taking into account both the spectral dissimilarity estimated on medium-resolution remote-sensing images acquired at various times of the year and the distance in space between each pixel and the target for the calculation of sub-pixel NDVIs of a target pixel.

As yield is always, the most crucial attribute among that desired, reliable yield assessment for lines of breeding is the key to the introduction of innovative crops in cultivation schemes. Two key steps are involved in typical breeding procedures: first, breeding populations with desired variants are deriving through combination and/or mutagenesis, and second, desired variants are selected through a series of field tests combined with necessary laboratory analyses. The accurate yield estimate for the lines of breeding or variations during the selection stage mostly rely on accurate field tests, which are frequently impacted by complex environmental factors (primarily soil homogeneity) acting on the field plots, particularly as the tested quantity of lines grew. To do that, it is necessary to develop new methods that will increase the accuracy of yield tests and improve the efficacy of output selection. The estimation of crop attributes, like height of plants, the amount of chlorophyll, index of leaf area, susceptibility to disease, moisture sensibility, nitrogen level, yield, and others, has made extensive use of remote sensing imagery [20].

Measuring the amount of material that enters through a location is a straightforward method for calculating the Biomass flow rate. With the quantitative method, feed sliders that run in pairs, with the top roller placed on an angle and the bottom roller fixed, move the wheat stems through the reaper. With an increase in the volume of material handled, the space between the two rollers will grow. Using a Duncan 9800 throttle position sensor also used, among other places, as wheel angle sensors on tractor auto-steer systems the open of the feed rolls, or roller opening up was calculated. By putting the detector on the pivoting point of the final feed roller, the sensor was able to measure an angle of 85 degrees [21].

Making all the images comparable or nearly similar so that they may be regarded as having been acquired in similar atmosphere and by the exact same sensors is one of the crucial steps in the picture pre-processing procedure. The image processing, analysis, and spatial information integration were carried out using ENVI 5.1, a digital image processing program, and ArcGIS 10.2. To match the provided land cover boundary pictures, all of the Landsat images in this study were geometrically matched to the WGS 1984 UTM zone 56N UTM projection system. To eliminate reflectance fluctuations across image dates due to air conditions and surface anisotropy, radiometric normalization was applied for the acquired images. Simple dark-object subtraction (DOS) approach was used for the bulk corrections of atmospheric effects. Notably, when processing the photos, no consideration was given to small fog, haze, shadow, or incorrect data pixels.

This study demonstrated how historical production patterns of wheat crops in the Bundaberg growing region may be plotted using time series Landsat images. In the simplest terms, these historical trends serve as a guide for the upcoming years' crop yield in terms of canopy GNDVI value. Thus, any departure from this trend can signal the beginning of a pervasive biotic or abiotic restriction. When it comes to predicting yields, the highest crop vigor GNDVI value was traditionally attained 145 days after planting, or in early April. It is recommended that this time frame be chosen as the ideal growth stage for the collecting of imagery from satellites for regional yield predictions. The ability to determine the maximum GNDVI from footage taken between the months of November and March has been made possible by the construction of a quadratic model that appropriately portrays the growth profile of sugarcane. In addition, does this allow forecasts to be made sooner in a growth season, but it also allows for the calculation of the highest GNDVI value for that season in the case that persistent cloud cover precludes the taking of a fresh image during the crucial April time [22].

A unique multi-year crop, sugarcane can be collected annually for up to 6-7 years before it needs to be replanted. The ratoons—the roots and lower parts of the plant—grow new stems that are harvested the next year after an annual harvest. For nine months, the majority of farms continuously harvest sugarcane. This is important because it makes ongoing crushing possible. As a result, monitoring the gathering and crushing happens almost daily. Crop yields must be routinely estimated due to the massive production flow. Storage needs, cash flow forecasts, fertilizer and water use, and crop yield forecasting are all extremely practical for farmers to make sensible judgements regarding crop insurance. Research on sugarcane breeding focuses on choosing genotypes that are appropriate for particular settings based on yield measurement [23].

Labelling of seeds, including those from wheat, maize, rice, rye, and other crops, is typically done to improve the preparation of food, preservation, advertising, health of crops, and resistance to disease. Each crop's seeds usually have a variety of anomalies, including rocks, invasive plants, detritus, broken kernels of corn and inert materials. These outliers must be separated in order to ensure quality. Quantifying these contaminants in grains is also crucial for grain grading. Traditionally, specialists use visual inspection to grade and identify different types of seeds. Human senses are susceptible to work load and fatigue when they are subjected to various characteristics, such as domain expertise, size, form, color, structures, etc. As a result, the method for identifying grains is variable, unreliable, and unbelievable. In this case, the research was carried out to create a reliable and effective system that is without all the restrictions mentioned previous [24].

For the reasons of crop cultivation and establishing policies, forecasting yields of crops and their geographical variation under changing circumstances is a difficult but crucial task. For participants that include private farmers to state analysts, the accessibility of knowledge on risks linked with the impact of climate change on farming results is crucial. This study was carried out as a test run in order to: (1) create remote sensing satellite based predicts of maize land in the usual maize developing region of Pakistan; (2) create a statistical-empirical approach to forecasting maize yields; and (3) evaluate the impact on temperature on seasonal variation in maize yields over a ten-year period. Using Landsat 8 visuals, an array of eight artificial intelligence methods were tested for locating maize-growing enterprises in Pakistan's Faisalabad district. 200 randomly chosen ground-verified locations were used to assess 200 models of classification within the study area. Utilizing Landsat-derived multi-temporal normalized difference vegetation index (NDVI) and LST (land surface temperature) data as indicators, the results of the maize mapping effort were utilized to estimate internal maize yields. A minimum relative shrinkage and selection regression model was used to determine the key screening-selected predictors for the yield prediction model [25].

One of among the most crucial aspects of agriculture, along with grain-based diets, is the cultivation of vegetables. About 33.9 million tonnes of commercial vegetables were produced in the US in 2018, worth $12.9 billion. Particularly, tomatoes have the most values of production that are actually used, and their market value rose by $1.9 billion (10% more) in 2018. As result of both abiotic and biotic variables like climate change, disease, and pests, which can significantly reduce productivity and quality of fruit, the cultivation of tomatoes has recently been under persistent strain. Modern characterization that can efficiently the map, observe, and forecast crop features is needed to identify the prospective yield potential in a tomato. Considering the significance of vegetable production, hand-sampling evaluations, which are damaging, time-consuming, costly, and used in conventional breeding efforts to create new varieties, track crop development and illness, and forecast yield, are still used today [25].

A great deal of data is necessary for precision farming in order to make sure that decisions about specific crops and plots are well-informed. Because inaccessible assessments using sensors have become harmless and more efficient in recent years, remote sensing data has been utilized to gather data in a rapid or close to real-time way for uses in agriculture. However, darkness, expenses, low spatial accuracy, and restricted time scale make it difficult for satellite-based and aerial imagery collection to get the data necessary for plant, plot-level assessment. Crop length is a crucial factor in determining watering needs, estimating final production, and evaluating overall crop health. In this work, a unique approach to tracking sugarcane crop growth with an unmanned aerial system (UAS) is put forth. Over the course of the growing period, UAS data were collected a total of seven times, every aerial collection consisting of over 200 photographs at a fifty-meter above sea level height that had substantial image overlapping. Using the Structures from Motion method to the photos produced an Orth-mosaic image and a 3D object cloud.

Effective agriculture is becoming more and more popular all around the world [26]. Crop and herb recognition, biomass assessment, and forecasting yields are the key goals. The accessibility of yield tracing devices, which are currently not very popular across farmers, is required for the evaluation of artificial intelligence approaches for remotely sensing-based yield forecasting. Convolutional neural network models

(CNNs), a deep learning technique that excels in recognizing images problems, are used in this study to develop a model for crop production prediction using NDVI and RGB data collected by UAVs. It is investigated how several CNN characteristics, such as the choice of the training process algorithm, the complexity of the network, the regularization technique, and the adjustment of the hyper parameters, all affect the accuracy of the predictions. For data collected early in the growing season with RGB colors data, a mean absolute error (MAE) in yield forecasting of 484.3 kg/ha and an average absolute percentage error of 8.8% were achieved using the AdaDelta training technique, regularization with early ceasing, and a CNN with 6 layers of convolution [26].

The ability of UAS-based remotely sensed information to assess crop features including canopy coverage, height of plants, and indices of vegetation over a wider area more often and reliably than measured by hand has been demonstrated in a number of studies. Numerous approaches to the interpretation of satellite- and airborne-derived imagery are still being put out and evaluated. In this study, author examine the use of support vector machines (SVMs), an exciting machine learning technique, as they are used for satellite detection. The rapidly rising volume of publications released in the past decade makes this assessment timely. Because of their capacity to generalize effectively even with limited training examples, a common restriction for satellite imaging software, SVMs are especially promising in the field of remote sensing [27].

The affordability of water, land, and even, lately, a global epidemic, are all putting increasing demands on current food and farming systems. We have made tremendous strides in genetic instruments over the past few years, but up until recently, we were sorely lacking in our capacity to accurately evaluate the state of crops at scale in the field using these tools.

Recent developments in satellite imagery and Artificial Intelligence (AI) have allowed us to accurately measure field scale genetic knowledge and incorporate big data into conservative and predictive systems for management. In this review examines the use of the latest developments in remote sensing and AI to enhance the durability of farming systems.

In this study, multi-temporal sensing images from an unmanned aircraft system (UAS) were used to construct a framework for machine learning for cotton production estimation. The suggested machine learning framework was based on an ANN (artificial neural network) and used three types of crop characteristics for predicting yield. Multi-temporal characteristics such as canopy coverage, canopy width, canopy quantity, normalized difference vegetation index (NDVI), excessive greenness index (ExG); non-temporal characteristics such as cotton boll measure, boll dimension, and boll quantity; and watering status as a subjective characteristic [26].

All wheat-breeding programs have the development of disease-resistant cultivars as one of their main objectives. Remote sensor-equipped Unmanned Aerial Vehicles (UAVs) can offer measurements of spectral characteristics that can used to gauge the severity of foliar disease. It has always been difficult to measure disease severity characteristics during genetic selection since it takes a lot of time, money, and labor to produce many reproductive lines. This study explores the possible application of low-cost UAVs with cameras that are digital as a field tool for assessing the seriousness of foliar disease in a wheat breeding program [28].

The development of a device built around a self-driving UAV was done to provide better distant sensing. The device was built on a lightweight, portable helicopter platform that weighed less than 14 kg. The UAV technology was able to acquire multi-spectral photographs at the specified locations and times thanks to its multi-spectral camera and independent system. Using combination of sensors methods, a UAV navigational system based on the Extended Kalman Filter (EKF) was created. The connection among a human controller and the UAV for operation organizing, activating flight commands, and keeping track of the flight in real-time called a base station. The UAV autonomously guided to the necessary checkpoints and hover over each waypoint to gather field picture data according to its navigational information and the points of interest provided by the base station.

In ecosystems on earth, plant dynamics and phenotype are crucial for predicting climate-vegetation conversations, changes in land use and land cover, and variability in from year to year plant production. Over the past few decades, variety of measurements and methods have been utilized extensively to track vegetation phenology across vast geographical domains utilizing satellite remote sensing data.

The study of time-series, multispectral in nature remote sensing imagery and emerging innovative methods for phenotype identification are cover in detail in this work. The goal and uses of identifying broad vegetation phenology stages often referred to as "land surface phenology" as well as more sophisticated techniques that predict species-specific phonological stages are outlined in this work [29]. The accuracy of yield prediction using satellite data to calculate rice yields [30], wheat [31], corn, soybean (Pandya *et al.*, 2023) and other crops. Classification is a technique that has been shown to evaluate remote sensing data and can classify the individual pixels in an image's spectral characteristics [32]. The yield of crops forecasts made early in the growing period are crucial for boosting output, controlling revenue losses from low yields, and protecting earth's resources. For enhanced oversight of agricultural inputs and the financial activities of the farm firm, short-range crop production estimates are crucial. These short-term cotton forecasts may also help with determining profits, gin workload, and value on the market.

Additionally, some cotton-producing nations, like the nation of Turkey, where water supply and handling are issues, may find benefit in short-range forecasting methods. In order to offer months-ahead estimates of cotton output in data-poor regions, this research uses an ANN model using simple or easily accessible climate and remote sensing data for the Moneymen region of Turkey. Cotton is a plant that requires a certain kind of climate to grow. Long duration without frost, a lot of heat, and lots of sunshine, and enjoys a hot, muggy climate. Weather and precipitation have an impact on other climatic factors that affect the growth and output of cotton, such as humidity levels and the temperature of the soil. Additionally, it was demonstrated that cotton production and quality are affected by the timing of planting, which is governed by both soil temperature and moisture, with late planting resulting in a lower yield and quality. ANNs need a variety of particular input parameters to train their models, and these parameters are typically only accessible at data point locations [33].

Two approaches for calculating the yield of various crops in Hungary using satellite remote sensing data. In order to adjust for geometry; radiometric, the atmosphere, and cloud dispersion, the procedures of reprocessing the remote sensing data are outlined. Reference crop fields were chosen in the first technique for field level estimate by using Landsat Thematic Mapper (TM) data for segmentation. The model's considerable distortion revealed information about the field's potential stress. Only publicly available county-level production statistics and NOAA AVHRR are used in the second method that is being discussed. For a period of eight years, the county-level yield statistics and the GYURRI derived index of vegetation were examined for eight various crops [34].

In this study, we focused on the latter prediction method, which used after the season has already ended. A reliable post-season forecast of yield is essential for a number of reasons, including:

a) Verifying the reported output and serving as a monitoring mechanism;
b) Identifying less productive farms and packages and using this information as a foundation to execute optimization strategies for the upcoming season;
c) Facilitating the preparation of collecting transport, such as estimating the necessary amount of vehicles for transport, accessible sugar mills, or places for storage
d) To evaluate the accuracy of farmer and farm systems for managing pre-harvest predictions
e) To adjust the cost of the commodity to reflect supply and collected quantities and,
f) If wheat planting intended to produce biofuel, energy availability and usage should be optimized.

Empirical models of regression based on vegetation index (VI) time series, such as (NDVI) and other indices composed of various spectrum band percentages have been applied to the particular instance of Wheat crop production estimation, Other methods for predicting sugarcane yields depend on three-dimensional in nature crop area modelling created by photogrammetric or LiDAR modelling using unmanned aerial vehicles (UAV) or airborne remote sensing [35]. Additionally, ecological systems make use of multimodal explanations from remote sensing data and other sources of agricultural data, such as soil and weather information factors for efficient light use, as well as phonological indicators.

Other methods, including systems with a Markovian and Probabilistic features, define the wheat cultivation area through different crop area categorization and predict the prospective yield based on the calculated area and last season yield data, without the need for actual data. In conclusion, empirical approaches produce straightforward solutions for specific crops when spectrum indices of vegetation and yields of crops exhibit acceptable consensus. However, scientific models necessitate reaffirmed local measurements to take into account harmful crop yield measures that limit their applicability.

Additionally, physiological models that are applicable to a variety of physiographic circumstances can be used. They are, dependent on expectations regarding harvest coefficients and photon usage efficiency, which need for regional and crop-specific validation libraries that are generally valid for 5 to 10 years. The most accurate of these various methods for estimating wheat crop yield are biological crop growth designs, but they also call for many different agronomical variables that are typically not available in practical use because of the scarcity of field campaigns and the difficulty of accessing the terrain [36].

In contrast to biological crop growth designs, EO-based regression equations are simple to deploy, just need remote sensing findings, and are thus appropriate for use inside a working agricultural tracking platform [36].

Many industries employ machine learning (ML) techniques, from supermarkets that analyze consumer behavior to businesses that forecast how customers will use their phones [37], this challenge necessitates the use of many datasets [38]. This shows that predicting agricultural yields involves a number of challenging processes and is not a simple operation.

Currently, crop yield prediction models can roughly forecast the actual production, but a higher yield accuracy for forecasting is still desired. Machine learning, a subset of Artificial Intelligence (AI) that focuses on learning, is a useful method that can estimate yields more accurately utilizing a variety of features. Machine learning (ML) may extract information from datasets by finding correlations, trends, and patterns. The models must be trained using datasets that depict the outcomes based on prior knowledge. Multiple features are used to build the prediction model, and as a result, the specifications of the predictions are established using previous information during the training stage. A portion of the previous data from the training stage is used for assessment of performance during the test stage [39].

Artificial intelligence-powered models can support sustainable agricultural practices by precisely Wheat yield harvests. They can aid in maximizing the use of resources like pesticides, fertilizers, and water, minimizing negative environmental effects. In summary, by providing fast, precise, and usable information, an AI-based remote sensing-based wheat yield prediction model has the potential to revolutionize wheat farming practices. It can increase productivity, reduce hazards, and support efficient and environmentally friendly crop management techniques. Overall plan that followed is shown in Figure 1.

Problem Statement: In the land of agriculture, wheat stands as a fundamental crop, providing food for a significant portion of the global population. However, the challenge of wheat yields estimation accurately and in a timely manner persists as a pressing concern. Conventional wheat yield estimation methods, which rely on outdated methods and historical data, frequently are unable to provide the accuracy and timeliness needed to handle the complexity of contemporary agriculture. The issue at hand has several facets. The challenges that regional farmers, agriculture consumers, and policymakers face are clear; without accurate and trustworthy yield predictions, they are forced to make crucial judgments about crop management, resource allocation, and market forecasts.

A reliable postseason production estimate is essential for a number of reasons, including the verification of the declared production, locating low-productivity lands and farms as a starting point for developing improved techniques for the upcoming season, the estimation of the necessary number of transport vehicles, therefore, to minimize expenses related to the commodity's storage and transportation, the verification of the reliability of the post-season yield estimate, and many others [40]. The estimation of sugarcane crop output in this specific case has been done using empirical regression models based on the vegetation index (VI) time series, including the Normalized Difference Vegetation Index (NDVI) and other indices made up of different spectral band ratios [41].

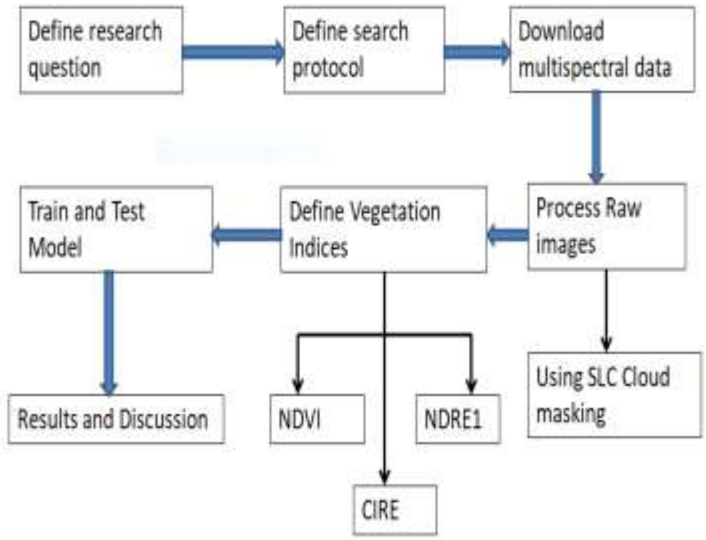


Figure 2. Review Protocol

The objectives of this research are:

- To acquire most prominent features using Multispectral dataset.
- To develop a preprocessing model for eliminating outlier in the Multispectral dataset.
- To develop a deep learning model for wheat yield estimation.

We choose various methods for generating predictor variables and assess them based on how well they estimate yield. The first technique uses NDVI and time series data from the other vegetation indices (VI). Many other studies have used as predictor factors, VI time series [41]. The second technique involves calculating metrics for phenology [42] [43]. Different markers, like the beginning and conclusion of the growing season, are among these metrics and are using to describe seasonally varying crop ontogenesis and crop growth.

Approaches for creation of predictor variables

There are various methods we can take into consideration while developing predictor variables for yield of Wheat Yield utilizing remote sensing data. Here are a few popular methods:

- **Vegetation Indices:** Calculate popular vegetation indices including the Enhanced Vegetation Index (EVI), Normalized Difference Red-Edge Index (NDRE), and Normalized Difference Vegetation Index (NDVI). These measurements of vegetation's greenness, vigor, and health serve as key markers of agricultural growth and productivity.
- **Spectral Bands:** Spectral bands from the remote sensing data that are known to be correlated with sugarcane yield should be used. For instance, the near-infrared (NIR) and red-edge bands are frequently useful for determining the biomass and chlorophyll content of vegetation (Zhao *et al.*, 2020b).
- **Radiometric Transformations:** Enhance certain characteristics or relationships with wheat yield by applying radiometric adjustments to the spectral bands, such as logarithmic or exponential transformations. Specific traits of the crop may be highlighted with the use of these modifications.
- **Texture Measures:** Calculate texture metrics to identify spatial patterns and heterogeneity within the wheat fields, such as standard deviation, entropy, or texture correlation. Texture measurements can reveal important details about the structure and variability of the crop, which can have an impact on production.
- **Temporal Features:** By examining numerous time points or seasonal fluctuations in the remote sensing data, take into account temporal aspects. To capture the temporal dynamics and growth patterns of Wheat across time, compute statistical metrics like mean, standard deviation, or rate of change.
- **Spatial Context:** Include variables that describe the relationship between or proximity to pertinent features in the spatial context. For instance, the distance to different land cover classes, roads, or water bodies that could affect Wheat development.

- **Topographic Factors:** Think about topographic elements like height, slope, or aspect because they might alter water availability and microclimates, which in turn affect Wheat productivity.
- **Weather Variables:** Include weather factors that have an impact on crop growth and development, such as temperature, precipitation, or sun radiation. Obtain past weather information for the study area and combine it with the data from remote sensing**.**
- **Data Fusion:** Investigate the integration of several data sources, such as geospatial datasets, field survey data, or remote sensing data, to produce more thorough predictor variables. A more comprehensive collection of information for yield prediction can be provided via data fusion techniques.

The most useful variables for the prediction model can be found using feature selection approaches like correlation analysis or feature importance measures from machine learning models. Illustration of the wheat yield predictor's method shown in Figure 2

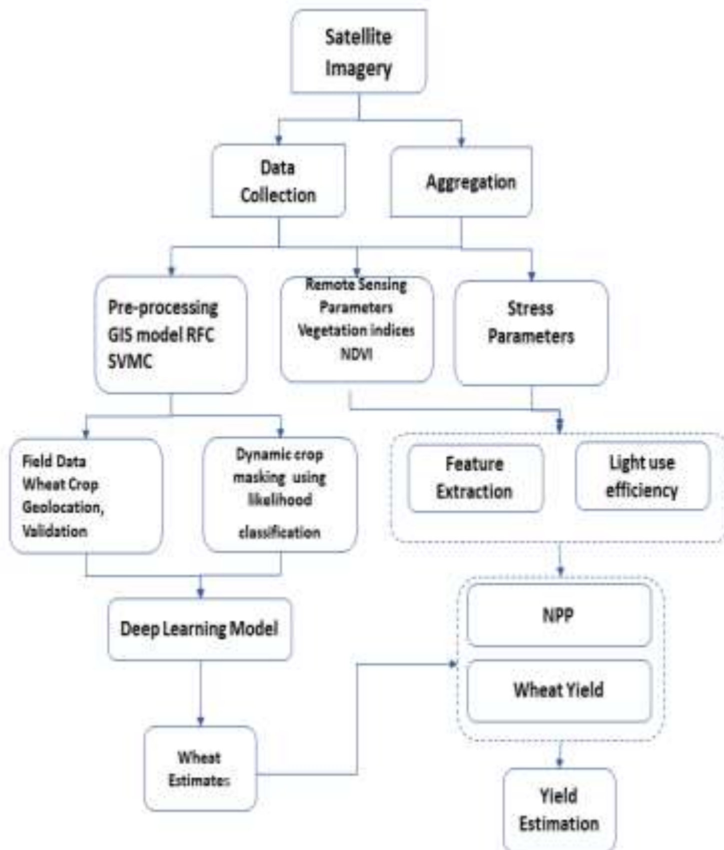


Figure 3. Illustration of the sugarcane yield predictor's method

3. Material and Methods

The district of Punjab Province chosen for this study is Dera Ghazi Khan. It contributes close to 30% of Punjab's total wheat yield. In southern Punjab, where the majority of the population works in agriculture, it is regarded as an agricultural district. It is a fertile region that generates a variety of crops, including wheat, cotton, maize, Sugarcane, and mangoes.

It is one of the significant crops that contribute significantly to agriculture in this area. In Figure 2.4 map of DG Khan District has been shows. The area under cultivation of wheat increased to 430,000 acres in 2020 from 310,000 acres in 2014-15. Six flour mills located in the district.

3.1. Data sources

It takes many resources to gather accurate in-situ crop data (yield, biomass, and other biophysical variables), but it is essential for trustworthy crop modelling. Figure 2.2 shows the research area's average annual sugarcane production over the previous five years. Nearly 36% of the respondents reported average sugarcane production over the previous five years of up to 600 pounds per acre, it was discovered. It shows that 601 to 800 mounds of sugarcane per acre on average were produced by nearly 45% of the farmers over the previous five years.

Year	Production (000 t)			Yield (kg/ha)		
	Forecast	95 % Limit		Forecast	95 % Limit	
		Lower	Upper		Lower	Upper
2013	55859	50103	61615	54689	49575	59803
2014	58287	52531	64043	54874	49760	59988
2015	57927	52171	63683	55103	49990	60217
2016	58302	52546	64058	55370	50256	60484
2017	59865	54110	65621	55667	50553	60781
2018	61060	55305	66816	55989	50875	61103
2019	62073	56317	67829	56332	51218	61445
2020	62995	57239	68751	56691	51577	61805
2021	63872	58116	69628	57065	51951	62179
2022	64727	58971	70483	57450	52336	62564
2023	65571	59815	71327	57845	52731	62959
2024	66410	60654	72166	58247	53134	63361
2025	67246	61490	73001	58657	53543	63770
2026	68080	62324	73836	59071	53957	64185
2027	68914	63158	74670	59490	54376	64604
2028	69747	63992	75503	59912	54798	65026
2029	70581	64825	76337	60337	55223	65451
2030	71414	65658	77170	60765	55651	65879

Figure 4. The forecast of sugarcane production and yield from 2013-2030 with 95% confidence interval

Figure 2.5 shows the sugarcane production in Pounds/Acre. This study uses the auto regressive moving average (ARMA) and auto regressive integrated moving average (ARIMA) models of forecasting to try and predict the production and yield of Pakistan's two primary cash crops, cotton and sugarcane. Productions and yields of both crops were projected for 18 years, from 2013 to 2030, using data from 1948 to 2012.For wheat production, wheat yield, and cotton production, respectively, ARMA (1, 4), ARMA (1, 1), and ARMA (0, 1) were determined to be suitable models, but ARIMA (2, 1, 1) was the best model for cotton yield forecasting. On fitted models, some diagnostic tests were also run and found to be well fitted [44].

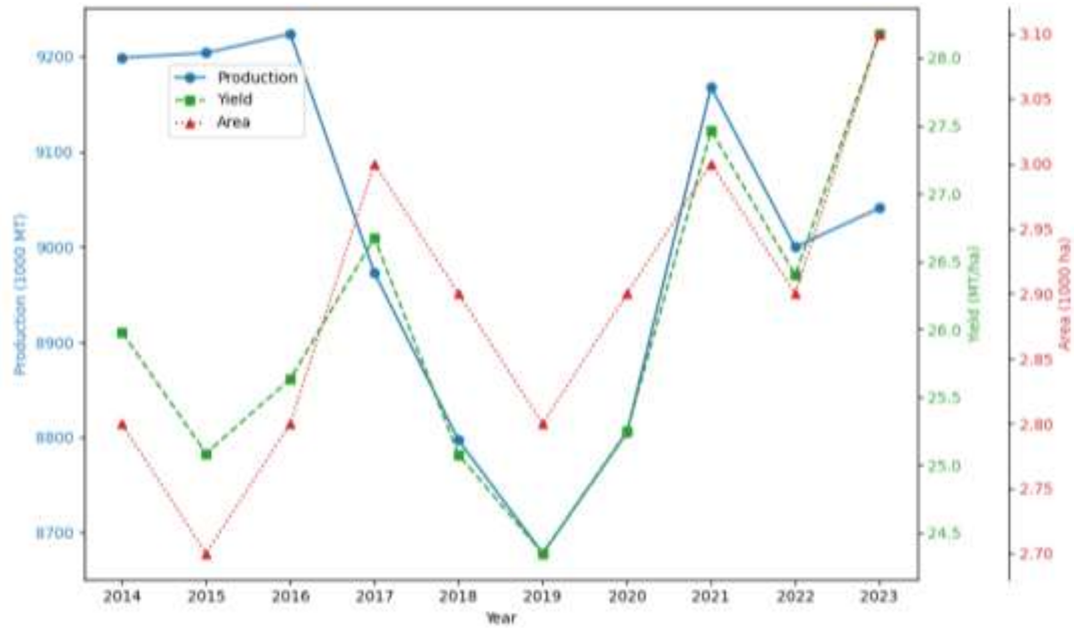


Figure 5. Production and Yield of Wheat Crop in Pakistan (2013-2023)

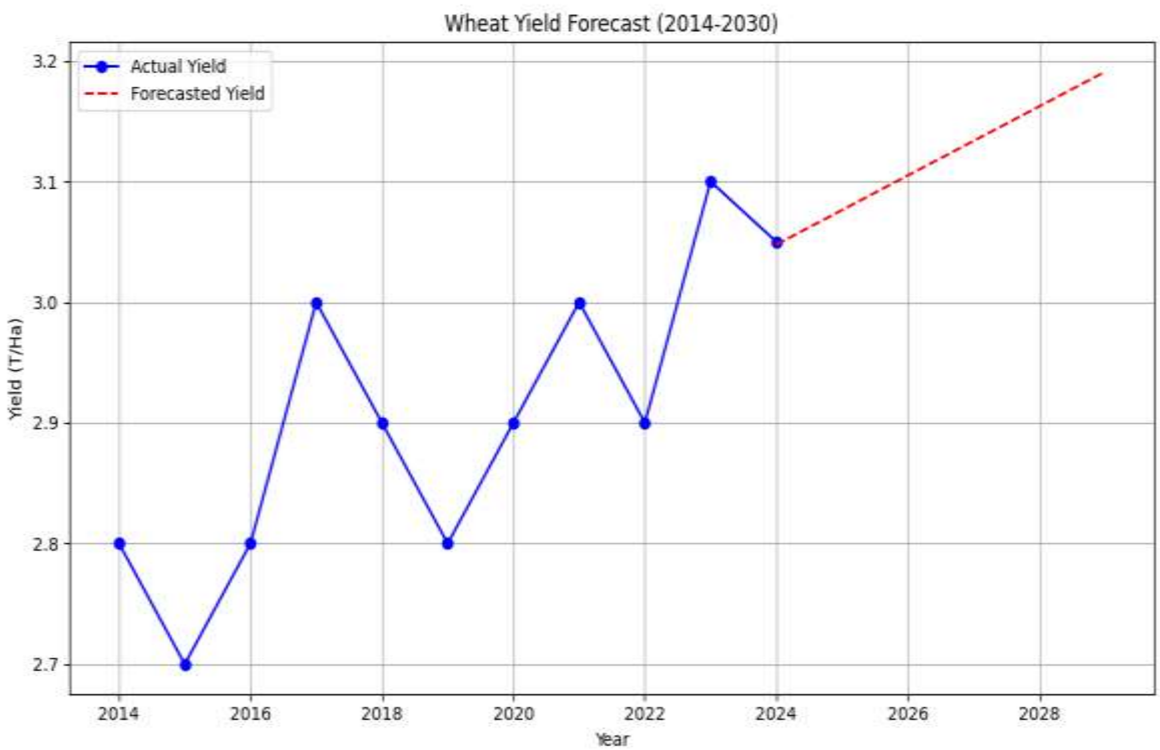


Figure 6. Yield Forecast of Wheat Crop in Pakistan (2013-2023)

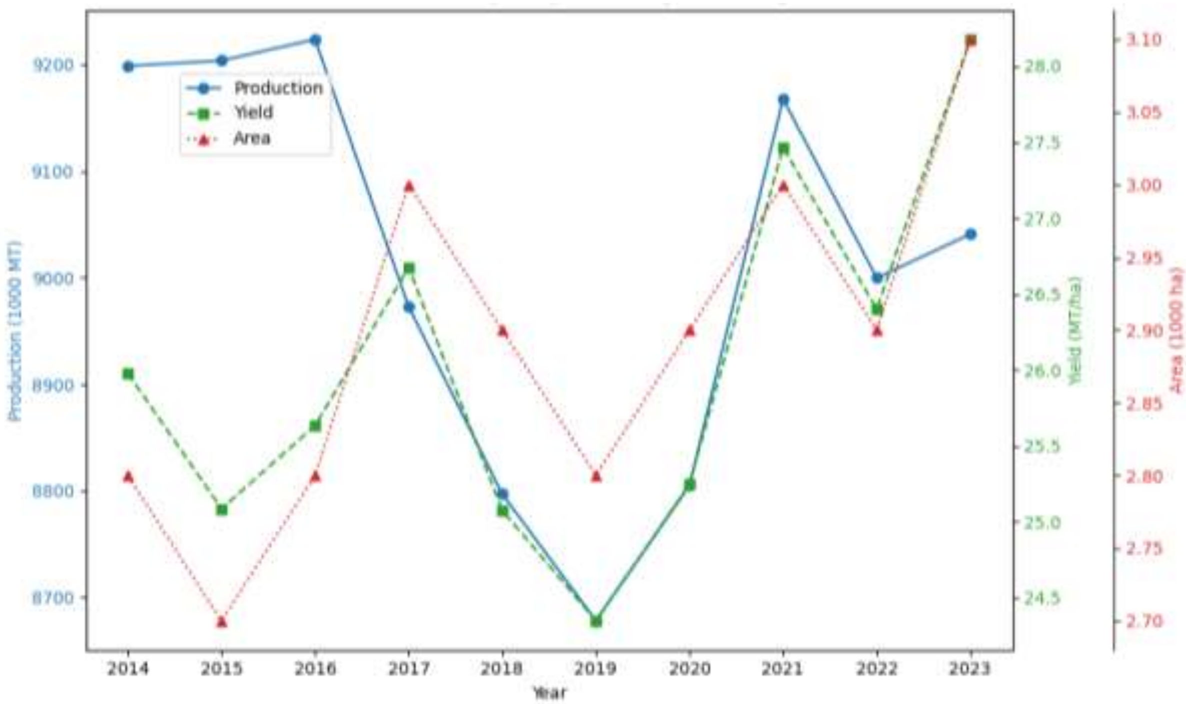


Figure 7. Wheat Production in Pounds/Acre (Raza and Amir, 2021)

3.2. Data from Remote Sensing

We captured the sugarcane growing season for the years 2022–2023 using multi-temporal Sentinel-2 Level-2A data from the time of ratooning (April 2023) through harvest (December 2022). The growing process of sugarcane took an average of 10 to 12 months in Pakistan. Due to the relatively small study area, meteorological data (such as air temperature, evapotranspiration, or precipitation data) was excluded from the analysis because of their low spatial resolution and limited utility for the regression models. Three different vegetation indicators (VIs) were computed: the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Red Edge 1 (NDRE1), and the Chlorophyll Index Red Edge (CIRE), using atmospherically corrected Sentinel-2 surface reflectance (Level-2A) data. The formulas for these indices are shown in Table 1.

Figure 8. Map of district Dera ghazi khan

Table 1. Selected Vegetation Indices for the Analysis

Vegetation Index	Description	Reference
NDVI	Normalized Difference Vegetation Index, sensitive for green biomass	(Tucker, 1979)
NDRE1	Normalized Difference Red Edge 1, less prone to saturation in denser plant canopies	(Gitelson *et al.*, 2005)
CIRE	Chlorophyll Index Red Edge, sensitive for plant leaf chlorophyll and mesophyll water stress	(Reed *et al.*, 1994)

3.3. Normalized Difference Vegetation Index

Based on the reflectance of near-infrared (NIR) and red light bands, the Normalized Difference Vegetation Index (NDVI), a popular vegetation index, measures the density and health of vegetation. Processing NDVI involves the following steps:

i. **Data Preparation:**
- Obtain the requisite near infrared and red bands in the satellite picture. The near infrared band corresponds to Band 8 (B08) in the case of Sentinel-2, while the red band belongs to Band 4 (B04).
- Make sure the imagery has been radio metrically calibrated and atmospherically corrected to take atmospheric fluctuations and sensor response into consideration.

ii. **Band Conversion:**
- Prepare the raw satellite imagery data for processing by converting it. Depending on the precise needs and the information at hand, this can entail converting the images to a radiance or reflectance format.

iii. **NDVI Calculation:**
- Compute the NDVI using the following formula:
  NDVI = (Band8 - Band4) / (Band8 + Band4)
- Red denotes reflectance values from the red band, while NIR denotes reflectance values from the near-infrared band. Depending on the data type, make sure the reflectance values are within the proper range, which is often 0-1 or 0-100.

iv. **Data Normalization:**
- To make NDVI readings easier to understand and compare, normalize them to a standardized range, like 0 to 1. Depending on the precise requirements, this stage may involve rescaling the NDVI data using min-max normalization or employing other normalization approaches.

v. **Visualization and Analysis:**
- Utilize an appropriate image visualization tool or program to visualize the NDVI data after processing. This will make it easier to interpret the vegetation health and density patterns in the research region. Higher NDVI values indicate healthier and denser vegetation. Colour scales can be used to illustrate various vegetation conditions.
- The NDVI data should be subjected to additional analysis, such as temporal analysis to evaluate the dynamics of the vegetation over time or spatial analysis to pinpoint regions with high or low vegetation productivity.

It is vital to remember that based on the dataset and tools available, the specific procedures and software utilized for processing NDVI may change. There are many software programs that offer the ability to analyse and analyse satellite images and calculate vegetation indices like NDVI, including ENVI, QGIS, and Python libraries like GDAL or Rasterio (Zhao *et al.*, 2020b).

3.4. Normalized Difference Red-Edge Index 1 (NDRE1)

The Standard Deviation Red-Edge Index 1 (NDRE1) is a vegetation index that evaluates plant chlorophyll content and vegetation health by using the red-edge band. Processing NDRE1 involves the following steps:

i. **Data Preparation:**

- Obtain satellite images with the necessary near infrared and red-edge bands. The precise band numbers may change depending on the sensor utilized. For instance, in Sentinel-2, Band 5 (B05) is the red-edge band that is closest to 705 nm, and Band 8 (B08) is the near-infrared band [45].
- To account for atmospheric effects and sensor characteristics, make sure the imagery is radio metrically calibrated and atmospherically corrected.

ii. Band Conversion:

- Transform the unprocessed satellite imaging data into a useful format, such as reflectance or radiance values, for processing. Think about the right data range for the particular dataset and requirements.

iii. NDRE1 Calculation:

- Compute the NDRE1 using the following formula:
  NDRE1 = (Band6 – Band5) / (Band6 + Band5)
- NIR and Red-edge are terms used to describe the reflectance values from the near-infrared band (e.g., Band 8) and the red-edge band (e.g., Band 5) respectively.

iv. Data Normalization:

- In order to make understanding and comparison easier, normalize the NDRE1 values to a standardized range, such as 0 to 1. Utilize the proper normalization techniques in accordance with the data set and specifications.

v. Visualization and Analysis:

- Utilizing the appropriate image visualization tools or software, visualize the processed NDRE1 data. Use colour gradations to indicate various plant states, with higher NDRE1 values denoting vegetation that is healthier and more robust.
- Use the NDRE1 data to conduct additional research, such as spatial analysis to pinpoint regions with high vegetation productivity or temporal analysis to evaluate changes in vegetation over time.

It is important to keep in mind that based on the dataset and accessible tools, the particular procedures and software utilized to analyse NDRE1 may change. For processing satellite images and computing vegetation indices like NDRE1, a variety of software packages can be used. These include remote sensing software like ENVI, open-source GIS software like QGIS, and programming libraries like Python's GDAL or Rasterio (Zhao *et al.*, 2020a).

3.5. Canopy Chlorophyll Index Red-Edge (CIRE)

An indicator of vegetation that offers data on the amount of chlorophyll and photosynthetic activity is the Canopy Chlorophyll Index Red-Edge (CIRE). Processing CIRE involves the following steps:

i. Data Preparation:

- Obtain satellite images with the relevant near infrared and red-edge bands. The precise band numbers may change depending on the sensor utilized. For instance, in Sentinel-2, Band 6 (B06) is the red-edge band that is closest to 710 nm, and Band 8 (B08) is the near-infrared band.
- To account for atmospheric influences and sensor characteristics, make sure the satellite imagery has been radio metrically calibrated and corrected for atmospheric conditions.

ii. Band Conversion:

- Transform the unprocessed satellite imaging data into a useful format, such as reflectance or radiance values, for processing. Think about the right data range for the particular dataset and requirements.

iii. CIRE Calculation:

- Compute the CIRE using the following formula:
  CIRE = (Band7 / Band5) - 1
- Red-edge and NIR refer to the reflectance values from the red-edge band (e.g., Band 6) and the near-infrared band (e.g., Band 8) respectively.

iv. Data Normalization:

- In order to make interpretation and comparison easier, normalize the CIRE values to a standardized range, such as 0 to 1. Utilize the proper normalization techniques in accordance with the data set and specifications.

v. Visualization and Analysis:

- Utilize the right image visualization software or tools to visualize the CIRE data that has been processed. Different vegetation conditions can be represented by colour scales, with greater CIRE values denoting more chlorophyll concentration and photosynthetic activity.
- Run further analyses on the CIRE data, such as geographical or temporal analyses, to determine locations with high chlorophyll content or to evaluate variations in vegetation vigour over the year.

It is significant to note that depending on the dataset and available resources, the specific procedures and software utilized for processing CIRE may change. To process satellite images and determine vegetation indices like CIRE, a variety of software packages can be used, including remote sensing tools like ENVI, open-source GIS software like QGIS, and programming libraries like Python's GDAL or Rasterio.

3.6. Method

We compared the accuracy of the yield estimation using two alternative methods for the development of predictor variables. First, NDVI, NDRE1, and CIRE time series of the three separate vegetation indicators are used. Many other studies have used VI time series as predictor variables, where some use time series descriptive statistics (Ronchetti *et al.*, 2023)in addition to temporal segments (Dimo Dimov) or dimensionality reduction strategies for time series.

The second technique involves computing phonological measures (Parida and Ranjan, 2019). These metrics include phenological indicators, such as season start and end dates, which are used to describe agricultural growth and seasonal crop ontogenesis.

3.7. Data preparation for Sentinel-2

Monitoring and evaluating wheat crops is one of several uses for the high-resolution multispectral imagery provided by the Sentinel-2 satellite program, which is run by the European Space Agency (ESA).

The data in the Sentinel-2 Level-2 collection have been pre-processed and corrected for atmospheric effects, making it suitable for additional processing and analysis. The following are the main procedures for handling the Sugarcane Sentinel Level-2 dataset (Parida and Ranjan, 2019):

Data Acquisition: The Sentinel-2 Level-2 dataset for the chosen region and time period of interest can be obtained.

Data Preparation: The downloaded dataset should be extracted, and it usually consists of compressed files. To make the dataset files, including the metadata and individual bands, easier to access during processing, arrange them in an organized directory manner.

Band Selection: Find the appropriate spectral bands for monitoring and predicting wheat yield. Several spectral bands, including the red, green, blue, near infrared (NIR), and shortwave infrared (SWIR) bands, are included in the Sentinel-2 dataset. The analysis's precise goals will choose which bands to use.

Image Registration and Mosaicking: This involved downloading the satellite photos and then masking them using the Level-2A product SLC cloud mask to only get pixels with the labels "land," "water," and "vegetation."

Vegetation Index Calculation: With the mean, minimum, maximum, and standard deviation of the pixel values for each observation date, we generated a number of vegetation indices and spatially aggregated them for each parcel object.

3.8. Time series calculations for the object-based vegetation index

Each multispectral Sentinel-2 raster band and the resultant spectral index raster's are trimmed to the parcel boundaries and saved in a database as part of the acknowledgment system. To assure the use of data points exclusively during the growing season, we constructed object-based time series for each parcel between the indicated planting and harvesting months based on the multi-temporal aggregation of the spectral indices mentioned above. Different saturation levels are of the respective VI are shown in Figure 2.6, Whereas the 10m NDVI saturates earlier than vegetation indices based on the Red Edge bands, it captures a higher amount of detail of the canopy structure [46].

Parametric metrics: The growth and development of wheat crops may be understood and predicted in large part because to phonological characteristics. Phenology is the study of the relationship between environmental conditions and the timing of recurrent processes in plants, such as blooming, fruiting, and senescence. The following phonological characteristics are important for predicting wheat yield: emergence, vegetative growth, flowering, ripening, senescence, and phonological timing and duration (Saini *et al.*, 2023).

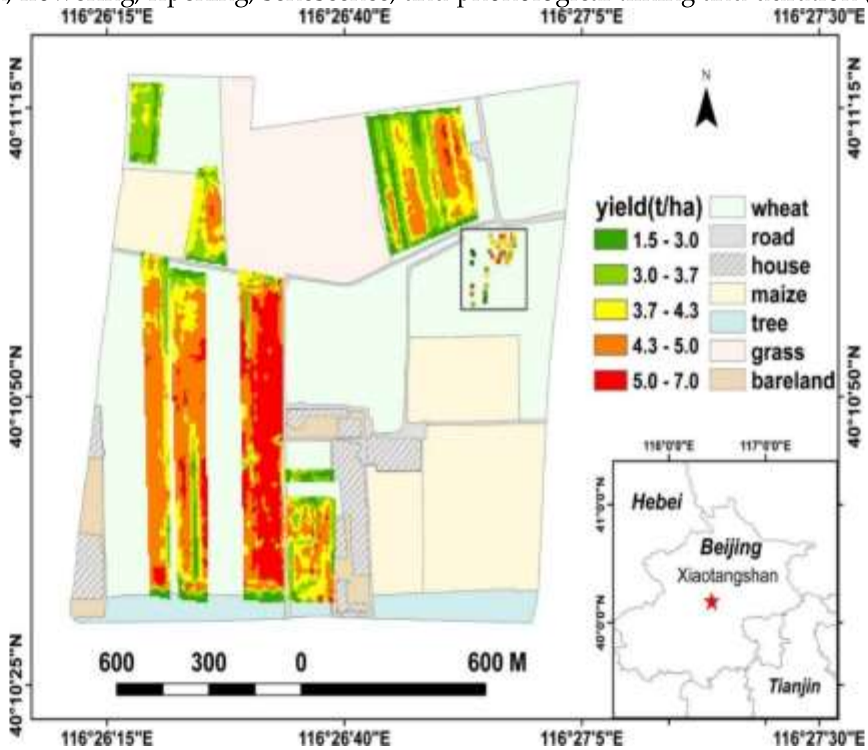


Figure 9. Visualization of the different saturation levels of the vegetation indices NDVI, NDRE1 and CIRE (on the same observation date)

We computed 12 distinct metrics used as input features based on the phonological metrics supplied by the USGS (Reed *et al.*, 1994) and extracted from EO data. As shown in comparable research studies (Gao and Zhang, 2021). NDVI time series are often used to extract phonological indicators and metrics. As NDVI signals become saturate with a larger canopy density, they make it possible to monitor vegetation types independently of biomass (Saad El Imanni *et al.*, 2022).

Utilizing ground observations, historical records, and remote sensing data, it is possible to monitor and quantify these phonological traits. Large-scale and regular measurements of phonological changes can be taken using remote sensing platforms like satellites or drones. These phonological traits can aid in the creation of reliable sugarcane yield prediction models when combined with machine learning and statistical modelling techniques. The following definitions apply to the generated phonological metrics:

(a) Phenotypic Indicators:
- The beginning of the growing season, including the date of crop emergence
- The day with the highest value recorded in the temporal sequence i.e., Seasonal apex
- Agricultural harvest date and absence of chlorophyll indicate the end of the season
- The highest NDVI value recorded throughout time

(b) Detailed Metrics:
- Total of the time series' average NDVI values
- Total of the time series' highest NDVI values
- The NDVI's range from lowest to highest

(c) Growth Indicators:
- Seasonal duration: the time between start and peak, expressed as the number of days
- First duration: the number of days between start and peak where green-up occurred
- Second duration: the number of days between peak and end during senescence

- Gradient between start and peak in growth rate
- Gradient between peak and end in growth rate

The phonological measures were calculated using daily, linearly interpolated NDVI time series as the input data. The gradients between start of season and peak of season and end of season, which are as the growth or senescence rates, are the sources of the growth and senescence rates. The NDVI time series' local maxima peaks, which sugarcane cutting occurrence, are the peaks. Unless this dimensionality reduced by temporal aggregation or by dimensionality reduction techniques like Principal Component Analysis, phonological markers are used in situations where various NDVI datasets have varying dimensionalities (Segarra *et al.*, 2020).

3.9. Discussion

Phonological measurements are computed in the given setting using daily NDVI (Normalized Difference Vegetation Index) series of data that have been linearly interpolated. An area's degree of greenness and the health of its plants are indicated by the widely used remote sensing indicator known as NDVI. To track changes in plant development over time, the NDVI measurements are gathered. The variations between these sites are computed to provide phonological measurements, such as the beginning, peak, and end of the growing season. These gradients describe the rates of growth and senescence, which reflect the rate of expansion or contraction of the vegetation. Researchers can comprehend the phonological processes of the vegetation by examining these rates (Shafique *et al.*, 2017).

The local peak peaks in the NDVI time series in the context of sugarcane cutting show the presence of sugarcane harvests. These peaks indicate the season's greatest level of vegetation greenness and point to the ideal time for sugarcane harvesting. Even so, dimensionality reduction methods like periodic aggregation or PCA (principal component analysis) can be used when working with numerous NDVI collections that may have varying dimensionalities (different amounts of data points). Reduced temporal resolution is achieved through temporal aggregation, for example, by averaging NDVI measurements over longer time periods. The statistical method known as PCA minimizes the number of dimensions in the data by focusing on its most significant patterns or components. Then, multiple NDVI datasets with various dimensionalities can be compared and analysed using phonological markers, which are obtained from the NDVI data series and determined growth and senescence rates. These markers offer important new perspectives on the phonological trends and cyclical dynamics of vegetation across time.

Figure 2.7 displays a graphical representation built on the VI time series collected from the cloud platform to give an overview of the deduced phonological features. The examination of phonological data is greatly aided by such graphics and client-based dashboards. They make it simple for academics and other interested parties to monitor and evaluate phonological patterns and trends, which improves our comprehension of how vegetation changes over time.

Regression Algorithms for Crop Yield Estimation: Many approaches are used to forecast yield at the regional and field levels as follows:

- Field surveys
- Crop models
- Remote sensing
- Statistical models

Regression methods can be used to forecast the yield of wheat depending on a variety of input variables when it comes to crop yield estimation.

Here are some commonly used regression algorithms for crop yield estimation: Linear Regression, Multiple Linear Regression, Decision Trees, Random Forest, Support Vector Regression (SVR), Gradient Boosting, and Neural Networks. Using historical yield data as the target variable and the pertinent input variables indicated earlier, several regression algorithms can be trained. Based on the input data, the models are then utilized to forecast the yield for fresh or next seasons.

It's crucial to evaluate the effectiveness of several algorithms using the right assessment measures and select the one that offers the best accuracy and generalization capabilities for sugarcane yield estimation (Singla *et al.*, 2020).

We experimented with the ordinary least squares (OLS) regression algorithm and the Random Forest (RF) regressor to compare linear and non-linear data correlations.

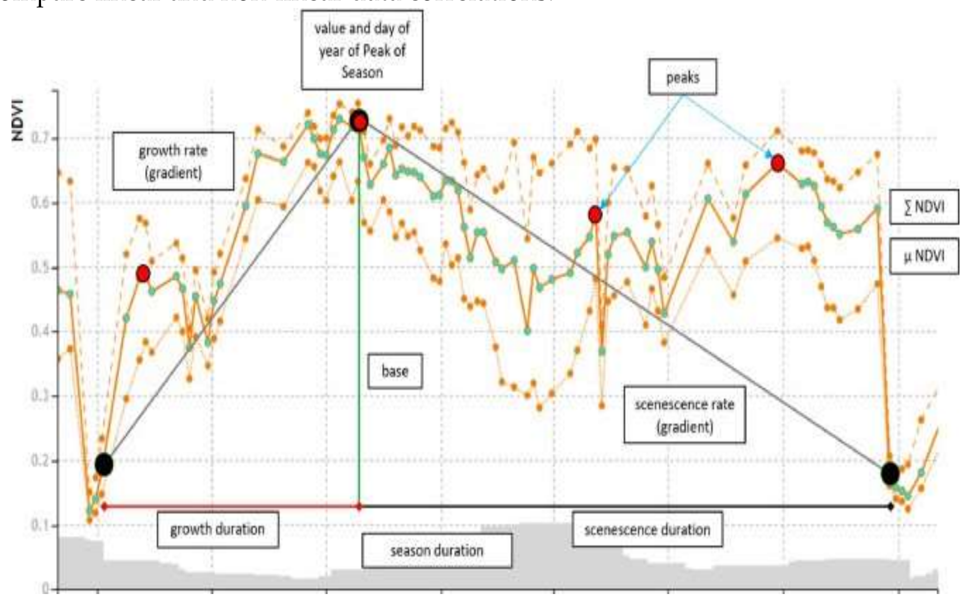


Figure 10. Graphical depiction of the derived phonological features from the NDVI time series

To reduce absolute error and handle more complex data patterns with high dimensionality (such as time series) and non-linear relationships than OLS, the RF method employs a large number of independent decision trees by averaging their predictions. Principal Component Analysis (PCA) was used to decrease the feature set's dimensionality to the components that accounted for the majority of variance and, thus, the majority of the information gains for the combinatorial usage of both types of predictor variables. By using an orthogonal transformation, the variables are linearly de-correlated, and the final components either include signal noise or the most de-correlated information about the variables (Feng *et al.*, 2021).

Random Forest (RF): Each tree in a random forest depends on the values of a random vector that was sampled randomly and with the same distribution for all the trees in the forest. As the number of trees in a forest increases, the generalization error converges as to a limit. The strength of each individual tree in the forest and the correlation between them determine the generalization error of a forest of tree classifiers. Each node is divided using a random selection of features, producing error rates that are comparable to the Adaboost but more resilient to noise. Internal estimates keep track of inaccuracy, strength, and correlation; they are used to demonstrate how the splitting process responds to an increase in the number of features. Internal estimations are another method for gauging variable significance. Regression can also use these concepts (Su *et al.*, 2019).

The out-of-bag (OOB) error experienced during training is used to estimate the random forest's error. Each tree is constructed using a unique bootstrap sample. Each bootstrap sample excludes roughly one-third of the observations at random. The OOB sample refers to these omitted observations for a specific tree. In many cases, finding parameters that would result in a low OOB error is a crucial factor in model selection and parameter tuning. Take note that the size of the subset of predictor variables is essential for determining the final depth of the trees in the random forest algorithm. We determine the so-called variable relevance of each variable in order to acquire some understanding of the complex model. This is determined by adding the improvement in the objective function specified in the splitting criterion for each predictor variable individually over all internal nodes of a tree and across all trees in the forest. The variable importance score in the Stata random forest implementation is normalized by dividing all scores by the maximum score; the most important variable's importance is always 100% (Su *et al.*, 2019).

The capacity of Random Forest can manage non-linear connections, a high number of input variables, and effectively handle missing data are only a few of its benefits for wheat yield estimation. Additionally, it offers insights into the significance of features, enabling you to pinpoint the primary factors influencing yield projections. It's crucial to remember that the quality and representativeness of the dataset, as well as the suitable choice of input variables, all affect how well the model performs (Su *et al.*, 2019).

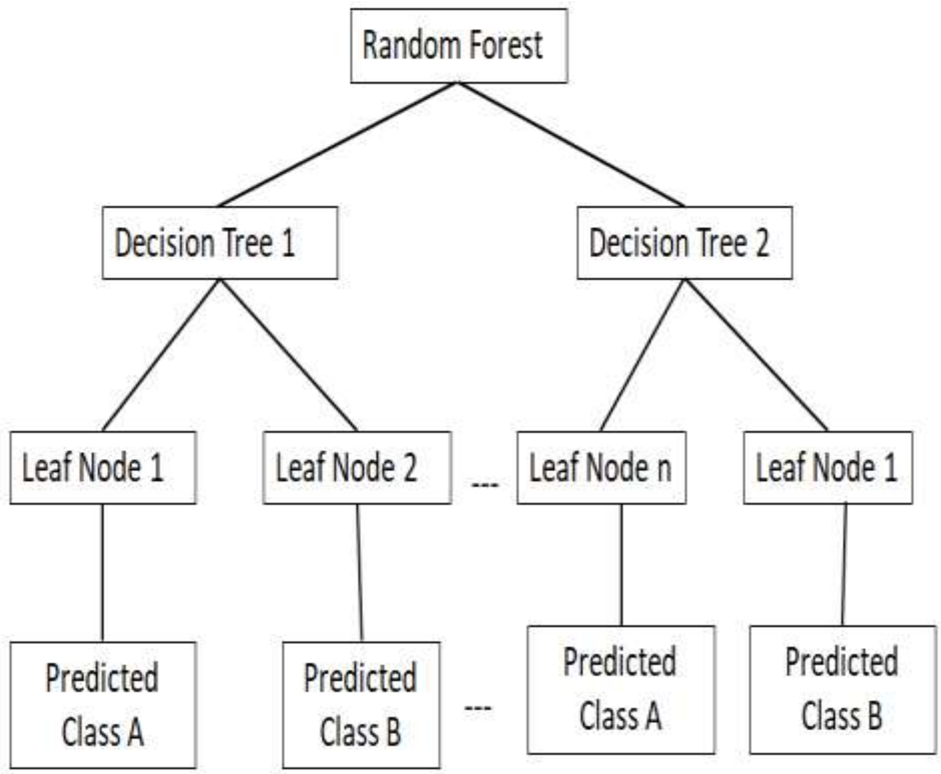


Figure 11. General Architecture of Random Forest

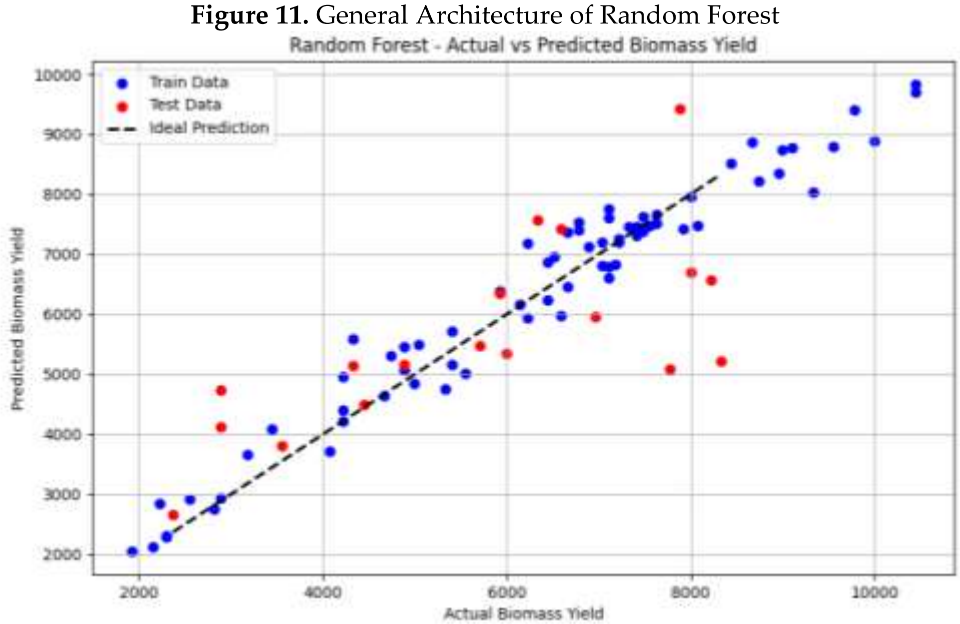


Figure 12. Results on Random Forest

Gradient Boosting: The widely used statistical technique of regression analysis has applications in a variety of fields, including social sciences and finance. We can represent the relationship between one or more independent variables and a dependent variable using this technique. Gradient Boosting stands out as one of the most popular and frequently used approaches among the various regression methodologies. By iteratively combining weak learners—typically decision trees—Gradient Boosting seeks to improve data analysis. This ensemble method builds models sequentially, with each new model correcting the errors made by the previous ones. The Gradient Boosting approach finds the model that best fits the given set of data points by minimizing the loss function through gradient descent optimization. This is accomplished by estimating a series of models to minimize the sum of the squared differences between the observed values of the dependent variable and the anticipated values from the ensemble of models (Tiruneh *et al.*, 2022).

The Gradient Boosting method is widely used because it is powerful and flexible. It offers a sophisticated and effective method for determining the link between variables and making predictions. The Gradient Boosting approach produces estimates that are ideal in terms of capturing complex patterns and minimizing errors in the model's predictions by focusing on the residuals of prior models. The Gradient Boosting approach also enables interpretation of the calculated feature importance, which sheds light on the significance and influence of each independent variable on the dependent variable. This makes it possible for researchers and analysts to evaluate the importance of each element and comprehend how it affects the desired outcome.

Although the Gradient Boosting approach is renowned for its robustness and high performance, it is crucial to remember that its efficacy depends on several parameters, including the learning rate, number of trees, and tree depth. The precision and reliability of the Gradient Boosting estimates may be impacted by improper tuning of these parameters. As a result, it is crucial to perform parameter tuning and, where appropriate, consider cross-validation. In conclusion, the Gradient Boosting method is a powerful regression technique that minimizes the loss function to find the model that best fits a specific set of data points. It offers a comprehensive and effective method for calculating a regression model's predictions and making forecasts. The Gradient Boosting method is preferred for its high accuracy and robustness, which enable researchers and analysts to understand the relationship between variables. However, it's crucial to carefully tune the parameters to ensure the model's applicability for the current investigation.

K-Nearest Neighbors (KNN): K-Nearest Neighbors (KNN) is a widely used non-parametric method in the estimation of wheat yield. KNN seeks to establish a relationship between one or more independent variables or predictor factors and a dependent variable, such as wheat yield. In order to do this, KNN uses the idea of proximity—predicting the value of the dependent variable based on the values of the k-nearest neighbors in the feature space. The method is straightforward: it calculates the distance between the point to be predicted and all other points, selects the k-nearest points, and predicts the value based on the average (or majority) of these neighbors.

The ease of use and interpretability of the KNN regression method make it a popular choice. It offers insightful information on the relationships between the predictor variables and wheat yield. KNN assesses the strength and direction of the correlations between the predictor variables and the dependent variable by considering the local patterns within the data.

But it is essential to understand the presumptions behind KNN regression. The linearity of the relationship between the predictor components and the dependent variable is not a requirement for KNN, making it suitable for capturing non-linear patterns. However, KNN assumes that the proximity in the feature space implies similarity in the output space, which might not always hold true. Additionally, the choice of k (number of neighbors) is crucial as it influences the model's bias-variance trade-off. A small k can lead to a highly flexible model (low bias, high variance), while a large k can result in a more stable model (high bias, low variance).

Homoscedasticity is another consideration, which states that the variance of the residuals should ideally be constant across all levels of the predictor variables. However, KNN does not directly address homoscedasticity as it is more focused on local proximities rather than global variance structures. In situations with heteroscedasticity, other techniques or transformations might be necessary to achieve more reliable results.

In conclusion, K-Nearest Neighbors (KNN) is a popular method for estimating wheat yield and analyzing the links between predictor variables and wheat yield. It offers a simple and understandable framework for performing regression analysis without requiring strong assumptions about the underlying data distribution. Researchers can get more precise and reliable estimates of wheat output and gain better understanding of the factors affecting its production by appropriately tuning the number of neighbors and considering the structure of the data.

Neural Networks: Neural Networks are a fundamental and frequently used method in the estimation of wheat yield. Neural Networks seek to establish a relationship between one or more independent variables or predictor factors and a dependent variable, such as wheat yield. This is achieved through a network of interconnected nodes (neurons) arranged in layers—input, hidden, and output layers. Each connection has an associated weight, and the network learns by adjusting these weights to minimize the loss function, typically through backpropagation.

The adaptability and high performance of Neural Networks make them a popular choice. They offer the capability to model complex non-linear relationships between the predictor variables and wheat yield. Neural Networks assess the strength and direction of the correlations by transforming the inputs through multiple layers of non-linear activation functions, thus capturing intricate patterns in the data.

It is essential to understand the structure and parameters behind Neural Networks. The architecture, including the number of layers and neurons, activation functions, learning rate, and regularization techniques, plays a crucial role in the model's efficacy. Neural Networks assume that sufficient data is available for training to avoid overfitting and to generalize well on unseen data.

The concept of homoscedasticity, which implies that the variance of the residuals should be constant, is less critical in Neural Networks as they inherently capture variance through their complex structure. However, the model's performance can still be affected by issues such as overfitting, under fitting, and convergence problems. Techniques like dropout, batch normalization, and early stopping are often employed to address these challenges and improve model robustness.

Residuals are further assumed to be normally distributed in many traditional regression techniques. However, Neural Networks do not make strong assumptions about the distribution of residuals, making them flexible and suitable for a wide range of applications. When deviations from normality occur, Neural Networks can still perform well due to their ability to learn directly from the data.

In conclusion, Neural Networks are a powerful technique for estimating wheat yield and analyzing the relationships between predictor variables and wheat yield. They offer a flexible and sophisticated framework for performing regression analysis, capable of modeling complex non-linear patterns. By appropriately tuning the network architecture and employing regularization techniques, researchers can achieve precise and reliable estimates of wheat output and gain a deep understanding of the factors affecting its production. Although the Neural Network approach is complex and computationally intensive, it provides robust and high-performing results that are invaluable in data analysis.

4. Results

Various regression models, including Random Forest (RF), K-Nearest Neighbors (KNN), Gradient Boosting, and Neural Networks, have been frequently used in the context of remote sensing-based wheat yield prediction models employing artificial intelligence. These models have unique traits and can provide important insights into how remote sensor data and wheat production relate to one another.

The Random Forest algorithm is a potent machine learning technique that creates an ensemble model by combining the predictions of various decision trees. It excels at managing nonlinear relationships and capturing intricate interactions between predictor factors. Due to its capability to handle the multidimensional and multi-temporal character of remote sensing data, the RF regressor has been used in models for predicting wheat yield based on remote sensing. The RF regressor's capacity to manage non-linear data is one of its benefits. It allows for more flexible modeling because it does not imply a linear connection between the predictor factors and the dependent variable. When dealing with complicated correlations between remote sensing parameters and wheat production, the RF algorithm's ability to capture non-linear trends and interactions can be useful. The RF regressor's capacity to deal with high-dimensional data is another benefit.

There are typically a lot of spectrum bands and generated indices in remote sensing data, which leads to a lot of predictors. The RF algorithm is capable of effectively handling this high-dimensional input and choosing the most pertinent aspects for prediction. It automatically selects features, lowering the possibility of overfitting and enhancing model generalization. However, the RF regressor has several restrictions. Due to the ensemble structure of the technique, it might be difficult to grasp the precise contribution of each individual predictor to the total forecast, which can be a disadvantage of the model's level of complexity and interpretability. Furthermore, compared to more straightforward models like OLS, the RF repressor can be computationally demanding, especially with large datasets, and may need more resources for training and prediction.

K-Nearest Neighbors (KNN) is a non-parametric method frequently employed in wheat yield estimation models using remote sensing data. The KNN algorithm predicts the value of the dependent variable based on the values of the k-nearest neighbors in the feature space. This method is straightforward. The ease of use and interpretability of the KNN regression method make it a popular choice. KNN does not assume a linear relationship, making it suitable for capturing non-linear patterns in the data. However, the choice of k (number of neighbors) significantly influences the model's performance, affecting the bias-variance trade-off. A small k can lead to a highly flexible model, while a large k can result in a more stable but potentially less accurate model. Additionally, KNN can be computationally intensive, particularly with large datasets, and may not perform well with high-dimensional data due to the curse of dimensionality.

Gradient Boosting excels in handling complex, non-linear relationships and interactions between predictor variables. It is particularly effective with high-dimensional remote sensing data, as it can focus on the most relevant features and improve model accuracy iteratively. The flexibility of Gradient Boosting in capturing non-linear trends and the ability to adjust its parameters, such as learning rate and number of trees, allow for fine-tuning the model to achieve optimal performance. However, Gradient Boosting can be computationally demanding and may require careful parameter tuning to avoid overfitting and ensure generalization.

Neural Networks are a fundamental and frequently used method in the estimation of wheat yield. These models establish a relationship between the predictor variables and wheat yield through a network of interconnected nodes (neurons) arranged in layers—input, hidden, and output layers. Each connection has an associated weight, and the network learns by adjusting these weights to minimize the loss function through backpropagation. KNN offer the capability to model complex non-linear relationships and can capture intricate patterns in remote sensing data. The architecture of Neural Networks, including the number of layers and neurons, activation functions, learning rate, and regularization techniques, plays a crucial role in the model's efficacy. While Neural Networks are highly flexible and powerful, they can be computationally intensive and require substantial amounts of data for training. Additionally, they may suffer from issues such as overfitting and convergence problems, which can be mitigated through techniques like dropout, batch normalization, and early stopping.

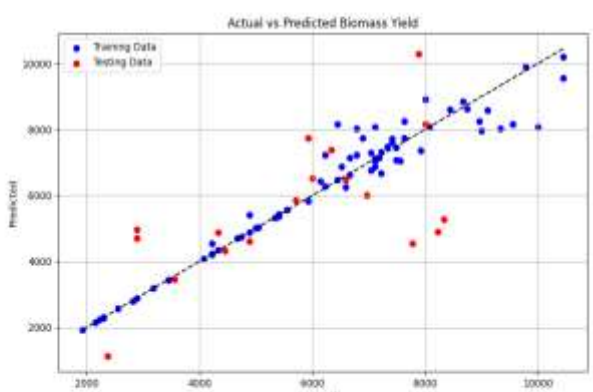
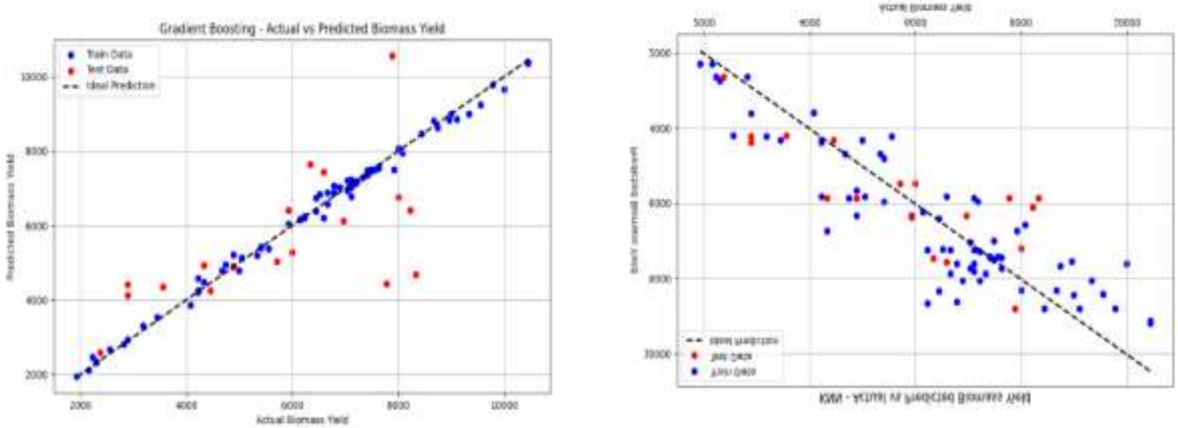


Figure 13. Results of Biomass estimation using Random forest

In remote sensing-based wheat yield prediction models, Random Forest (RF) regressor, K-Nearest Neighbors (KNN), Gradient Boosting, and Neural Networks each offer unique advantages and disadvantages. The RF regressor is proficient with non-linear connections and high-dimensional data, making it a good choice for capturing intricate interactions among remote sensing variables. KNN provides simplicity and interpretability, but its performance is highly dependent on the choice of k and may struggle with high-dimensional data. Gradient Boosting offers high accuracy and flexibility in modeling complex patterns, although it requires careful parameter tuning and can be computationally demanding. Neural Networks provide a sophisticated framework for capturing complex non-linear relationships, but they require significant resources and large amounts of data for effective training.



(a) **(b)**

Figure 13. Results of yield prediction using Gradient Boosting

The above figure 15 scatter plot illustrates the neural network model's actual versus predicted biomass yield. The x-axis denotes actual biomass yield, while the y-axis shows predicted biomass yield. Training data points are in blue, and test data points are in red. The black dashed line signifies ideal predictions, where actual yield matches predicted yield. Points close to this line indicate high accuracy, reflecting the model's effectiveness for both training and test data (Singh *et al.*, 2023).

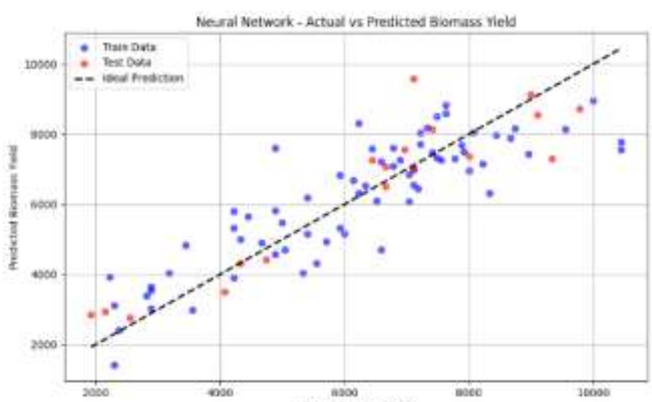


Figure 14. Results of biomass estimation using neural network

The bar chart fig. 16: compares the R^2 scores of different machine learning models used for wheat yield estimation. The models and their respective R^2 scores are:

- Decision Tree: R^2 score of 0.21, shown in blue.
- Random Forest: R^2 score of 0.67, shown in orange.
- Neural Network: R^2 score of 0.85, shown in green.

K-Nearest Neighbors (KNN): R^2 score of 0.61, shown in red. The Neural Network model has the highest accuracy.

4.1. Examining Time Series Variables

Because the CIRE (Canopy Index Recovery) features performed better than other vegetative indices (VIs) at forecasting wheat yield, we specifically focused on them in our study. In order to highlight their importance in the model findings, the following figures exclusively show the CIRE features. We used linear regression analysis, more specifically the ordinary least squares (OLS) method, for each time step of the CIRE to determine the ideal dates for sugarcane yield estimation. According to the findings, the most useful times for yield estimation were just before and during the season's peak, which corresponded to the phase from the beginning of the period (or ratooning) to point at which the canopy is fully closed (or NDVI saturation).

We produced multitemporal images presenting the R^2 (coefficient of determination) and RMSE (root mean square error) scores to visually represent the effectiveness of the regression model utilizing various predictor variables. With the help of these numbers, we were able to determine the ideal satellite observation dates that produced the best model performance as measured by the highest R2 scores and lowest RMSE values. We were able to identify the precise times during the growing season that provided the most precise and dependable predictions by examining the multitemporal visualization of R2 and RMSE scores. These dates showed the strongest correlations between the measured CIRE levels and wheat yield, as shown by the highest R2 scores. The acceptability of these dates for the regression model was further supported by the lowest RMSE values, which indicated smaller prediction errors (Gillard *et al.*, 2021).

In order to concentrate our data collection efforts on the most useful time periods, it is essential to identify the best satellite observation dates. By collecting remote sensing data during these particular dates, we may improve the accuracy and precision of our projections of sugarcane output. In conclusion, our investigation demonstrated how the CIRE characteristics outperformed other VIs in terms of predicting wheat yield. The data in our analysis demonstrated the best times to estimate yields, highlighting the value of satellite measurements made before and at the height of the season. These dates demonstrated the strongest model performance, as indicated by the highest R2 scores and lowest RMSE values. These results allow us to shorten data collection processes and enhance the precision of our regression model.

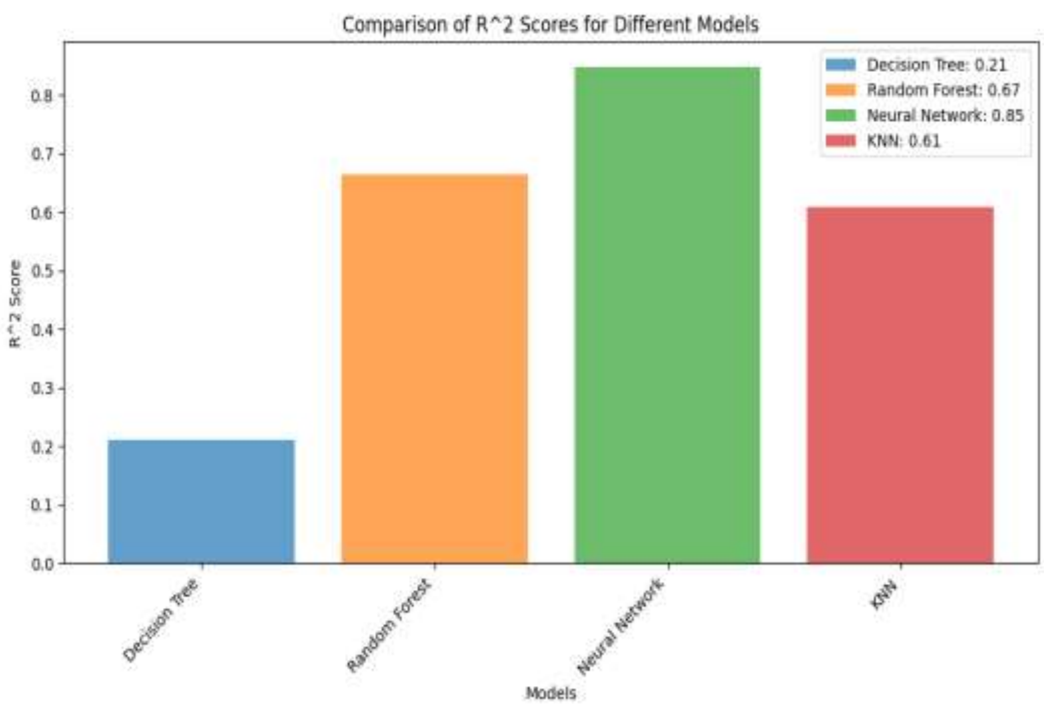


Figure 15. Comparison of R^2 scores of different models

To further optimize the choice of ideal satellite observation dates and improve the forecasting capabilities of the wheat yield prediction model, future research can study the integration of new factors and advanced modelling techniques.

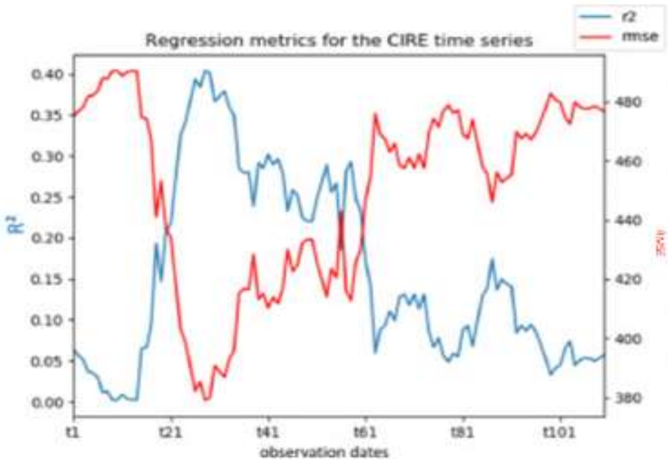


Figure 16. Evolvement of R² and RMSE scores for CIRE

We identified a distinct pattern in the R2 scores across the various stages of the wheat development cycle through our examination of the CIRE period sequence. The R2 marks regularly rose during the green-up or evolution level, peaked at their highest levels at the end of the season, and then started to fall during the senescence period. As the vegetative vigor and biomass accumulation rise throughout the green-up phase, the rising R2 values suggest a greater link among CIRE and wheat yield estimation. The CIRE values capture changes in the wheat flour canopy that correspond to this phase's active growth and development.

The R2 scores peaked during the height of the growing season, showing the largest link mid of CIRE session and wheat yield. The CIRE values, which offer important information on the health and productivity of the wheat crop, coincide with the time of maximum canopy closure and biomass buildup during this period. The R2 scores rapidly fell after the season's peak, during the senescence period. This drop in scores represents the wheat crop's declining production and vigor as it moves closer to maturity and eventual harvest. The drop in vegetative activity that characterizes the senescence phase is captured by the CIRE values, leading to reduced correlations with sugarcane yield.

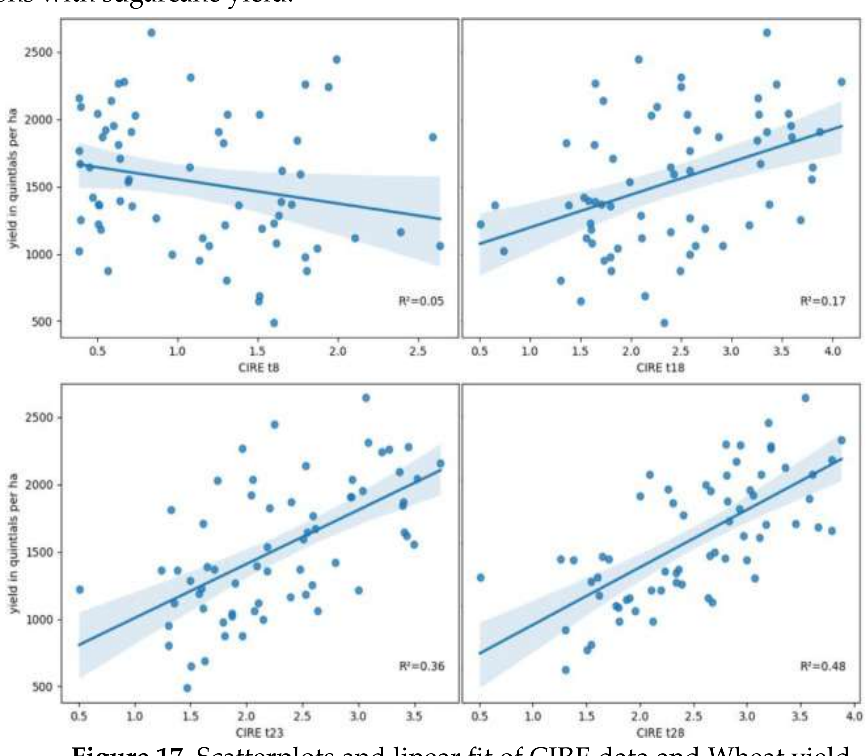


Figure 17. Scatterplots and linear fit of CIRE data and Wheat yield

We noticed a stabilization of the linear model after merging the CIRE time series data with the regression model. This suggests that the capacity of the model to reflect the general trend and dynamics of sugarcane growth is improved by the inclusion of several observations from various phases of the growing season. A more thorough description of the crop's phonological variations is provided by the combined use of the CIRE time series, which enhances model performance and raises R2 scores. This shows that the complexity and variety of sugarcane development patterns and their connection to yield may not be fully captured by relying exclusively on isolated observations. The time series analysis allows for a more thorough comprehension of the dynamics of growth and more accurate forecasting of sugarcane production.

In conclusion, our study of the CIRE time series revealed that the R2 scores progressively rise during the green-up phase, peak at the maximum canopy closure of the season, and then gradually fall during the senescence period. Relying just on a single observation results in poorer R2 scores, while combining the CIRE time series results in a stabilized linear model and improved performance. These results underline how crucial it is to anticipate sugarcane output accurately by taking into account the complete time series and the dynamics of wheat growth.

To improve the wheat estimation accuracy of wheat yield based on the CIRE time series, future study can investigate more sophisticated modelling techniques and incorporate further variables. A further important insight into the underlying mechanisms and dynamics of wheat development and productivity can be gained by examining the precise elements impacting the patterns in the R2 scores.

4.2. Exploring the independent variables

In conclusion, our study showed a distinct relationship between Wheat yield and the predictor factors (mean CIRE, senescence rate, and growth gradient). The potential of these variables as measures of wheat productivity is suggested by the positive correlation between mean CIRE and yield as well as the negative correlations seen for senescence rate and growth gradient. These discoveries advance our knowledge of the variables affecting sugarcane output and can guide agricultural management practices meant to maximize crop productivity and guarantee long-term Wheat production.

In this research, we used the Random Forest (RF) algorithm to conduct a SHAP (Shapley Additive Explanations) analysis of the phonological variables. We were able to comprehend the significance of the different factors in forecasting wheat output through this study and acquire a clearer understanding of their contribution to the overall model performance. Several phonological factors were found to have substantial effects on the model's performance by the SHAP analysis. The highest NDVI value, which corresponds to the season's peak (POS), has become a key predictor. During the growing season, a higher POS value indicated a stronger and vigorous growth stage, which benefited wheat yield. According to this finding, an important element in accurately forecasting Wheat output is capturing the maximum vegetative vigor realized during the season.

The season duration, which is the total of days between the start of the time period (SOS) and the conclusion of the time period (EOS), was another significant variable discovered by the SHAP study. The length of the growing season and its effect on wheat yield were well-understood thanks to the season's duration. Sugarcane yield was shown to be correlated with season length, suggesting that a longer window of ideal growth circumstances can result in more biomass building up. Additionally, a key factor in the effectiveness of the model was the growth indicated by the gradient among SOS and POS. This variable measured how quickly and vigorously plants grew throughout the early part of the season. A faster growth rate suggested a quick and active green-up process, which boosted wheat production (Khan *et al.*, 2020).

One benefit of employing SHAP analysis over traditional feature importance metrics is its capacity to calculate and quantify each sample's contribution to overall model performance. This indicates that the analysis offers insights into how particular samples affect the model's predictions in addition to highlighting the significance of particular variables. Understanding the precise causes of the wheat yield changes within the dataset can be facilitated by this knowledge. In conclusion, we were able to pinpoint the major factors that contributed to the act of the whole typical through SHAP analysis of phonological variables utilizing the RF method. The three factors that were shown to have the most bearing on wheat yield prediction were the highest

NDVI value (POS), the length of the season, and growth pace (gradient between SOS and POS). These discoveries help us better understand the underlying dynamics of wheat growth and offer useful information for making agricultural decisions (Hamada *et al.*, 2021).

We acquire a thorough grasp of not just the significance of variables but also how each unique sample contributes to the overall model predictions by utilizing the power of SHAP analysis. This information can be used to improve the model, spot potential outliers or significant examples, and direct future study and use of the wheat production forecast model. Future research may investigate how SHAP analysis might be combined with other cutting-edge machine learning algorithms and remote sensing methods to increase the precision and interpretability of sugarcane yield estimates (Rodríguez-Pérez and Bajorath, 2020).

We represented the SHAP feature importance together with their corresponding intensity in order to fully comprehend the effect of phonological variables on the sugarcane yield prediction model. We were able to evaluate the contributions of many variables and the impact of their values on the effectiveness of the model thanks to this visualization. The three most important variables— the summed NDVI values, growth rate, and senescence rate —were highlighted using the SHAP feature importance plot. Each variable's intensity, as shown by it's high and low values, gave a sense of how much of an impact it had on the model. Greater contributions to the model's predictions were correlated with higher levels of these variables.

The visualization for the total NDVI values revealed that higher values had a more significant favorable effect on the model. This suggests that the NDVI values, which measure the overall vegetation vigor throughout the growing season, had a beneficial impact on the forecast of wheat yield. On the other hand, lower values had a more negligible effect on the model, indicating that regions with weaker vegetation had less of an impact on the calculated yield. A similar association between intensity and impact could be seen in the growth rate variable. The model's predictions performed better when growth rates were higher, which suggested a quick and active green-up process. This suggests that regions with greater growth rates early in the growing season were also associated with greater wheat yields. Lower growth rates, on the other hand, had a comparably less impact on the model, indicating a smaller contribution to wheat yield estimation.

A similar pattern was also seen in the senescence rate variable. The model was more negatively impacted by higher senescence rates, which represent a more rapid drop in vegetative vigor towards the end of the season. This implies that regions where the senescence process occurred more quickly had lower wheat yields. Lower senescence rates, on the other hand, had a less significant negative effect on the model, indicating a less significant influence on wheat yield estimation. It's important to note that several factors, such EOS (end of the season) and season length, showed a more complicated correlation between intensity and impact. The model's predictions were improved by higher EOS values and longer season lengths, but exceptionally high values had the opposite effect. This shows that these factors have an ideal range within which additional increases may not always result in higher yields. These results underline how crucial it is to take into account the ideal timing and length of the wheat growth cycle in order to maximize production.

5. Conclusion

In conclusion, understanding the contribution and influence of phonological variables on the wheat yield estimation model was made possible through the visualization of the importance of SHAP feature intensity. Three important variables that have a significant impact on yield estimation are growth rate, senescence rate, and the sum of NDVI values. The visualization emphasized the significance of greater values in these variables for influencing yield forecasts favorably, while extreme values in other variables, such EOS and season duration, had the opposite effect. These results can guide management and decision-making techniques for maximizing wheat yield and agricultural practices.

Future studies can investigate additional visualizations and cutting-edge methodologies to examine the connections between phonological factors, their intensities, and their effects on forecasts of wheat production in greater detail.

The early estimation capabilities of the suggested framework should be improved and validated through additional research that types use of leading-edge remote sensing tools, machine learning algorithms, and extensive field observations. We can pave the way for more effective and efficient sugarcane production and support overall food security and agricultural sustainability by continually improving our understanding of and capacity to predict wheat yield dynamics. The OLS regressor consistently produced the best R2 scores, which indicate a better fit between the predicted and observed values—and the lowest root mean squared error (RMSE), which indicates smaller discrepancies between the anticipated and observed values.

This shows that, based on the chosen feature sets, the OLS regression model delivers the most exact and accurate yield of both wheat estimation and wheat quantity.

The Random Forest (RF) regression model, which showed the peak R2 scores for the phonological indicators in wheat yield and wheat estimates, respectively, was also put up against the performance of the OLS regressor in our comparison. The ability of the RF regression model to handle data's non-linearity and provide accurate estimates of the dependent variables demonstrated promise. A cross-validation strategy was not used when fitting the data, which resulted in overfitting, which was particularly noticeable in the RF regressor's R2 scores above 0.9.

We stress the value of using cross-validation methods and model evaluation on a separate test dataset to reduce overfitting. This guarantees that the model's performance is not biased towards the training data and that it can make accurate predictions for data that has not yet been observed. We were able to accurately evaluate the regression models' performance by simple the data into training and test sets of 80% and 20%, respectively.

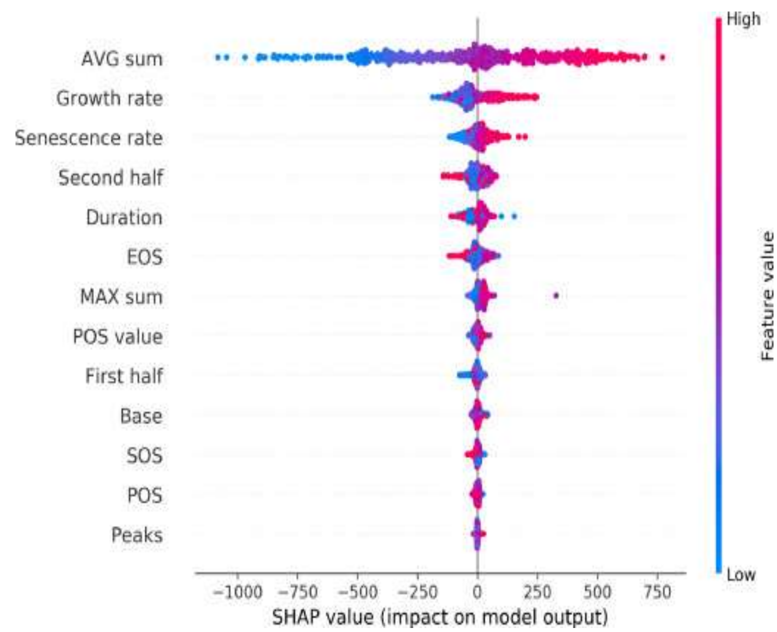


Figure 18. SHAP features position for the phonological variables

The trade-offs among model complexity and performance are highlighted by our findings. The OLS repressor fared better in terms of overall model fit and error estimation than the RF regression model, which showed the capability to capture non-linear correlations in the image data. The OLS regression model is an appealing option for sugarcane yield and sugar amount prediction jobs due to its clarity and interpretability. Overall, our results showed that the RF model of regression performed worse than the OLS regressor in terms of fitting and error estimation. In comparison to the RF model, the OLS model produced lower error estimates and offered an improved overall fit to the data. This suggests that the OLS model did a better job of capturing the correlations between the remote sensing-derived predictor variables and both wheat estimation and wheat dealer. Although the RF regressor demonstrated promise in addressing non-linearity, care must be taken to prevent overfitting. We make sure that model evaluation is more thorough and that predictions for data that hasn't yet been observed are accurate by using cross-validation methods and independent test datasets. The precise objectives of the study and the intended balance between model complexity and interpretability ultimately determine which regression algorithm is used.

However, we were also aware of the RF (Random Forest) regressor's capability to handle non-linear correlations. The RF model demonstrated potential for detecting intricate patterns and relationships in the data. Overfitting, when the model overly conforms to the train data and performs badly on unobserved data, must be avoided at all costs. We used cross-validation techniques and separate test datasets to provide a thorough assessment of the models and precise predictions for unobserved data. By repeatedly dividing the information into validation and training subsets and analyzing performance across various data partitions, cross-validation aids in determining the model's capacity for generalization. We can validate the model's performance on brand-new observations thanks to a distinct test dataset from the model's training and validation data (Cooper *et al.*, 2019).

The choice of a regression technique is based on the unique goals of the investigation and the preferred ratio of model complexity to interpretability. OLS regression offers simple explanations of the connections between variables because it is a linear model. RF regression, on the contrary, can capture non-linear trends but could be more complicated and difficult to understand. In the end, the regression algorithm selected should be in line with the goals of the study, taking into account the compromise between complexity of the model and interpretability, and assuring strong evaluation methods to generate correct predictions for unobserved data (Szulczewski *et al.*, 2018).

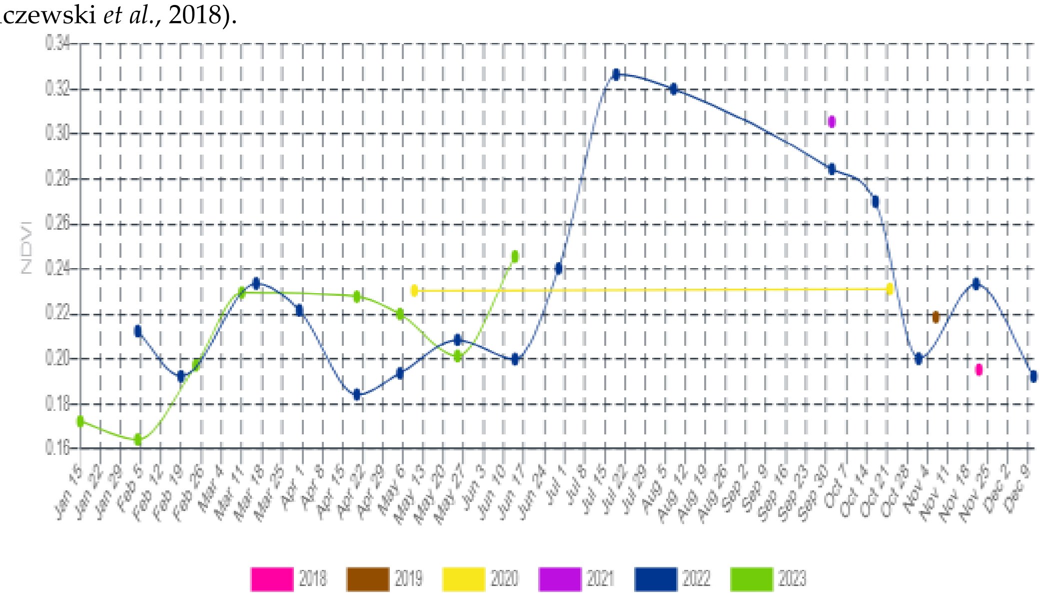


Figure 19. Corresponding Graph of last five year

The figure is a scatter plot titled "Random Forest - Actual vs Predicted wheat Biomass Yield." The x-axis shows the actual biomass of wheat yield, and the y-axis shows the predicted biomass of wheat yield. Blue dots represent all training data, and red dots represent test data. A black dashed line indicates ideal predictions (where actual equals predicted). The plot evaluates the model's accuracy, with data points closer to the dashed line indicating better predictions (Szulczewski *et al.*, 2018).

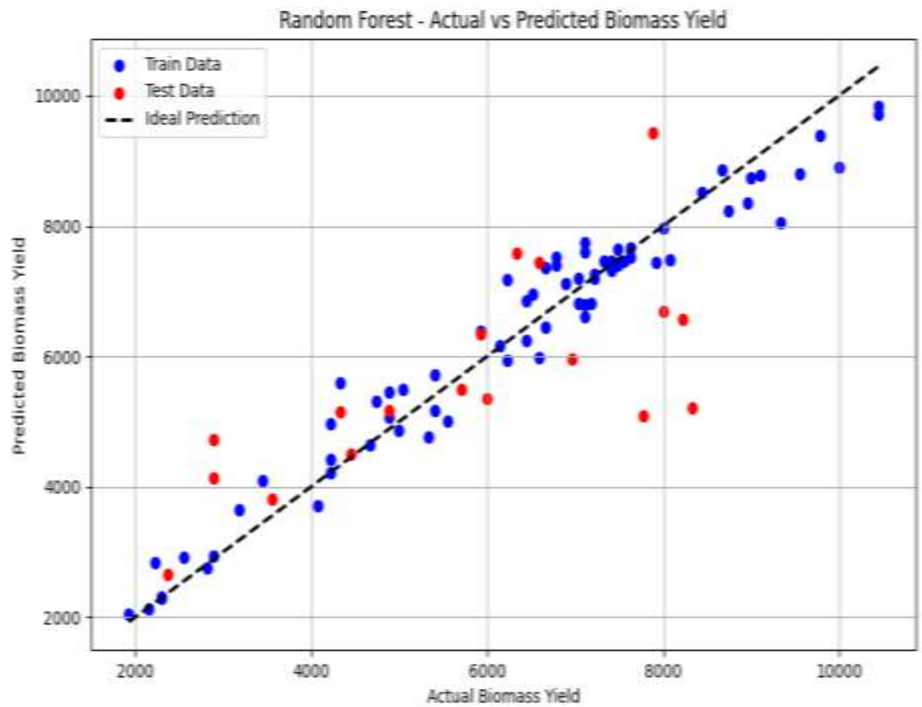


Figure 20. Validation for RF-based regression

The data points are scattered around the ideal prediction line, indicating the performance of the Random Forest model. The closer data points are to be dashed line, and better the model's predictions. This plot visually assesses the Random Forest model's ability to predict biomass yield by comparing the predicted values against the actual values for both training and test datasets (Li *et al.*, 2020).

Author Contributions:

Muhammad Rashid: Writing– original draft – review & editing, Software, Methodology, Conceptualization and Funding. Prof. Dr. Salman Qadri, Dr. Abdul Razzaq, Dr Sarfraz Hashim, Dr Habib-ur-Rehman and Ali Hamza: Writing Visualization, Software, Investigation, Data curation, Conceptualization.

Data availability: The complete dataset will be shared by the corresponding authors on request.

Acknowledgments: We like to acknowledge MNS University of Agriculture, Multan for supporting developing of this research.

Statements and Declarations

Competing Interests: We declare that we do not have any commercial or associative interest that represents a Conflict of interest in connection with the work and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously.

References

1. Ahmad, I., A. Singh, M. Fahad, M.M.J.C. Waqas and E.I. Agriculture. 2020. Remote sensing-based framework to predict and assess the interannual variability of maize yields in Pakistan using Landsat imagery. 178:105732.
2. Arab, S.T., R. Noguchi, S. Matsushita, T.J.R.S.a.S. Ahamed and Environment. 2021. Prediction of grape yields from time-series vegetation indices using satellite remote sensing and a machine-learning approach. 22:100485.
3. Aranguren, M., A. Castellón and A.J.R.S. Aizpurua. 2020. Wheat yield estimation with NDVI values using a proximal sensing tool. 12:2749.
4. Arshad, S., J.H. Kazmi, M.G. Javed and S.J.E.J.O.A. Mohammed. 2023. Applicability of machine learning techniques in predicting wheat yield based on remote sensing and climate data in Pakistan, South Asia. 147:126837.
5. Ashapure, A., J. Jung, A. Chang, S. Oh, J. Yeom, M. Maeda, A. Maeda, N. Dube, J. Landivar, S.J.I.J.O.P. Hague and R. Sensing. 2020. Developing a machine learning based cotton yield estimation framework using multi-temporal UAS data. 169:180-194.
6. Callier, P. and O.J.a.P. Sandel. 2021. Introduction to Artificial Intelligence. 60:18-20.
7. Canata, T.F., M.C.F. Wei, L.F. Maldaner and J.P.J.R.S. Molin. 2021. Sugarcane yield mapping using high-resolution imagery data and machine learning technique. 13:232.
8. Cao, J., Z. Zhang, Y. Luo, L. Zhang, J. Zhang, Z. Li and F.J.E.J.O.A. Tao. 2021. Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. 123:126204.
9. Chen, Y., W.S. Lee, H. Gan, N. Peres, C. Fraisse, Y. Zhang and Y.J.R.S. He. 2019. Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages. 11:1584.
10. Cooper, N., A. Panteli and N.J.a.E. Shah. 2019. Linear estimators of biomass yield maps for improved biomass supply chain optimisation. 253:113526.
11. Deng, Q., M. Wu, H. Zhang, Y. Cui, M. Li and Y.J.R.S. Zhang. 2022. Winter Wheat Yield Estimation Based on Optimal Weighted Vegetation Index and BHT-ARIMA Model. 14:1994.
12. Dubey, S.K., A. Gavli, S. Yadav, S. Sehgal and S.S.J.J.O.T.I.S.O.R.S. Ray. 2018. Remote sensing-based yield forecasting for sugarcane (Saccharum officinarum L.) crop in India. 46:1823-1833.
13. Feng, H., H. Tao, Y. Fan, Y. Liu, Z. Li, G. Yang and C.J.R.S. Zhao. 2022. Comparison of winter wheat yield estimation based on near-surface hyperspectral and UAV hyperspectral remote sensing data. 14:4158.
14. Feng, L., Y. Wang, Z. Zhang and Q.J.R.S.O.E. Du. 2021. Geographically and temporally weighted neural network for winter wheat yield prediction. 262:112514.
15. Gillard, X., V. Coppé, P. Schaus and A.A. Cire. Improving the filtering of branch-and-bound MDD solver. International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research, 2021. Springer, 231-247.
16. Hai, C., L. Wang, X. Chen, X. Gui, X. Wu and J.J.R.S. Sun. 2023. Estimating Maize Yield from 2001 to 2019 in the North China Plain Using a Satellite-Based Method. 15:4216.
17. Hamada, Y., C.R. Zumpf, J.F. Cacho, D. Lee, C.-H. Lin, A. Boe, E. Heaton, R. Mitchell and M.C.J.L. Negri. 2021. Remote sensing-based estimation of advanced perennial grass biomass yields for bioenergy. 10:1221.
18. Jahromi, M.N., S. Zand-Parsa, F. Razzaghi, S. Jamshidi, S. Didari, A. Doosthosseini and H.R.J.E.J.O.A. Pourghasemi. 2023. Developing machine learning models for wheat yield prediction using ground-based data, satellite-based actual evapotranspiration and vegetation indices. 146:126820.
19. Jamali, M., E. Bakhshandeh, B. Yeganeh and M.J.a.I.S.R. Özdoğan. 2023. Development of machine learning models for estimating wheat biophysical variables using satellite-based vegetation indices.
20. Karimli, N. and M.O.J.I.I.J.O.G.-I. Selbesoğlu. 2023. Remote Sensing-Based Yield Estimation of Winter Wheat Using Vegetation and Soil Indices in Jalilabad, Azerbaijan. 12:124.
21. Khan, M.S., M. Semwal, A. Sharma and R.K.J.P.A. Verma. 2020. An artificial neural network model for estimating Mentha crop biomass yield using Landsat 8 OLI. 21:18-33.
22. Li, B., X. Xu, L. Zhang, J. Han, C. Bian, G. Li, J. Liu, L.J.I.J.O.P. Jin and R. Sensing. 2020. Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. 162:161-172.
23. Li, R., D. Wang, B. Zhu, T. Liu, C. Sun, Z.J.F. Zhang and E. Security. 2023. Estimation of grain yield in wheat using source–sink datasets derived from RGB and thermal infrared imaging. 12:e434.

24. Liu, S., Z. Hu, J. Han, Y. Li, T.J.C. Zhou and E.I. Agriculture. 2022. Predicting grain yield and protein content of winter wheat at different growth stages by hyperspectral data integrated with growth monitor index. 200:107235.
25. Luo, S., X. Jiang, W. Jiao, K. Yang, Y. Li and S.J.A. Fang. 2022. Remotely Sensed Prediction of Rice Yield at Different Growth Durations Using UAV Multispectral Imagery. 12:1447.
26. Meraj, G., S. Kanga, A. Ambadkar, P. Kumar, S.K. Singh, M. Farooq, B.A. Johnson, A. Rai and N.J.R.S. Sahu. 2022. Assessing the yield of wheat using satellite remote sensing-based machine learning algorithms and simulation modeling. 14:3005.
27. Mountrakis, G., J. Im, C.J.I.J.O.P. Ogole and R. Sensing. 2011. Support vector machines in remote sensing: A review. 66:247-259.
28. Naqvi, S., M. Tahir, G. Shah, R. Sattar, M.J.a.E. Awais and E. Research. 2019. Remote estimation of wheat yield based on vegetation indices derived from time series data of Landsat 8 imagery. 17.
29. Nihar, A., N. Patel, R. Singh and A.J.J.O.T.I.S.O.R.S. Danodia. 2023. Satellite-Based Light Use Efficiency Model for Sugarcane Yield Estimation at Field Scale. 51:1357-1369.
30. Pandey, S., N. Patel, A. Danodia, R.J.T.I.a.O.T.P. Singh, Remote Sensing and S.I. Sciences. 2019. Discrimination of sugarcane crop and cane yield estimation using Landsat and IRS resourcesat satellite data. 42:229-233.
31. Pandya, P., N.K.J.J.O.W. Gontia and C. Change. 2023. Early crop yield prediction for agricultural drought monitoring using drought indices, remote sensing, and machine learning techniques.jwc2023386.
32. Parida, B.R. and A.K.J.R.S.I.E.S.S. Ranjan. 2019. Wheat acreage mapping and yield prediction using Landsat-8 OLI satellite data: a case study in Sahibganj Province, Jharkhand (India). 2:96-107.
33. Ronchetti, G., G. Manfron, C.J. Weissteiner, L. Seguini, L.N. Scacchiafichi, L. Panarello, B.J.C. Baruth and E.I. Agriculture. 2023. Remote sensing crop group-specific indicators to support regional yield forecasting in Europe. 205:107633.
34. Saad El Imanni, H., A. El Harti and L.J.A. El Iysaouy. 2022. Wheat Yield Estimation Using Remote Sensing Indices Derived from Sentinel-2 Time Series and Google Earth Engine in a Highly Fragmented and Heterogeneous Agricultural Region. 12:2853.
35. Saini, P., B. Nagpal, P. Garg, S.J.B.a.O.B. Kumar and Technology. 2023. Evaluation of Remote Sensing and Meteorological parameters for Yield Prediction of Sugarcane (Saccharum officinarum L.) Crop. 66:e23220781.
36. Segarra, J., J. González-Torralba, Í. Aranjuelo, J.L. Araus and S.C.J.R.S. Kefauver. 2020. Estimating wheat grain yield using Sentinel-2 imagery and exploring topographic features and rainfall effects on wheat performance in Navarre, Spain. 12:2278.
37. Shafique, M., M.A. Khan, W.S. Khan, W. Ahmad and S.J.J.O.N. Khan. 2017. Fabrication, characterization, and in vivo evaluation of famotidine loaded solid lipid nanoparticles for boosting oral bioavailability. 2017:7357150.
38. Sharma, N., S. Saxena, S. Dubey, K. Choudhary, S. Sehgal, S.J.T.I.a.O.T.P. Ray, Remote Sensing and S.I. Sciences. 2019. Analysis of Sugarcane Acreage and Yield Estimates Derived from Remote Sensing Data and Other Hybrid Approaches Under Fasal Project. 42:157-163.
39. Singh, B., S. Kumar, A. Elangovan, D. Vasht, S. Arya, N.T. Duc, P. Swami, G.S. Pawar, D. Raju and H.J.F.I.P.S. Krishna. 2023. Phenomics based prediction of plant biomass and leaf area in wheat using machine learning approaches. 14:1214801.
40. Singla, S.K., R.D. Garg and O.P.J.R.D.I.A. Dubey. 2020. Ensemble Machine Learning Methods to Estimate the Sugarcane Yield Based on Remote Sensing Information. 34.
41. Su, J., C. Liu, X. Hu, X. Xu, L. Guo, W.-H.J.C. Chen and E.I. Agriculture. 2019. Spatio-temporal monitoring of wheat yellow rust using UAV multispectral imagery. 167:105035.
42. Szulczewski, W., A. Żyromski, W. Jakubowski, M.J.R. Biniak-Pieróg and S.E. Reviews. 2018. A new method for the estimation of biomass yield of giant miscanthus (Miscanthus giganteus) in the course of vegetation. 82:1787-1795.
43. Tiruneh, G.A., D.T. Meshesha, E. Adgo, A. Tsunekawa, N. Haregeweyn, A.A. Fenta and J.M.J.P.O. Reichert. 2022. A leaf reflectance-based crop yield modeling in Northwest Ethiopia. 17:e0269791.
44. Wolanin, A., G. Mateo-García, G. Camps-Valls, L. Gómez-Chova, M. Meroni, G. Duveiller, Y. Liangzhi and L.J.E.R.L. Guanter. 2020. Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. 15:024019.

45. Zhao, S., Q. Lin, J. Ran, S.S. Musa, G. Yang, W. Wang, Y. Lou, D. Gao, L. Yang and D.J.I.J.O.I.D. He. 2020a. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. 92:214-217.
46. Zhao, Y., A.B. Potgieter, M. Zhang, B. Wu and G.L.J.R.S. Hammer. 2020b. Predicting wheat yield at the field scale by combining high-resolution Sentinel-2 satellite imagery and crop modelling. 12:1024.