ANNUAL RESEARCH REPORT 2023-2024

Program Leader

Dr. Syed Md. Abdullah Shiblee, CSO & Head
Chief Scientific Officer & Head

Dr. M. A. Monayem Miah
Ex. Chief Scientific Officer & Head

Agricultural Statistics and ICT Division
Bangladesh Agricultural Research Institute
Gazipur-1701

November 2024

Annual Research Report 2023-2024

Compiled &Edited by

Kowshik Kumar Saha, Agrl. Engineer
Dr. Mohammad Mukhlesur Rahman, SSO
Nur Mohammad, SSO
Mohammad Rasel, SO
Jamila Khatun Prioty, SO

Agricultural Statistics and ICT Division

Bangladesh Agricultural Research Institute
Gazipur-1701, Bangladesh

How to cite

Agricultural Statistics and ICT Division (2023). *Annual Research Report 2022-2023*. Bangladesh Agricultural Research Institute, Gazipur-1701.

Editorial Committee

Dr. Syed Md. Abdullah Shiblee, CSO & Head
Kowshik Kumar Saha, Agrl. Engineer
Dr. Mohammad Mukhlesur Rahman, SSO
Nur Mohammad, SSO
Mohammad. Rasel, SO
Jamila Khatun Prioty, SO

Compilation

Kowshik Kumar Saha, Agrl. Engineer
Nur Mohammad, SSO

Cover Design

Mohammad Mukhlesur Rahman, SSO
Nur Mohammad, SSO
Kazi Saidur Rahman, SSO
Mohammad Rasel, SO

Published by

Dr. Syed Md. Abdullah Shiblee
Chief Scientific Officer & Head
Agricultural Statistics and ICT Division
Bangladesh Agricultural Research Institute,
Gazipur-1701, Phone: +88-0249270129
E-mail: cso.asict@bari.gov.bd
Website: www.bari.gov.bd
www.baripmis.org/web/asict

Published on

November 2024

List of Contributors

Dr. Syed Md. Abdullah Shiblee, CSO & Head
Kowshik Kumar Saha, Agrl. Engineer
Dr. Mohammad Mukhlesur Rahman, SSO
Nur Mohammad, SSO
Dr. Zobaer Akond, SSO
Kazi Saidur Rahman, SSO
Istiak Ahmed, SO
Mohammad Rasel, SO
Jamila Khatun Prioty, SO
Md. Shakil Hossain, SO

PREFACE

The application of Information and Communications Technology (ICT) in agriculture is increasingly becoming important. The ICT revolution has brought huge implications for both social and economic development in the world and obviously in Bangladesh as well. The use of ICT in agriculture ranges from advanced modern technologies, such as GPS navigation, satellite communication, earth observation satellites (EOSs), unmanned aerial vehicles (UAVs), agricultural robots, internet of things (IoTs), data science, machine learning, cloud computing, and mobile &web-based apps to older technologies such as radio and television. However, the stakeholders still lack basic communication infrastructure in accessing crucial information to make timely decisions. The application of ICT in agriculture generates possibilities to resolve problems of stakeholders and also to promote agricultural production by providing scientific information timely and directly to farmers.

The importance of ICT was rightly recognized by BARI and a dedicated cell was constituted with competent scientists from different disciplines. Later, the Agricultural Statistics section was separated from the Agricultural Economics Division and renamed as the Agricultural Statistics and ICT (ASICT) section to take overall responsibilities of the ICT Cell. Finally, ASICT Division evolved in 2012 through the implementation of the project entitled "strengthening of information and communication technology (ICT) and biometrical facilities at BARI".

ASICT Division has been active both in support services and research dimensions since its inception. For ICT service and modern ICT-related research and development, the ASICT division is maintaining partial network connectivity, e-governance, e-agricultural, and different laboratories (MIS Lab, biometry lab, etc.) to ensure clients' (scientists and end-users) services. The agro-environmental remote sensing and modeling (ARSAM) laboratory of this division is performing applied research to harness the power of modern remote sensing and geospatial technologies in promoting sustainable cropping intensification. The award-winning BARI mobile and interactive web application **কৃষি প্রযুক্তি ভাণ্ডার** has been playing a vital role in disseminating modern agricultural technologies to the doorsteps of farmers. GeoMango, a semi-automated satellite-based mango orchard mapping system for the greater Rajshahi region developed at ARSAM Lab shows the power and potential of satellite-based crop mapping and monitoring opportunities for Bangladesh.

Indeed, I am happy to see that the Annual Research Report 2023-2024 of the ASICT Division has been published. This report covers the activities of research, outreach, and ICT services rendered inside BARI and to other stakeholders.

I gratefully acknowledge the contribution of all the scientists and staff in conducting research and other activities and successfully preparing the report.

Dr. Syed Md. Abdullah Shiblee
Chief Scientific Officer and Head
Agricultural Statistics and ICT Division

Contents

Agricultural Statistics and Information & Communication Technology (ASICT) Division at a Glance

The ASICT division initially served as a Statistical Section of the Agricultural Economics Division of Bangladesh Agricultural Research Institute (BARI) since 2007. Afterward, the section worked as a new name titled "Information and Communication Technology (ICT) section under the Training and Communication Wing up to February 18, 2008. Once more, BARI authority changed this section to Agricultural Statistics and Information & Communication Technology (ASICT) Section by issuing an order on 19th February 2008. Finally, the ASICT section had been updated to a new division as per the decision of the 49th Board of Management meeting of BARI on 9th April 2012. Currently, the division performs its activities under the Training and Communication Wing of BARI.

ASICT division comprises two parts: Agricultural Statistics and ICT. The division is headed by a Chief Scientific Officer (CSO) who is assisted by a number of Senior Scientific Officers (SSO), Scientific Officers (SO), System Analyst, Programmer, Assistant Programmer, Assistant Maintenance Engineer, and supporting staff. Concerned scientists of both parts conduct different types of researches relevant to Agricultural Statistics and ICT. In addition, the scientists of both parts are providing agricultural statistics and ICT-related support services to BARI scientists in order to facilitate agricultural research and development. However, ASICT is constantly keeping its vital role for BARI as a whole.

DETECTION OF DIFFERENCES IN VEGETATION AND CHLOROPHYLL CONTENT IN AGRICULTURAL FIELD USING UNMANNED AERIAL VEHICLES

M. MUKHLESUR RAHMAN[1], NUR MOHAMMAD[1], ISTIAK AHMED[1], M. A. MONAYEM MIAH[1] AND M. J. ALAM[2]

[1]ASICT Division, BARI, Gazipur and [2]Soil Science Division, BARI, Gazipur

Abstract

This reports a field-scale study to detect differences in the amount of vegetation and chlorophyll content of crops using an unmanned aerial vehicle (UAV) fitted with a multispectral camera. The purpose of this study, on the experimental farm of TCRC, Bangladesh Agricultural Research Institute. The normalized difference vegetation index (NDVI) and normalized difference red edge (NDRE) were obtained from five spectral band images (red, green, blue, infrared (NIR), and red edge (REDGE)) that were processed by software into a full image map. We used the image map obtained to analyze the farmland and identify variations in the greenness of plants. We compared two layers with different indices and indicated differences in vegetation activity for NDVI and NDRE. NDVI showed a visible green color wherever vegetation was present. With NDRE and NDCI we observed crops with low chlorophyll content, indicating nitrogen limitation in the leaves. These observations demonstrate the efficacy of using NDRE and NDCI as a sensitive index for monitoring chlorophyll content. Therefore, we propose that different indices may be most useful for different crops, plant density, and growth stages.

Introduction

Until recently, unmanned aerial vehicles (UAVs), unmanned aerial systems, and remotely piloted vehicles also often called drones, were mostly developed and used for military purposes. These systems are aircraft with remote controls. They are equipped with precision sensors, for example, inertial navigation systems, such as an accelerometer and a gyroscope, for recognizing the location of an aircraft in airspace. A microcomputer helps automate the flight process and simplifies navigation in piloting. Moreover, the control is accompanied by special software, which makes it easy to manage all systems. For a specific type of tasks, auxiliary sensors and cameras of different ranges are used (Boiarskii and Hasegawa, 2017).

The necessity to improve the quality and profitability of agricultural production stimulates farmers to introduce new technologies. In recent years, unmanned aerial vehicle (UAV) technology has become a routine tool for farm management. For example, UAV monitoring can be used to produce map data for early-season soil analysis, which is useful in planning seed planting. It is a helpful step in the evolution of agricultural management systems toward increased efficiency of land use (Whelan and McBratney, 2012). The use of NDVI in agriculture is beginning to develop rapidly, and the issue of introducing these technologies into the agriculture sphere is becoming urgent. Modern hardware, such as multispectral cameras, makes remote analysis more informative and significantly expands their range of applications (Boiarskii and Hasegawa, 2017). A multispectral camera is a useful tool for field monitoring, as it can provide data for use in crop forecasting, yield predictions, crop status mapping, and detecting weeds, diseases, and nutrient deficiency (Berni et al., 2009). Indices are defined as a ratio of the difference between the reflectance of different spectral bands, which provide different data layers. Decreasing photosynthesis rates and changes in leaf mesophyll are usually associated with decreasing reflectance of wavelengths within the NIR spectral range (Carter and Knapp, 2001).

This study describes the use of a UAV to analyze crop vegetation on an experimental farm. We compared two indices, the NDVI and NDRE and examined the indices for potential application in agronomy and further research. The purpose of this study was to identify poorly growing areas of vegetation that might require additional soil fertilizer and to visualize foliar nitrogen to indicate plant health.

Objectives

1. To identify crop classification and develop a digital map for selected crop fields
2. To generate a chlorophyll content map for selected crops
3. To identify particular chlorophyll deficiencies for selected crops

Materials and Methods

Study area was selected at Tuber Crop Research Center (TCRC) at BARI, Gazipur potato experiment field. This experiment field lies between 23°59'13.23"N latitude and 90°24'41.35"E longitude.

Fig. 1. The study site at TCRC experimental Field

Data collection procedure: We used the UAV model Phantom 4 multispectral, manufactured by DJI, China (Fig. 2a). The DJI Phantom 4 Multispectral consolidates the process of capturing data that gives insight into crop and vegetative management. This system is completely integrated with the same powerful performance that DJI is known for, this includes 27 minutes of flight time and ocusync transmission system.

The UAV had a multispectral RedEdge camera that captured five spectral bands (red: 650 ± 16 nm, green: 560 ± 16 nm, blue: 450 ± 16 nm, NIR: 840 ± 26 nm, and red edge: 730 ± 16 nm). The additional bands of the camera are REDGE and NIR. The NIR band is in the spectral region behind the red band and is useful in detecting the state of plant health. Another constituent spectral region is the REDGE. This band is located between the red band and the NIR band. Plants increase the reflection coefficient between the red and NIR region, which leads to a sharp increase in the reflection coefficient through the REDGE band. Different combinations of the bands allowed us to observe different analytical layers (Fig. 2b).

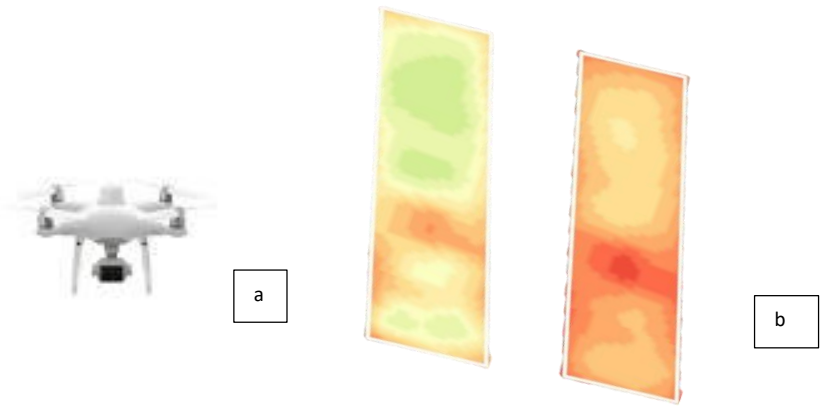


Fig. 2. UAV and examples of data layers (NDVI and NDRE)

In this study we used NDVI, NDRE and NDCI indices (see Results for definitions of the indices) to evaluate crop vegetation activity and chlorophyll content. We tested the UAV at various altitudes and speeds. In this study, the UAV flew at 60 m altitude at 18 km/h. Each UAV mission was planned using DJIGS Pro software on an Apple iPad (Fig. 3). Mission time was 05 minutes.

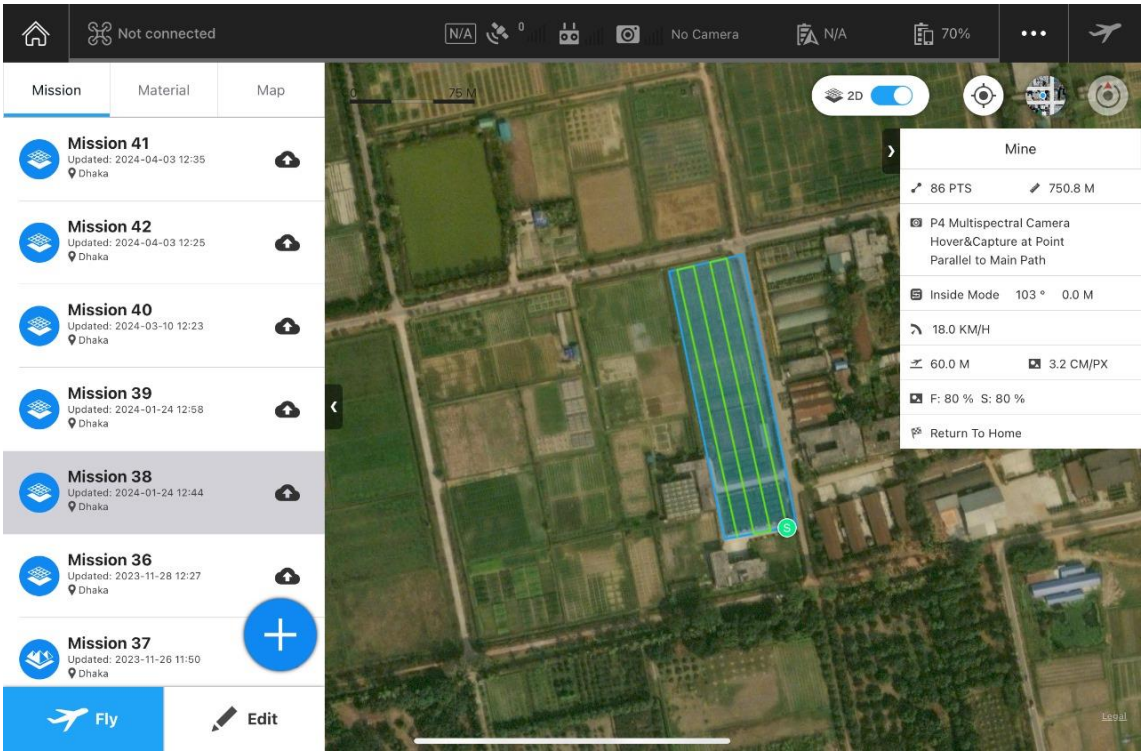


Fig. 3.Example of mission planning software display

The camera focal size was 5.5 × 4.8 mm, and the ground sampling distance from the 60 m altitude flight was 3.2 cm/pixel. Front overlap and side overlap of successive flight paths were set at 80% and 80%, respectively. We obtained 516 raw images in five-band or 86-point images (Fig. 4).We selected high overlap to increase the precision and quality of the output data.

There are numerous options for DJIGS Pro software, many of which are license vision applications that can be used. The preferred application will vary depending on the characteristics of the drone and the sensors, but, as a rule, the user has several compatible options, between used tools and software.

The UAV images were processed in Agisoft software into stereo pairs using photogrammetry algorithms.

The stereo pairs consisted of images with different geo-locations. The stereo image pairs were used to generate a point cloud (2D points), from which we derived an orthophoto plan. There are various indices and area calculations for which we use QGIS and ArcGIS software.

Fig. 4.Example of raw image of study site

We processed the captured images using Agisoft software to obtain five-band maps. Red, green and blue (RGB) colors cannot be recognized by the human eye from the mixed five-band map. The raw five-band map is represented as a gray-scale image (Fig. 5).

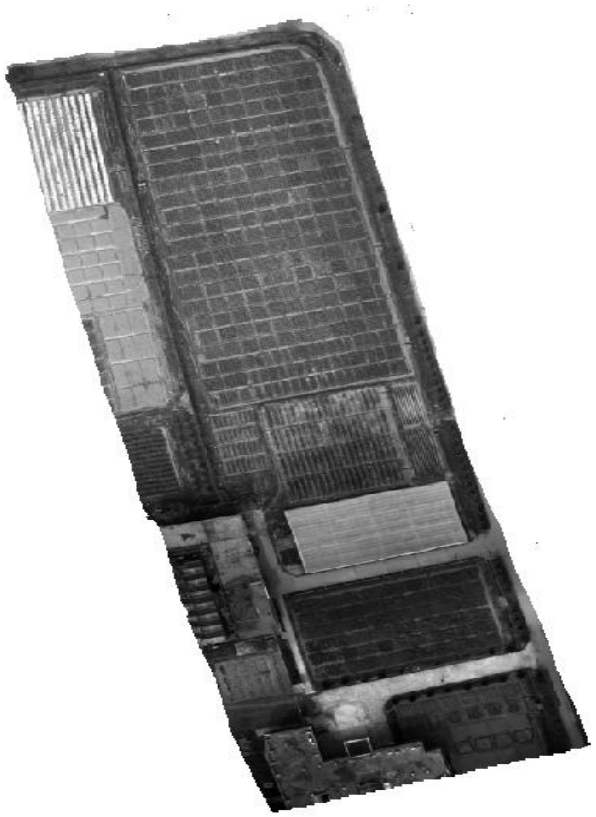


Fig. 5.Example of five-band map of Study area TCRC Field at BARI, Gazipur

Results and Discussion

Several different index values have been used in agriculture to analyze crop health at certain growth stages, the presence of weeds and crop moisture status. The use of remote sensing data offers the potential to improve the prediction of crop diseases and weeds (Homolova et al., 2013). Indices are used to contrast the stronger chlorophyll absorption of red wavelengths with the higher reflectance of NIR wavelengths for NDVI, and the red edge (REDGE) wavelength to indicate chlorophyll content (Carlson and Ripley, 1997).

A formula must be applied to the raw data to obtain agronomically useful information. We used QGIS (free version) software to calculate NDVI and NDRE indices. NDVI is defined as a ratio of the difference between the NIR band and the red band versus the sum of the two bands (Eq. 1). NDRE is defined as the same ratio as for NDVI, but using REDGE instead of the red band (Eq. 2) (Tucker, 1979). Another index NDCI is also used Red and REDGE both bands (Eq. 3).

$$\mathrm{NDVI} = \frac{\mathrm{NIR}-\mathrm{RED}}{\mathrm{NIR}+\mathrm{RED}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

$$\mathrm{NDRE} = \frac{\mathrm{NIR}-\mathrm{REDGE}}{\mathrm{NIR}+\mathrm{REDGE}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

$$\mathrm{NDCI} = \frac{\mathrm{REDGE}-\mathrm{RED}}{\mathrm{REDGE}+\mathrm{RED}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

We compared three layers with different indices and indicated differences in vegetation activity for NDVI and NDRE. We found a uniform distribution of NDVI wherever vegetation was present.

NDVI provides an indicator of the amount of healthy vegetation, measured on a scale from 0 to 1, where the closer to the index 1, the higher the amount of vegetation (Fig. 6). On our field we observed crops without visible depressed greenery. This index is most useful at early and middle growth stages: it indicates the health of crops (such as poor growth or crop diseases) as measured by the reflectance of NIR and red bands, which are both commonly used to indicate the greenness of actively growing leaves.

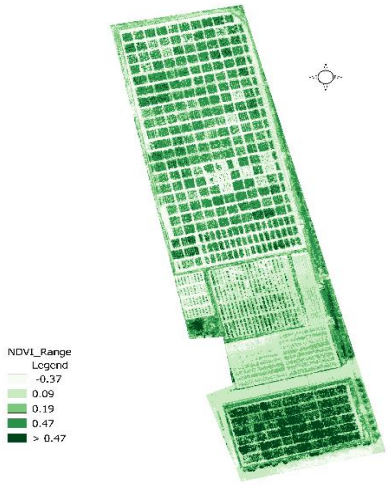


Fig 6: NDVI Layer

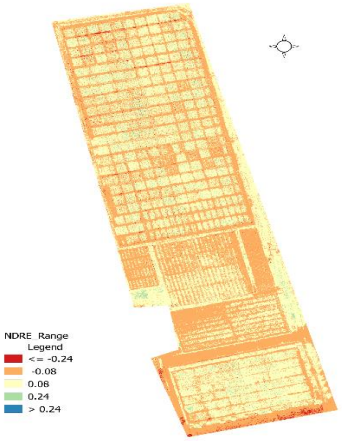


Fig 7: NDRE Layer

The REDGE spectral band, as provided by the Mica Sense Company in their Red Edge camera, was used in our research to indicate chlorophyll content (Fig.7). NDRE is considered the better version of NDVI as it is much more sensitive. This remote sensing technique is ideal for crops in the middle and later stages of growth. The basis of this sensor is an accumulation of chlorophyll. The REDGE wavelength of 730 ± 16 nm provides a sensitive indicator of chlorophyll content in leaves (how green a leaf appears), the density of leaves, and soil background effects. NDRE image we can see lower readings where less Nitrogen was applied. Chlorophyll has maximum absorption in the red band wave, and so red light cannot penetrate beyond a few layers of leaf cells. Leaves are more translucent to red-edge light than red light: the REDGE wavelength penetrates a leaf much more deeply than red or blue wavebands. Therefore, NDRE is more suitable for middle and late growth stages, when crops have accumulated high concentrations of chlorophyll in the leaves and red light will penetrate poorly.

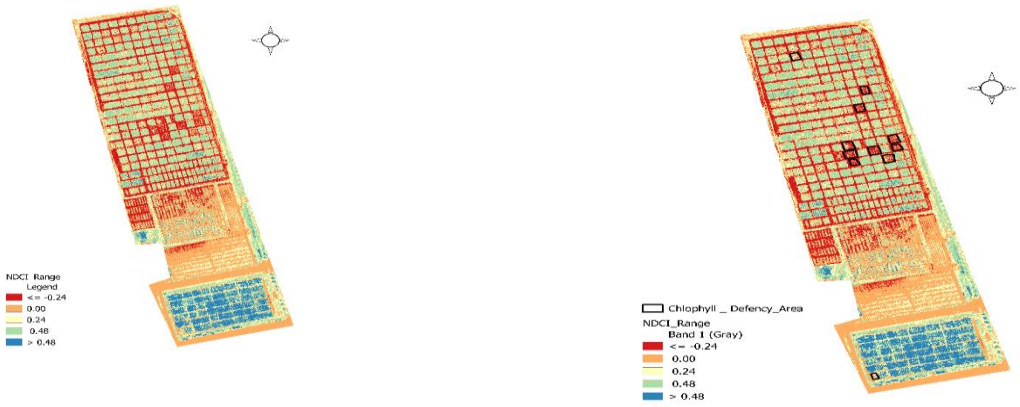


Fig 8: NDCI Layer and Chlorophyll deficiency area ma

Normalized difference chlorophyll index (NDCI) a novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. The chlorophyll index is a key metric for understanding crop health and assessing nitrogen status, making it an important tool for growers and researchers alike. Data findings from chlorophyll index measurements can be used to improve crop yields, identify plant stressors, and compare different phenotypes and varieties (Boiarskii & Hasegawa, 2019). We used the NDCI index in the study area and identified the chlorophyll deficiency area (Fig 8).

In this study, we compared three useful indices for detecting differences in the observed data. NDVI is most effective when used to analyze large areas of land for vegetation density and how green a crop appears. It enables evaluation of the health of crops and the effectiveness of cultivation. NDRE and NDCI enable visualization of chlorophyll content in leaves. It allows identification of areas that might benefit from chemical analysis of the soil or that may have additional fertilizer requirements (especially nitrogen) or may have had improper application of fertilizer. In future studies, we will compare data obtained by UAV with a handheld chlorophyll meter (SPAD) at ground level.

Table 1: Polygon ID-wise chlorophyll deficiency area

Sl.no	Shape	Area (Sft)
1	Polygon	113.20
2	Polygon	97.67
3	Polygon	71.38
4	Polygon	64.65
5	Polygon	44.02
6	Polygon	73.28
7	Polygon	90.76
8	Polygon	82.18
9	Polygon	117.88
10	Polygon	116.48
11	Polygon	111.94

We calculated the Normalised Difference Chlorophyll Index (NDCI) and found the eleven chlorophyll deficiency polygon. The maximum chlorophyll deficiency area was found at 117.88 sft at polygon no. 9, and the minimum was found at 44.02 sft at polygon no. 5. The total chlorophyll deficiency area was found at 983.44 sft.

Conclusion

UAVs are useful for field surveying and high-resolution monitoring. They produce precise map data for early chlorophyll deficiency areas. In this study, we used NDVI, NDRE, and NDCI to analyze an experimental site and detect areas of poorly growing vegetation. We compared the three indices to determine the effectiveness of using different wavelengths for agronomic monitoring. The data layers we obtained and processed will be communicated to farmers to help them evaluate areas where it may be necessary to conduct a chemical analysis of the soil and additional application of fertilizers. We found the chlorophyll deficiency area through the NDRE and NDCI indices. In the future, we will use the spry drone where chlorophyll deficiency is presented. For those areas, nutrients will be used.

References

1.Berni J, Zarco-Tejada PJ, Suarez L and Fereres E (2009). Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *IEEE Trans Geosci Remote Sens, 47(3):722–7.*

2. Boiarskii B and Hasegawa H (2017). Technologies of cartography and field monitoring using unmanned aerial vehicles (UAV). Actual problems of agroindustrial complex: a view of young researchers, *Smolensk State Agricultural Academy, pp. 213–216.* See also URL https://elibrary.ru

3. Carlson TN and Ripley DA (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sens. *Environ., 62: 241–252.*

4. Carter GA and Knapp AK (2001). Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *Am. J. Bot., 88: 677–684*.

5. Homolova L, Maenovsky Z, Clevers J, Garcia-Santos G and Schaepman ME (2013). Review of optical-based remote sensing for plant trait mapping. *Ecol. Complexity, 15: 1–16.*

6. Tucker CJ (1979). Red and photographic infrared linear combinations for monitoring vegetation.

7. Whelan BM and McBratney, AB (2012). Downscaling for site-specific crop management needs? In: B. Minasny, BP Malone and AB McBratney (eds), Digital Soil Assessments and Beyond, *Taylor and Francis, pp. 353–35.*

8. Boiarskii, B., & Hasegawa, H. (2019). Comparison of NDVI and NDRE indices to detect differences in vegetation and chlorophyll content. *J. Mech. Contin. Math. Sci, 4, 20-29.*

YIELD PREDICTION OF MUSTARD CROP BY USING SATELLITE BASED REMOTE SENSING TECHNIQUE IN BANGLADESH

M. MUKHLESUR RAHMAN[1], NUR MOHAMMAD[1], ISTIAK AHMED[1], M. A. MONAYEM MIAH[1] AND SUMAN BISWAS[2]

[1]ASICT Division, BARI, Gazipur and [2]Islamic University, Kushtia

Abstract

Mustard (*Brassica spp*.) is one of the important oilseed crops which has potential demand as the preferred edible oil for the majority of people of Bangladesh. The accurate estimation of both harvested area and yield of mustard are equally important in ensuring the accurate determination of their product. The traditional measurement of these statistics is time-consuming, tedious, and costly. Whereas remote sensing techniques are being used to easily measure these statistics at high spatial and temporal resolutions. Therefore, an attempt was made to predict the mustard yield through satellite-based remote sensing techniques before its harvesting. To get this done, the high spatial-temporal resolution Satellite imageries of Sentinel 2A (~10m) and Landsat 8 (~30m) were acquired for the three study locations after setting the experiment in a farmer's field for three consecutive mustard growing seasons of 2022-23, 2023-24 and 2024-25. The mean Normalized Difference Vegetation Index (NDVI) was extracted from the maximum NDVI-produced temporal satellite imageries within the growing season from 20 farmer's mustard fields of each study location. The first year's results revealed that in most cases the yield was maximum for the field where the NDVI values were not maximum and vice-versa. However, the relationships of the extracted mean NDVI and yields will be established using the classical linear regression model where the model will be developed using the first two years' data and will be validated using the data of the last study period.

Introduction

Mustard is the most dominant oilseed crop in Bangladesh and has experienced an expansion in area, production, and yield over time while facing the fierce competition for land for the production of cereals, e.g., rice, wheat, and maize. Mustard is a cold-loving *Rabi* crop that grows during October-February usually under rain-fed and low-input conditions in Bangladesh. Crop production estimates are generally portrayed as the product of two components: area (to be) harvested and (expected) yield per unit area. The accurate forecasting/estimation of both harvested area and yield are equally important in ensuring the accurate determination of their product.

Remote sensing images endow with entrance to spatial information at a global scale; of features and phenomena on Earth on an almost concurrent basis (El-Telbany et al., 2019). It has the capability of the recognition of crop classification, crop growth monitoring, and crop yield estimation (Mohd et al., 1994). In terms of how well it works for field research, it can find and give data on spatial variation and approval (Schuler, 2002). Remote sensing technology can play a significant role in the agricultural sector to provide timely and accurate information (Atzberger, 2013).

The agricultural application of satellite RS technology requires quantitative processing of satellite RS data with high accuracy and reliability. For yield prediction and estimation of crops, it is necessary to achieve a very high accuracy and reliability. This is the reason why even after a relatively long time (more than 20 years) no routine yield estimation method for a wide range of operational applications has been developed. However, a large amount of work produced important steps forward in this field. Naturally, most of the experiments and research concentrated on obtaining quantitative relation between satellite (or airborne) RS data and crop yields and used two main types of possible general strategies (Hamar et al., 1988a,b).

These indices are commonly used for drought detection, monitoring excessive soil wetness, assessment of weather impacts on vegetation, and evaluation of vegetation health and productivity (Unganai and Kogan, 1998; Kogan, 2001; Kogan, 2002; Kogan et al., 2003; Singh et al., 2003). The NDVI data were used extensively in vegetation monitoring, crop yield assessment, and forecasting (Hayes et al., 1982; Benedetti and Rossinni, 1993; Quarmby et al., 1993).

In Bangladesh researchers and extension personnel collect most of its crop statistics through the compilation of representative field sample data, which takes a lot of time and leaves out information about the spatial distribution of field variability. Nevertheless, this traditional technique is time-consuming, tedious, and costly. Currently, remote sensing techniques are being used to measure these statistics at high spatial and temporal resolutions. The application of remote sensing in estimating agricultural performance indicators is increasing as it offers a time and cost-effective reproducible method for measurement that can cover larger physical areas as compared to in-situ methods (Sadras et al., 2015). Therefore, an attempt was made to predict the mustard yield through satellite-based remote sensing techniques before its harvesting. The specific objectives of this study are given below.

Objectives

1. To create a map of selected mustard fields using satellite images; and
2. To estimate mustard yield by satellite based remote sensing technique.

Materials and Methods

Study area selection: Three promising mustard growing areas namely Sirajganj, Tangail, and Manikganj districts were purposively selected for this study. These three districts covered 15.40%, 8.17%, and 5.97% of the mustard area and shared 15.28%, 7.03%, and 5.53% of the total mustard production in Bangladesh, respectively (BBS, 2022). The Ullabara Upazila of Sirajganj district lies between 24°19′ N latitude and 89°34′ E longitude, Kalihati Upazila of Tangail district lies between 24°22′ N latitude and 89°00′ E longitude, and Daulatpur Upazila of Manikganj district lies between 23°57′ N latitude and 89°50′ E longitude (Figure 1).

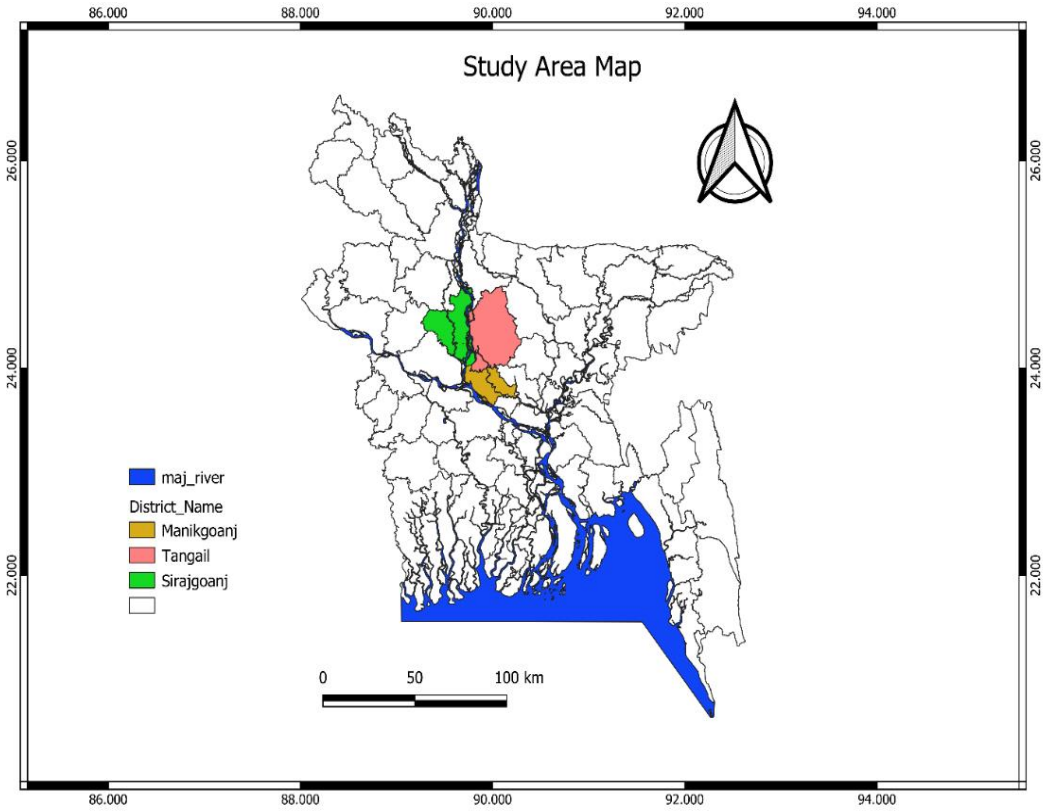


Figure 1: Study area map of Bangladesh

Data collection procedure: In this study, a total of 60 farmers' mustard fields taking 20 fields from each selected location were identified for the three mustard growing seasons 2022-2023, 2023-2029, and 2024-2025 by interviewing the sampled farmers. The identified mustard fields of Ullapara, Kalihati, and Daulatpur Upazila have been indicated in Figure 2, Figure 3, and Figure 4 respectively. It is important to mention here that each of the sampled farmers is committed to growing the same variety of mustard on the same identified plots for the next two succeeding years for the sack of this research. First year's (2022-23) yield data were acquired through direct interviewing sampled farmers from the selected mustard fields of the aforesaid locations.

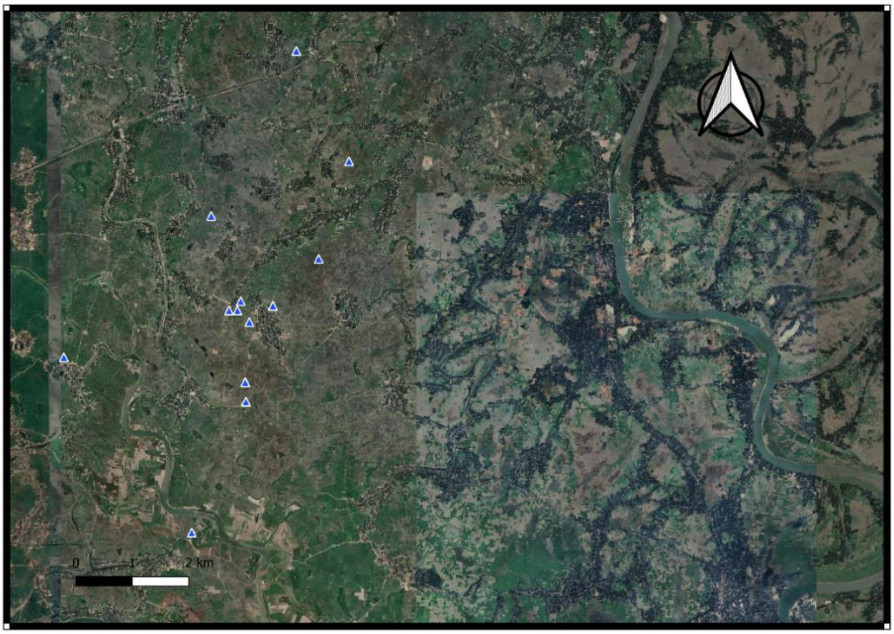


Figure 2: Locations of the selected mustard fields (blue triangles) over Ullapara Upazila, Sirajganj district

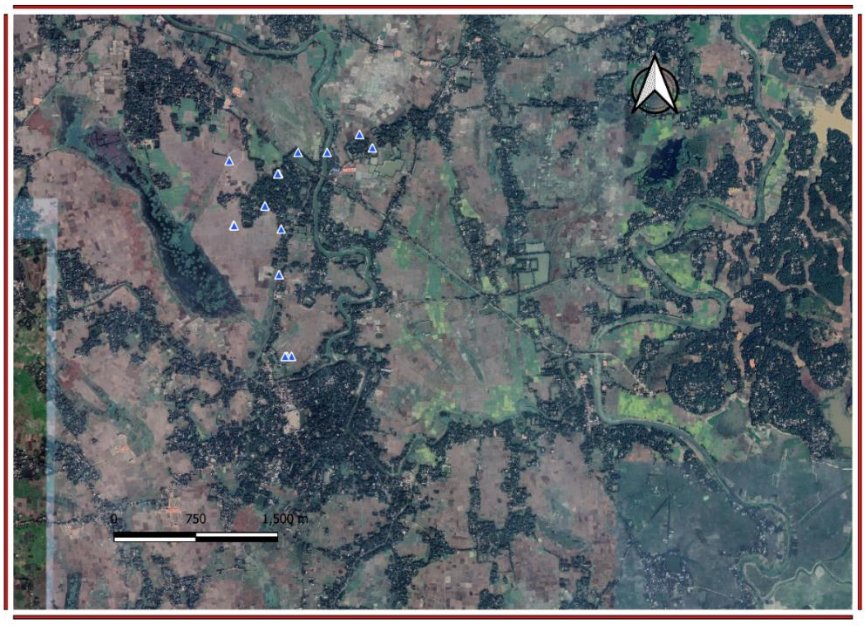


Figure 3: Locations of the selected mustard fields (blue triangles) over Kalihati Upazila, Tangail district

Figure 4: Locations of the selected mustard fields (ping triangles) over Daulatpur Upazila, Manikganj district

Use of satellite images: In this study, the satellite images of Sentinel 2A and Landsat 8 (OLI) were used, and those were compiled from ESA (European Space Agency) and the USGS Earth Explorer website. At Sentinel 2A, the span of 13 spectral bands, from the visible and the near-infrared to the shortwave infrared at different spatial resolutions ranging from 10 to 60 meters on the ground, takes global land monitoring to an unprecedented level (Melillos *et al.,* 2019). MSI (Multispectral Imager) covers 13 spectral bands (443–2190 nm), with a swath width of 290 km and a spatial resolution of 10 m (four visible and near-infrared bands), 20 m (six red edge and shortwave infrared bands) and 60 m (three atmospheric correction bands). Whereas, Landsat 8 (OLI) are Sun-synchronous satellites staying at an altitude of 705 km above the earth with a 16-day repeat cycle (Claverie *et al.*, 2018). Landsat 8 has two types of sensors, namely the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS). The OLI sensor provides nine spectral bands, including a pan band, and TIRS provides two spectral bands. However, only Red and NIR spectral bands were used for the study purposes.

Image preprocessing technique: For Landsat data, the Raw Digital Numbers (DN) were changed to Top-of-Atmosphere (TOA) reflection values after reference (Simonetti *et al.*, 2015). Generally, remote sensing data adopted from satellite sensors is greatly influenced by several factors, such as atmospheric scattering and absorption, sensor calibration, and also by data processing procedures (Teillet, 1986). For that, two techniques were used to preprocess the satellite images: (1) radiometric calibration and (2) atmospheric correction (Wang *et al.,* 2019). Radiometric calibration means a set of correction techniques that are associated with correction for the sensitivity of satellite sensors, topography and sun angle, atmospheric scattering, and absorption (Kim and Elman, 1990). The open source-based Quantum Geographic Information System (QGIS) software 3.22.10 version allows a plugin, and the plugin gives a tool for atmospheric correction, which is known as dark object subtraction (DOS-1) level 1. In this study, this tool was employed on the radiometrically calibrated images to minimize its atmospheric scattering effect. DOS-1 searches for every pixel in a band and finds the dark value. Scattering is done by subtracting this value from each pixel in the band. In this study, two techniques were followed to

download and process the Sentinel-2A images. The satellite images of Sentinel 2A and Landsat 8 (OLI) were compiled by the ESA European Space Agency and the USGS Earth Explorer website. Then, the free and open-source QGIS 2.18.10 version software together with the Semi-automatic Classification Plugin (SCP) was used to preprocess the imagery (Congedo, 2016). The benefit of using the SCP is that the user may preview, download, and correct Sentinel-2 images using the same interface for each date and each tile. The vegetation indices can then be calculated, saved, and compared to other dates, all in the same QGIS environment.

Table 1: Model development using Landsat 8 (OLI) single image of the growing season

Location	Ullapara	Kalihati	Daulatpur
Growing Season	2022-23	2022-23	2022-23
IAD	03/01/2023	03/01/2023	01/03/2023
DAP	67	60	63
Growing Season	2023-24	2023-24	2023-24
IAD	28/01/2024	28/01/24	28/01/24
DAP	69	71	68

Note: IAD= Image acquisition date; DAP= Days after plantation

Table 2: Model development using Sentinel 2A single image of the growing season

Location	Ullapara	Kalihati	Daulatpur
Growing Season	2022-23	2022-23	2022-23
IAD	30/12/2022		
DAP	63		

Note: IAD= Image acquisition date; DAP= Days after plantation

Calculation of Normalized Difference Vegetation Index (NDVI)

The NDVI is a dimensionless ratio between surface reflectance from the near-infrared and red bands of the spectrum which can be calculated as the following equations of Table 3.

Table 3: Vegetation index using Sentinel-2 and Landsat-8 satellite images

Index	Sentinel 2A	Landsat 8
NDVI	$\frac{B_8 - B_4}{B_8 + B_4}$	$\frac{B_5 - B_4}{B_5 + B_4}$
	For Sentinel 2A, B_8=NIR Band, B_4=RED Band	For Landsat 8, B_5=NIR Band, B_4=RED Band

Results and Discussion

Mustard Yield of Farmers' Fields and Its Corresponding NDVI

Mustard yield data and NDVI values from Sentinel 2A and Landsat 8 satellite images (Figure 5) were collected from selected 20 farmers' fields of three different locations, i.e. Ullapara, Kalihati, and Daulatpur Upazila for the mustard growing season 2022-23. Yield data of the farmers' fields were collected through direct inspection of the fields. The results have been presented in Table 4, Table 6, and Table 8 along with their corresponding NDVI values of satellite images for the mustard growing season of 2022-23.

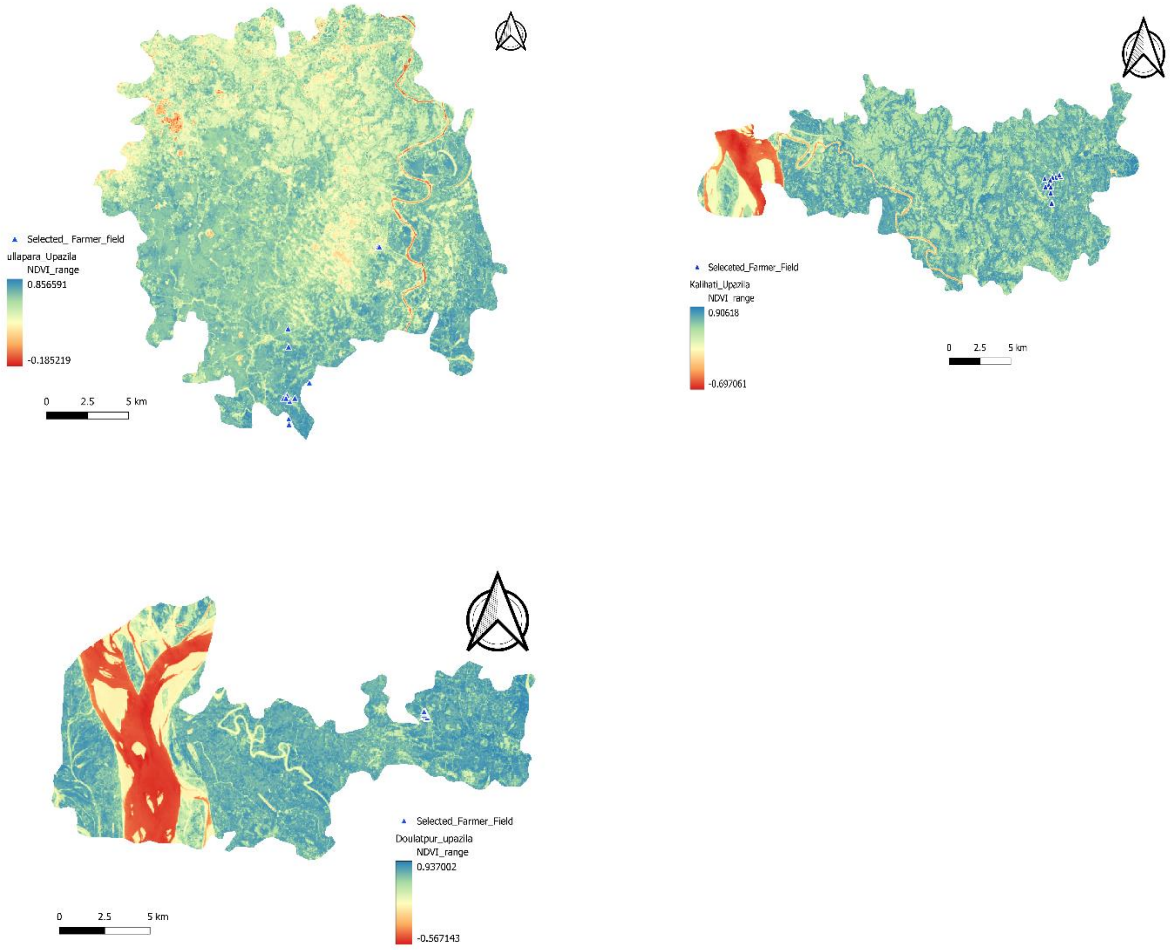


Figure 5: Spatial distribution of the NDVI from satellite images of Ullapara, Kalihati, and Daulatpur Upazilas for 2022-23 mustard growing seasons respectively

Mustard yield data and NDVI values from Sentinel 2A and Landsat 8 satellite images (Figure 6) were collected from selected 20 farmers' fields in three different locations, i.e. Ullapara, Kalihati, and Daulatpur Upazila for the mustard growing season 2023-24. Yield data of the farmers' fields were collected through direct inspection of the fields. The results have been presented in Table 5, Table 7, and Table 9 along with their corresponding NDVI values of satellite images for the mustard growing season of 2023-24.

Figure 6: Spatial distribution of the NDVI from satellite images of Ullapara, Kalihati, and Daulatpur Upazilas for 2023-24 mustard growing seasons respectively

Table 4: NDVI values of satellite images and yields of corresponding Farmer's fields at Ullapara, Sirajganj during the season of 2022-23

Farmer's Field No.	Longitude	Latitude	NDVI	Observed Yield (t/ha)
1	89.5671900	24.31072	0.797	1.98
2	89.5024165	24.19310	**0.851**	2.14
3	89.5679702	24.31071	0.672	1.73
4	89.5117804	24.21701	0.658	1.77
5	89.5117383	24.21701	0.711	1.98
6	89.5118608	24.21394	0.697	1.98
7	89.5024165	24.19320	0.751	1.81
8	89.5089597	24.22843	0.710	1.85
9	89.5024265	24.19320	0.562	1.48
10	89.5089097	24.22842	0.810	2.17
11	89.5109755	24.22991	0.751	1.74

12	89.5124510	24.22657	**0.478**	1.07
13	89.5103853	24.22842	0.718	1.80
14	89.4135483	24.26348	0.698	1.88
15	89.5679720	24.31071	0.694	1.98
16	89.5679720	24.31072	0.754	2.14
17	89.5124347	24.25667	0.789	2.29
18	89.5123454	24.26657	0.740	1.85
19	89.5156795	24.22833	0.532	1.50
20	89.5245481	24.23664	0.678	1.80

Table 4 shows that the NDVI values were a maximum of 0.851 for the 2nd field and a minimum of 0.478 for the 12th field at Ullapara Upazila of Sirajganj district. On the contrary, the yield was maximum at 2.29 t/ha for the 17th field, and minimum at 1.07 t/ha for the 12th field. That means the yield was maximum for the field where the NDVI values were not maximum. But the minimum yield was found for the field where the NDVI values were the lowest.

Table 5: NDVI values of satellite images and yields of corresponding Farmer's fields at Ullapara, Sirajganj during the season of 2023-24

Farmer's Field No.	Longitude	Latitude	NDVI	Observed Yield (t/ha)
1	89.5671900	24.31072	0.756	1.50
2	89.5024165	24.19310	0.793	1.45
3	89.5679702	24.31071	0.785	1.53
4	89.5117804	24.21701	0.731	1.34
5	89.5117383	24.21701	0.731	1.40
6	89.5118608	24.21394	-	0.00
7	89.5024165	24.19320	0.792	1.48
8	89.5089597	24.22843	0.708	1.46
9	89.5024265	24.19320	0.792	1.36
10	89.5089097	24.22842	0.708	1.50
11	89.5109755	24.22991	0.752	1.42
12	89.5124510	24.22657	0.686	1.44
13	89.5103853	24.22842	0.701	1.42
14	89.4135483	24.26348	0.766	1.48
15	89.5679720	24.31071	0.685	1.40
16	89.5679720	24.31072	0.385	0.93
17	89.5124347	24.25667	-	0.00
18	89.5123454	24.26657	-	0.00
19	89.5156795	24.22833	-	0.00
20	89.5245481	24.23664	0.867	1.46

Table 5 shows that the NDVI values were a maximum of 0.867 for the 20th field and a minimum of 0.385 for the 16th field at Ullapara Upazila of Sirajganj district. On the contrary, the yield was maximum at 1.50 t/ha for the 1st and 10th field, and minimum at 0.93 t/ha for the 12th field. That means the yield was maximum for the field where the NDVI values were not maximum. But the minimum yield was found for the field where the NDVI values were the lowest.

Table 6: NDVI values of satellite images and yields of corresponding Farmer's fields at Kalihati, Tangail during the season of 202-23

Farmer's Field No.	Longitude	Latitude	NDVI	Observed Yield (t/ha)
1	24.3489000	90.0266000	0.775	1.98
2	24.3488700	90.0266000	0.597	0.99
3	24.3488828	90.0266054	0.728	1.51
4	24.3365546	90.0290074	**0.792**	**2.01**
5	24.3469805	90.0280701	0.632	1.24
6	24.3432559	90.0278810	0.782	1.78
7	24.3432573	90.0278850	0.653	1.24
8	24.3432391	90.0278713	0.623	1.48
9	24.3365547	90.0285911	0.617	0.89
10	34.3473045	90.0237759	0.642	1.15
11	24.3472667	90.0238250	0.759	1.56
12	24.3365546	90.0290071	0.591	0.99
13	24.3365547	90.0284241	0.695	1.48
14	24.3365547	90.0284342	0.682	1.46
15	24.3525869	90.0233823	0.753	0.78
16	24.3532672	90.0322156	0.457	**0.33**
17	24.3536377	90.0362991	**0.421**	0.44
18	24.3536377	90.0363994	0.681	1.28
19	24.3532678	90.0295907	0.507	0.60
20	24.3547522	90.0351324	0.638	1.20

In the Kalihati Upazila, the NDVI values were a maximum of 0.792 for the 4th field and a minimum of 0.421 for the 17th field. On the other hand, the yield was maximum at 2.01 t/ha for the 4th field, and the yield was minimum at 0.33 t/ha for the 16th field (Table 6). That means the yield was maximum for the field where the NDVI values were maximum. But the minimum yield was not found for the field where the NDVI values were the lowest.

Table 7: NDVI values of satellite images and yields of corresponding Farmer's fields at Kalihati, Tangail during the season of 2023-24

Farmer's Field No.	Longitude	Latitude	NDVI	Observed Yield (t/ha)
1	24.3489000	90.0266000	0.502	**0.67**
2	24.3488700	90.0266000	0.602	0.91
3	24.3488828	90.0266054	0.572	0.86
4	24.3365546	90.0290074	0.592	0.99
5	24.3469805	90.0280701	0.761	1.47
6	24.3432559	90.0278810	0.558	0.99
7	24.3432573	90.0278850	0.548	1.03
8	24.3432391	90.0278713	0.683	1.19
9	24.3365547	90.0285911	0.677	1.19
10	34.3473045	90.0237759	0.586	1.03
11	24.3472667	90.0238250	**0.786**	**1.48**
12	24.3365546	90.0290071	0.652	1.15

13	24.3365547	90.0284241	0.596	0.95
14	24.3365547	90.0284342	0.682	0.94
15	24.3525869	90.0233823	0.642	1.04
16	24.3532672	90.0322156	**0.407**	0.74
17	24.3536377	90.0362991	0.549	0.88
18	24.3536377	90.0363994	0.581	**0.67**
19	24.3532678	90.0295907	0.647	1.16
20	24.3547522	90.0351324	0.653	1.12

In the Kalihati Upazila, the NDVI values were a maximum of 0.786 for the 11th field and a minimum of 0.407 for the 16th field. On the other hand, the yield was maximum at 1.48 t/ha for the 11th field, and the yield was minimum at 0.67 t/ha for the 1st and 18th field (Table 7). That means the yield was maximum for the field where the NDVI values were maximum. But the minimum yield was not found for the field where the NDVI values were the lowest.

Table 8: NDVI values of satellite images and yields of corresponding Farmer's fields at Daulatpur, Manikganj during the season of 2022-23

Farmer's Field	Longitude	Latitude	NDVI	Yields (t/ha)
1	89.89617	23.96939	0.748	1.73
2	89.89623	23.96851	0.768	1.84
3	89.89586	23.96874	0.553	1.37
4	89.89582	23.96914	0.839	**2.24**
5	89.89614	23.96795	0.793	1.98
6	89.89599	23.96781	0.779	1.74
7	89.89561	23.96835	0.775	1.83
8	89.89476	23.96858	0.846	1.85
9	89.89457	23.96891	**0.877**	1.98
10	89.89493	23.96909	0.755	1.73
11	89.89591	23.96835	0.833	1.80
12	89.89551	23.96846	0.829	1.98
13	89.89641	23.9681	0.741	1.76
14	89.89619	23.96942	0.748	1.65
15	89.89603	23.96978	0.738	1.78
16	89.89474	23.96957	0.628	1.59
17	89.89518	23.97038	0.675	1.76
18	89.89464	23.97102	**0.545**	**0.88**
19	89.89537	23.97092	0.764	1.76
20	89.89479	23.97112	0.848	1.98

Table 8 shows that the NDVI values were a maximum of 0.877 for the 9th field and a minimum of 0.545 for the 18th field in the Daulatpur, Manikganj district. In contrast to that, the yield was maximum at 2.24 t/ha for the 4th field, and the yield was minimum at 0.88 t/ha for the 18th field. That means the yield was maximum for the field where the NDVI values were not maximum. But the minimum yield was found for the field where the NDVI values were the minimum.

Table 9: NDVI values of satellite images and yields of corresponding Farmer's fields at Daulatpur, Manikganj during the season of 2023-24

Farmer's Field	Longitude	Latitude	NDVI	Yields (t/ha)
1	89.89617	23.96939	0.841	1.73
2	89.89623	23.96851	**0.848**	1.71
3	89.89586	23.96874	0.834	1.65
4	89.89582	23.96914	0.649	1.49
5	89.89614	23.96795	0.654	1.48
6	89.89599	23.96781	0.838	1.74
7	89.89561	23.96835	0.631	1.60
8	89.89476	23.96858	0.832	1.85
9	89.89457	23.96891	0.820	**1.98**
10	89.89493	23.96909	0.551	1.24
11	89.89591	23.96835	0.632	1.50
12	89.89551	23.96846	0.831	1.98
13	89.89641	23.9681	0.845	1.76
14	89.89619	23.96942	**0.541**	1.35
15	89.89603	23.96978	0.845	**1.98**
16	89.89474	23.96957	0.759	1.59
17	89.89518	23.97038	0.784	1.76
18	89.89464	23.97102	0.825	1.76
19	89.89537	23.97092	0.559	**1.32**
20	89.89479	23.97112	0.625	1.58

Table 9 shows that the NDVI values were a maximum of 0.848 for the 2nd field and a minimum of 0.541 for the 14th field in the Daulatpur, Manikganj district. In contrast, the yield was maximum at 1.98 t/ha for the 9th and 15th field, and the yield was minimum at 0.1.32t/ha for the 19th field. That means the yield was maximum for the field where the NDVI values were not maximum. However, the minimum yield was not found for the field where the NDVI values were the minimum.

Model Estimation: Determining the Relationship of Mustard Yield and Image-Driven NDVI

To recognize the relationship between mustard yield and satellite image-driven NDVI the simple linear regression model was employed with the single-season data i.e. 2022-2023 and 2023-2024 individually as well as with the combined mean data. For Landsat-8 images the relationships were established for the study locations of Ullapara upazila with the data 2022-2023 and 2023-2024 individually (Figure 7a-b) and with the combined data (Figure 7c).

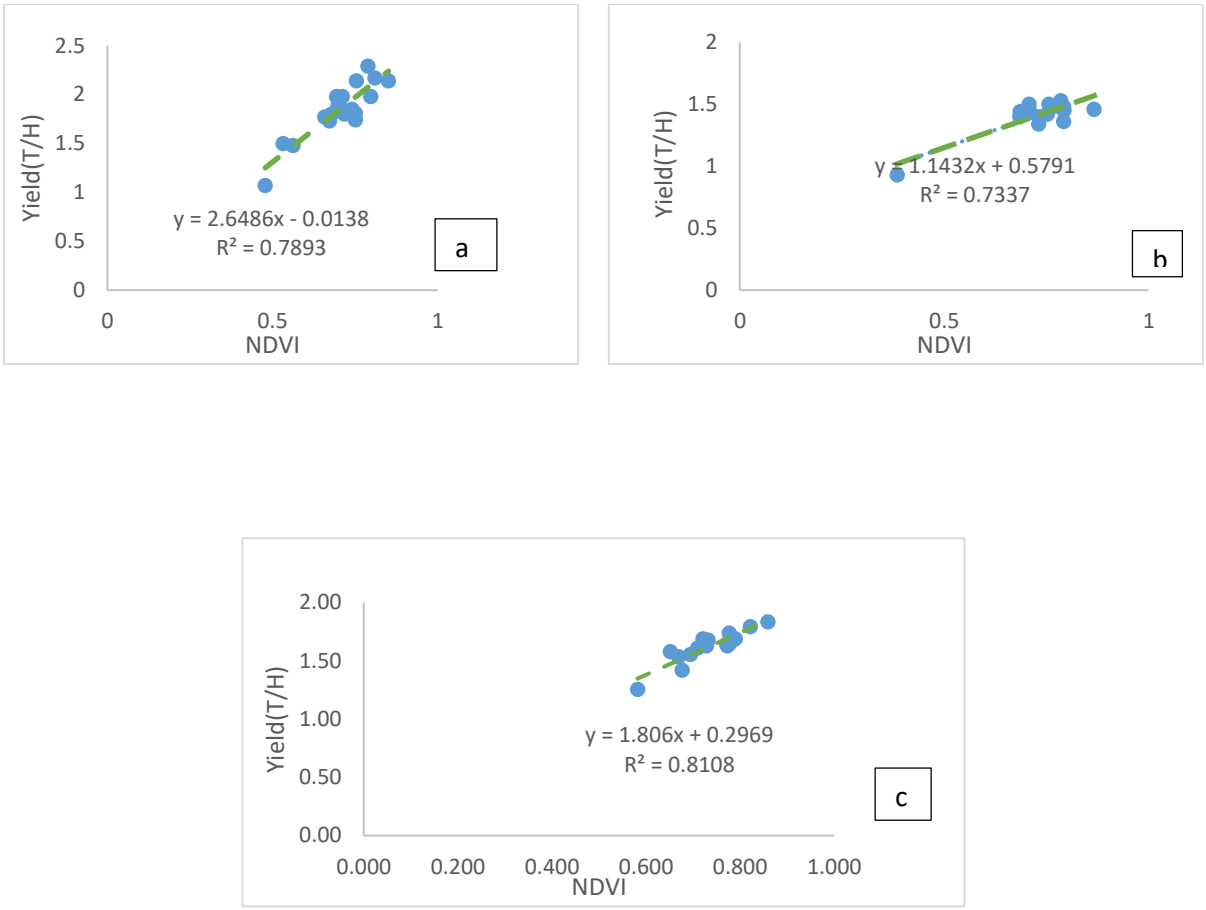
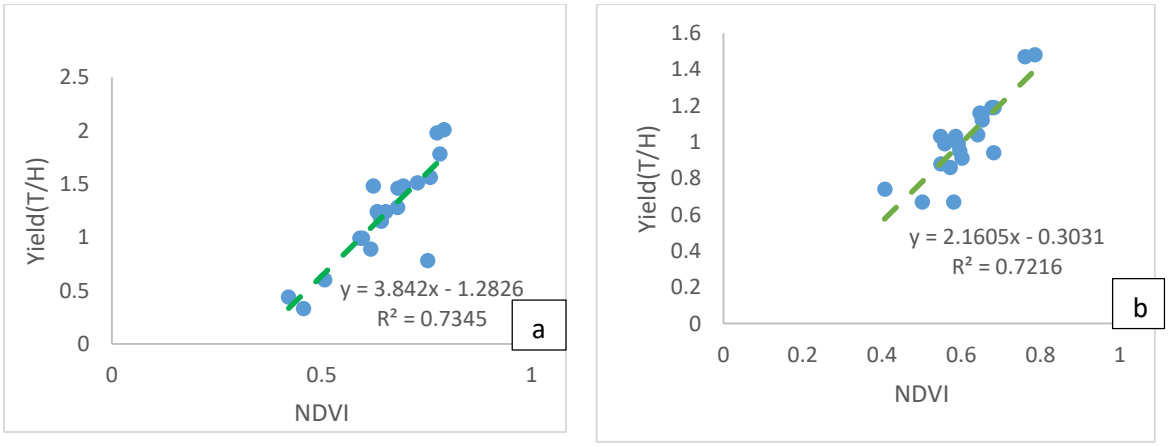


Figure 7: Model estimation from regression analysis of yield and Landsat-8 images driven NDVI for Ullapara Upazila

Similar relationships were determined for Kalihati Upazila. For Landsat-8 images, the relationships were established for the study locations with the data 2022–2023 and 2023–2024 individually (Figure 8a–b) and with the combined data (Figure 8c).



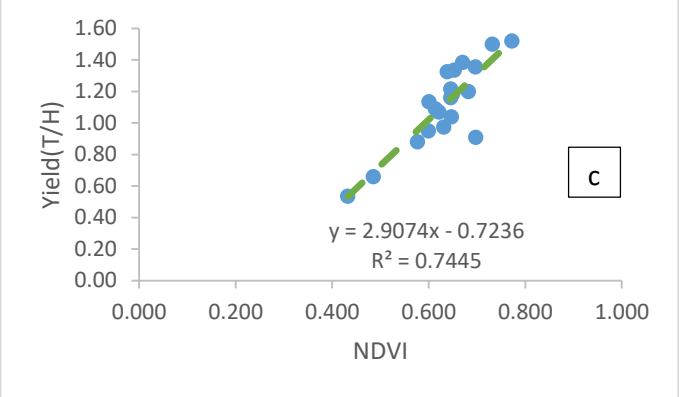
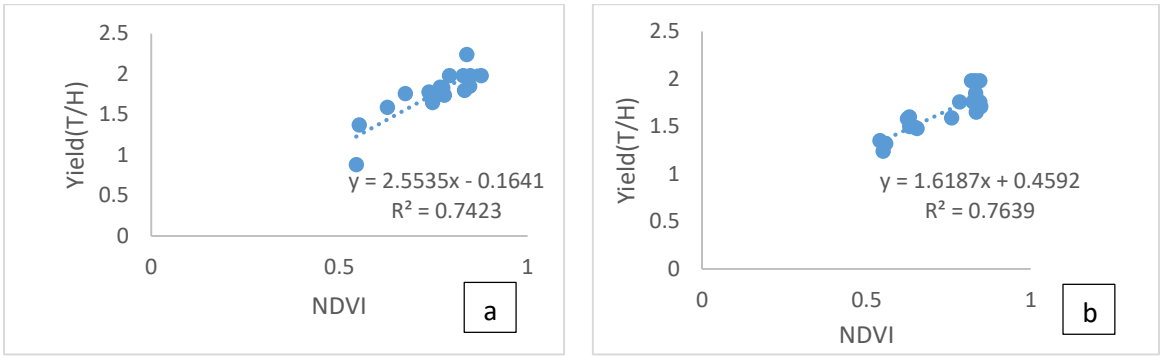


Figure 8: Model estimation from regression analysis of yield and Landsat-8 images driven NDVI for Kalihati Upazila

Accordingly, relationships were determined for Daulatpur Upazila. For Landsat-8 images, the relationships were established for the study locations with the data 2022–2023 and 2023–2024 individually (Figure 9a–b) and with the combined data (Figure 9c). All of the Upazilas combined data performed better than the single-season data.



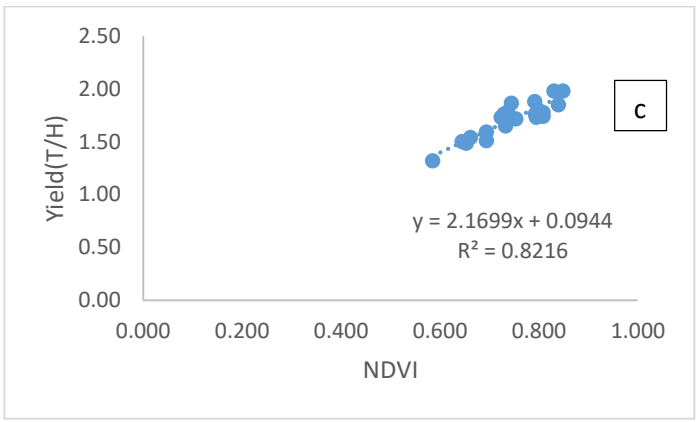


Figure 8: Model estimation from regression analysis of yield and Landsat-8 images driven NDVI for Daulatpur Upazila

Conclusion

This research aims to provide an operational technique with adequate technological components for monitoring and forecasting mustard yield in Bangladesh. In the farmers' fields of the selected locations, the developed system investigates the combined use of satellite remote sensing (RS) and Geographic Information System (GIS) technology. The objective of the study is to construct a remotely sensed yield prediction model that used the high spatial resolution of Sentinel 2A and Landsat 8 satellite images to forecast mustard yield one month ahead of harvest. The first year's results revealed that in most cases the yield was maximum for the field where the NDVI values were not maximum and vice-versa and the second-year result was the same as the first-year result. However, the relationships of the extracted mean NDVI and yields will be established using the classical linear regression model where the model will be developed using the first two years' data and will be validated using the data of the last study period.

References

Hamar, D., Ferencz, C.S., Lichtenberger, J., Tarcsai,G.Y. and Ferenczne A.R.I. (1988a) The use of remotely sensed data in yield forecasting I. Ground based experiments. *No ̈ve ́nytermele ́s*, 37:165–179 (in Hungarian).

Hamar, D., FERENCZ, C.S., Lichtenberger, J., Tarcsai, G.Y. and Ferenczne A.R. I. (1988b) The use of remotely sensed data in yield forecasting II. *Satellite experiments. No ̈ve ́nytermele ́s*, 37: 357–367 (in Hungarian)

National Research Council (1989). Diet and Health: Implications for Reducing Chronic Disease Risk. Washington, DC: The National Academies Press, 1989. doi:10.17226/1222

Kogan, F.N. (2001). Operational space technology for global vegetation assessment. *Bull. Am. Meteorol. Soc.* 82 (9):1949–1964.

Kogan, F. N. (2002). World Droughts in the new millennium from AVHRR-based vegetation health indices. *Eos Trans. Am. Geophys. Union* 48 (557): 562–563.

Kogan, F.N., Gitelson, A., Zakarin, E., Spivak, L. and Lebed, L. (2003). AVHRR-based spectral vegetation index for quantitative assessment of vegetation state and productivity: calibration and validation. *Photogrammetric Eng. Remote Sens.* 69 (8): 899–906.

Singh, R.P., Roy, S. and Kogan, F. (2003). Vegetation and temperature condition indices from NOAA AVHRR data for drought monitoring over India. *Int. J. Remote Sens*. 24: 4393–4402.

Unganai, L.S. and Kogan, F.N. (1998). Drought monitoring and Corn yield estimation in Southern Africa from AVHRR data. *Remote Sens. Environ*. 63: 219–232.

Hayes, J.T., O'Rourke, P.A., Terjung, W.E., and Todhunter, P.E. (1982). YIELD: A numerical crop yield model of irrigated and rainfed agriculture. *Publications in Climatology*, p. 35

Quarmby, N.A., Milnes, M., Hindle, T.L. and Silicos, N. (1993). The use of multitemporal NDVI measurements from AVHRR data for crop yield estimation and prediction. *Int. J. Remote Sens*. 14: 199–210

Benedetti, R. and Rossinni, P. (1993). On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in Emilia Romagna. *Remote Sens. Environ*. 45: 311–326.

Mohd, D.M., Ahmad, S. and Abdullah, A. (1994). Agriculture application of remote sensing: paddy yield estimation form Landsta-5 thematic mapper data.

El-Telbany, M.E. and Maged, M.A. (2019). Wireless sensor networks for extreme environments: remote sensing and space industry. *International Journal of Hybrid Intelligence* 12(1): 41-54.

Atzberger, C. (2013). Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sensing* 5(2): 949-881.

Sadras, V. O., Cassman, K., Grassini, P., Bastiaanssen, W. G. M., Laborte, A. G., Milne, A. E., ... & Steduto, P. (2015). Yield gap analysis of field crops: Methods and case studies.

Melillos, G., Agapiou, A., Themistocleous, K., Michaelides, S., Papadavid, G. and Hadjimitsis, D.G. (2019). Field spectroscopy for the detection of underground military structures. *European Journal of Remote Sensing* 52(1): 385-399.

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.C., Skakun, S.V. and Justice, C. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment* 219: 145-161.

Simonetti, D., Marelli, A. and Eva, H.D. (2015). IMPACT: Portable GIS Toolbox for image processing and land cover mapping. *Publication Office of the European Union* 10: 143497.

Teillet, P.M. (1986). Image correction for radiometric effects in remote sensing. *International Journal of Remote Sensing* 7(12): 1637-1651.

Wang, R.F., Zhang, Y., Li, J.G., Zhao, W., Wang, F.Z., Cao, H.J. and Duan, Y.P. (2019). Development of green tide monitoring with satellite images. *Journal of Coastal Research* 90(SI): 104-111.

Congedo, L, (2016). Semi-automatic classification plugin documentation. *Release* 4(0.1): 29.

End Note:

Characteristics of Landsat 8 Sensors

Landsat 8, also known as the Landsat Data Continuity Mission (LDCM), was sent into space on an Atlas-V rocket on February 11, 2013, from Vandenberg Air Force Base in California (https://www.usgs.gov/tools/landsat-8-overview). With the launch of Landsat 8, NASA has launched the most recent Landsat satellite, which is equipped with the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) sensors (https://www.usgs.gov/landsat missions/ landsat-8).

Table 1: Characteristics of Landsat 8 sensors

Satellite	Sensors	Band types	Wavelength (µm)	Resolution (m)
Landsat 8	Operational Land Imager (OLI)	Band 1 – Coastal aerosol (deep blue)	0.43-0.45	30
		Band 1 - Blue	0.45-0.51	30
		Band 3 – Green	0.53-0.59	30
		Band 4 – Red	0.64-0.67	30
		Band 5 – NIR	0.85-0.88	30
		Band 6 – Shortwave infrared (SWIR) 1	1.57-1.65	30
		Band 7 – Shortwave infrared (SWIR) 2	2.11-2.29	30
		Band 8 –Panchromatic	0.50-0.68	15
		Band 9 – Cirrus	1.36-1.38	30
	Thermal infrared sensor (TIRS)	Band 10 – TIRS1	10.60-11.19	100
		Band 11 – TIRS2	11.50-12.51	100

Characteristics of Sentinel-2 sensors

Copernicus Sentinel-2 consists of two sun-synchronous polar-orbiting satellites phased 180° apart. Its huge sweep width (290 km) and high revisit duration (10 days at the equator with one satellite and 5 days with two satellites in cloud-free circumstances, resulting in 2-3 days at mid-latitudes) will aid in monitoring of Earth's surface changes. (https://sentinel.esa.int/ web/sentinel/missions/sentinel-2). Detailed descriptions of the mission goals, satellite description, and ground section are provided in this Sentinel-2 Mission Guide at a high level. Also included are discussions on relevant legacy missions, thematic areas, and Copernicus services, as well as orbital parameters and coverage, instrument payload,

and data outputs. This document describes the data products available to users at the Level-1B, Level-1C, and Level-2A levels, as well as the Level-1C tiling grid.

Table 2: Spectral bands for the Sentinel-2 sensors

Sentinel-2 bands	Sentinel-2A		Sentinel-2B		Spatial resolution (m)
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)	
Band 1 – Coastal aerosol	442.7	21	442.2	21	60
Band 2 – Blue	492.4	66	492.1	66	10
Band 3 – Green	559.8	36	559.0	36	10
Band 4 – Red	664.6	31	664.9	31	10
Band 5 – Vegetation red edge	704.1	15	703.8	16	20
Band 6 – Vegetation red edge	740.5	15	739.1	15	20
Band 7 – Vegetation red edge	782.8	20	779.7	20	20
Band 8 – NIR	832.8	106	832.9	106	10
Band 8A – Narrow NIR	864.7	21	864.0	22	20
Band 9 – Water vapour	945.1	20	943.2	21	60
Band 10 – SWIR – Cirrus	1373.5	31	1376.9	30	60
Band 11 – SWIR	1613.7	91	1610.4	94	20
Band 12 – SWIR	2202.4	175	2185.7	185	20

FORECASTING OF ONION PRODUCTION IN BANGLADESH USING DIFFERENT MODELS AND MACHINE LEARNING ALGORITHM

NUR MOHAMMAD[1], MOHAMMAD MUKHLESUR RAHMAN[1] AND ISTIAK AHMED[1]
[1]ASICT Division, BARI Gazipur-1701

Abstract

Agriculture is important to the economy and employment of Bangladesh. Onion is the largest area coverage and maximum production capture among all the spices crops in Bangladesh. Onion is the most demandable crop for multiple uses especially in daily consumption in our country. Here, we have considered time series data for the years from 1970-71 to 2022-23 published by the Bangladesh Bureau of Statistics (BBS). Two classical linear model like Autoregressive Integrated Moving Average (ARIMA) and Mixed-model approach (Dynamic regression model) and also two machine learning algorithm (MLA) like Artificial Neural Network (ANN) and Autoregressive neural Network (ARNN) have been considered in this study. Both of these models have been considered to forecast the productions of onion in Bangladesh. The forecasting performances of those models have been compared by using RMSE, MAE, and MAPE. Result revealed that the mixed-model with ARIMA (0, 1, 1) was outperformed than ARIMA, ANN and ARNN for considering area of onion. The length of the 95% confidence interval of the forecast values of the mixed-model is smaller than that of the rest of model indicating its better predictive performance. These forecast values are helpful for collecting information and planning resources for the government, researchers, and businesspeople, as well as farmers making key decisions about onion crop production prior to the harvesting in Bangladesh.

Keywords: ARIMA, ANN, ARNN, Mixed-model, Forecast and Production

Introduction

Onion (*Allium cepa L.*) is one of the most significant profitable spice crops, in Bangladesh. It has also huge demand all over the world (Anjum and Barmon, 2017). It can be used in both mature and immature level as vegetable and spices, and increases the taste of food by its flavor that contains protein, calcium, carbohydrates and vitamin C. It is also used as medicine to recover from insect injury and raw throat (Bose and Som, 1990). In Bangladesh, onion ranks top among all the spice crops in production (BBS, 2023). It has evolved into a complicated scientific study aiming at generating the highest amount of agricultural output with the lowest amount of time, space, and energy in order to meet the demands of an expanding population and economy (Paswan et al., 2022). Higher onion production with the available area is required due to exploding population increase and rising per capita onion consumption. Onion is cultivated all over the country extensively in winter season. Although production of onion is increasing day by day, but in a land hungry country like Bangladesh it may not be possible to meet the domestic demand due to increase in population. Sometimes Government faced some unusual political problems for price hike of onion. So, each year our Government has to import a big amount of onion from neighboring and other countries to meet up its demand (Haque et al., 2011). For these reasons, the accurate forecasting of onion yield is highly important in ensuring the accurate determination of their product. This generally demands a believable and valid apriori prediction of the production of crops. For future planning and policymaking, accurate forecasting of such key commercial crops is required. An attempt has been made to construct an appropriate forecasting model because onion production forecasting could help in estimating production and making decisions about export and import policies, distribution, price policies, storage and marketing strategies. The forecasting methodology can help the farmers and other stakeholders to make it sustain for a larger duration (Mahto et al., 2021).

Auto Regressive Integrated Moving Average (ARIMA) model is a forecasting extrapolation approach that like any other requires simply historical time series data on the variables to be forecasted. ARIMA model is the most popular accepted time series model. It has numerous applications in various domains due its high mathematical precision, fexible nature, and greater reliable results (Kaur et al., 2023). It can handle any data pattern with simplicity. The advantage of ARIMA modelling over other univariate time series models is that it delivers the fewest mean squared forecast error variations in addition to revealing the intrinsic behaviour (producing process) in the time series variables (Rankja et al., 2017). Several researchers from home and abroad were applied ARIMA model for forecasting different crops production (Sarika et al., 2011; Suresh et al., 2011; Mishra and Singh, 2013; Kumari et al., 2014; Naveena et al. ,2014; Hossain and Abdulla, 2015; Rathod et al., 2017; Rathod et al., 2018, Kumar et al., 2020 and Devaki and Mohideen, 2021). The ARIMA is a more complex approach, requiring steps such as model identification, estimation and validation. Moreover, mixed model approach is dynamic in nature that can be used to determine how independent variables impact the response variable of interest (Pankratz, 1991). The forecasting performance of mixed model for four crops e.g., Aus, Aman, Boro and Potato is performed better than the univariate econometric model ARIMA in Bangladeshduring the season 2015-16 considering area as predictor (AMIS, 2017). Very few researchers were used mixed model approach for crop production forecast in Bangladesh. Mohammad et al. (2022) used mixed model for maize production forecast which was performed better than ARIMA model. The production forecast of potato, mixed model was suited better than ARIMA (Rahman et al., 2022). Each ARIMA and mixed model are based on a linear combination of past values and/or errors.

Artificial neural networks (ANN) are nonlinear model that are able to capture various nonlinear structures present in the data set. ANN model specification does not require prior assumption of the data generating process, instead it is largely depend on characteristics of the data (Vijay and Mishra, 2018). ANN have been applied in a range of situations where traditional statistical approaches have been used. Artificial Neural Networks (ANN) have the ability to discover complicated nonlinear correlations between dependent and independent variables implicitly, therefore they can be utilized to more accurately predict future demand. These are currently preferred tool in predicting future crop yield and production. As a result, it's past time for ANN to be recognized as a powerful data analysis tool (Paswan et al., 2022). Neural Networks have the advantage that can approximate any nonlinear functions without any apriori information about the properties of the data series (Hossain et al., 2017). In order to assess the stability and long-term viability of sugarcane production in the Indian state of Bihar, Paswan et al. (2022) employed ARIMA and Artificial Neural Networks (ANN) models. Hossain et al. (2017) used ANN for predicting jute crop which was better suited than ARIMA. Sultana and Khanam (2020) conducted a study using the yearly rice production data from 1972 to 2013 to predict the rice production of Bangladesh and found that the ARIMA model is superior to ANN model.

The artificial neural networks allow emulating the processing of information that the brain performs and allow it to be approximated to any function (Velásquez *et al*., 2008). Autoregressive Neural Network (ARNN) takes advantage of autoregressive (AR) models and multilayer perceptron (MLP) to capture the complex dynamics. Several experiments with ARNN were performed using three different configurations by modifying the number of neurons. ARNN was outperformed than ARIMA for the prediction of glucose levels in diabetic patients in Mexico (Vera et al., 2020). Although production of onion is increasing day by day, but in a land hungry country like Bangladesh, it may not be possible to meet the domestic demand due to increase in population. To meet the demand of domestic consumption of onion, it is too much essential to estimate the production of onion in Bangladesh which leads us to do this research. Due to its importance, several time series forecasting methods have been proposed, such as the ARIMA, Mixed model and machine learning algorithm (Hossain et al., 2017). Thus, this paper attempts to compare the accuracy of ARIMA, Mixed model approach, ANN and ARNN for forecasting the onion production in Bangladesh

Objectives

1. To identify the best fitted model for onion production.
2. To forecast the production of onion in Bangladesh using best fitted model.

Materials and Methods

Data

This study considered the published secondary data of yearly onion area (Hectare) and production (Metric tons) in Bangladesh which was collected over the 52 (fifty two) financial period 1970-71 to 2021-22 from Bangladesh Bureau of Statistics (BBS).

ARIMA Model

Historical observations and random mistakes (errors) are combined to form a linear relationship that can be used to forecast future variables in ARIMA. Box-Jenkins for forecasting the future value of a time series, ARIMA modelling employs a three-step iterative technique: model identification, parameter estimation and residual diagnostics testing (Box, et al., 1967 and Box, et al., 2009). The global model is based on a linear combination of past values (AR components) and errors (MA components), being named Auto Regressive Integrated Moving-Average (ARIMA). The non-seasonal model is denoted by the form ARIMA (p,d,q) and is denoted by the equation:

$$\emptyset_p\,(L)(1-L)^d y_t = \theta_q(L)e_t$$

Where y_t is the series, e_t is the error, L is the lag or backshift operator (e.g.$L^3 y_t = y_{t-3}$), $\emptyset_p = 1 - \emptyset_1 L - \emptyset_2 L^2 - \cdots - \emptyset_p L^p$ is the AR polynomial of order P, d is the differencing order and $\theta_q = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q$ is the MA polynomial of order q. The ARIMA model is constructed using those iterative steps. Firstly, we convert a non-stationary series into a stationary one by applying the difference of order d. Then, once a stationary series is obtained, we select the model parameters p and q from the ACF plot and PACF plot, respectively. Finally, we obtain the "best fitted" model by analyzing the residuals (Panja et al., 2023).

Mixed Model approach

Mixed model (dynamic regression model) is a regression model which allows lagged values of the independent variable(s) to be included. This model is applied to predict what will occur to predict variable if the independent variable changes. Dynamic regression model is a mixed model approach which can be used to detect the roll of independent variables in determining the variable of interest i.e., response variable (AMIS, 2017).

The appropriate dynamic regression model is to fit a multiple regression model of the form:

$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots\cdots\cdots + \beta_k X_{t-k} + N_t$

Where, Y_t refers the predicted variable;

X_{t-k} means the independent variable with the time-lag k=0,1….K;

$\beta_0, \beta_1, \ldots\ldots\beta_k$ the regression coefficient

And N_t is an ARIMA process.

Artificial Neural Network (ANN)

Artificial neural networks (ANN) model specification does not require prior assumption of the data generating process, instead it is largely depend on characteristics of the data. Neural models are innate candidates for forecasting due to their nonlinear and noise tolerance capabilities. In essence, neural networks use a nonlinear modelling technique to approximate any function with reasonable accuracy (Kumar et al., 2023). The basic idea is to train an ANN with past data and then use this network to predict future values in Figure 1. The model capacity and complexity increase when more hidden layers and layer nodes are added to the model architecture (Kontopoulou et al., 2023).

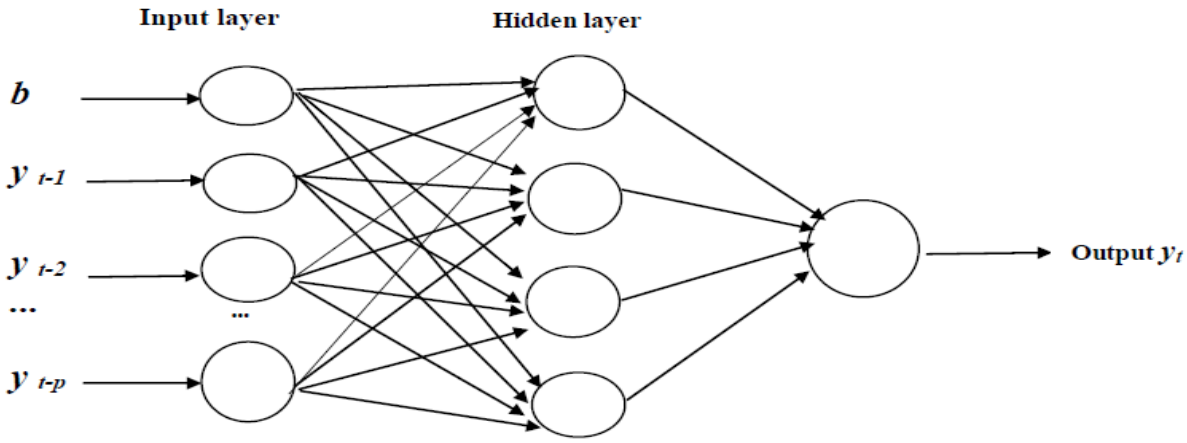


Figure 1: Artificial Neural Network system

ANN connect nerve cells in numerous layers of networks, analogous to how neurons in the human brain connect to each other (Remus, W., and O'Connor, M. 2001). Nodes are the names given to the neurons. Without a lot of data or knowledge, ANN is a self-adaptive nonlinear nonparametric data-driven statistical method for regulating nonlinearity and approximating complex relationships. It is employed in the forecasting of time series. The functional form of ANN is as follows if the input nodes have lagged values:

$y_t = h(y_{t-1}, y_{t-2}, y_{t-3}, \ldots, y_{t-1}) + \varepsilon_t$ or above functional form can be expressed mathematically

$$y_t = w_{ij} \sum_{j=1}^{q} w_j \, . \, g\left(w_{oj} + \sum_{i=1}^{p} w_{ij} \, y_{t-1} \right) + \varepsilon_t$$

The bias of input is defined as y_{t-1} transformed data (input), w_{ij} is weights associated with input nodes, and w_j is weights associated with hidden nodes, where p is the number of input nodes and q is the number of hidden nodes.

Autoregressive Neural Network (ARNN)

Autoregressive neural networks (ARNN) derived from the artificial neural network is specifically designed for modeling nonlinear time series data (Faraway et al., 1998). ARNN we developed is a straightforward multi-layer auto-regressive artificial neural network implemented in a very intuitive and interactive simulator coded in Perl. ARNN can optionally be extended with hidden layers to achieve greater forecasting accuracy (Ranaldi et al., 2022). ARNN combines an autoregressive linear model (AR) and multilayer perceptron (MLP) that contains a hidden layer (Vera et al., 2020). ARNNs (Trapletti et al., 2000) as a nonlinear generalization of AR(p) models. The ARNN is a model that allows using the

advantages of the AR and MLP to capture complex dynamics (Velásquez *et al*., 2008; Velásquez *et al*., 2009). The architecture of an ARNN is shown in Figure 2.

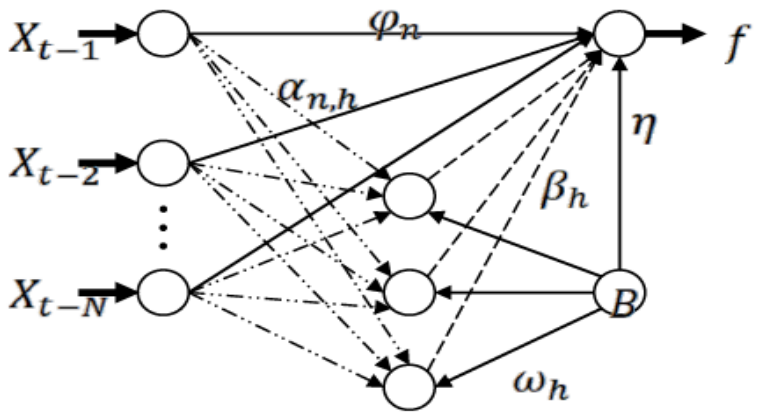


Figure 2: Autoregressive Neural Network Architecture (ARNN)

The ARNN model has a dependent variable f, that is obtained from applying a non linear function to N previous values, $X_{t-n}\ for\ n = 1, \ldots\ldots, N$:

$$f = \eta + \sum_{n=1}^{N} \varphi_n X_{t-n} + \sum_{h=1}^{H} \beta_h\, G(\omega_h + \sum_{n=1}^{N} \alpha_{n,h}\, X_{t-n})$$

Where, N is the total number of previous values. η is weight values and X_{t-n} is the input values.

Where $\mathbb{G}G$ is the sigmoid adaptive function defined as:

$$G(u) = \left[\frac{1}{1 + \exp(-u)}\right]^{M}$$

The model parameters $\eta, \varphi_p, \beta_h, \omega_h, \alpha_{p,h}$ and M for $i = 1, \ldots\ldots, N$ and $h = 1, \ldots\ldots, H$ which are estimated by minimizing the regularization error: λE, where λ is user defined parameter (Breu et al., 2011).

Ljung-Box Test

The Ljung-Box test is a test for autocorrelated errors. The functional form of the test statistic is as follows

Q= n (n+2)$\sum_{k=1}^{h}(n-k)^{-1}\,\rho_k^2$, which follows χ^2 with the degree of freedom h-m.

Where n, k, h, m and ρ_k refer the number of observations, the number of lag, the maximum lag, the number of parameters and the autocorrelation function respectively. Ljung-Box test can be applied to test the hypothesis that all of the autocorrelations are zero, that is, the series is a white noise.

The correlogram executes of the residuals (forecast errors) shows the autocorrelations of the residuals. Autocorrelation of the residual series falling outside the boundaries of the correlogram suggest there may be few additional information in the series which is not being seized by predicting model.

Evaluation of Forecasting Model

The accuracy of the models has been compared by using three statistics regarding the forecasting approach of the models.

The Root Mean Square (RMSE) can be defined as:

$$RMSE = \sqrt{E(f_i - Y_i)^2}$$

The Mean Absolute Error (MAE) used to measure how closely the prediction goes to the outcome. It is denoted as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - Y_i|$$

The goodness of fit of a model has been evaluated by the Mean Absolute Percentage Error (MAPE). It can be defined as:

$$MAPE = \frac{\frac{1}{n}\sum_{i=1}^{n}|f_i - Y_i|}{Y_i} * 100$$

Where, Y_t =actual observation and f_t =fitted observation

Results and Discussion

Stationarity checking of onion production

A time series plot is a two-dimensional plot of time series data. The production of onion is fluctuating over the time (Figure 3). The fluctuation with time denotes the data series may not be stationary. Specially, onion production is increased frequently in two times during the financial year during 2003-04 to 2009-10 and 2011-12 to 2019-20 (Figure 3). Increasing and decreasing trends of onion production present that the variance is not stable which leads the onion production data series not stationary because it's mean, variance and autocorrelation depends on time. It is also perceivable that the mean, variance and autocorrelation do not remain constant from time to time i.e., time independent, so we can express that the production of onion is not stationary (Figure 3).

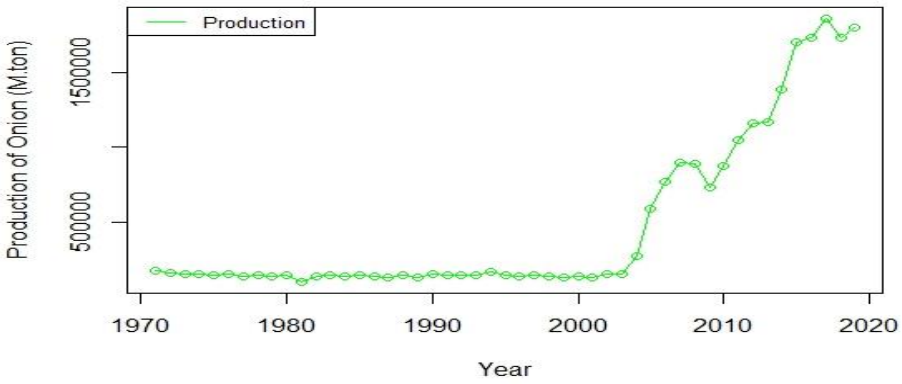


Figure 3: Onion production during the year 1970-71 (1971) to 2021-22 (2022)

The stationarity of onion (in M.Tons) from financial year 1970-71 to 2021-22 has been checked on the basis of Augmented Dickey-Fuller and Phillips-Perron test. Augmented-Dickey-Fuller (ADF) unit root test and Phillips-Perron (PP) unit root test are applied to identify whether the data series is stationary or not. At first difference, onion productions in distinct year's data are stationary which is presented in Table 1.

Table 1: The stationary checking onion production (M. Ton)

Difference	Augmented Dickey-Fuller test		Phillips-Perron test		Comments
	Statistic	P-value	Statistic	P-value	
No	0.79	0.99	1.30	0.99	Not Stationary
First	-4.27	0.01	-31.4	0.01	Stationary

Best model assessment for onion and validation

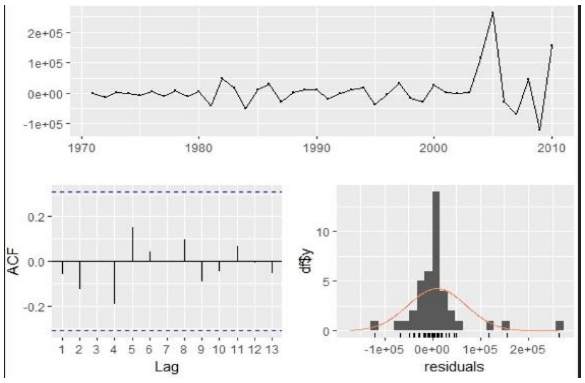
The best model is selected on the basis of the lowest value of root mean squared error (RMSE), mean absolute error (MAE) and mean absolute Percentage error (MAPE). The Box-Ljung test also used to examine the residuals because the coefficients were found to be significant. The values of Ljung-Box Q statistic for all selected models were found to be non-significant i.e., the residuals follows white noise, i.e., autocorrelations are zero. ARIMA, mixed-model, ANN and ARNN are presented in Table 2 for predicting onion production on the basis of separate model selection criteria.

Table 2: Model selection criteria for checking the best fitted model

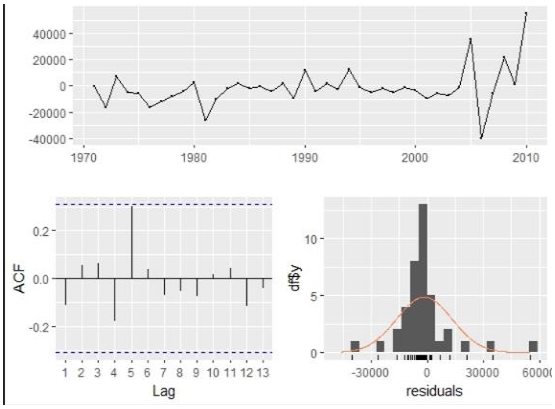
	RMSE	MAE	MAPE	Ljung-Box Q	
				Statistic	P-value
ARIMA (0, 1, 2)	84.7850.62	740332.22	40.034	4.22	0.64
Mixed-model with ARIMA (0, 1, 1)	521770.98	473815.55	27.089	7.24	0.40
ANN (3, 2)	1039881.16	944609.71	52.829	12.60	0.13
ARNN (3, 2)	786483.15	705717.60	39.297	4.89	0.76

From Table 2 we observed that, mixed-model with ARIMA (0, 1, 1) is the best selected model for predicting onion production comparing to ARIMA, ANN ARNN. ARIMA model depends on crop production time series data (response variable) but does not include crop area data. On the other hand, mixed model approach is a dynamic regression model which make out the role of independent variables in determining response variable. So mixed-model depends on crop production time series data (response variable) including crop area time series data (independent variable). On the other hand, ANN with 3 laged values and 2 hidden layers was not satisfied the model evaluation criteria but ARNN (3,2) was better than ARIMA and ANN. Here, the mixed-model is the best fitted model for predicting onion production on the basis of different model selection criteria.

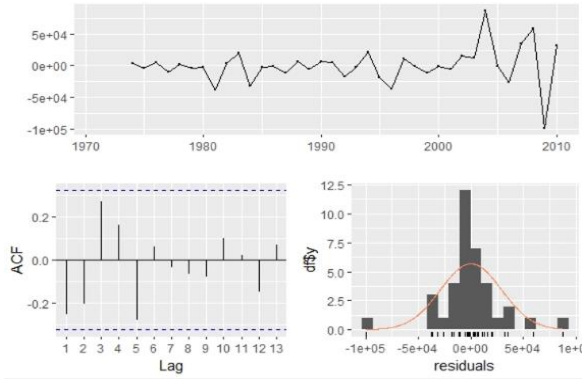
The diagnostic check required determining the residuals from selected models were not autocorrelated and normally distributed simply white noise. An evaluation of the selected Models based on a residual plot for autocorrelations functions was shown in Figure 4, clearly indicate that all autocorrelations ware lies in the 95 percent confidence interval, indicate the residuals behavior like white noise and histogram plots indicate the normal behavior in residuals.



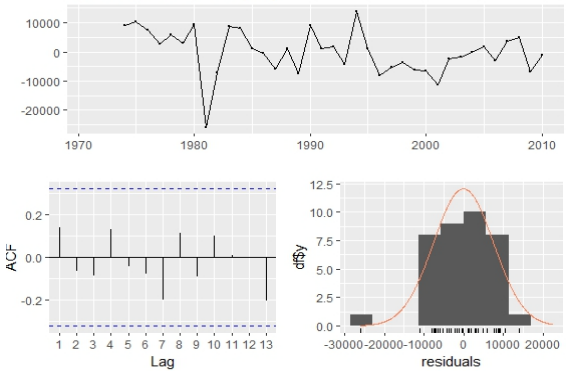
Residuals from ARIMA (0,1, 2)



Residuals from Mixed with ARIMA (0,1, 1)



Residuals from ANN (3, 2)



Residuals from ARNN (3, 2)

Figure 4: Plots of Residuals from ARIMA, Mixed, ANN and ARNN algorithms

Here, the onion data are divided into two segments i.e. training data set and testing data set. 75% data are referred as training data and the rest are treated as testing data. The first 39 years of data (training data set) were used to develop the model, and the final thirteen years of data were used to ensure that the models chosen were accurate From Figure 5, we observed that the interval variation of forecast onion production data for mixed-model approach is closer than ARIMA. But for ANN and ARNN, interval estimation of forecast onion production is not found. Finally, it may be concluded that the mixed-model approach will give more appropriate interval estimation rather than ARIMA for onion production (Figure 5).

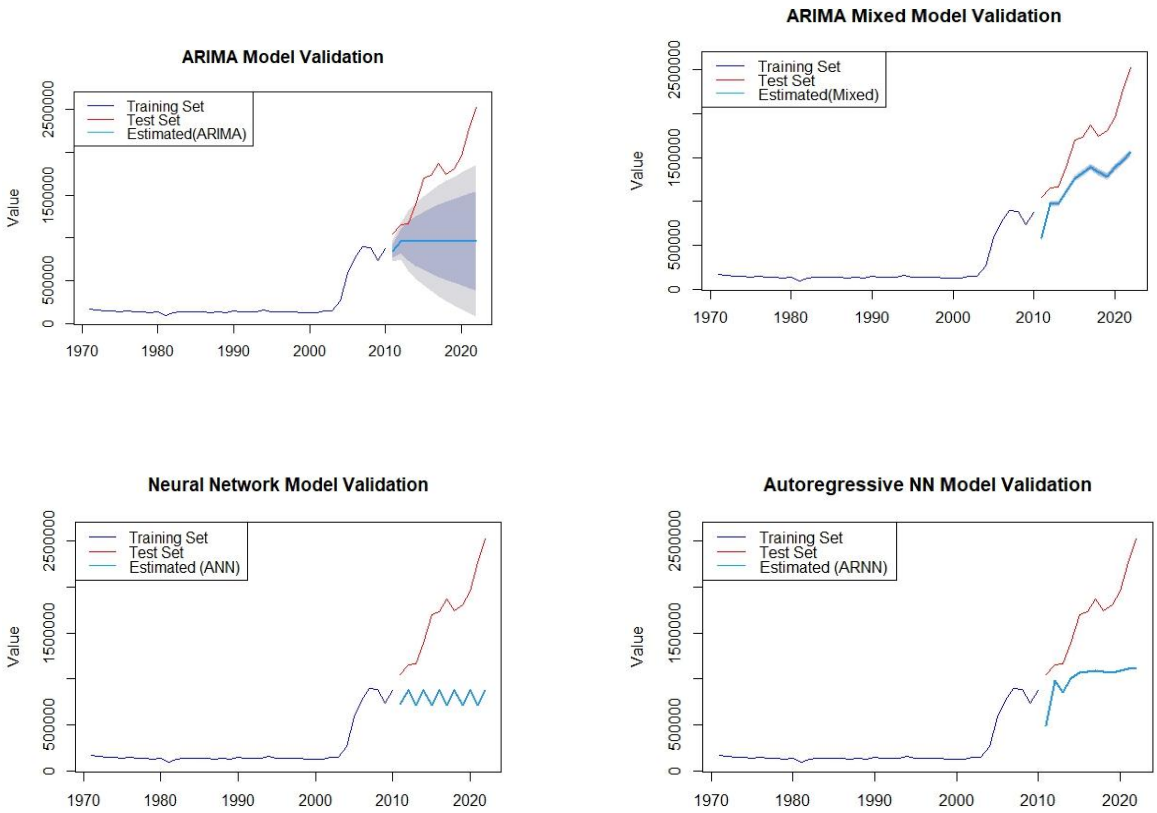


Figure 5: Validation graph for ARIMA, Mixed, ANN and ARNN model

Forecast onion production on the basis of best fitted model

The forecasted onion productions in Bangladesh by the best fitted model considered in this study are given in Table 3. We obtained forecasted values for the next 5 years for onion production by mixed-model approach in Bangladesh. Furthermore, mixed-model may be used when the information on area under onion cultivation are available. On the basis one of the most effective predicting indicators, forecast value of mixed model approach as well as 95% confidence level for five financial seasons for production of onion is presented in Table 3.

Table 3: Forecast short term production of onion (in M.Ton)

Year	Mixed-model with ARIMA (0, 1, 1)		
	Forecast	Lower limit	Upper limit
2022-23	2573535	2405150	2741921
2023-24	2575082	2300238	2849927
2024-25	2606023	2224760	2987285
2025-26	2701949	2162879	3241019
2026-27	2732202	2109145	3355259

But the mixed-model approach exposes that the onion production of 2022-23 will be reach up to in 2026-27 considering land area and lag of land area as the independent variables. The length of the confidence interval of the predicted value for mixed-model approach is closer than the other model.

Conclusion

We have found the best forecasting model through ARIMA, Mixed-model, ANN and ARNN model to forecast the productions of onion in Bangladesh. We compare the forecasting performances of these models by using RMSE, MAE and MAPE of prediction. Because the values of RMSE, MAE and MAPE are the lowest for the Mixed-model comparing with the ARIMA, ANN and ARNN model. It has been come out that the Mixed-model model with ARIMA (0, 1, 1) was the best forecasting model for onion production in Bangladesh. The proposed Mixed-model for forecasting for information and resource planning should be used by researchers, businesses, and farmers in Bangladesh decisions about onion crop yield. The specific policy recommendation is that forecasting of the onion production should be made immediately after plantation when area information will be available to facilitate decisions regarding import, export and storage management.

References

Akpan, E. A., and Moffat, I. U. (2016). Dynamic Time Series Regression: A Panacea for Spurious Correlations. *International Journal of Scientific and Research Publications*, Volume 6, Issue 10, October 2016. ISSN 2250-3153.

Anjum, A. and Borman, B. K. (2017). Profitability and Comparative Advantage of Onion (Allium cepa) Production in Bangladesh: An Economic Study in Some Selected Areas. *The Agriculturists* 15(2): 66-78 (2017).

AMIS. (2017). Chapter VI. Model based forecasting. Strengthening Agriculture Market Information System (AMIS) in Bangladesh project. *Analytical report on methodologies of crop estimation and forecast and private stock of food grain survey 2016-17*. Statistics and Informatics Division (SID), Bangladesh Bureau of Statistics (BBS). Ministry of Planning. 2017, 38-59.

BBS (2023). *Yearbook of Agricultural Statistics of Bangladesh*. Bangladesh Bureau of Statistics, Ministry of Planning, Government of the People's Republic of Bangladesh, Dhaka.

Bose, T. K. and Som, M. G. 1990. *Vegetable crops in India*. Naya Prokash, Calcutta, 6:545-582.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (1967). *Time Series Analysis: Forecasting and Control*, third ed., Prentice Hall, Englewood Clifs.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (2009) *Time Series Analysis: Forecasting and Control*, third ed., Holden-Day, San Francisco, USA.

Breu, F., Guggenbichler, S., & Wollmann, J. (2011). ARNN: *A packages for time series forecasting using autorreresive neural networks*. Vasa, 8(2).

Devaki, C., Mohideen, D. A. K. (2021). Forecasting analysis for sugarcane production in Tamil Nadu using ARIMA Model. *Journal of Xi'an University of Architecture & Technology*. 2021;13(6):778-790.

Faraway, J., Chatfield, C. (1998). Time series forecasting with neural networks: a comparative study using the air line data. *Journal of the Royal Statistical Society*: Series C (Applied Statistics) 47(2), 231–250.

Haque, M. A., Miah, M. A. M., Hossain, S., Rahman, M. S. and Moniruzzaman, (2011). Profitability of Onion Cultivation in Some Selected Areas of Bangladesh, Bangladesh Journal of Agricultural Research, 36(3):427-435.

Hossain, M. M. and Abdulla, F. (2015). Forecasting the Tea Production of Bangladesh:

Application of ARIMA Model. *Jordan Journal of Mathematics & Statistics*, 8(3): 257-270.

Hossain, M. M., Abdulla, F. and Hossain, Z. (2017). Comparison of ARIMA and neural network model to forecast the jute production in Bangladesh. *Jahangirnagar University Journal of Science*.Vol. 40, No. 1, pp.11-18

Kaur, J., K. S. Parmar and S. Singh. (2023). Autoregressive models in environmental forecasting time series: a theoretical and application review. *Environmental Science and Pollution Research* (2023) 30:19617–19641.

Kontopoulou, V. I., A. D. Panagopoulos, I. Kakkos and G. K. Matsopoulos. (2023). A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. Future Internet 2023, 15, 255. https://doi.org/10.3390/fi15080255.

Kumar, K. S., T. Gowthaman and B. Bhattacharyya. (2023). Comparison of Arima and Ann for Forecasting the Annual Rainfall of Nadia District, West Bengal, India. Eco. Env. & Cons. 29 (August Suppl. Issue) : 2023; pp. (S24-S29)

Kumar, K. S. and Shitap, M. (2020). Statistical evaluation of stepwise regression method and autoregressive integrated moving average method for forecasting of groundnut (arachis hypogaea l.) productivity in Junagadh district of Gujarat.*Int.J.Curr.Microbiol.App.Sci*9(11): 84-93.https://doi.org/10.20546/ijcmas.2020.911.009.

Kumari, P., Mishra, G. C., Pant, A. K., Shukla, G. and Kujur, S. N. (2014). Autoregressive Integrated Moving Average (ARIMA) approach for prediction of rice (*Oryza sativa* L) yield in India. *The Bioscan*, 9(3): 1063-1066.

Mahto, A.K.,M. A. Alam, R. Biswas, J. Ahmed and S. I. Alam. (2021). Short-Term Forecasting of Agriculture Commodities in Context of Indian Market for Sustainable Agriculture by Using the Artificial Neural Network. *Hindawi Journal of Food Quality* Volume 2021, Article ID 9939906, 13 pages. https://doi.org/10.1155/2021/9939906.

Makridakis, S., Wheelwright, S. C., Wiley & Sons, R. J. (1998). *A review of: "Forecasting: Methods and Applications"*. pp.642. Third edition, ISBN 0-471-53233-9." IIE TRANSACTIONS, 31(3), p. 282.

Mishra, G. C. and Singh, A. (2013). A study on forecasting, prices of groundnut oil in Delhi by ARIMA methodology and Artificial Neural Networks. *AGRIS online papers in Economics and Informatics*, 5: 25-34.

Mohammad, N., M. A. Islam, M. M. Rahman, I. Ahmed, and M. G. Mahboob. (2022). Forecasting of Maize Production in Bangladesh using Time Series Data. *The Bangladesh Journal of Agricultural Economics, 43(2) (2022):18-32.*

Moraye, K., Pavate, A., Nikam, S., and Thakkar, S. (2021). Crop yield prediction using random forest
Algorithm for major cities in Maharastra state. *International Journal of Innovative Research in Computer Science & Technology (IJIRCST),*Volume-9, Issue-2, March 2021.

Naveena, K., Rathod, S., Shukla, G. and Yogish, K. J. (2014). Forecasting of coconut production in India: A suitable time series model. International Journal of Agricultural Engineering, 7(1): 190–193. 162

Panja, M, T. Chakraborty, U. Kumar, and A. Hadid. (2023). Probabilistic Auto Regressive Neural Networks for Accurate Long-range Forecasting.

Pankratz, A. (1991). *Forecasting with dynamic regression models.Forecasting with Dynamic Regressions Models*, 3nd Ed., New York, John Wiley and Sons, 1991DOI: 10.1016/0169-2070(92)90081.

Paswan, S., A. Paul, A. Paul and A. S. Noel. (2022). Time series prediction for sugarcane production in Bihar using ARIMA & ANN model. *The Pharma Innovation Journal* 2022; SP-11(4): 1947-1956.

Rahman, N. A. F., Aziz, M. A., Rahman M. M. and N. Mohammad. (2013). Modeling on Grass Pea and Mung Bean Pulse Production in Bangladesh Using ARIMA Model. *IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS)*

Rathod, S. and Mishra, G. C. (2017). Weather based modeling for forecasting area and production of mango in Karnataka. *International Journal of Agriculture, Environment and Biotechnology*. IJAEB: 10(1): 149-162, February 2017.

Rathod, S. and Mishra, G. C. (2018). Statistical Models for Forecasting Mango and Banana Yield of Karnataka, India. *J. Agr. Sci. Tech*. (2018) Vol.20: 803-816.

Rahman, M. M., M. A. Islam, M. G. Mahboob, N. Mohammad and I. Ahmed. (2022). Forecasting of Potato Production in Bangladesh using ARIMA and Mixed Model Approach. *Journal of Agriculture and Veterinary Science.* 9(10): 136-145.

Ranaldi, L., M. Gerardi, and F. Fallucchi. (2022). Using Auto-Regressive Multi-Layer Artificial Neural Networks to Predict Financial Time Series. Information 2022, 13, 524. https://doi.org/10.3390/info13110524.

Remus W, O'Connor M. *Neural Networks for Time-Series Forecasting*, Springer, New York, 2001.

Sarika, Iquebal, M. A. and Chattopadhyay, C. (2011). Modelling and forecasting of pigeonpea (Cajanus cajan) production using autoregressive integrated moving average methodology. Ind. J. Agric. Sci., 81(6): 520-523.

Suresh, K. K. and Priya, S. R. K. (2011). Forecasting sugarcane yield of Tamilnadu using ARIMA models. Sugar Tech., 13(1): 23-26.

Sultana A. and M., Khanam. (2020). Forecasting Rice Production of Bangladesh Using ARIMA and Artificial Neural Network Models. *Dhaka University Journal of Science*, 68(2), 143-147.

Trapletti, A., Leisch, F., and Hornik, K. (2000). *Stationary and Integrated Autoregressive Neural Network Processes*. Neural Comput., 12(10):2427-2450.

Velásquez, J D, Dyner, I., & Souza, R. C. (2008). Electricity spot price modelling in Brasil using an autoregressive neural network. *Ingeniare*, 16(3), 394–403.

Velásquez, Juan David, Franco, C. J., & García, H. A. (2009). Un modelo no lineal para la predicción de la demanda mensual de electricidad en colombia. *Estudios Gerenciales*, 25(112), 37–54. DOI: https://doi.org/10.1016/S0123-5923(09)70079-8

Venugopal, A., Aparna, S., Mani, J., Mathew, R., and Williams, V. (2021). Crop Yield Prediction using Machine Learning Algorithms. *International Journal of Engineering Research & Technology* (IJERT).

Vera, D. A. O., D. A. Gutiérrez-Hernández, M. A. Escobar-Acevedo C. Lara-Rendón and D.

A. Velázquez-Vázquez. (2020). Comparison of algorithms for the prediction of glucose levels in patients with diabetes. Nº 26, Vol. 13 (2), 2020. ISSN 2007 – 0705, pp.: 1 – 19.

Vijay, N. and G.C. Mishra. (2018). Time Series Forecasting Using ARIMA and ANN Models for Production of Pearl Millet (BAJRA) Crop of Karnataka, India. International Journal of Current Microbiology and Applied Sciences (2018) 7(12): 880-889.

FORECASTING ONION YIELD BY USING SATELLITE-BASED REMOTE SENSING TECHNIQUE IN BANGLADESH

NUR MOHAMMAD[1], MOHAMMAD MUKHLESUR RAHMAN[1], ISTIAK AHMED[1], MOHAMMAD RASEL[1], AND MD. ABDUL MONAYEM MIAH[1]

[1]ASICT Division, BARI Gazipur-1701

Abstract

Onion is one of the major vegetables as well as spices crops with the largest production worldwide. Onion plays as a major contribution as spices crop which is used in daily meal in Bangladesh. Pre-harvest prediction of a crop yield may prevent a disastrous situation and help decision-makers to apply more reliable and accurate strategies regarding food security. Remote sensing can be used for yield estimation prior to harvest at the field level to provide helpful information for agricultural decision making. Remote sensing images are capable of identifying crop health, as well as predicting its yield. The Normalized Difference Vegetation Index (NDVI) calculated from remotely sensed data have been widely used to predict crop yield. Yield prediction models based on a time series of satellite images and high-density yield data, and to indicate the best phenological stage of onion crop to obtain satellite images for this purpose. The study used 16-day (~ 30 m) Landsat 8/9 OLI (Operational Land Imager) high resolution reflectance data for the two years viz. 2022-2023; 2023-2024 at three different locations viz. Sujanagar, Pabna; Baliakandi, Rajbari and Durgapur, Rajshahi in Bangladesh. The single date of cloud free image acquisition based on maximum NDVI for Landsat 8/9 OLI satellite image was used for 2022-2023 and 2023-2024 onion growing period to develop the yield prediction model. Regression model was performed between NDVI values and 35 farmers filed level onion yields for all locations. The yield vs. NDVI relationship for Landsat 8/9 image exposed that the multiple determination of coefficient (R^2) along with MAPE, RMSE and NRMSE for second year onion growing season i.e. 2023-2024 is fitted well than first year onion growing season i.e. 2022-2023 for all selected locations.

Key words: NDVI, Landsat 8/9, Onion, Prediction, Yield and Satellite image

Introduction

Onions (*Allium cepa L.*) of various cultivars and forms are used for consumption all over the world, and their growth depends on numerous environmental factors, such as the climate and the soil's nutrient and moisture content (Newton et al., 2011). Over 140 countries grow onion owing to its nutritional and medicinal properties. Onions can be consumed as raw, fresh, or processed and contain antioxidants and anti-inflammatory compounds that reduce cholesterol levels, diabetes, high blood pressure, and the risk of cardiovascular diseases (Lee et al., 2012; Akash et al., 2014 and Yoshinari et al., 2012). The global production of 98 million Mt of onion makes it the second most important vegetable crop after tomato (Recciardi et al., 2020; Venâncio et al., 2014 and FAO, 2021). Bangladesh is basically an agricultural country. Onion is one of the most significant profitable vegetable crops, which is used as spices in our daily meal in Bangladesh. Besides Bangladesh, it has also huge demand all over the world. It can be used in both mature and immature level as vegetable and spices, and increases the taste of food by its flavor that contains protein, calcium, carbohydrates and vitamin C. It is also used as medicine to recover from insect injury and raw throat (Bose and Som, 1990). Onion is cultivated all over the country extensively in winter season. In Bangladesh, onion ranks top among all the spice crops in production and almost 47% area covered in all spices crops (BBS, 2022). Although production of onion is increasing day by day, but in a land hungry country like Bangladesh it may not be possible to meet the domestic demand due to increase in population. There is an acute shortage of onion in relation to its requirement. Every year, Bangladesh has to import a big amount of onion from neighboring and other countries to meet up its demand (Haque et al., 2011). For these reasons, the accurate forecasting of onion yield is hugely important in ensuring the accurate determination of their product.

The precise and timely monitoring of potential yields is crucial for decision making as it influences markets, export–import decisions and farm income budgeting (Zhao et al., 2020). Currently crop statistics in Bangladesh are mostly gathered by accumulation of representative field sampling data which is time consuming and missing with information from spatial distribution of field variability. On the other hand, the prediction of crop yield before the harvest is one of the most significant concerns in agriculture since variations in crop yield from year to-year impact international trade, food supply, and market prices. However, both crop statistics and yield estimation are estimated usually through conventional means (field experiments or surveys). Currently, the use of satellite based remote sensing has become an effective way of making yield predictions due to its advantages of simple data acquisition, low cost, efficiency, wide spatial coverage, and short operating cycles (Peng et al., 2014; Zhang et al., 2019; Wang et al., 2020). The RS technology makes it possible to improve the performance of the prediction models and monitor the crop's growth status in a non-destructive and timely manner at a specific regional scale (Cai et al., 2019). Satellite based remote sensing is one of the best tools to provide vital information about the distribution of crops and its growing conditions over large areas, it can be applied for onion growth monitoring and yield forecast.

Pre-harvest crop yield prediction is essential for planning and making various policy decisions. Traditional methods are time consuming, subjective, and costly. Developing empirical models that are using weather data is also determined by several problems. Enormous types of sensors such as aerial photogrammetry, multi-spectral scanners airborne, satellite imagery, high and low spectral and spatial resolution and field-based spectrometer analysis were used to collect electromagnetic radiation information (Zhou et al., 2017). During the last few years, many empirical models have been developed to predict crop yield before the harvesting, but many of them have become unpractical, especially those are depending on filed data collection. As the satellite based remote sensing is one of the best tools to provide vital information about the distribution of crops and its growing conditions over large areas, it can be applied for onion growth monitoring and yield forecast. Nogueira et al. (2018) reported the use of vegetation indices obtained with images from the Landsat-8 satellite's OLI sensor to estimate yield. Evaluating two seasons considered low- and high-yielding, they found that NDVI had the strongest yield correlation during the dormancy and flowering stages. Zhang et al. (2019) predicted rice paddy productivity from a vegetation index extracted from Landsat 8 satellite imagery and Sari et al. (2013) utilized Landsat 8 satellite imagery to estimate rice paddy yield. The combination of data acquired by Landsat 8 and Sentinel 2 remote sensing satellites can provide a high temporal resolution (3–5 days) which is crucial for several applications requiring dense satellite data time series (Li et al., 2017). Several studies have been carried out by the correlation of normalized difference vegetation index (NDVI) with yield (Liu W.T et al., 2002). Recent studies took benefit of Landsat 8 and Sentinel 2 data to approach crop yield forecasting at a moderate spatial resolution. For example, Lai et al. (2018) applied time-integrated Landsat NDVI for wheat yield estimation in Australia. They applied an asymmetric bell-shaped growth model to fit NDVI against time. Skakun et al. (2019) applied the combination of Landsat 8 and Sentinel 2 high frequency of observations (3–5 days) at moderate spatial resolution (10–30 m), which is important for crop yield studies which were executed for the model with near infrared (NIR) and red spectral bands and derived AUC, constant, quadratic and linear coefficients of the quadratic model.

In Bangladesh, Bala & Islam (2009) expanded potato yield estimations models by using NDVI, LAI (leaf area index), and fraction of photosynthetically active radiation (fPAR) vegetation indices for Munshiganj District of Bangladesh by applying Moderate Resolution Imaging Spectroradiometer (MODIS) (with lowest resolution greater than 250m) 8-day composite surface reflectance data and noticed that an average error of estimation is about 15% for the study location. Newton et al. (2018) improved a potato yield prediction model by applying 16-day high resolution (~ 30 m) Landsat surface reflectance data to identify the maximum normalized difference vegetation index (NDVI) value of a potato growing season in Munshiganj District of Bangladesh. The maximum coefficient of determination (R^2) of yield forecasting equation was found to be 0.81 between the mean NDVI and potato yield and and the result revealed that the difference between predicted and actual filed yield is about 10.4%. However, very few studies have been conducted on the relationship between high resolution (~ 30 m) Landsat 8 satellite data and crop yield in Bangladesh. Even though, this study made an attempt to construct a onion yield prediction model

based on NDVI at Sujanagar Upazila, Pabna; Baliakandi Upazila, Rajbari and Durgapur Upazila, Rajshahi districts of Bangladesh respectively using high-resolution Landsat 8 Operational Land Imager (OLI) surface reflectance data. The use of high resolution Landsat 8/9 image has been applied in this study to improve yield assessment model for the onion crop for all selected locations in Bangladesh.

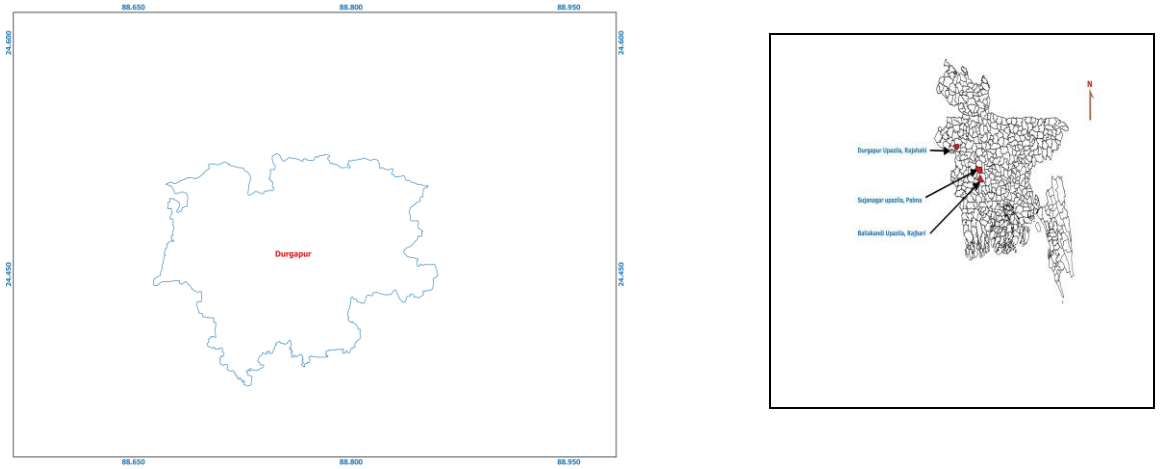
Objectives

1. To develop yield prediction model of onion crop using remote sensing technique
2. To forecast onion yield by satellite based remote sensing technique

Materials and Methods

Study Area

The study was conducted at Sujanagar, Baliakandi and Durgapur upazila in Pabna, Rajbari and Rajshahi district respectively which are the major and promising onion growing areas of Bangladesh. Three districts viz. Pabna, Rajbari and Rajshahi were covered almost 43 % area coverage as well as ranked first, third and fifth position for cultivation onion in Bangladesh. These three districts covered 20.86%, 13.94% and 7.60% onion cultivated area all over Bangladesh respectively (BBS, 2022). The three upazilas Sujanagar, Baliakandi and Durgapur lie between 23° 85′ to 24° 0′ N latitude and 89° 40′ to 89° 70′ E longitude, 23° 55′ to 23° 70′ N latitude and 89° 40′ to 89° 70′ E longitude and 24° 30′ to 24° 60′ N latitude and 89° 45′ to 89° 95′ E longitude, respectively (Figure 1). All upazilas like Sujanagar, Baliakandi and Durgapur cover 334.34 km^2 and 225.0 km^2 and 195.03 km^2 areas where 75%, 70% and 80% areas are cultivable land respectively. The climate condition of those areas are hot and humid from April to October (summer) and cool and dry from November to March (winter). The agricultural pattern of those areas are categorized by two growing seasons, one is Rabi and the rest is Kharif. Rabi is the main growing season, which is dominated by onion, oilseeds and vegetables that starts in late October or the beginning of early November and ends in April for those upazilas. On the other hand, Kharif is dominated by rice and jute, which starts in May and ends until September. Other foods which include potato, pepper, pulses, sugarcane, and wheat are also cultivated in those areas.



Durgapur, Rajshahi

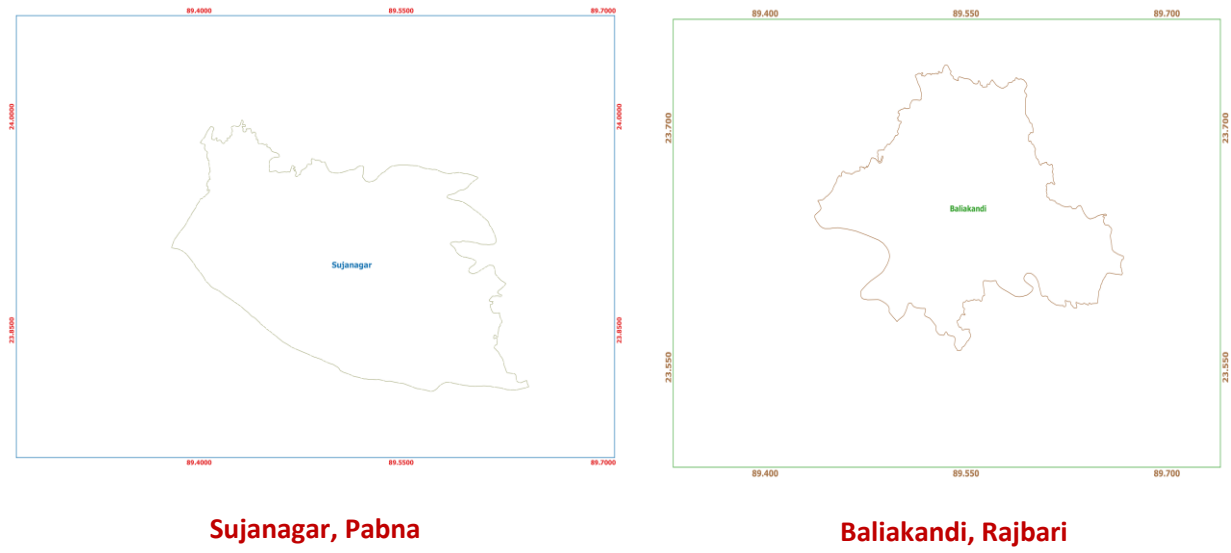


Figure 1: Study area map of selected locations

Yield data collection from farmer's fields

35 farmer's onion fields were selected for the three onion growing seasons 2022-2023, 2023-2024 and 2024-2025 respectively with agreement of the farmers at three selected upazilas (Figure 2, 3 and 4). A total of 35 farmer's different onion field data were collected from those selected locations for each season. Crop information data such as field GPS locations, planting and harvesting time and yield were collected from those selected upazilas farmers' fields.

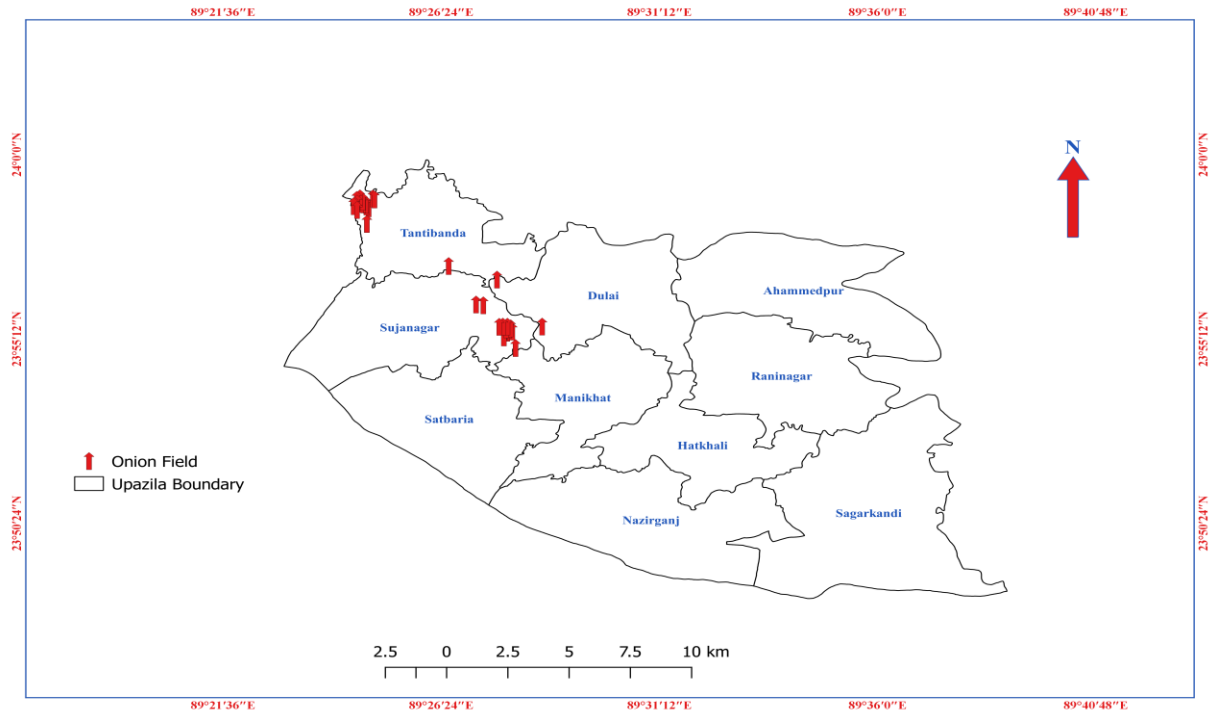


Figure 2: Location of selected 35 onion fields (red arrows) over Sujanagar, Pabna.

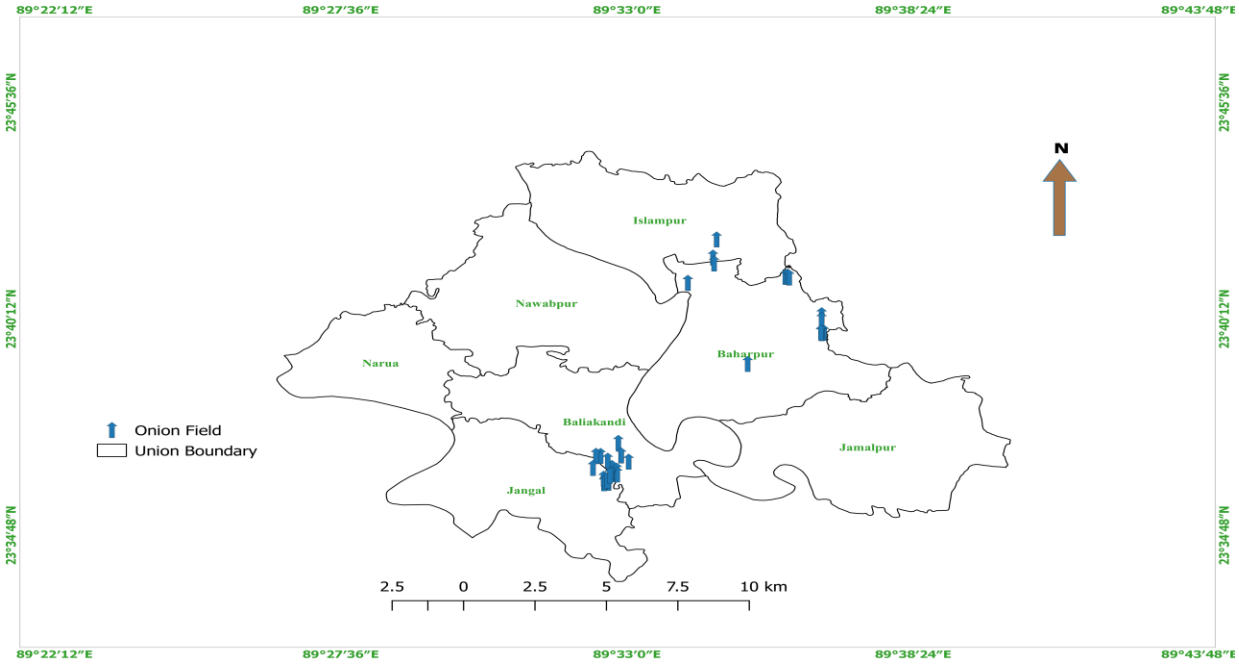


Figure 3: Location of selected 35 onion fields (blue arrows) over Baliakandi, Rajbari.

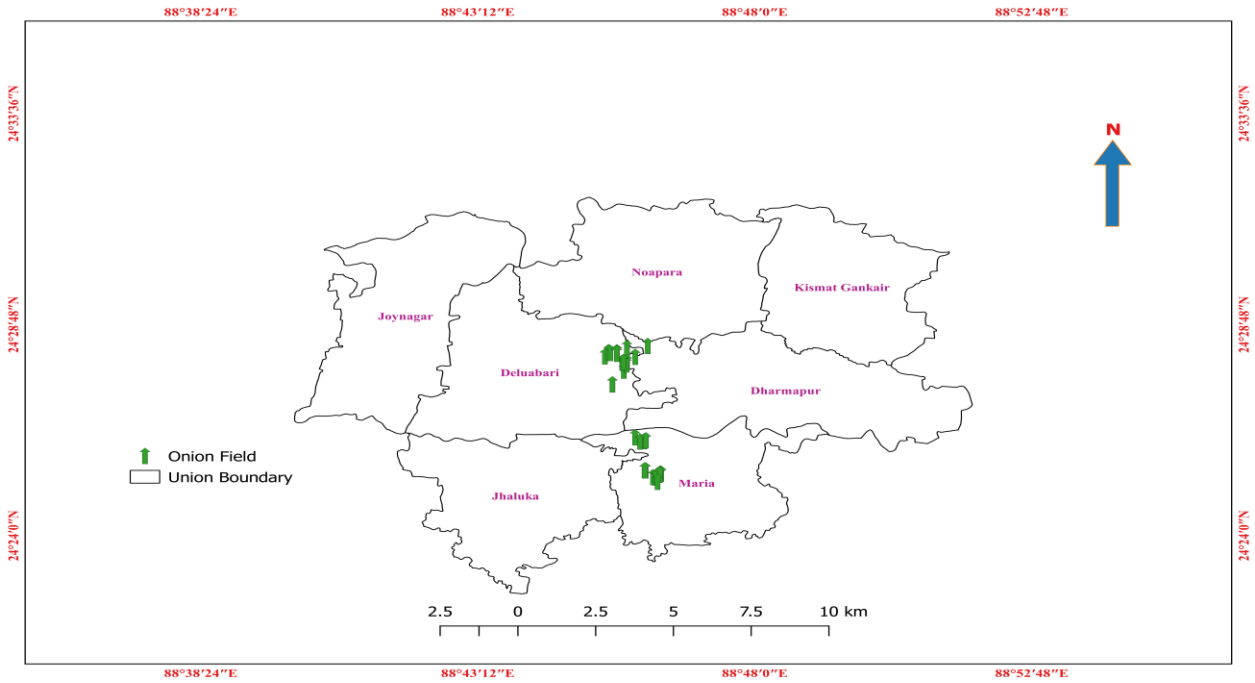


Figure 4: Location of selected 35 onion fields (green arrows) over Durgapur, Rajshahi.

Satellite Data

Landsat 8/9 Satellite Image

Satellite image Landsat 8 (OLI) was used for the yield estimation analysis, which is available on the USGS Earth Explorer website (https:// earthexplorer.usgs.gov/). Landsat 8 (OLI) is a sun-synchronous satellite staying at an altitude of 705 km above the earth with a 16-day repeat cycle. Landsat 8 has two types of sensors, especially the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS). The

OLI sensor equips nine spectral bands, including a pan band, and TIRS produces two spectral bands. The spectral and technical characteristics of Landsat 8 are presented in Table 1.

Table 1: Spectral and technical characteristics of Landsat 8

Band Number	Wavelength Range (nm)	Resolution
1	0.433–0.453	30 m
2	0.450–0.515	30 m
3	0.525–0.600	30 m
4	0.630–0.680	30 m
5	0.845–0.885	30 m
6	1.560–1.660	30 m
7	2.100–2.300	30 m
8	0.500–0.680	15 m
9	1.360–1.390	30 m
10	10.6-11.2	100 m
11	11.5-12.5	100 m

Landsat 9 was launched September 27, 2021, joins Landsat 8 in orbit; the satellite orbits are 8 days out of phase. The combined Landsat 8 + Landsat 9 revisit time for data collection with be every 8 days. Landsat 9 will collect data at the same rate as Landsat 8. Landsat 9, like Landsat 8, has a higher imaging capacity than past Landsats which is available from USGS (Figure 5). Landsat 9 have eleven spectral bands acquired by the OLI/TIRS and OLI-2/TIRS-2 instruments, respectively like Landsat 8. Landsat 9, like Landsat 8, is designed to simultaneously image four visible spectral bands, one near-infrared spectral band, three shortwave-infrared spectral bands at 30 m (98 ft) spatial resolution, plus one panchromatic band at 15 m (49 ft) spatial resolution, and two thermal bands at 100 m (328 ft) spatial resolution.

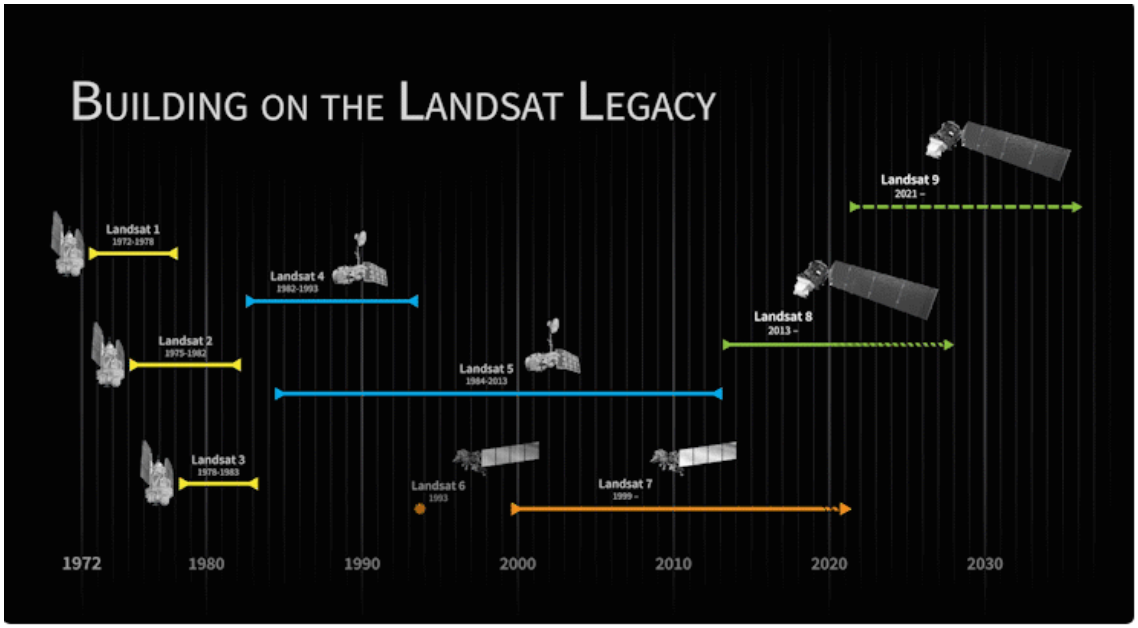


Figure 5: Timeline of the Landsat program

We downloaded total 6 images which was maximum cloud free were collected from Landsat 8 OLI satellite data for Sujanagar, Baliakandi and Durgapur upazila in Pabna, Rajbari and Rajshahi district respectively in the two years 2022-2023, 2023-2024 respectively. The single date of image acquisition based on maximum greenness was used for each growing period i.e. 2022-2023 and 2023-2024 for onion cultivation. The onion sowing date was considered to be last week of November and third week of January for those upazilas for a growing season 2022-2023 and 2023-2024 for the entire study site based on the information taken from the location visits. Every single image was calculated from the starting day

of the plantation. The dates of image acquisition of Landsat 8/9 OLI for the onion growing season 2022-2023 and 2023-2024, for those locations used for this study are presented in Table 2.

Table 2: Model development of onion using Landsat 8/9 for selected locations

Locations	Sujanagar, Pabna		Baliakandi, Rajbari		Durgapur, Rajshahi	
Growing Season	2022-2023	2023-2024	2022-2023	2023-2024	2022-2023	2023-2024
IAD	10/02/2023	28/01/2024	26/02/2023	28/01/2024	14/03/2023	08/03/2024
DAP	72	75	73	73	77	72

IAD=Image acquisition date; DAP= Days after plantation

Satellite Image Pre-Processing

For Landsat data, raw digital numbers (DN) were adjusted to top-of-atmosphere (TOA) reflectance values following reference (Simonetti et al., 2015). Two techniques were used to preprocess the satellite images: (1) radiometric calibration and (2) atmospheric correction. Remote sensing data adopted from satellite sensors are influenced by several factors, such as atmospheric scattering and absorption, sensor-target-illumination geometry, sensor calibration, and also by the data processing procedures (Teillet, 1986). For that, radiometric calibration is needed. Radiometric calibration means a set of correction techniques that are associated to correction for the sensitivity of satellite sensor, topography and sun angle, atmospheric scattering, and absorption (Kim et al., 1990). The radiometric calibration was done by transforming the digital numbers (DNs) to surface reflectance by radiance conversion. Open source-based Quantum Geographic Information System (QGIS) software 3.28.2 version allows a plugin, and the plugin gives a tool for atmospheric correction, which is known as dark object subtraction (DOS-1) level 1. In this study, this tool was used in the radiometrically calibrated images to minimize atmospheric scattering effect. DOS-1 searches each pixel of a band to find the darkest value. The scattering is eliminated by subtracting this value from every pixel in the band.

Normalized Difference Vegetation Index (NDVI)

The NDVI index is indicated in crops where at maximum vigour the bare soil component is missing in the measured reflectance from the satellite, so it comes from vegetation reflectance only (Fuente et al., 2023). Differences in the phenological growth stages of different plants are reflected in the temporal NDVI profiles, since NDVI can measure growth conditions (greenness of vegetation) (Belgiu & Csillik, 2018; Croitoru et al., 2012). It is a dimensionless index, which is performed from the ratio between the surface reflectance of the NIR and RED bands of the spectrum as follows (Equation 1) (Rouse et al., 1974).

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

Where RED (Visible red) and NIR (Near infrared) are reflectance measurements for RED and NIR bands, respectively. The factors like strong reflectance in NIR and strong absorption in Visible Red of specific vegetation distinguish the vegetation from bare soil. NDVI for a given pixel can always output in a number that ranges from - 1 to +1; however, for natural surfaces NDVI values are within the 0 to +1 range. Negative values of NDVI i.e. values approaching -1 correspond to water. An NDVI close to 0 corresponds to no vegetation, while Values lies between -0.1 to 0.1 generally corresponded to barren areas of rock, sand or snow.

Yield prediction model

The final step is to determine the relationship between NDVI and onion yield form farmers' field with the equation below:

$$y = f(x) \quad (2)$$

Where y and x are onion yield data collected from farmers' field and NDVI, respectively. The relationship between NDVI and crop like onion yield have been observed through the linear regression model, where the response variable denoted by onion yields and the explanatory variables by NDVIs. To develop the onion yield estimation model for those fields, the data of onion yield and Landsat 8/9 (OLI) image was used for first year 2022-2023 and 2023-2024 respectively.

Performance evaluation of yield prediction model

The coefficient of multiple determinations (R^2), the root mean square error (RMSE), mean absolute percentage error (MAPE) and the normalized RMSE (nRMSE) were selected as performance evaluation metrics. These metrics are mathematically defined as:

$$R^2 = \frac{\sum_{i=1}^{n} E_i \times O_i - \sum_{i=1}^{n} E_i \times \sum_{i=1}^{n} O_i}{\sqrt{\sum_{i=1}^{n} E_i{}^2 - (\sum_{i=1}^{n} E_i)^2) \times \sqrt{\sum_{i=1}^{n} O_i{}^2 - (O_i)^2}}} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(E_i - O_i)^2}{n}} \quad (4)$$

$$MAPE = \frac{\frac{1}{n}\sum_{i=1}^{n}|E_i - O_i|}{O_i} \times 100 \quad (5)$$

$$nRMSE = \sqrt{\frac{\sum_{i=1}^{n}(E_i - O_i)^2}{n}} \times \frac{100}{\bar{O}} \quad (6)$$

Where, E_i is the estimated values, O_i is the corresponding observed values, $\bar{O}$ is the mean value of O_i and n is the number of samples.

Generally, the larger R^2 is, the smaller the RMSE and MAPE is, and the better the model fit is. The range of the nRMSE metric generally defines the model accuracy. A value of nRMSE < 10% indicates that the estimated and measured values are highly consistent, the range 10% < nRMSE < 20% indicates good consistency, while the range 20% ≤ nRMSE < 30% indicates medium consistency, and finally the range nRMSE ≥ 30% indicates poor consistency.

Results and Discussion

Onion yield from farmers' field and corresponding NDVI values for different locations

Onion yield data and NDVI values from Landsat 8/9 data for corresponding farmers' fields have been collected from different locations viz. Sujanagar, Pabna; Baliakandi, Rajbari and Durgapur, Rajshahi during the two onion growing seasons 2022-2023 and 2023-2024 respectively. 35 farmers' field yield data collected from those upazilas and their corresponding NDVI values for Landsat 8/9 satellite image have been presented in a Table 3 and 4 for onion growing seasons 2022-2023 and 2023-2024 respectively.

Table 3: NDVI values of Landsat 8/9 satellite image and yields of corresponding Farmer's fields at Selected locations during the season of 2022-2023.

Farmer's Field	Sujanagar, Pabna		Baliakandi, Rajbari		Durgapur, Rajshahi	
	NDVI	Yield (t/ha)	NDVI	Yield (t/ha)	NDVI	Yield (t/ha)
1	0.63	15.2	0.67	16.5	0.46	14.8
2	0.70	17.5	**0.77**	**19.3**	0.55	15.5
3	0.70	17.6	0.35	10.2	0.32	11.2
4	**0.79**	**19.8**	**0.31**	**8.90**	0.55	16.7
5	0.71	18.2	0.67	16.8	0.47	14.5

6	0.73	18.8	0.60	15.4	0.48	15.2
7	0.73	18.5	0.53	11.4	0.37	12.8
8	0.74	18.8	0.54	12.5	0.39	13.5
9	0.70	17.3	0.58	14.5	0.61	17.2
10	0.76	19.3	0.61	14.8	0.33	13.1
11	0.74	18.5	0.72	17.6	0.41	13.8
12	0.72	18.8	0.51	10.5	0.45	14.5
13	0.68	14.0	0.54	11.8	0.62	17.5
14	0.68	14.5	0.72	17.8	0.49	15.1
15	0.70	16.4	0.55	9.88	**0.67**	**21.5**
16	0.47	12.5	0.56	13.5	0.65	20.2
17	0.50	11.0	0.74	18.5	0.43	14.3
18	0.65	15.7	0.51	12.8	0.33	11.7
19	0.51	13.1	0.56	14.8	0.48	15.5
20	0.45	12.2	0.60	15.6	**0.30**	**8.46**
21	**0.41**	**9.15**	0.74	17.8	0.33	12.5
22	0.64	15.2	0.52	12.5	0.47	14.1
23	0.65	15.6	0.55	13.5	0.46	10.5
24	0.63	15.5	0.74	18.2	0.47	11.5
25	0.72	15.0	0.54	13.8	0.45	10.3
26	0.61	13.5	0.60	14.7	0.52	15.2
27	0.60	14.5	0.55	13.7	0.58	17.2
28	0.63	15.0	0.57	14.1	0.46	12.8
29	0.61	13.0	0.54	14.2	0.45	12.1
30	0.55	12.5	0.54	13.8	0.63	18.4
31	0.64	15.5	0.53	12.8	0.48	14.5
32	0.64	15.3	0.67	15.3	0.39	14.8
33	0.70	17.2	0.62	15.8	0.59	18.2
34	0.62	14.5	0.65	14.8	0.53	17.5
35	0.59	14.2	0.54	11.5	0.54	18.3

Table 4: NDVI values of Landsat 8/9 satellite image and yields of corresponding Farmer's fields at Selected locations during the season of 2023-2024.

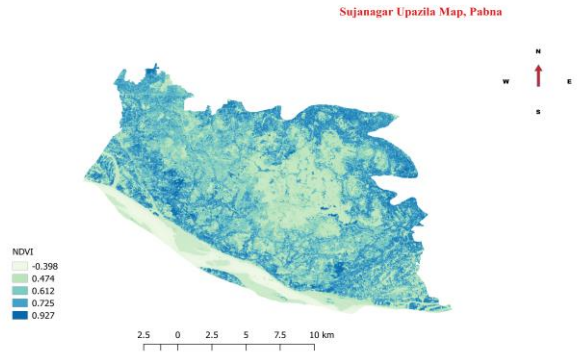
Farmer's Field	**Sujanagar, Pabna**		**Baliakandi, Rajbari**		**Durgapur, Rajshahi**	
	NDVI	**Yield (t/ha)**	**NDVI**	**Yield (t/ha)**	**NDVI**	**Yield (t/ha)**
1	0.65	16.5	0.61	16.2	0.53	15.82
2	0.75	18.2	**0.79**	**19.4**	0.37	11.2
3	0.72	17.4	0.32	10.7	0.37	11.2
4	**0.84**	**19.5**	**0.3**	**9.6**	0.64	17.8
5	0.69	17.2	0.59	15.2	0.6	17.2
6	0.76	18.6	0.63	15.8	0.61	17.5
7	0.65	17.3	0.63	16.1	0.58	16.2
8	0.75	18.3	0.61	15.5	0.52	15.45
9	0.71	17.4	0.64	16.4	0.67	18.1
10	0.82	18.9	0.58	14.2	0.39	11.8
11	0.74	17.8	0.72	17.2	0.69	18.8
12	0.73	18.5	0.6	15.8	0.57	15.62
13	0.64	14.5	0.55	13.6	0.55	14.95
14	0.61	14.8	0.69	16.9	0.51	14.35
15	0.51	11.2	0.57	12.8	0.63	17.45
16	0.42	11.8	0.58	13.9	0.67	18.25
17	0.45	12.7	0.71	17.1	0.58	16.35

18	0.62	15.6	0.58	13.8	0.42	14.25
19	0.48	13.4	0.59	14.8	0.75	21.25
20	0.42	12.2	0.61	16.2	**0.35**	**10.85**
21	**0.38**	**11.1**	0.71	17.6	0.42	13.95
22	0.68	16.1	0.56	12.6	0.52	13.9
23	0.62	15.6	0.55	13.1	0.52	14.25
24	0.61	15.2	0.71	17.8	0.52	14.3
25	0.68	16.5	0.6	15.7	0.7	15.6
26	0.58	15.3	0.55	13.5	0.7	20.85
27	0.61	15.4	0.57	13.8	0.7	19.98
28	0.64	15.3	0.56	13.7	0.52	13.95
29	0.6	13.5	0.53	12.6	**0.78**	**21.65**
30	0.51	11.3	0.56	13.2	0.72	20.85
31	0.65	15.8	0.56	13	0.75	14.6
32	0.66	15.6	0.67	15.8	0.63	17.75
33	0.68	16.4	0.68	16.2	0.77	21.4
34	0.66	15.8	0.65	15.3	0.73	20.55
35	0.52	12.8	0.58	14.4	0.78	21.5

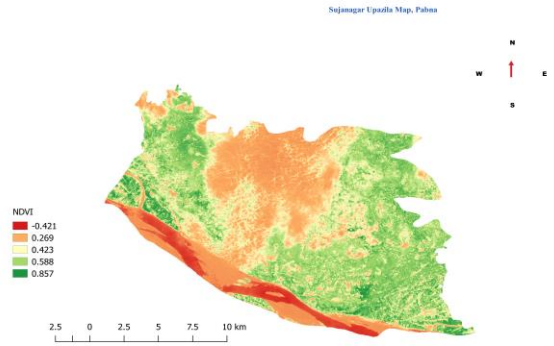
Table 3 shows that the highest NDVI value was 0.79 and the lowest was 0.41 and the maximum yield was 19.80 t/ha and the minimum was 9.15 t/ha for farmer's field 4 and 21 respectively for Sujanagar upazila, Pabna. On the other hand, the maximum and the minimum NDVI value and yield for Baliakandi upazila, Rajbari were 0.77 and 0.31 and 19.30 t/ha and 8.90 t/ha for the farmer's field 2 and 4 respectively; the highest and the lowest NDVI value and yield for Durgapur upazila, Rajshahi were 0.67 and 0.30 and 21.50 t/ha and 8.46 t/ha for the farmer's field 15 and 20 respectively for the onion growing season 2022-2023. Consequently, the maximum and minimum NDVI were 0.84 and 0.38 at the farmer's field 4 and 21 for Sujanagar, Pabna; 0.79 and 0.30 at the farmer's field 2 and 4 for Baliakandi, Rajbari and 0.78 & 0.35 at the farmer's field 29 and 20 for Durgapur, Rajshahi for the onion growing season 2023-2024 (Table 4)

Regression analysis of the NDVI values over the field locations

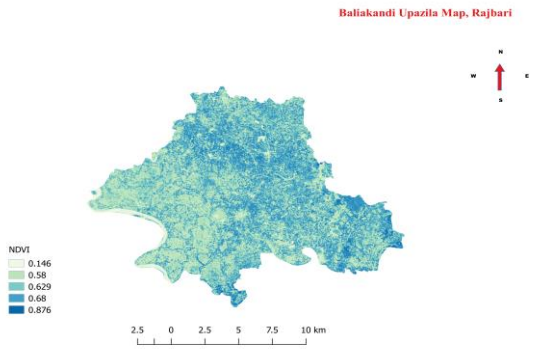
A total of three Landsat 8/9 satellite images during the onion growing season 2022-2023 and 2023-2024 were selected from different locations. Based on available images, the maximum NDVI were found 72th, 73th and 77th days after plantation for first year and 75th, 73th and 72th days after plantation for second year for Sujanagar, Pabna; Baliakandi, Rajbari and Durgapur, Rajshahi respectively during the onion growing season 2022-2023 and 2023-2024 respectively. The spatial distribution of the NDVI varies from year to year. Spatial distribution of the NDVI over the selected location for selected distinct satellite images against growing season 2022-2023 and 2023-2024 were presented in Figure 6.



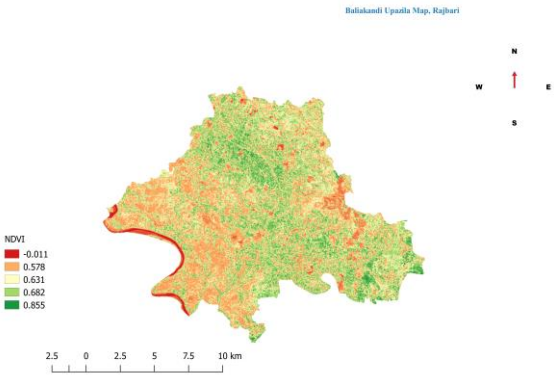
a. NDVI Map Sujanagar, 2022-2023



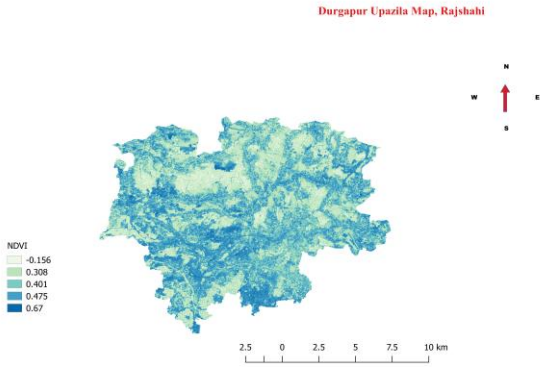
d. NDVI Map Sujanagar, 2023-2024



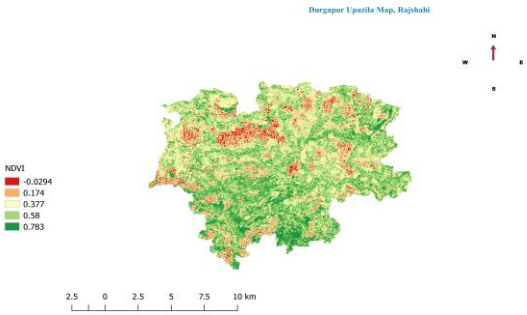
b. NDVI Map Baliakandi, 2022-2023



e. NDVI Map Baliakandi, 2023-2024



c. NDVI Map Durgapur, 2022-2023



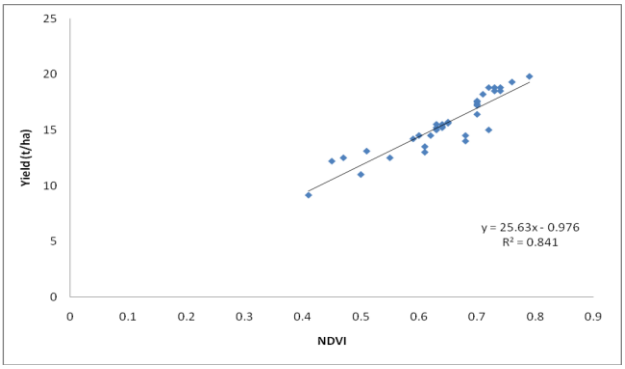
f. NDVI Map Durgapur, 2023-2024

Figure 6: Spatial distribution of the NDVI for Landsat 8/9 images during growing season 2022-2023 and 2023-2024. a+d. 72th & 75th days after plantation at Sujanagar, Pabna; b+e. 73th & 73th days after plantation at Baliakandi, Rajbari; c+f. 77th & 72th days of plantation at Durgapur, Rajshahi.

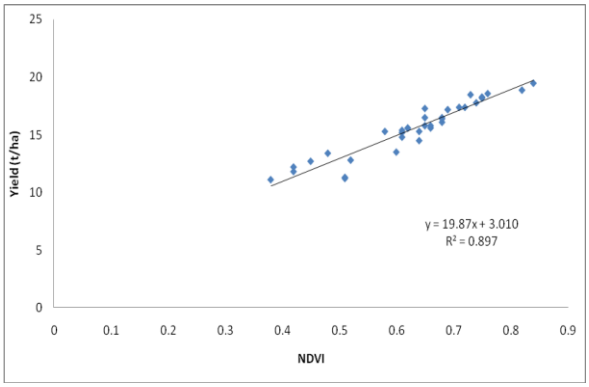
Maximum NDVI distribution from Landsat 8 satellite data was found for Suganagar, Pabna and followed by Baliakand Rajbari and Durgapur, Rajshahi respectively during the onion growing season 2022-2023 and 2023-2024 respectively (Figure 5).

Onion yield and NDVI relationship using regression model

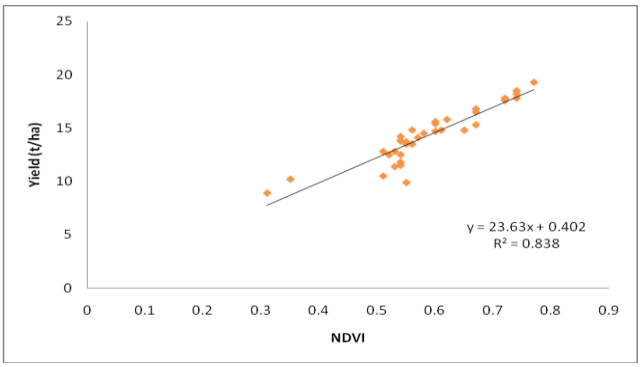
Regression analysis of onion yield against the single season basis (2022-2023) NDVI for different selected locations like Sujanagar, Pabna; Baliakandi, Rajbari and Durgapur, Rajshahi was performed for Landsat 8/9 satellite image was graphically presented in Figure 6.



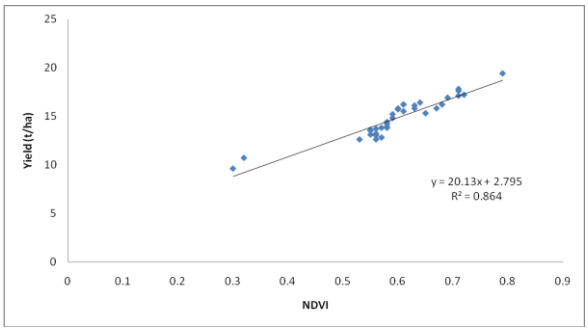
a. Yield vs NDVI; Sujanagar, 2022-2023



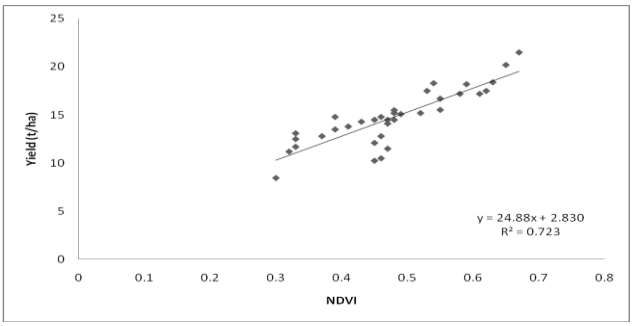
d. Yield vs NDVI; Sujanagar, 2023-2024



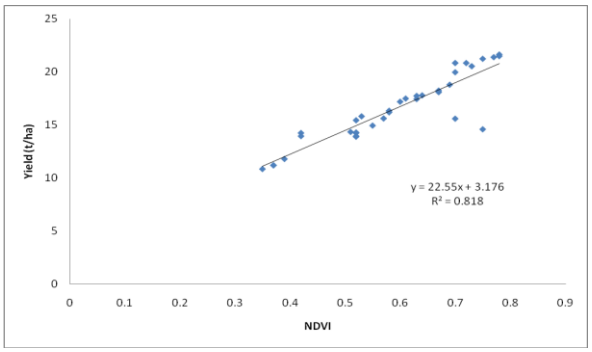
b. Yield vs NDVI; Baliakandi, 2022-2023



e. Yield vs NDVI; Baliakandi, 2023-2024



c. Yield vs NDVI; Durgapur, 2022-2023



f. Yield vs NDVI; Durgapur, 2023-2024

Figure 6: Yield prediction model established from regression analysis between yield data collected from 35 farmers' onion field for different images for Landsat 8/9 during the onion growing season 2022-2023 and 2023-2024.

The relationship between onion yield and NDVI are showed positive correlation among all selected locations for the two years growing season 2022-2023 and 2023-2024 respectively. Here, the regression coefficients of all the fitted model of Landsat 8/9 image are showing highly significant effect for selected locations. The yield vs. NDVI relationship for Landsat 8/9 satellite image are shown in Figure 6 revealed that R^2 which are highest and other accuracy criterion like MAPE, RMSE and **N**RMSE of NDVI for second year onion growing season i.e., 2023-2024 is outperformed than the first year onion growing season i.e., 2022-2023 for all selected locations. The parameters of the regression analysis estimated from yield vs. NDVI relationship for 2022-2023 and 2023-2024, together with the value of R^2, MAPE, RMSE and NRMSE are presented in Table 5.

Table 5: Regression parameter and model criterion for two growing season at Farmers' onion field

Location	Growing Seasons	Regression parameter				Best Model Criteria			
		β_0	β_1	SE(β_1)	P-Value	MAPE	RMSE	NRMSE	R^2
Sujanagar	2022-2023	-0.98	25.63	1.940	0.0000	0.051	0.998	6.43	0.841
	2023-2024	**3.01**	**19.87**	**1.172**	**0.0000**	**0.040**	**0.744**	**4.79**	**0.898**
Baliakandi	2022-2023	0.40	23.63	1.808	0.0000	0.061	1.025	7.18	0.838
	2023-2024	**2.79**	**20.13**	**1.386**	**0.0000**	**0.045**	**0.741**	**4.99**	**0.864**
Durgapur	2022-2023	2.83	24.88	2.681	0.0000	0.087	1.49	10.15	0.723
	2023-2024	**3.18**	**22.55**	**1.852**	**0.0000**	**0.052**	**1.33**	**8.03**	**0.817**

The yield vs. NDVI relationship for Landsat 8/9 image for all selected locations like Sujanagar, Pabna; Baliakandi, Rajbari and Durgapur, Rajshahi revealed that the estimated model along with the performance evaluation of second year onion growing season (2023-2024) was performed well than first year onion growing season (2022-2023) in Table 5.

Conclusions

The study has investigated the prediction capacity of remote sensing NDVI data for onion yield in selected three locations viz. Sujanagar, Pabna; Baliakandi, Rajbari and Durgapur, Rajshahi of Bangladesh which is known for onion dominating district. In the farmers' fields of the selected locations, the developed system investigates the combined use of satellite remote sensing (RS) and Geographic Information System (GIS) technology. The study has also investigated the relationship between NDVI and yield for the study region. Here the high spatial resolution satellite image Landsat 8/9 (OLI) was used in this study. The single date of cloud free image acquisition based on maximum NDVI was used for 2022-2023, 2023-2024 cropping period for onion cultivation to develop the yield prediction model. The two year's results revealed that in most cases the yield was maximum for the field where the NDVI values were not maximum and vice-versa. Nevertheless, the relationships of the extracted NDVI and yields of onion crop will be established using the classical linear regression model where the model will be developed using the first two years' data and will be validated using the data of the last study period for all selected locations.

References

Akash MSH, Rehman K, Chen S (2014) Spice plant Allium cepa: Dietary supplement for treatment of type 2 diabetes mellitus. Nutrition 30:1128–1137. https://doi.org/10.1016/j.nut.2014.02.011

Bala, S. K., Islam, A. S. (2009). Correlation between potato yield and MODIS-derived vegetation indices. *Int J Remote Sens* 30(10): 2491–2507

BBS, (2022). Bangladesh Bureau of Statistics, Statistical Yearbook of Bangladesh, Statistics Division, Ministry of Planning, Government of the People's Republic of Bangladesh, Dhaka.

Belgiu, M., and Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote sensing of environment*, 204, 509–523.

Bose, T. K. and Som, M. G. (1990). Vegetable crops in India. Naya Prokash, Calcutta, 6:545-582.

Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Peng, B. (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agric. For. Meteorol. 274, 144–159.

Croitoru, A.-E., Holobaca, I.-H., Lazar, C., Moldovan, F., & Imbroane, A. (2012). Air temperature trend and the impact on winter wheat phenology in Romania. *Climatic Change*, 111, 393–410.

FAO (2021) Crop Statistics, FAOSTAT. In: Food Agric. Organ. United Nations. http://faostat.fao.org. Accessed 15 Jul 2022

Fuente, D., Rivilla, E., Tena, A., Vitorino1, J., Navascués, E., and Tabasco, A. (2023). Yield estimation using machine learning from satellite imagery. BIO Web of Conferences 68, 01013 (2023) https://doi.org/10.1051/bioconf/20236801013

Haque, M. A., Miah, M. A. M., Hossain, S., Rahman, M. S. and Moniruzzaman, (2011). Profitability of Onion Cultivation in Some Selected Areas of Bangladesh, Bangladesh Journal of Agricultural Research, 36(3):427-435.

Kim, H. H. and Elman, G. C. (1990). Normalization of satellite imagery. *Int J Remote Sens* 11(8):1331–1347.

Lai, Y. R.; Pringle, M. J.; Kopittke, P. M.; Menzies, N. W.; Orton, T. G. and Dang, Y. P. (2018). An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI. *Int. J. Appl. Earth Obs. Geoinf*. 2018, 72, 99–108.

Lee B, Jung J-H, Kim H-S (2012) Assessment of red onion on antioxidant activity in rat. Food Chem Toxicol 50:3912–3919. https://doi.org/10.1016/j.fct.2012.08.004

Li, J.; Roy, D.P. (2017). A Global Analysis of Sentinel-2A, Sentinel-2B and Landsat-8 Data Revisit Intervals and Implications for Terrestrial Monitoring. *Remote Sens*. **2017**, 9, 902.

Liu, W. T. and Kogan, F. (2002). Monitoring Brazilian soybean production using NOAA/AVHRR based vegetation condition indices. *Int J Remote Sens* 23(6):1161–1179

Newton, A.C., Johnson, S.N., Gregory, P.J. (2011). Implications of climate change for diseases, crop yields and food security. Euphytica 179 (1), 3–18.

Newton, I. H, Islam, A. F. M T., Islam, A. K. M. S., Tarekul, G. M., Anika, T. and Razzaque, S. (2018). Yield Prediction Model for Potato Using Landsat Time Series Images Driven Vegetation Indices. *Remote Sensing in Earth Systems Sciences* (2018) 1:29–38.

Nogueira, S. M. C., Moreira, M. A., Volpato, M. M. L. (2018). Relationship between Coffee Crop Yield and Vegetation Indexes Derived from Oli/Landsat-8 Sensor Data with and without Topographic Correction. Eng. Agric. 2018, 38, 387–394.

Peng, D., Huang, J., Li, C., Liu, L., Huang, W., Wang, F., et al. (2014). Modelling paddy rice yield using MODIS data. Agric. For. Meteorol. 184, 107–116. doi: 10.1016/j.agrformet.2013.09.006

Ricciardi L, Mazzeo R, Marcotrigiano AR, et al (2020) Assessment of Genetic Diversity of the "Acquaviva Red Onion" (Allium cepa L.) Apulian Landrace. Plants 9:260. https://doi.org/10.3390/plants9020260

Sari, D.K., Ismullah, I. H., Sulasdi, W. N. et al. (2013). Estimation of water consumption of lowland rice in tropical area based on heterogeneous cropping calendar using remote sensing technology. Procedia Environ. Sci. 2013; 17: 298–307.

Simonetti, D.; Marelli, A.; Eva, H. (2015). IMPACTool Box: Portable GIS Toolbox for Image Processing and Land Cover Mapping; Publications Oce of the European Union: Luxembourg, 2015; ISBN 978-92-79-50115-9.

Skakun, S., Vermote, E., Franch, B., Roger, J. C., Kussul, N., Ju, J. and Masek, J. (2019). Winter Wheat Yield Assessment from Landsat 8 and Sentinel-2 Data: Incorporating Surface Reflectance, Through Phenological Fitting, into Regression Yield Models. *Remote Sens*. 2019, 11, 1768:1-19; doi:10.3390/rs11151768. www.mdpi.com/journal/remotesensing.

Teillet, P. M. (1986). Image correction for radiometric effects in remote sensing. *Int J Remote Sens* 7(12):1637–1651.

Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ*. 1979, 8, 127–150.

Venâncio JB, Dias N da S, de Medeiros JF, et al (2022) Yield and Morphophysiology of Onion Grown under Salinity and Fertilization with Silicon. Sci Hortic (Amsterdam) 301:111095. https://doi.org/10.1016/j.scienta.2022.111095

Wang, X., Huang, J., Feng, Q., and Yin, D. (2020).Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches. Remote Sens. 12 (11), 1744. doi: 10.3390/rs12111744

Yoshinari O, Shiojima Y, Igarashi K (2012) Anti-Obesity Effects of Onion Extract in Zucker Diabetic Fatty Rats. Nutrients 4:1518–1526. https://doi.org/10.3390/nu4101518

Zhang, B., Chen, Z., Peng, D., Benediktsson, J., Liu, B., Zou, L., et al. (2019). Remotely sensed big data: Evolution in model development for information extraction. Proc. IEEE. 107 (12), 2294–2301. doi: 10.1109/JPROC.2019.2948454

Zhang, K., Ge, X., Shen, P. et al. (2019). Predicting rice grain yield based on dynamic changes in vegetation indexes during early to mid-growth stages. Remote Sens. 2019; 11(4): 1–24.

Zhao, Y., Potgieter, A. B., Zhang, M., Wu, B., & Hammer, G. L. (2020). Predicting wheat yield at the field scale by combining high-resolution sentinel-2 satellite imagery and crop modelling. *Remote Sensing*. https:// doi. org/ 10. 3390/ rs120 61024

Zhou, X., Zheng, H.B., Xu, X.Q., He, J.Y., Ge, X.K., Yao, X., Cheng, T., Zhu, Y., Cao, W.X., Tian, Y.C. (2017). Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. ISPRS Journal of Photogrammetry and Remote Sensing 130, 246–255.

IN SILICO GENOME-WIDE IDENTIFICATION, CHARACTERIZATION, AND PHYLOGENETIC ANALYSIS OF THE DICER-LIKE (DCL), ARGONAUTE (AGO), AND RNA-DEPENDENT RNA POLYMERASE (RDR) GENE FAMILIES IN BRASSICA SPECIES (Brassica napus L.)

[1]ZOBAER AKOND, [2] SHEIKH HASNA HABIB, [3] NURUL HAQUE MOLLAH
[1]ASICT, BARI, GAZIPUR; [3]ORC, BARI; [3]BIOINFORMATICS LAB, DEPT. OF STATISTICS, UNIVERSITY OF RAJSHAHI.

Abstract

Three major gene families including Dicer-like (DCL), Argonaute (AGO), and RNA-dependent RNA polymerase (RDR) are popularly known as the RNAi-related genes in eukaryotes to control gene expression at the transcriptional or post-transcriptional level and often maintains plant growth and development. These gene families also control gene expression in response to various pathogens and abiotic factors. This mechanism is typically continued by the 19-24 nt size non-coding small RNA molecules (miRNA and siRNA). However, detailed and updated bioinformatics investigation including phylogeny analysis, domain and gene structures, GO enrichment, subcellular localization, etc. have not been done yet. Our study finally identified 8 *DCL*, 27 A*GO*, and 17 *RDR* RNAi-related genes in *Brassica napus*. The phylogenetic analysis of the total 52 BnRNAi proteins with the 20 AtRNAi proteins showed that protein members of BnDCL, BnAGO, and BnRDR groups clustered into four, nine, and five subgroups. Domain and exon-intron composition demonstrated that the protein/gene members also hold the same characteristics similar to their AtRNAi counterparts. Some replacements (R-arginine, S-serine-S, and P-proline) in 14 BnAGO proteins were also found in DDH/H motifs of the PIWI domain which signifies their role in yield production and different pathogenic responses. GO enrichment analysis for BP and MF also concluded that a pool of biological and molecular functions is significantly linked (p-value<0.05) to the RNAi pathway mechanism in *B.napus* like other crops. Subcellular localization analysis illustrated that the maximum BnRNAi proteins were the nucleus, chloroplast, and cytoplasmic-oriented. Thus, our results would deliver an excellent source for deep molecular investigation of these genes/proteins to explore their functional potentiality for future rapeseed-mustard crop improvements against different environmental factors in Bangladesh.

Keywords: RNAi genes, *Brassica napus*, Bioinformatics Analysis

1. Introduction

Crops by nature create some particular molecular mechanisms during their life span to survive in adverse environments. Gene silencing called RNA interference (RNAi) is one of them. This is a powerful mechanism that is highly conserved in most multicellular eukaryotic organisms. Typically, some non-coding small RNA molecules (miRNA and siRNA) take part in this mechanism to control gene expression at the transcription and post-transcription (PTS) level by targeting complementary RNAs and defending cells from harmful exogenous and endogenous genetic materials (Yun and Zhang, 2023). Notably, RNAi plays a highly important role in plants' various biological, molecular, and cellular processes such as growth and development, epigenetic modifications, maintaining genomic stability, and different pathogenic and environmental stress responses(Qian et al., 2011a; Cui et al., 2020a).

Three key gene groups viz., Dicer-like (DCL), Argonaute (AGO), and RNA-dependent RNA polymerase (RDR) take part in the RNAi mechanism(Cao et al., 2016; Krishnatreya et al., 2021). RNAi cycle is first initiated by the double-stranded RNA (dsRNA) that is sliced by RNaseIII-type DCL proteins into small RNA (sRNA) molecules of length 21-24 nucleotide termed as micro-RNA (miRNA) or short-interfering

RNA (siRNA)(Cui et al., 2020a; Yun and Zhang, 2023). Next, one strand of these miRNAs/siRNAs is combined with AGO protein to create a pre-RNA-induced silencing complex(pre-RISC) that needs the molecular chaperone Heat shock protein 70 (Hsp70)/90(Hsp90) (Iwasaki et al., 2010; Yun and Zhang, 2023). The sRNA is merged by the function of the N-domain of AGO and only the guide RNA strand exists in the complex to produce a complete RISC(Mirzaei et al., 2014; Akond et al., 2020; Yun and Zhang, 2023). The RISC binds with complementary miRNAs led by single-stranded sRNA to restrict translation during post-transcriptional gene silencing (PTGS), RNA degradation, and creation of heterochromatin during transcriptional gene silencing (TGS), which leads to particular gene silencing. These sRNAs are produced from the targeted RNA by the host cell encoded RNA-dependent RNA polymerases (RDRs) to generate more dsRNAs to process miRNAs/siRNAs and to start a new round of RNA silencing.

The proteins of these three families contain some particular domains that measure their important characteristics (structure and function). The proteins of the DCL family are identified for the presence of six domains, viz., a DEAD domain, a helicase conserved C-terminal (Helicase C) domain, a Dicer dimerization domain (Dicer dimer/Duf283), a PAZ (Piwi Argonaute and Zwille) domain, a Ribonuclease III domain (Ribonuclease 3), and a double-stranded RNA-binding domain (DSRM)(Cui et al., 2020a). The proteins of the AGO family are characterized and identified for the presence of variable MID and N-terminal domains and well-conserved PAZ and PIWI (P-element-induced wimpy testis) domains(Cui et al., 2020a; Akond et al., 2022a). The proteins of the RDR family are crucially important to continue the RNA silencing process. The proteins of this gene group contain a conserved RNA-dependent RNA polymerase (RDRP) catalytic domain(Cui et al., 2020a; Jing et al., 2023).

Investigations so far of these three gene families have been carried out on more than two dozen plant species. Results showed a varying number of gene/protein members of these three groups are present such as the Arabidopsis, rice, maize, foxtail millet, grapevine, tomato, wheat, soybean, pepper, cucumber, barley, sugarcane, sweet orange, tea, sorghum, quinoa, tobacco, coffee, sunflower, banana, common bean, walnut, and date palm possess four, five, five, eight, four, seven, seven, seven, four, five, five, four, four, five, five, eight, four, nine, five, three, six, five and four DCL genes; 10, 19, 18, 19, 13, 15, 39, 21, 12, seven, 11, 21, eight, 18,14, 21, nine,11, 15, 13, 13, 15 and seven AGO genes; six, eight, five, 11, five, six, 16, seven, six, eight, seven, 11, four, nine, seven, 11, three, eight, 10, five, five, 13 and three RDR genes, respectively(Cui et al., 2020a; Faysal et al., 2021; Akond et al., 2022a; Alvarez-Diaz et al., 2022; Fang et al., 2023; Podder et al., 2023; Naim et al., 2024).

The previous studies provided sufficient evidence that these gene families are strongly conserved in plant species with diversified characteristics and functions, but broad-scale analyses of these genes have not yet been done in *Brassica napus*. For example, the Arabidopsis thaliana DCL1(*AtDCL1*) shows the properties of miRNA biogenesis whereas *AtDCL2*/*DCL3*/*DCL4* mediate siRNA processing(Wang et al., 2016). *AtDCL3*/*DCL4* are also important for RNA-directed DNA methylation of the FWA transgene, which is related to histone H3 lysine 9 (H3K9) methylation(Qian et al., 2011a; Wang et al., 2016). Although *AtDCL2* generates siRNAs to form the defensive process against viral infection and *cis*-acting antisense transcript, *AtDCL4* controls the vegetative period changed by the generation of siRNAs from the trans-acting transcript(Deleris et al., 2006a). Interestingly, *AtDCL3* chooses short dsRNAs and *AtDCL4* cuts the long dsRNA substrates(Wang et al., 2016; Akond et al., 2022a). DCL genes in plants such as *AtDCL1* and *AtDCL3* increase flowering(Schmitz et al., 2007). Recently a study has identified that three proteins viz., *Pt-DCL1*, *Pt-DCL2*, and *Pt-DCL3* are strongly related to the leaf rust pathogen (*Puccinia triticina*) that is known as a very destructive disease of wheat(Dubey et al., 2020).

Argonaute proteins also play key activities during the RNA-guided gene-silencing process for plants' growth and development(Carbonell et al., 2012a; Meng et al., 2013a). For example, the *AtAGO1* is connected to the transgene-silencing pathways (Fagard et al., 2000) and the *AtAGO4* to the epigenetic silencing (Zilberman et al., 2003). The *AtAGO7* and *AtAGO10* proteins control growth (Hunter et al., 2003) and meristem functions in plant cells (Moussian et al., 1998). A recent study showed that some AGO genes (*TaAGO2a/2b*) of wheat are predicted to play a role in wheat growth and development specifically in grain quality improvement(Akond et al., 2022a). Two AGO genes (*TaAGO1b*/*TaAGO4*)

identified from wheat are predicted to play a key role in the vegetative and reproductive phase in cold environments at the vernalizing stage(Meng et al., 2013b). RDR proteins in plants (*Arabidopsis*, maize) are also engaged in RNAi-related processes viz., co-suppression, protection of pathogen infection, chromatin modification, and post-transcriptional gene silencing actions(Dalmay et al., 2000a; Mourrain et al., 2000a; Jovel et al., 2007).

Brassica is the second-largest oilseed crop after soybean (*Glycine max* L.) in the world's oilseed production (Styles, 2003). Among the 37 species in the *Brassica* genus, the four most widely cultivated species for oilseed are *Brassica rapa* L., *B. juncea* L., *B. napus* L., and *B. carinata A. Braun*. The two species *B. napus* and *B. rapa* are the third leading source of vegetable oil in the world after soybean and palm and the world's second-leading source of protein meal(Gupta, 2015). *Brassica napus*(rapeseed) (AC; n=19) is allopolyploid and originated probably several times independently from hybridization between the diploid *B. rapa* (A-genome; n = 10) and *B. oleracea* (C-genome; n=9) genome donors about 7500 years ago (Chalhoub et al., 2014). *B. napus* is cultivated and bred in several regions of the world and developed globally into one of the major oil crops (oilseed rape) (Basunanda et al., 2010). In Bangladesh, according to the commerce ministry yearly 18-20 lac metric tons of oil are being imported to fulfill the demand which costs about Tk. 20000-25000 crore and only 2 lac metric tons are produced domestically (01 May 2023, Prothom-Alo). Production of oilseeds from these *Brassica* species is frequently experiencing yield loss in Bangladesh in the field condition to different environmental factors mostly bacterial, viral, and fungal diseases along with insects, and pests.

To overcome these situations new transgenic approaches (RNAi/RNA silencing and CRISPR/Cas9) can be initiated to develop transgenic oilseed crops against such diseases and different environmental stresses. Studies suggested that the RNA silencing genes (DCL, AGO, and RDR) in plants play numerous roles in regulating growth and development as well as important biotic and abiotic stress responses(Gan et al., 2017; Prakash et al., 2017; Qin et al., 2018a; Akond et al., 2022a; Yun and Zhang, 2023). The RNA-interference (RNAi) mechanism not only helps plants protect themselves from different pathogenic diseases but also keeps transposons and endogenous genes silenced epigenetically(Prakash et al., 2017).The in-depth analysis of the genomic functions of these genes and their associated regulatory elements in *B. napus* is little known. The current study will be carried out for identification, characterization, and bioinformatic analyses of DCL, AGO, and RDR gene families in the *B. napus* genome. These results will provide important clues to molecular plant breeders and biotechnological researchers for the implementation of gene silencing pathways and RNAi-mediated host immunity in different BARI-developed rapeseed-mustard varieties for the development of improved nutrient-enriched climate-change-ready cultivars against different biotic and abiotic factors that will ultimately contribute significantly to the increased oil seed crop production for the sustainability of agricultural development as a whole in Bangladesh.

Objectives: The main objectives however of this study include:

- To identify and characterize the RNA silencing machinery/RNAi-related genes in *Brassica napus* genome;
- To analyze the evolutionary relationship with the RNAi proteins in *Arabidopsis thaliana*;
- To analyze protein domain structure and gene structure;
- To analyze the gene ontology and sub-cellular localization;

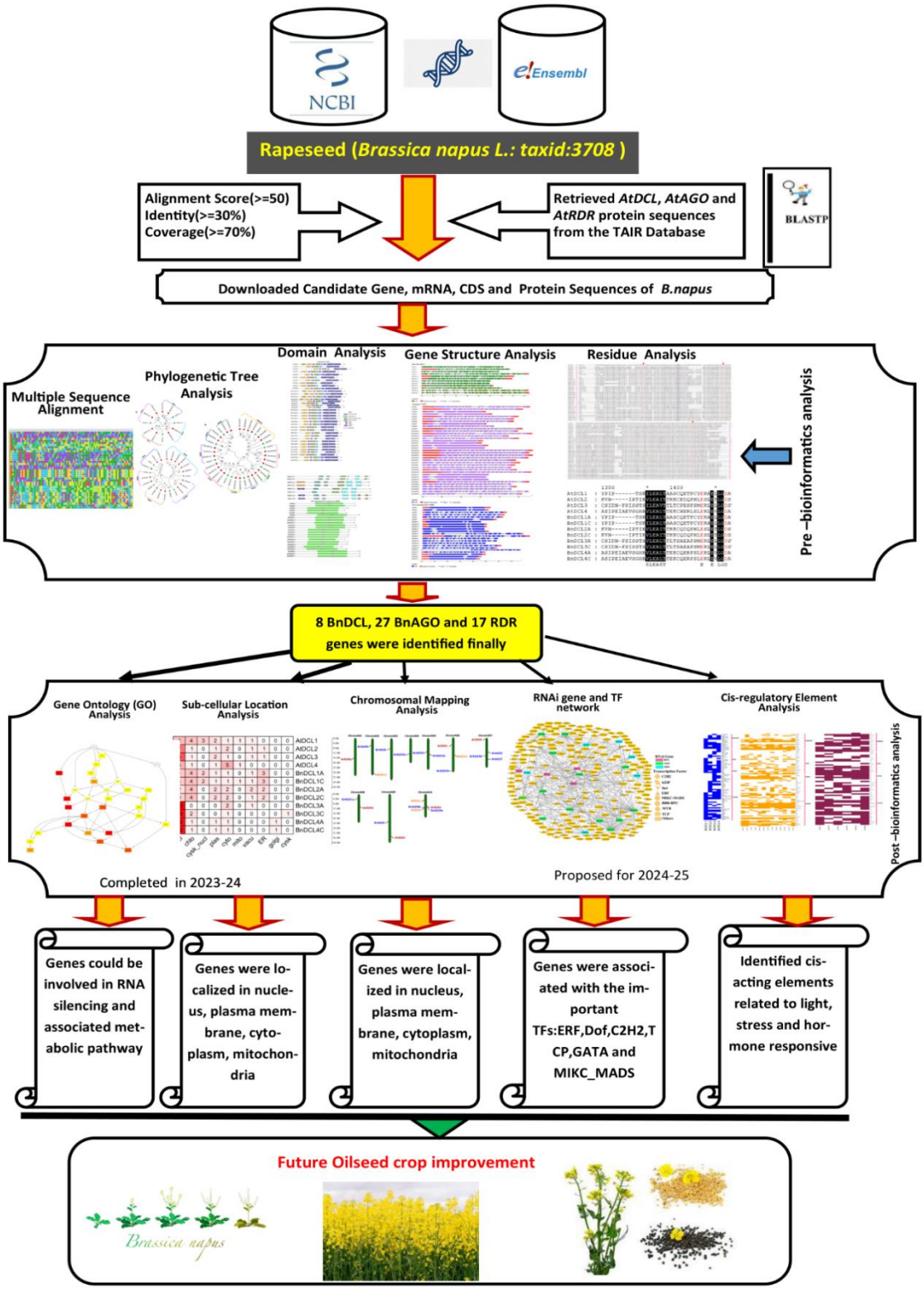


Figure 1. Flowchart of the study

2. Materials and Methods

2.1 Data Sources

First of all, the TAIR database (http://www.arabidopsis.org/) was used to collect the protein sequences of DCL, AGO, and RDR gene families of *Arabidopsis thaliana*, and then these sequences were used for BLASTP 2.11.0+(Stephen F. Altschul, Thomas L. Madden et al., 1997) search in the NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) for the genome of *Brassica napus* L.(taxid:3708) and cross-checked was also performed in the EnsemblePlants database(http://plants.ensembl.org/index.html) for more confirmation to predict the RNAi protein sequences of *B. napus*. Default algorithm parameter settings were used (E-value=0.05 and Matrix= BLOSUM62) during the BLAST process. All putative protein sequences were used to find the conserved domains that are also possessed by the AtRNAi proteins and redundant sequences were discarded. Finally, 52 best possible RNAi-related genes (8 BnDCL, 27 BnAGO, and 17 BnRDR) were selected. Important genomic info of these genes/proteins such as accession number, DNA, mRNA, CDS, and encoded protein length, exon and intron numbers, chromosomal location were recorded. All identified proteins were designated as BnDCL, BnAGO, and BnRDR using taxonomy based on the phylogenetic relationship with the proteins of the *Arabidopsis thaliana* counterpart. Prediction of the theoretical pI (isoelectric point) and Mw (molecular weight) of BnDCLs, BnAGOs, and BnRDRs was also recorded using the online tool ExPASyComputepI/Mwtool (http://au.expasy.org/tools/pito ol.html)(Duvaud et al., 2021).

2.2 Bioinformatics Analysis Approaches

Different bioinformatic analysis techniques including multiple sequence alignment, phylogenetic tree construction, gene (exon-intron) structure, domain arrangement, gene ontology, and subcellular localization analysis were performed for genome-wide identification, characterization, and diversity analysis of the RNAi genes in *Brassica napus*. These methods are described in the following sub-sections 2.2.1-2.2.5

2.2.1 Multiple Sequence Alignment, Phylogenetic Tree Construction and Nomenclature

Protein sequences of DCL, AGO, and RDR genes collected from *Arabidopsis thaliana* and *Brassica napus* genomes from the TIAR and NCBI databases respectively were used first for multiple sequence alignment using the Clustal W method(Rédei, 2008) in MEGA 11(Kumar et al., 2016a). A phylogenetic tree was constructed using MEGA (v.11) following the Neighbor-Joining (NJ) method to investigate the evolutionary closeness of DCL, AGO, and RDR genes in *B. napus* with those of their *Arabidopsis* counterpart, The phylogeny test was performed using the bootstrap method with 1000 replications to measure statistical support for nodes (Felsenstein, 1985) and the evolutionary distances were computed using the p-distance substitution method with complete deletion of the gaps/missing data(Kumar et al., 2016b). The candidate genes were given new names under the group names *BnDCL*, *BnAGO*, and *BnRDR* in line with the phylogenetic likeness and sequence homologies with the analogous *A. thaliana* homologs. More statistics of the selected proteins used for phylogenetic tree construction were recorded in Table 1.

2.2.2 Prediction of Protein Domain Structure

Analysis of the conserved functional domains of the selected *BnDCL*, *BnAGO*, and *BnRDR* proteins was made using the EnsemblePlants database (http://plants.ensembl.org/Triticum_aestivum/Info/Index) following the HMMER(Eddy, 2009) method to justify whether the domain compositions of all BnRNAi proteins are consistent with their *Arabidopsis* orthologs. The protein members of *BnDCL*, *BnAGO* and *BnRDR* groups with a higher number of functional domains of their *Arabidopsis* counterparts (*AtDCL*, *AtAGO*, and *AtRDR*) were finally selected for succeeding analyses. Six functional domains similar to that of *A. thaliana* namely DEXDc, HELICc, PAZ, RIBOc, and DSRM were considered for final identification of *BnDCL* genes. The protein members of the group *BnAGO* having DUF1785, PAZ, and Piwi functional domains like *AtAGO* are termed the AGO proteins. The *BnRDR* family was also identified for the presence of two common RDRP and RRM domains for protein structures like *AtRDR*. Additionally, all identified *BnAGO* and *BnRDR* proteins were investigated to find the most potentially

conserved metal-chelating catalytic triad regions or residues in the PIWI and RDRP domains. The alignment profiles were done using the ClustalW program in MEGA and GENEDOC software. All *AtAGO* and *AtRDR* proteins were considered during alignment.

2.2.3 Gene Structure Analysis

The gene/exon-intron structure of the 8 *BnDCL*, 27 *BnAGO* and 17 *BnRDR* genes were studied by the web tool Gene Structure Display Server (GSDS 2.0: http://gsds.cbi.pku.edu.cn/index.php). All putative coding sequences (CDS) and DNA/genomic sequences of *B.napus* with the CDS and DNA sequences of RNAi genes of *A.thaliana* were analyzed to justify the gene structural resemblance.

2.2.4 Gene Ontology Analysis

Gene Ontology (GO) analysis has been made by the online database Plant Transcription Factor Database (PlantTFDB, https://planttfdb.cbi.pku.edu.cn//)(Jin et al., 2017) to investigate the engagement of the BnRNAi genes in the biological processes (BPs), molecular functions (MFs), and cellular components (CCs). The *p*-values have been calculated using Fisher's exact test with Benjamini-Hochberg adjustments. GO terms with a p-value < 0.05 were considered statistically significant.

2.2.5 Prediction of the Sub-cellular Localization

To predict the cellular locations of the identified genes a web-based tool called WoLF PSORT (https://wolfpsort.hgc.jp/) T(Horton et al., 2007) was used. The protein sequences of the corresponding BnRNAi genes were used for the prediction of the cellular location. After that, the results were visualized by constructing a heatmap using TBtools(Chen et al., 2020).

3.0 Results and Discussion

3.1 *In silico* genome-wide identification of the RNAi genes from the *B. napus* genome

The finest set of RNAi-related genes/proteins from the *B.napus* genome were identified by taking the protein sequences of *AtDCL*, *AtAGO*, and *AtRDR* as query sequences for BLASTP search against the *B.napus* genome from the NCBI database. In total, 52 RNAi-related (8 *BnDCL*, 27 *BnAGO*, and 17 *BnRDR*) genes/proteins were selected for different bioinformatics analyses. Several genomic/proteomic info were also recorded for all putative genes/proteins (Table 1). Generally, in plants and animals, several important functional domains in the protein sequences such as DExD-helicase/DEAD, helicase-C, Dicer-dimmer/Duf283, PAZ, RNase III (Ribonuclease-3), and double-stranded RNA-binding (dsRB/DSRM) are highly preserved in the DCL proteins (Liu et al., 2014a). By HMM analysis (Johnson et al., 2010)8 DCL proteins in *B.napus* were also recognized based on these six domains which are analogous to *AtDCL*s. These six domains are known as the complete characteristics of plant DCL protein structure from the DCL family (class 3 RNase III family) (MacRae and Doudna, 2007). All *BnDCL* possessed DEAD/ResIII and this helicase-type DExH-box domain of the endoribonuclease dicer holds the ATP-binding region(Marchler-Bauer et al., 2017).

The genomic (DNA) and protein coding sequence (CDS) lengths of the selected *BnDCL*s varied from 9444bp and 4926bp (*BnDCL4C*, BnaC09g37430D) to 6398bp and 4170bp (*BnDCL2C*, BnaC05g47910D) with their corresponding encoded peptide length 1389aa to 1641aa. (Table 1). The model plant *Arabidopsis*, a dicot-type plant, preserves four DCL gene members that were reported to produce small RNAs(Margis et al., 2006a; Kapoor et al., 2008a). Eight DCL genes were recognized in the *B.napus* genome in our study. The members *BnDCL1* and *BnDCL4* may inclined to generate small RNAs nearly a size of 21- to 24 nt from plant-encoded dsRNAs. In *Arabidopsis*, *AtDCL1* and *AtDCL4* are also crucial for the biogenesis of 21 nt small RNAs that correspond to miRNAs and siRNAs(Kapoor et al., 2008a). *AtDCL3* also plays a function in producing the 24 nt RNAs that mediate de novo DNA methylation, gene silencing, and chromatin modification(Henderson et al., 2006). Earlier broad studies implied that the viral resistance mechanism in plants is highly maintained by the DCL family (Qin et al., 2017a). DCL4 is known as the main producer of 21-nt siRNAs related to viral defense(Fusaro et al., 2006; Moura et al., 2019). Nevertheless, DCL4 and DCL2 exhibited functional properties against viral infections, several studies showed that DCL3 genes also presented high significance in virus defenses through the

introduction of numerous 24-nt siRNAs from invading viruses(Fusaro et al., 2006; Llave, 2010; Moura et al., 2019). In *A. thaliana*, the DCL3 gene also performs direct cleavage of positive-sense RNA viruses such as *Cucumber mosaic virus* (a cucumovirus), *Oilseed rape mosaic virus* (a tobamovirus), and *Turnip crinkle virus* (a carmovirus)(Blevins et al., 2006; Deleris et al., 2006a). A study by Bozorov et al. (Ahmadovich Bozorov et al., 2012) showed the influence of *Nicotiana attenuate* DCL on the performance of *Manduca sex* larvae, and DCL3 and DCL4 were engaged in the regulation of several genes driving signaling and defense metabolite activation against herbivore attack(Moura et al., 2019). A study on the DCL genes of cotton showed that the DCL3 gene had a role in viral defense(Moura et al., 2019).

Argonaute proteins are very important RNA-binding proteins that are characterized by the presence of PAZ and PIWI domains (Qin et al., 2018b). Based on these two domain characteristics, 27 *BnAGO*s were also recognized from the *Brassica napus* genome in this study (Table 1). Conserved functional domain analysis from the EnsemblePlants and NCBI exhibited that all *BnAGO* proteins tend to contain an N-terminus PAZ domain like the *AtAGO*s. The DNA and CDS lengths of the *BnAGO*s range from 6740bp and 3261bp (*BnAGO1A2*, BnaA05g17460D) to 3169bp and 2664bp (*BnAGO2C*, BnaCnng68320D) with corresponding peptide length 1086aa and 887aa.

AGO proteins in the *B. napus* genome are assumed to engage in RNA-directed gene-suppressing functions and are notably associated with the growth and development of various organs or tissues in *B. napus* like other plants(Cenik and Zamore, 2011; Ji et al., 2011; Carbonell et al., 2012b). Moreover, previous studies showed that the PIWI domain in AGO proteins exhibits comprehensive homology to RNase H that binds the siRNA 5′ end to the target RNA (Höck and Meister, 2008a) and engage to slice the target mRNAs that show sequences complementary to siRNA or miRNAs(Rivas et al., 2005a). AGOs play as catalytic proteins that are supposed to possess three conserved metal-chelating residues/regions in the PIWI domain such as aspartate (D), aspartate (D), and histidine (H) called DDH(Kapoor et al., 2008b) that function as the catalytic triad. This triad was first explored in AtAGO1(Akond et al., 2022b).

Finally, our HMM analysis helped to identify that all elements of the *BnRDR* group have a definite RdRP well-preserved domain. The DNA and CDS lengths of the members of the family differ from 6789bp (BnRDR3C, BnaC08g36490D) and 3597bp (BnRDR6A, XP_048632623; BnRDR6C1, XP_022554911; BnRDR6C2, XP_013720258) to 3702bp (BnRDR1A, BnaA06g09600D) and 2586bp (BnRDR5C3, BnaC07g01120D). The peptide length varied from 1198aa (BnRDR6A, XP_048632623; BnRDR6C1, XP_022554911; BnRDR6C2, XP_013720258) to 861aa (BnRDR5C3, BnaC07g01120D).

The pI values in the table implied that all putative *BnDCL* genes hold acidic (5.91~7.64) characteristics. On the other hand, the predicted *BnAGO*s possess higher basic characteristics according to their pI values (8.70~10.18). The pI analysis of the *BnRDR*s showed that the 14 proteins are more likely to be basic while the BnRDR3A and BnRDR4A have the highest pI values that displayed the basic characteristics. The isoelectric point is the pH at which a particular molecule or surface conveys no net electrical charge(Schuurmans Stekhoven et al., 2008). Plant protein pI values are also significant in post-translational changes and biochemical activities in the RNAi gene family(Zhu et al., 2005). The pI values in many plant species commonly range from 1.99 to 13.96(Mohanta et al., 2019).

3.2 Classification of the BnRNAi Genes and Phylogenetic Analysis.

To classify all identified genes/proteins and to study the evolutionary relatedness of the 52 RNAi genes of the allotetraploid plant *Brassica napus* with the 10 RNAi genes of the dicotyledonous model plant *A.thaliana* , all protein sequences for both these species were first aligned by Clustal W using MEGA 11.0. After that, by using these aligned protein sequences three separate phylogenetic trees were constructed by MEGA 7.0 using the neighbor-joining (NJ) method with 1000 bootstrap replications (Kumar et al., 2016b). The phylogeny analysis exhibited that the *BnDCL* proteins showed a very high sequence similarity with that of the *AtDCL*s. The 8 *BnDCL*s however grouped into four distinct clades with the corresponding AtDCL subgroups (AtDCL1, AtDCL2, AtDCL3, and AtDCL4) (**Figure 1A**). Each clade contains two gene elements that clustered with the corresponding AtDCL orthologs. These results implied the extreme conservation of the DCL genes in this dicot-type species. Sequence similarity with the 4 AtDCLs and the phylogenetic tree output finally helped to name the DCL proteins BnDCL1A, BnDCL1C, BnDCL2A, BnDCL2C, BnDCL3A, BnDCL3C BnDCL4A and BnDCL4C (**Figure 1A**).

In the second tree (**Figure 1B**), the 27 BnAGO proteins clustered into nine groups. The groups were named after the AtAGOs (AtAGO1-AtAGO10). Of the nine BnAGO subgroups, BnAGO1 (BnAGO1A1 and A2, BnAGO1C1 and C2) and BnAGO4 (BnAGO4A1 and A2, BnAGO4C1 and C2) each subgroup contains four gene components, BnAGO2 (BnAGO2A and 2C), BnAGO3 (BnAGO3A and 3C), BnAGO5 (BnAGO5A and 5C), BnAGO6 (BnAGO6A and 6C) and BnAGO10 (BnAGO10A and 10C) each subgroup possesses two gene elements; BnAGO7 subfamily contains three gene members (BnAGO7A, C1, C2), BnAGO5 contains five members (BnAGO9A1, A2, BnAGOC9C1, C2, C3) and BnAGO8 possesses only one gene member BnAGO8A which is located in the chromosome A (**Table 1**).

In the third tree, we observe that like DCLs and AGOs, the RDRs of *B. napus* were also named after the *AtRDR* orthologs that were very close in the phylogeny tree with higher sequence similarity. The BLASP search results and domain composition helped to identify 17 RDR genes/proteins finally. These protein members in the phylogeny analysis clustered ultimately into five groups. The first subfamily BnRDR1 contains three members (BnRDR1A, C1, C2); the second (BnRDR2) subgroup possesses two members (BnRDR2A and 2C) and third (BnRDR3/4) subfamily possesses three members (BnRDR3A, 3C and BnRDR4A) (**Figure 1C**). The fourth subgroup BnRDR5 contains the highest six protein members (BnRDR5A1, A2, C1, C2, C3, C4) whereas the subfamily BnRDR6 contains three members (BnRDR6A, BnRDR6C1, BnRDR6C2).

Table 1. Basic genomic info about *DCL,AGO,* and *RDR* protein families in *B.napus*

Gene no.	Gene name	Accession ID	DNA length (bp)	mRNA length (bp)	CDS length (bp)	Length (aa)	pI	Mw (kDa)	No. of exons /Intron	Chromosomal location
Dicer-like (DCL)										
1	*AtDCL1*	AT1G01040	7705	5877	5733	1910	6.16	213.67	20/19	chr1:23416...31120
2	*AtDCL2*	AT3G03300	8289	4824	4167	1388	6.77	156.86	22/21	chr3:767926...776214
3	*AtDCL3*	AT3G43920	7283	4743	4743	1580	6.33	177.42	24/23	chr3:15753548...15760830
4	*AtDCL4*	AT5G20320	10055	5666	5109	1702	6.74	191.27	25/23	chr5:6859212...6869266
1	*BnDCL1A*	BnaA10g00800D	7484	5921	5439	1812	6.28	156.70	18/17	chrA10:390521...398004
2	*BnDCL1C*	BnaC05g00860D	7318	5608	5430	1809	6.01	201.54	20/19	chrC05:435258...442575
3	*BnDCL2A*	BnaA05g32540D	6470	4249	4167	1388	7.16	156.70	22/21	chrA05:22308269...22314738
4	*BnDCL2C*	BnaC05g47910D	6398	4170	4170	1389	7.64	156.94	22/21	chrC05:42659693...42666090
5	*BnDCL3A*	XP_013656716[a]	6726	4890	4572	1523	5.91	141.60	24/23	A6 NC_063439.1(15638285...15645010)
6	*BnDCL3C*	BnaC03g54010D	7070	4705	4596	1531	6.41	172.25	25/24	chrC03:40673425...40680494
7	*BnDCL4A*	BnaA10g15080D	9388	5177	4929	1642	6.63	185.01	25/24	chrA10:11827194...11836581
8	*BnDCL4C*	BnaC09g37430D	9444	5155	4926	1641	6.63	184.48	24/24	chrC09:40709757...40719200
Argonaute (AGO)										
1	*AtAGO1*	AT1G48410	6597	3615	3153	1050	9.68	116.46	21/20	chr1:17886036...17892632
2	*AtAGO2*	AT1G31280	3893	3602	3045	1014	9.97	113.42	3/2	chr1:11181504...11185396
3	*AtAGO3*	AT1G31290	4281	3841	3585	1194	9.78	129.18	3/2	chr1:11188293...11192573
4	*AtAGO4*	AT2G27040	5890	3236	2775	924	9.19	102.84	22/21	chr2:11536502...11542391
5	*AtAGO5*	AT2G27880	5612	3381	2994	997	9.92	111.08	20/19	chr2:11871378..11876989
6	*AtAGO6*	AT2G32940	5606	3123	2637	878	8.59	98.68	22/21	chr2:13971916..13977512
7	*AtAGO7*	AT1G69440	3922	3443	2973	990	9.59	113.39	3/2	chr1:26101407..26105328
8	*AtAGO8*	AT5G21030	4381	2565	2565	854	8.99	96.03	21/20	chr5:7139892..7144272
9	*AtAGO9*	AT5G21150	6143	3075	2691	896	9.45	100.52	22/21	chr5:7192239..7198381
10	*AtAGO10*	AT5G43810	6164	3668	2967	988	9.64	110.86	19/18	chr5:17610744..17616907
1	*BnAGO1A1*	BnaA08g03260D	5847	3823	3135	1044	9.66	115.85	23/22	chrA08:2681072...2686918
2	*BnAGO1A2*	BnaA05g17460D	6740	4346	3261	1086	9.61	120.33	23/22	chrA05:12290150...12296889
3	*BnAGO1C1*	BnaC08g46720D	5291	3381	3159	1052	9.74	116.78	21/20	chrC08_random: 953598...958888
4	*BnAGO1C2*	BnaC05g25730D	5910	3251	2943	980	9.27	109.15	24/22	chrC05:21116236...21122145
5	*BnAGO2A*	BnaA09g25290D	3712	3364	3072	1023	10.18	113.91	4/3	chrA09:18324937...18328648
6	*BnAGO2C*	BnaCnng68320D	3169	2930	2664	887	9.90	100.70	4/2	chrCnn_random: 67905947...67909115
7	*BnAGO3A*	BnaA05g14760D	3501	3362	3033	1010	9.97	112.74	2/1	chrA05:9228612...9232112
8	*BnAGO3C*	BnaC06g41790D	3340	3108	3108	1035	10.01	114.35	2/1	chrC06_random: 1066401...1069740
9	*BnAGO4A1*	BnaA04g15560D	5887	4033	2769	922	8.91	103.06	22/21	chrA04:12884285...12890171
10	*BnAGO4A2*	BnaA07g13010D	5716	3963	2772	923	8.7	103.36	22/21	chrA07:11653377...11659092
11	*BnAGO4C1*	BnaC04g38560D	5890	4031	2769	922	8.80	103.11	22/21	chrC04:39739228...39745117

12	*BnAGO4C2*	BnaC04g54830D	5652	3800	2772	923	8.80	103.36	22/21	chrC04_random: 2230442...2236093
13	*BnAGO5A*	BnaA07g13430D	4997	3199	2874	957	9.91	106.71	20/19	chrA07:11907998...11912994
14	*BnAGO5C*	BnaC04g16450D	5220	3294	2859	952	10.06	106.03	21/20	chrC04:14487678...14492897
15	*BnAGO6A*	BnaA03g15180D	4948	2892	2604	867	9.10	97.20	22/21	chrA03:7005357...7010304
16	*BnAGO6C*	BnaC03g18310D	4736	2696	2604	867	8.99	97.37	22/21	chrC03:9391663...9396398
17	*BnAGO7A*	BnaA07g24280D	3461	2990	2955	984	9.71	112.47	3/2	chrA07:18160385...18163845
18	*BnAGO7C1*	BnaC02g19190D	3334	2931	2931	976	9.58	111.58	3/2	chrC02:15451981...15455314
19	*BnAGO7C2*	BnaC06g43420D	3447	2733	2700	899	9.74	102.67	6/5	chrC06_random: 2865213...2868659
20	*BnAGO8A*	BnaA02g05290D	5571	3666	2721	906	9.6	101.34	21/20	chrA02:2403187...2408757
21	*BnAGO9A1*	BnaA10g14450D	4627	2715	2715	904	9.60	102.04	22/21	chrA10:11492941...11497567
22	*BnAGO9A2*	BnaA10g14440D	4902	2922	2748	915	9.77	102.52	21/20	chrA10:11481378...11486279
23	*BnAGO9C1*	BnaCnng35060D	4647	2721	2721	906	9.79	102.57	22/21	chrCnn_random: 33265084...33269730
24	*BnAGO9C2*	BnaC09g36780D	5520	2763	2763	920	9.44	103.80	24/23	chrC09:40119310...40124830
25	*BnAGO9C3*	BnaC09g36860D	5190	2790	2649	882	9.51	99.37	20/19	chrC09:40226286...40231475
26	*BnAGO10A*	BnaA06g36540D	4921	3500	2928	975	9.70	109.27	17/16	chrA06:23915363...23920283
27	*BnAGO10C*	BnaC07g17330D	5667	4191	2946	981	9.70	109.75	17/16	chrC07:23533982...23539648

Gene no.		Accession ID	DNA length (bp)	mRNA length (bp)	CDS length (bp)	Length (aa)	pI	Mw (kDa)	No. of exons /introns	Chromosomal location
RNA-dependent RNA polymerase (RDR)										
1	*AtRDR1*	AT1G14790	4310	4045	3324	1107	7.88	126.21	2	chr1:5093961...5098270
2	*AtRDR2*	AT4G11130	4184	3717	3402	1133	6.62	129.32	3	chr4:6780449...6784632
3	*AtRDR3*	AT2G19910	5327	3368	2979	992	7.78	112.75	17	chr2:8595605...8600931
4	*AtRDR4*	AT2G19920	5195	3310	2823	940	7.78	106.81	17	chr2:8601858...8607052
5	*AtRDR5*	AT2G19930	5102	3127	2934	977	7.10	110.88	17	chr2:8607340...8612441
6	*AtRDR6*	AT3G49500	4693	4015	3591	1196	7.22	136.92	1	chr3: 18348981 - 18353673
1	*BnRDR1A*	BnaA06g09600D	3702	3517	3330	1109	7.56	126.63	3/2	chrA06:5134289...5137990
2	*BnRDR1C1*	XP_013696632[a]	4061	3775	3336	1111	5.93	84.40	3/2	C5 NC_063448.1 (7288528...7292588)
3	*BnRDR1C2*	BnaC05g10980D	4051	3828	3282	1093	7.00	124.83	4/3	chrC05:6358607...6362657
4	*BnRDR2A*	BnaA09g22040D	3920	3390	3390	1129	6.51	128.25	4/3	chrA09:14659131...14663050
5	*BnRDR2C*	BnaCnng57100D	4193	3677	3378	1125	6.40	127.52	4/3	chrCnn_random: 56892147...56896339
6	*BnRDR3A*	BnaA09g43930D	6357	3180	3003	1000	8.18	113.41	17/16	chrA09:30278290...30284646
7	*BnRDR3C*	BnaC08g36490D	6789	4680	3000	999	7.92	113.88	17/16	chrC08:33743646...33750434
8	*BnRDR4A*	BnaA10g12910D	4765	2982	2982	993	8.48	112.64	18/17	chrA10:10516020...10520784
9	*BnRDR5A1*	BnaA07g00800D	4703	2928	2928	975	7.4	110.64	20/19	chrA07:581670...586372
10	*BnRDR5A2*	BnaA07g00770D	3994	2757	2757	918	6.08	103.58	17/16	chrA07:561541...565534
11	*BnRDR5C1*	BnaC07g01150D	4993	3198	2934	977	6.51	110.60	19/17	chrC07:1764972...1769964
12	*BnRDR5C2*	BnaC07g01170D	4711	3077	2877	958	7.03	108.36	17/16	chrC07:1805711...1810421
13	*BnRDR5C3*	BnaC07g01120D	7736	2701	2586	861	5.8	97.84	17/16	chrC07:1738692...1746427
14	*BnRDR5C4*	XP_048620374	4999	3340	2853	950	6.32	84.22	17/16	C7 NC_063450.1 (2586680...2591678)
15	*BnRDR6A*	XP_048632623	5022	4076	3597	1198	6.52	137.04	2/1	A1 NC_063434.1(13593651...13598672)
16	*BnRDR6C1*	XP_022554911	4682	3988	3597	1198	6.93	137.07	2/1	C4 NC_063447.1(33425330...33430011)
17	*BnRDR6C2*	XP_013720258	4555	3998	3597	1198	6.78	137.03	2/1	C1 NC_063444.1 (22534074...22538628)

[a] *The noted sequences were from the NCBI BioProject: PRJNA844685, while all the others were from the NCBI BioProject: PRJEB5043.*

3.2 Classification of the BnRNAi Genes and Phylogenetic Analysis.

To classify all identified genes/proteins and to study the evolutionary relatedness of the 52 RNAi genes of the allotetraploid plant *Brassica napus* with the 10 RNAi genes of the dicotyledonous model plant *A.thaliana* , all protein sequences for both these species were first aligned by Clustal W using MEGA 11.0. After that, by using these aligned protein sequences three separate phylogenetic trees were constructed by MEGA 7.0 using the neighbor-joining (NJ) method with 1000 bootstrap replications (Kumar et al., 2016b). The phylogeny analysis exhibited that the *BnDCL* proteins showed a very high sequence similarity with that of the *AtDCL*s. The 8 *BnDCL*s however grouped into four distinct clades with the corresponding AtDCL subgroups (AtDCL1, AtDCL2, AtDCL3, and AtDCL4) (**Figure 1A**). Each clade contains two gene elements that clustered with the corresponding AtDCL orthologs. These results implied the extreme conservation of the DCL genes in this dicot-type species. Sequence similarity with the 4 AtDCLs and the phylogenetic tree output finally helped to name the DCL proteins BnDCL1A, BnDCL1C, BnDCL2A, BnDCL2C, BnDCL3A, BnDCL3C BnDCL4A and BnDCL4C (**Figure 1A**).

In the second tree (**Figure 1B**), the 27 BnAGO proteins clustered into nine groups. The groups were named after the AtAGOs (AtAGO1-AtAGO10). Of the nine BnAGO subgroups, BnAGO1 (BnAGO1A1 and A2, BnAGO1C1 and C2) and BnAGO4 (BnAGO4A1 and A2, BnAGO4C1 and C2) each subgroup

contains four gene components, BnAGO2 (BnAGO2A and 2C), BnAGO3 (BnAGO3A and 3C), BnAGO5 (BnAGO5A and 5C), BnAGO6 (BnAGO6A and 6C) and BnAGO10 (BnAGO10A and 10C) each subgroup possesses two gene elements; BnAGO7 subfamily contains three gene members (BnAGO7A, C1, C2), BnAGO5 contains five members (BnAGO9A1, A2, BnAGOC9C1, C2, C3) and BnAGO8 possesses only one gene member BnAGO8A which is located in the chromosome A (**Table 1**).

In the third tree, we observe that like DCLs and AGOs, the RDRs of *B. napus* were also named after the *AtRDR* orthologs that were very close in the phylogeny tree with higher sequence similarity. The BLASP search results and domain composition helped to identify 17 RDR genes/proteins finally. These protein members in the phylogeny analysis clustered ultimately into five groups. The first subfamily BnRDR1 contains three members (BnRDR1A, C1, C2); the second (BnRDR2) subgroup possesses two members (BnRDR2A and 2C) and third (BnRDR3/4) subfamily possesses three members (BnRDR3A, 3C and BnRDR4A) (**Figure 1C**). The fourth subgroup BnRDR5 contains the highest six protein members (BnRDR5A1, A2, C1, C2, C3, C4) whereas the subfamily BnRDR6 contains three members (BnRDR6A, BnRDR6C1, BnRDR6C2).

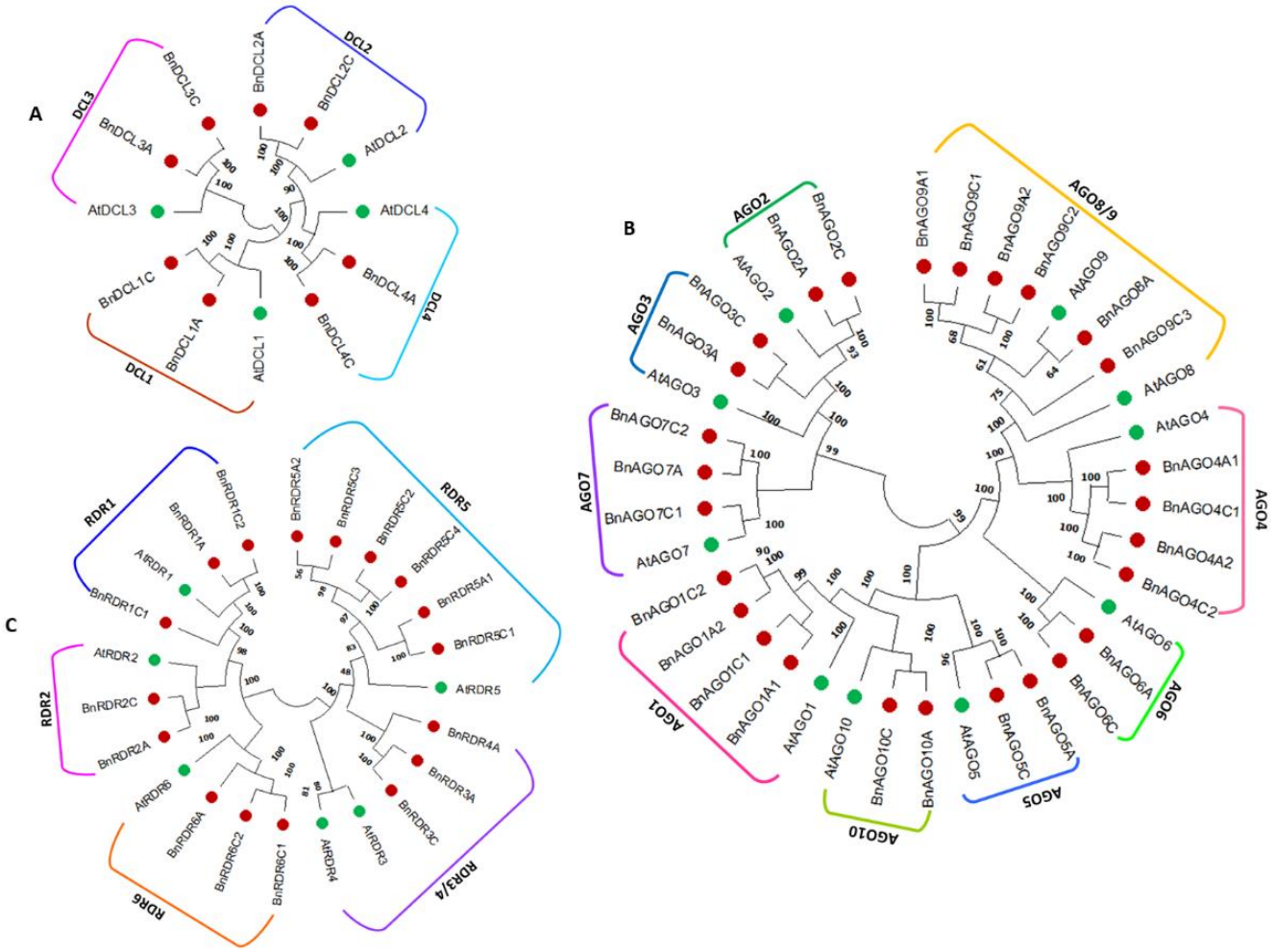


Figure 1. Phylogenetic trees of all *A. thaliana* and *B.napus* RNAi-related gene families. (A) *BnDCL* (B) *BnAGO* and (C) *BnRDR*. The trees were produced by MEGA 11.0 using the Neighbor-Joining (NJ) method with a bootstrap of 1000. The proteins of *A.thaliana* and *Brassica rapa* are indicated with green and red color circles before the corresponding taxon names.

3.3 Multiple sequence alignment of DCL, AGO, and RDR proteins in *B.napus* and *A. thaliana*

Multiple sequence alignment (MSA) of the identified 8 BnDCL, 27BnAGO and 17 BnRDR protein sequences with the 4 AtDCL, 10AtAGO, and 6 AtRDR protein sequences were carried out using MEGA 11(Figure 2-4). The alignment profile exhibited that the RNase III catalytic sites of the putative BnDCLs in the two RNase III domains at the glutamate (E), aspartate (D), aspartate (D), and glutamate (E) position with the orthologous AtDCL proteins (Figure 2). The alignment showed that there is no replacement of the amino acids for any predicted proteins. The amino acids are well-preserved with their AtDCLs counterpart.

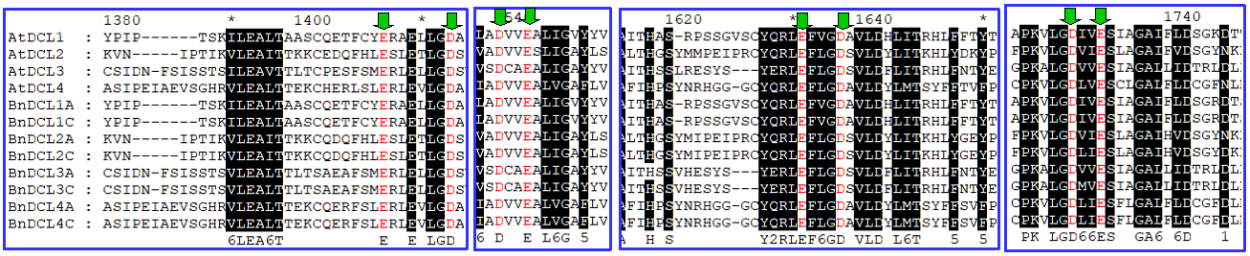


Figure 2. The multiple sequence alignment of *BnDCL* and *AtDCL* proteins for RNase III domains (RIBOc I and II) following the ClustalW method by MEGA 11. The downward green arrows indicate the position of conserved two RNase III domains at the glutamate (E), aspartate (D), aspartate (D), and glutamate (E) (EDDE) positions.

From the figure 3 it has been showed that the three metal-chelating well-preserved catalytic residues (D = aspartate, D = aspartate, and H = histidine) are identified in the PIWI domain that was initially found in *AtAGO1*(Kapoor et al., 2008b). This type of region or triad is characterized for endonuclease activity(Baumberger and Baulcombe, 2005a; Arribas-Hernández et al., 2016)**.**The alignment profile of the 27 BnAGO proteins with the 10 AtAGOs displayed that the most of the proteins conserved DDH triad regions in the PIWI domains (**Figure 3 and Table 2**). On the other hand, the sequence alignment of AtRDRs and BnRDRs demonstrated that DLDGD catalytic motif was present in the RdRP domain (**Figure 4**)

Figure 3. Alignment profile of PIWI domain of *BnAGO* proteins. The protein sequences were aligned following the ClustalW method by MEGA 11.0. The conserved Asp, Asp, and His (DDH) triad residues corresponding to D762, D848, and H988 of AtAGO1 are indicated with downward arrows, whereas the conserved H residue corresponding to H800 of AtAGO1 is boxed with black color. Amino acid positions corresponding to each protein are indicated at the start and end of each line.

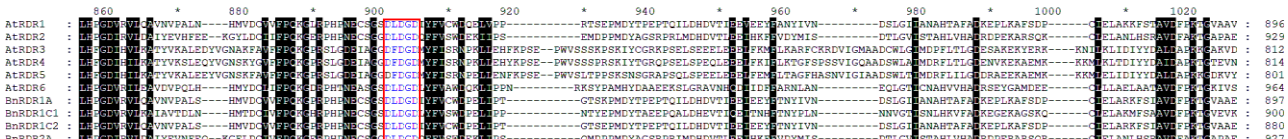


Figure 4. The alignment composition of the RDR proteins of the *B.napus* and *Arabidopsis* by the Clustal W method using MEGA 11. The conserved DLDGD catalytic motif is boxed with a blue color.

The catalytic DDH/H region was found in the 11 proteins (BnAGO1A1, BnAGO1A2, BnAGO1C1, BnAGO1C2, BnAGO2A, BnAGO2C, BnAGO5A, BnAGO5C, BnAGO7A, BnAGO10A and BnAGO10C) out of the 27 proteins that are similar to AtAGO1, AtAGO3, AtAGO5, AtAGO7 and AtAGO10 (**Figure 3 and Table 2**). The DDH/R triad residue was found in the nine BnAGO proteins (BnAGO3A, BnAGO3C, BnAGO8A, BnAGO9A1, BnAGO9A2, BnAGO9C1, BnAGO9C2, BnAGO9C3) like the AtAGO9 (**Figure 3 and Table 2**). For these nine proteins, the fourth amino acid H in the 800th position is replaced by the arginine (R). The amino acid arginine (R) plays an important function in plants for various physiological and biochemical activities, growth, and development, and also helps plants adapt to diverse environmental conditions(Yang and Gao, 2007). DDD/S region was found in the four proteins (BnAGO4A1, BnAGO4A2, BnAGO4C1, and BnAGO4C2) unlike AtAGO4 (DDH/S). For these four proteins, the third

amino acid is replaced by the aspartate (D). Several research showed that the higher rate of this amino acid (D) ensures the nutritional properties in maize like other plants particularly in seeds, due to its significant role in the synthesis of four important amino acids viz., lysine/Lys (K), threonine/Thr(T), methionine/Met(M), and isoleucine/Ile(I)(Gibbon and Larkins, 2005; Galili, 2011; Wang et al., 2018). Furthermore, the DDD/H motif was found in the two BnAGO7C1 and BnAGO7C2 similar to AtAGO2. Hence the presence of such DDD residues in the PIWI domain of predicted BnAGO proteins is supposed to have a significant contribution to healthy and sufficient seeds in the seed capsule of the oilseed crops. The motif region DDH/P was observed in the BnAGO6A and BnAGO6C similar to AtAGO6 and AtAGO8 (Table 2).

In this study, we interestingly observe that two proteins of subgroup BnAGO2 hold the DDH/H triad like AtAGO7 whereas the proteins of subgroup BnAGO7 possess the DDD/H triad residue like AtAGO2. On the other hand, the amino acid H in the 988th position and the 800th position is replaced by aspartate(D) and serine(S) in the four protein members of the BnAGO4 subgroup though the H is present in the 988th position in AtAGO4. Serine (S) is an important amino acid in plants that keeps a role in metabolism and different signaling pathways, for example, the phosphorylated pathway which removes the defects in embryo, male gametophytes, and roots(Ros et al., 2014). Additionally, the serine(S) rich proteins play a pivotal role in plant responses to both biotic and abiotic stresses(Kishor et al., 2020). Consequently, the alteration/mutational changes of the D762, D848, H988, and H800 (DDH/H) catalytic residues in the BnAGO proteins significantly implied that these protein blocks deliver crucial role in metabolism, growth, and development along with addressing the different stresses during their life span.

Table 2. **Comparison of the AGO proteins in the PIWI domains between *B.napus* and *A. thaliana*.**

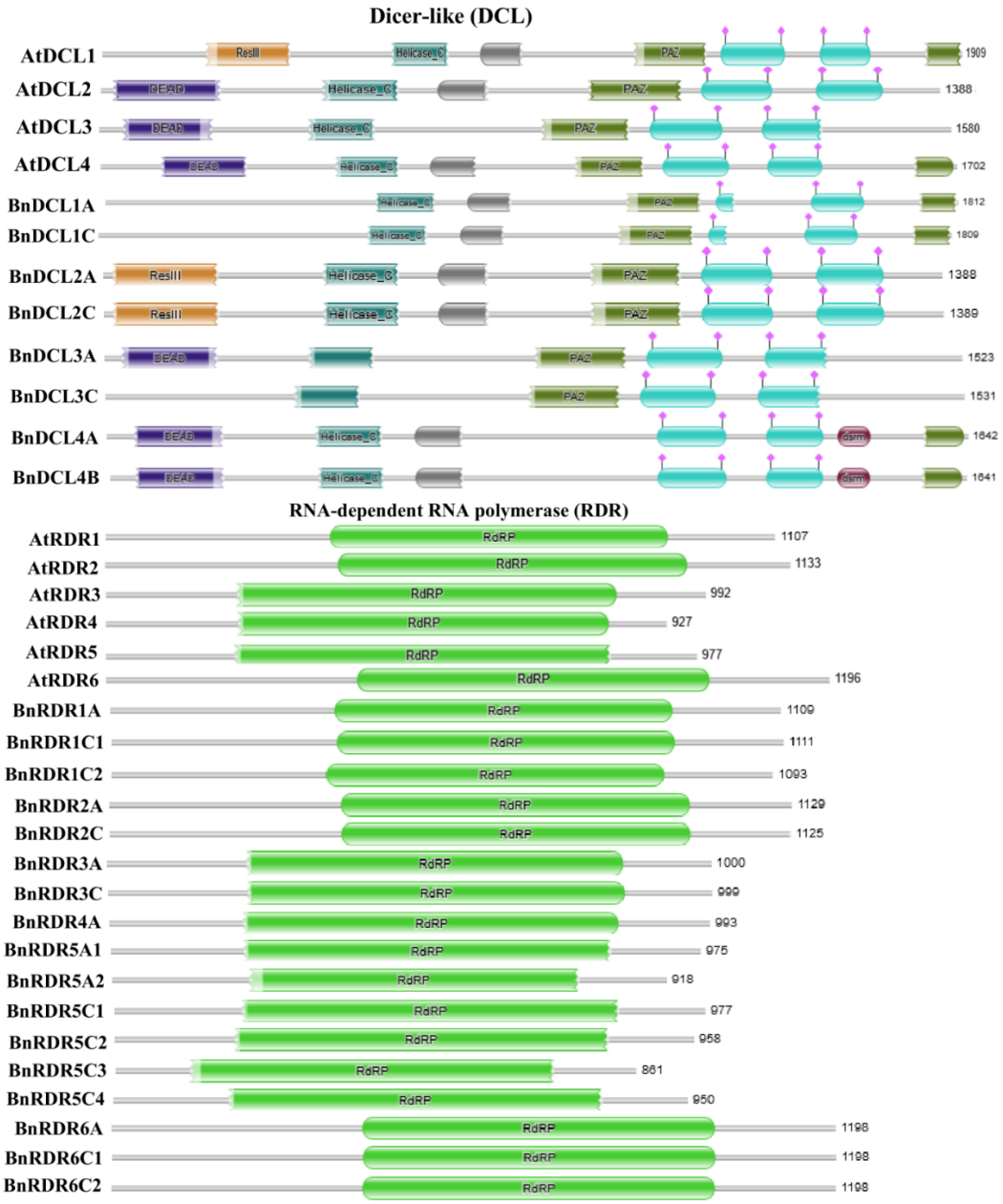
Serial No.	*Brassica napus*		*Arabidopsis thaliana*	
	Argonaute (AGO)	**Motif**	**Argonaute (AGO)**	**Motif**
1	*BnAGO1A1*	DDH/H	*AtAGO1*	DDH/H
2	*BnAGO1A2*	DDH/H	*AtAGO2*	DDD/H
3	*BnAGO1C1*	DDH/H	*AtAGO3*	DDH/H
4	*BnAGO1C2*	DDH/H	*AtAGO4*	DDH/S
5	*BnAGO2A*	DDH/H	*AtAGO5*	DDH/H
6	*BnAGO2C*	DDH/H	*AtAGO6*	DDH/P
7	*BnAGO3A*	DDH/R	*AtAGO7*	DDH/H
8	*BnAGO3C*	DDH/R	*AtAGO8*	DDH/P
9	*BnAGO4A1*	DDD/S	*AtAGO9*	DDH/R
10	*BnAGO4A2*	DDD/S	*AtAGO10*	DDH/H
11	*BnAGO4C1*	DDD/S		
12	*BnAGO4C2*	DDD/S		
13	*BnAGO5A*	DDH/H		
14	*BnAGO5C*	DDH/H		
15	*BnAGO6A*	DDH/P		
16	*BnAGO6C*	DDH/P		
17	*BnAGO7A*	DDH/H		
18	*BnAGO7C1*	DDD/H		
19	*BnAGO7C2*	DDD/H		
20	*BnAGO8A*	DDH/R		
21	*BnAGO9A1*	DDH/R		
22	*BnAGO9A2*	DDH/R		
23	*BnAGO9C1*	DDH/R		
24	*BnAGO9C2*	DDH/R		
25	*BnAGO9C3*	DDH/R		
26	*BnAGO10A*	DDH/H		
27	*BnAGO10C*	DDH/H		

Motifs are corresponded to conserved D762, D848, H988/H800 of Arabidopsis AGO1; D(aspartate), H(Histidine), P(Proline), R(Arginine), S(Serine);

3.4 Functional domain composition analysis and structural characterization of BnDCL, BnAGO, and BnRDR genes

For the functional domain analysis, first of all, the well-characterized DCL, AGO, and RDR protein sequences of *Arabidopsis* were used as query sequences for BLASTp search against the *B.napus* genome from the NCBI. The predicted protein sequences were then studied for domain composition. Some physio-chemical characteristics and domain alignment of the BnDCL BnAGO and BnRDR. were nearly very close to their *Arabidopsis* orthologs. Lastly, 8 DCL, 27 AGO, and 17 RDR genes were identified in the *B.napus* genome based on their domain structure those are similar to *A. thaliana* (**Figure 5**). We observe that the *B.napus* genome possesses twice as many RNAi genes/proteins as *A. thaliana* with the exception that the BnAGO and BnRDR gene groups possess seven and nine more AGO and RDR gene members respectively compared to AtAGO and AtRDR (**Table 1**). Eight dicer-like (DCL) genes were named as BnDCL in the *B.napus* genome from the NCBI due to the existence of six specific conserved domains viz. DEXDc (DEAD-like helicases superfamily), HELICc (helicase superfamily C-terminal domain), PAZ (Piwi, Argonaute, Zwille), RIBOc (Ribonuclease III family), and DSRM (Double-stranded RNA binding motif) consistent with AtDCLs. Moreover, prior investigations revealed that the maximum plant DCL genes (class 3 RNase III family) remained well-identified for the presence of these six functional domains (MacRae and Doudna, 2007; Kapoor et al., 2008a; Bai et al., 2012; Yadav et al., 2015; Cao et al., 2016; Gan et al., 2016; Zhao et al., 2016; Qin et al., 2018b). The length of these eight proteins varies from 1812aa to 1388aa whereas the length of the proteins of AtDCLs ranges from 1910 aa to 1388 aa (**Table 1**). The eight BnDCLs showed a molecular weight potentiality between 141.60–201.54 kDa (**Table 1**). Domain structure analysis of BnDCL proteins showed that the members of the BnDCL2 and BnDCL3 subgroups lack a double-strand RNA binding (DSRM) domain from DEAD ENDPOINT 1 like AtDCL2 and AtDCL3 (**Figure 5**). Some earlier investigations showed that this DSRM domain is present in Soybean (*Glycine max*) and Shorgum (*Sorghum Bicolor*)(Kang et al., 2010; Liu et al., 2014a; Hamar et al., 2020) and absent in and Pepper (*Capsicum Annuum*)(Cao et al., 2016; Qin et al., 2018b).

The BnAGO proteins hold the molecular weight between 97.20 ~120.33 kDa. AGO proteins typically consist of four domains: DUF1785 (N-terminal function-unknown domain), PAZ, MID, and C-terminal PIWI domains(Qian et al. 2011; Liu et al. 2014). A total of 27 BnAGO proteins were extracted finally based on two specific domain characteristics viz., PAZ and PIWI following the HMMER technique(Johnson et al., 2010; Eddy, 2011) from the NCBI and EnsemblePlants database. The peptide chain of the BrAGO proteins ranges from 1086 aa to 867 aa (**Table 1**). Protein domain detection by NCBI and EnsemblePlants exhibited that all BnAGOs shared an N-terminus PAZ domain and a C-terminus PIWI domain that are vibrant characteristics of the plant argonaute proteins. Furthermore, previous research revealed that the PIWI domain shows complete homology to RNaseII that binds the siRNAs 5´ end to the target RNA (Höck and Meister, 2008b; Qian et al., 2011a) and is involved in the cleavage activity of the target RNAs that have sequence complementarity to the small RNAs(Baumberger and Baulcombe, 2005b; Rivas et al., 2005b; Qian et al., 2011a). The active site of one PIWI domain is liable to cleave RNA and typically holds a strongly conserved metal-chelating Asp–Asp– His (DDH) region/motif(Kapoor et al., 2008b; Liu et al., 2014b). This catalytic residue was first revealed in Arabidopsis AGO1, and a conserved histidine (H) at position 800 (H800) was also detected to be vital for AGO1 for in vitro endonuclease activity(Baumberger and Baulcombe, 2005a; Qian et al., 2011b).


Dicer-like (DCL)
AtDCL1
AtDCL2
AtDCL3
AtDCL4
BnDCL1A
BnDCL1C
BnDCL2A
BnDCL2C
BnDCL3A
BnDCL3C
BnDCL4A
BnDCL4B
ResIII
DEAD
Helicase_C
PAZ
dsrm
1909
1388
1580
1702
1812
1809
1388
1389
1523
1531
1642
1641
RNA-dependent RNA polymerase (RDR)
AtRDR1
AtRDR2
AtRDR3
AtRDR4
AtRDR5
AtRDR6
BnRDR1A
BnRDR1C1
BnRDR1C2
BnRDR2A
BnRDR2C
BnRDR3A
BnRDR3C
BnRDR4A
BnRDR5A1
BnRDR5A2
BnRDR5C1
BnRDR5C2
BnRDR5C3
BnRDR5C4
BnRDR6A
BnRDR6C1
BnRDR6C2
RdRP
1107
1133
992
927
977
1196
1109
1111
1093
1129
1125
1000
999
993
975
918
977
958
861
950
1198
1198
1198

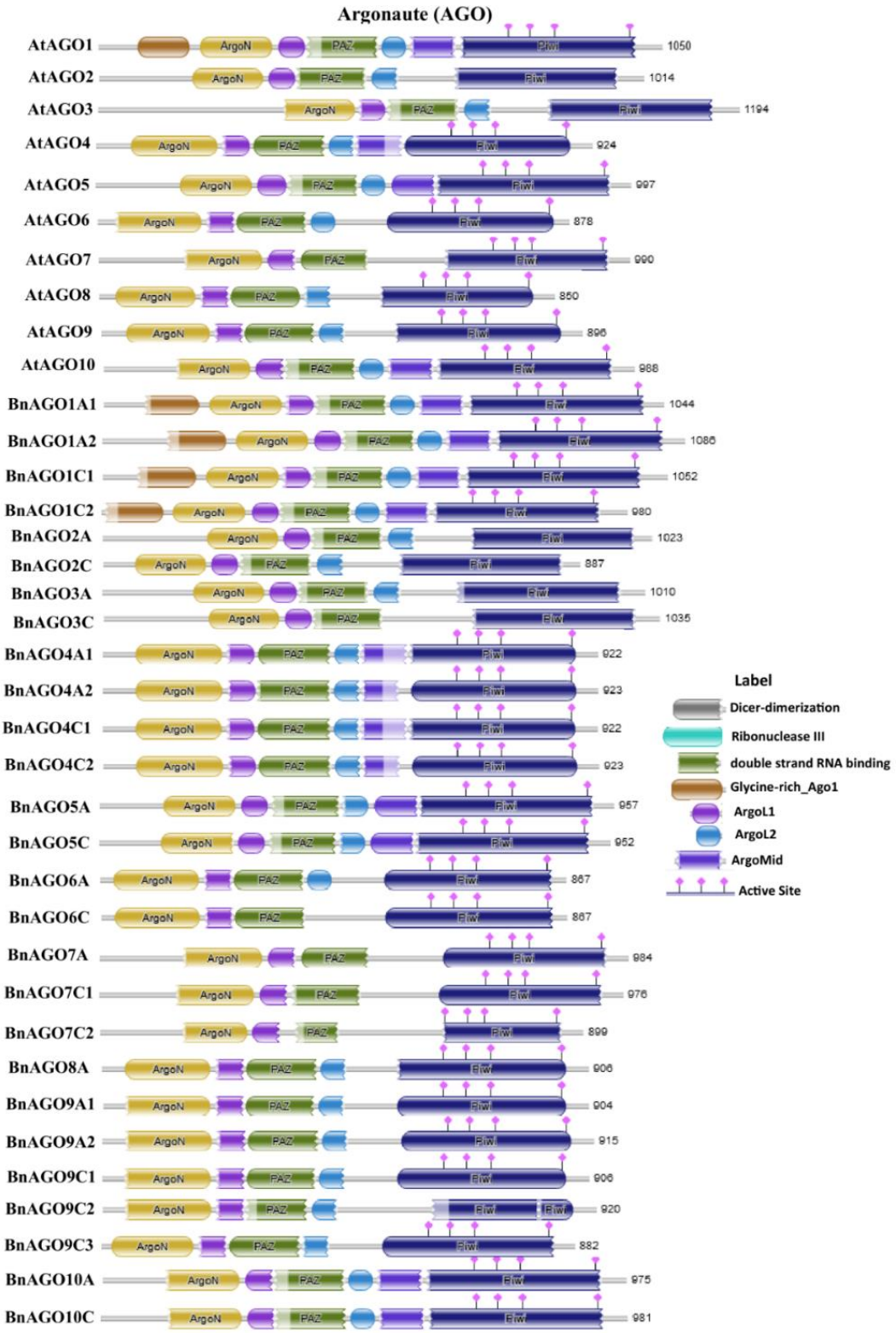


Figure 5. Domain composition of BnDCL, BnAGO, and BnRDR protein families. Conserved functional domains were predicted from the Pfam database. Domains are indicated in different color boxes with their corresponding names inside the box.

3.5 Gene(exon-intron) structure analysis of BnDCL, BnAGO, and BnRDR genes

Exons are built with nucleic acid blocks present in the mRNA and well-preserved whereas the introns are non-coding nucleic acid chains that are removed by the RNA splicing before translation. To study the gene/exon-intron structure of the predicted RNAi genes from *B.napus*, we used the GSDS v2.0 web tool to visualize their possible structural pattern and development and also to check the rate of similarity with the structure of exon-introns of the corresponding RNAi-related genes in *Arabidopsis thaliana*. The analysis demonstrated that intron and exon numbers were generally conserved corresponding to the gene members of the AtRNAi amongst the members of the BnDCL BnAGO and BnRDR groups (**Figure 6 and Table 1**). The intron numbers however varied from 17-24 among the BnDCLs which is almost similar to AtDCLs (19-23) (**Table 1**). On the other hand, the maximum BnAGO genes also had nearly the same number of introns (16-23) as their AtAGO counterparts (18-21) with some exceptions that the members (seven members) of the subgroup BnAGO2/3/7 had the lower range (1-5) like AtAGO2/3/7 (2). Thirdly among the 17 BnRDR proteins, the gene members of the subgroups BnRDR3/4/5 contained 16-19 introns like AtRDR3/4/5 (16-17) while the range for the other genes was 2-4 (**Table 1** and **Figure 6**). Earlier studies on different crops also showed the same trend of the presence of exons and introns in the predicted RNAi genes(Cui et al., 2020b; Faysal et al., 2021; Mosharaf et al., 2021a; Akond et al., 2022b; Podder et al., 2023). A study suggested that RDR genes show important roles in plants for gene silencing against viral infection, for example, CaRDR1 had an important role in pepper resistance against TMV(Qin et al., 2017b). This analysis however suggested that the exon-intron patterns are analogous within the members of each subgroup of DCL, AGO, and RDR, and hence, the gene (exon-intron) structures of the BnDCL, BnAGO, and BnRDR genes were very similar to those of the AtRNAi genes too.

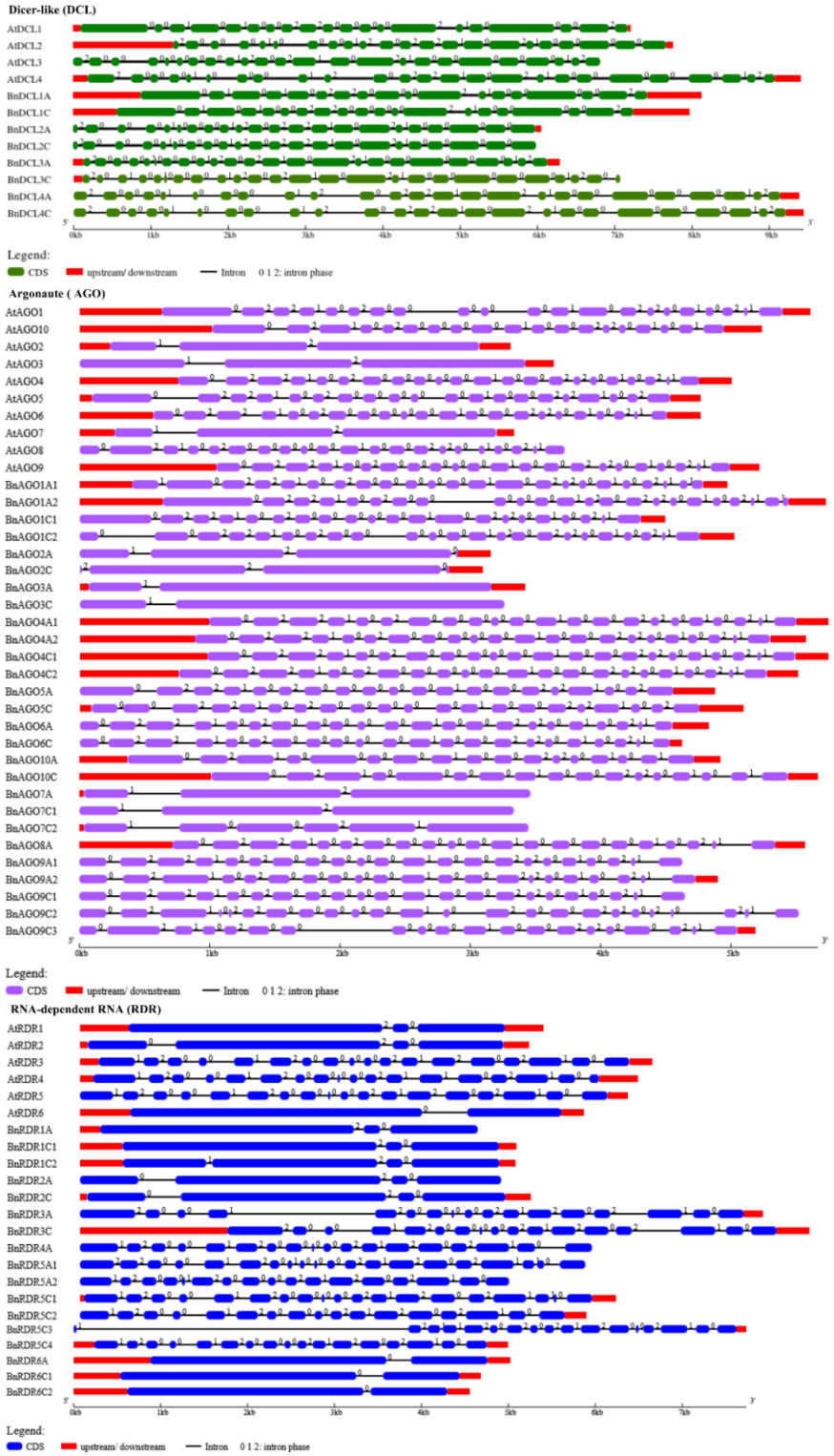


Figure 6. Gene(exon-intron) structure of the *A. thaliana* and *B.napus* DCL, AGO, and RDR genes. The exons (green, violet, and blue), introns (black lines), and intron phases (0, 1, and 2) are mentioned in the figure. The gene structure was analyzed using the online GSDS 2.0 server by considering their full-length genomic/DNA and coding sequences (CDS).

3.6 Functional enrichment analysis of BnDCL, BnAGO, and BnRDR genes

To elucidate the involvement of the particular RNAi genes of *B. napus* in performing various biological processes (BPs) and molecular functions (MFs) and also to predict their cellular locations, we made an enrichment analysis using GO. The results showed that a pool of important BPs and MFs are linked with the identified BnRNAi genes (Supplementary File 1). The top 10 significantly enriched BP, MF, and CC terms with their associated genes are recorded in Table 3. GO BPs presented significant enrichment of BnRNAi genes in processes related to transcription, response to the virus/defense response to the virus, immune effector process, post-transcriptional gene silencing, modification of host morphology by symbiont, and dsRNA fragmentation (**Table 3**). Concerning predicted BPs, the virus response (GO:0009615, p-value:8.00E-29), defense response to the virus process (GO:0051607, p-value:1.40E-28), and immune effector process (GO:0002252, 2.10E-26) are connected to as many as 11 BnAGO, 3 BnDCL and BnRDR1A genes and the association of these BPs were also suggested by the previous studies (Mourrain et al., 2000b; Deleris et al., 2006b; Margis et al., 2006b; Faysal et al., 2021; Mosharaf et al., 2021a). GO analysis for BP exhibited that 5 BnDCL, 5 BnAGO, and 2 BnRDR (BnRDR1A, 2A) genes are required for posttranscriptional gene silencing (GO:0016441, p-value:5.60E-26) and gene silencing by RNA (GO:0035194, 2.80E-24) (**Table 3**) that is also supported by some earlier investigation (Dalmay et al., 2000b; Sabbione et al., 2019; Mosharaf et al., 2021a). The function of the posttranscriptional gene silencing in plants is known as one of the important BP processes that frequently occurs in plants(Fire et al., 1998; Mosharaf et al., 2021b). For the fragmentation of the dsRNA substrate (GO:0031050, p-value:1.50E-22) 5 BnDCL, 3 BnAGO (BnAGO4C1,6C,7C1), and 2 BnRDR genes (BnRDR1A, 2A) were predicted to link during RNAi process and this connection had also been proved with the GO results of an earlier analysis on the banana and sweet orange (Faysal et al., 2021; Mosharaf et al., 2021b). Hence it suggests that the process of dsRNA fragmentation is one of the crucial parts of the RNAi mechanism in crops. The modification of the host morphology (GO:0044003, p-value:8.10E-23) was related to 4 BnDCLs and 5 BnAGOs.

In the situation of MFs, RNA polymerase activity, nucleic acid binding, dsRNA-specific ribonuclease, siRNA binding, endoribonuclease activity, etc., were predicted to be largely linked with several RNAi genes in *B.napus* (**Table 3**). GO analysis about MF showed that only RDR genes (12 BnRDR genes out of 17) were connected to two polymerase actions viz., RNA-directed RNA polymerase (GC:0003968, p-value:2.48E-27) and RNA polymerase (GO:0034062, p-value: 9.90E-16). In nucleic acid binding (GO:0003676, p-value:9.78E-17) activity, all 27 predicted BnAGO genes, 7 BnDCL out of 8 and 3 BnRDR (BnRDR 1C2, 2A, 2C) out of 17 were linked and this was also supported by the GO analysis of RNAi genes in sweet orange and date palm (Mosharaf et al., 2021b; Naim et al., 2024). To perform cleavage action two GO terms ribonuclease III activity (GO:0004525; 4.01E-14) and double-stranded RNA-specific ribonuclease activity (GO:0032296, p-value: 4.01E-14) are significantly associated with 7 BnDCL genes except BnDCL3A. Previous reports also suggested that most of the RNAi genes belonging to the DCL family hold this enzymatic characteristic to cleave the dsRNA to start the process of RNAi-mediated gene silencing pathway in different crops (Qian et al., 2011b; Akond et al., 2022b). Some 5 BnAGO genes like AtAGO and OsAGO were also responsible for the activity of siRNA binding (GO:0035197, p-value: 2.10E-13). The GO analysis also demonstrated that the endoribonuclease activity including the term (GO:0004521, p-value: 3.06E-11) was associated with almost all dicer-like and one BnAGO1C1 genes in *B.napus* (Table 3) and this results also supported by a previous study on banana (Faysal et al., 2021). In the case of CCs, the BnRNAi genes are mainly enriched in the nuclear lumen, nucleolus, organelle lumen, lumina of intracellular organelle, membrane-enclosed lumen, nuclear chromatin, etc. (**Table 3**). In this GO analysis, we also studied the similarities and differences of the predicted GO terms belonging to the RNAi genes of date palm (Phoenix dactylifera), *A. thaliana,* and rice (*Oryza sativa*) (**Table 3**). Our results therefore suggested that in *B. napus*, a large number of RNAi genes are associated with some important biological processes and molecular actions.

Table 3. The top 10 significantly (p<0.05) enriched GO functions. The analysis results of BnRNAi with the AtRNAi and OsRNAi proteins

GO ID	Annotated Terms	p-value	Annotated Genes
Biological Processes (BP)			
GO:0001172	transcription, RNA-templated	1.00E-30	BnRDR-1A, 1C2, 2A, 2C, 3A, 3C, 4A, 5A1, 5A2, 5C1, 5C3
GO:0009615	response to virus	8.00E-29	BnAGO-2A, 5A, 8A, 10A, 6C, BnDCL-1A, 2A, 4A, 3C, BnRDR1A, BnAGO7C1, BnAGO4C1, BnAGO1C1 AtRDR-1, 6, AtAGO-1, 2, 3, 4, 5, 6, 7, 9, 10, AtDCL-2, 3, 4 SHL2, OsAGO-1a, 2, 4b, OsRDR1, OsDCL-1a, 2a, 3b, SHL4, OsMEL1, SHO1, OsPNH1
GO:0051607	defense response to virus	1.40E-28	BnAGO-1C1, 2A, 4C1, 5A, 8A, 10A, 6C, 7C1, BnDCL-1A, 2A, 4A, 3C
GO:0002252	immune effector process	2.10E-26	BnAGO-2A, 4A, 5A, 8A,10A, 1C1, 3C, 4C1, 6C, 7C1, BnDCL-1A, 2A, 4A
GO:0016441	posttranscriptional gene silencing	5.60E-26	BnDCL-1A, 2A, 4A, 3C, 4C, BnRDR-1A, 2A, BnAGO-1C1, 5A, 6C, 4C1,7C1,
GO:0030422	production of siRNA involved in RNA interference	1.00E-24	BnDCL-1A, 2A, 4A, 3C, 4C, BnRDR-1A, 2A, BnAGO-6C, 4C1, 7C1 AtDCL-1,2,3,4, AtRDR-1, AtAGO-4,6,7, AtRDR-2,6 SHL2, OsRDR-1, 2, SHL4, OsDCL2a, OsAGO4b, SHO1, OsDCL-1a, 1b, 1c, 2a, 2b, 3b
GO:0016246	RNA interference	2.10E-24	BnDCL-1A, 2A, 4A, 3C, 4C, BnRDR-1A, 2A, BnAGO-4C1, 6C, 7C1 AtDCL-1, AtRDR-1, AtAGO-7, AtAGO-4, AtAGO-6, AtDCL-2, AtDCL-3, AtRDR-6, AtRDR-2, AtDCL-4, SHL2, OsRDR-1,2, SHL4, OsAGO4b, SHO1, OsDCL-1a, 1b, 1c, 2a, 2b
GO:0035194	posttranscriptional gene silencing by RNA	2.80E-24	BnDCL-1A, 2A, 4A, 3C, 4C, BnRDR-1A, 2A, BnAGO-1C1, 6C, 7C1, 4C1
GO:0044003	modification by symbiont of host morphology or physiology	8.10E-23	BnDCL-1A, 2A, 4A, 3C, BnAGO-2A, 5A, 1C1, 4C1, 7C1
GO:0031050	dsRNA fragmentation	1.50E-22	BnDCL-1A, 2A, 4A, 3C, 4C, BnRDR-1A, 2A, BnAGO-4C1, 6C, 7C1 AtDCL-1, 2, 3, 4, AtRDR-1, 2, 6, AtAGO-4, 6, 7 SHL2, OsRDR-1, 2, SHL4, OsAGO4b, SHO1, OsDCL-1a, 1b, 1c, 2a, 2b, 3b
Molecular Functions (MF)			
GO:0003968	RNA-directed RNA polymerase activity	2.48E-27	BnRDR-1A, 1C2, 2A, 2C, 3A, 3C, 4A, 5A1, 5A2, 5C2, 5C1, 5C3 AtRDR-1,2, 6, OsRDR-1, 2, 3, 4, SHL2
GO:0003676	nucleic acid binding	9.78E-17	BnAGO-1A1, 1A2, 1C1, 1C2, 2A, 2C, 3A, 3C, 4A2, 5A, 6A,7A, 7C2, 8A, 9A2, 9A1, 9C1, 9C2, 9C3, 4A1, 4C1, 4C2, 10A, 10C, 5C, 6C, 7C1, BnDCL-1A, 1C, 2A,2C, 3C, 4A, 4C, BnRDR-1C2, 2A, 2C
GO:0034062	RNA polymerase activity	9.90E-16	BnRDR-1A, 1C2, 2A, 2C, 3A,3C, 4A, 5A1, 5A2, 5C1,5C2, 5C3
GO:0004525	ribonuclease III activity	4.01E-14	BnDCL-1A, 1C, 2a, 2C, 3C, 4A, 4C AtDCL-1,2 3, 4, OsDCL-1a, 1b, 1c, 2a, 2b, 3a, 3b, SHO1
GO:0032296	double-stranded RNA-specific ribonuclease activity	4.01E-14	BnDCL-1A, 1C, 2A, 2C, 3C, 4A, 4C, AtDCL-1, 2, 3, 4 OsDCL-1a, 1b, 1c, 2a, 2b, 3a, 3b, SHO1
GO:0035197	siRNA binding	2.10E-13	BnAGO-1C1, 2A, 8A, 4C1, 6C, AtAGO-1, 2, 4, 6, 9 OsAGO-1a,2, 4b, OsMEL1
GO:0016779	nucleotidyltransferase activity	3.04E-12	BnRDR-1A, 1C2, 2A, 2C, 3A, 3C, 4A, 5A1, 5A2, 5C1, 5C2, 5C3
GO:0004521	endoribonuclease activity	3.06E-11	BnDCL-1A, 1C, 2A, 2C, 3C, 4A, 4C, BnAGO1C1
GO:0016891	endoribonuclease activity, producing 5'-phosphomonoesters	2.72E-10	BnDCL2A, BnDCL1A, BnDCL4A, BnDCL3C, BnDCL1C, BnDCL2C, BnDCL4C
GO:0004540	ribonuclease activity	2.72E-10	BnDCL-1A, 1C, 2A, 2C, 4A, 3C, 4C, BnAGO1C1
Cellular Components (CC)			
GO:0031981	nuclear lumen	8.70E-06	BnRDR2A, BnDCL-1A, 3C, 4C1, 5C
GO:0005730	nucleolus	1.50E-05	BnRDR2A, BnDCL-3C, BnAGO-4C1, 5C
GO:0043233	organelle lumen	1.70E-05	BnRDR2A, BnDCL-1A, 3C, BnAGO-4C1, 5C
GO:0070013	intracellular organelle lumen	1.70E-05	BnRDR2A, BnDCL-1A, 3C, BnAGO-4C1, 5C
GO:0031974	membrane-enclosed lumen	1.80E-05	BnRDR2A, BnDCL-1A, 3C, BnAGO-4C1, 5C
GO:0044428	nuclear part	3.40E-05	BnRDR2A, BnDCL-1A, 3C, BnAGO-4C1, 5C
GO:0000790	nuclear chromatin	5.80E-05	BnAGO-4C1, 5C
GO:0005634	nucleus	0.00019	BnAGO-1C1, 4C1,5C, 8A, 6C, BnRDR2A, BnDCL-1A, 3C, 4A
GO:0044454	nuclear chromosome part	0.00025	BnAGO5C, BnAGO4C1
GO:0016604	nuclear body	0.00038	BnDCL1A, BnAGO4C1

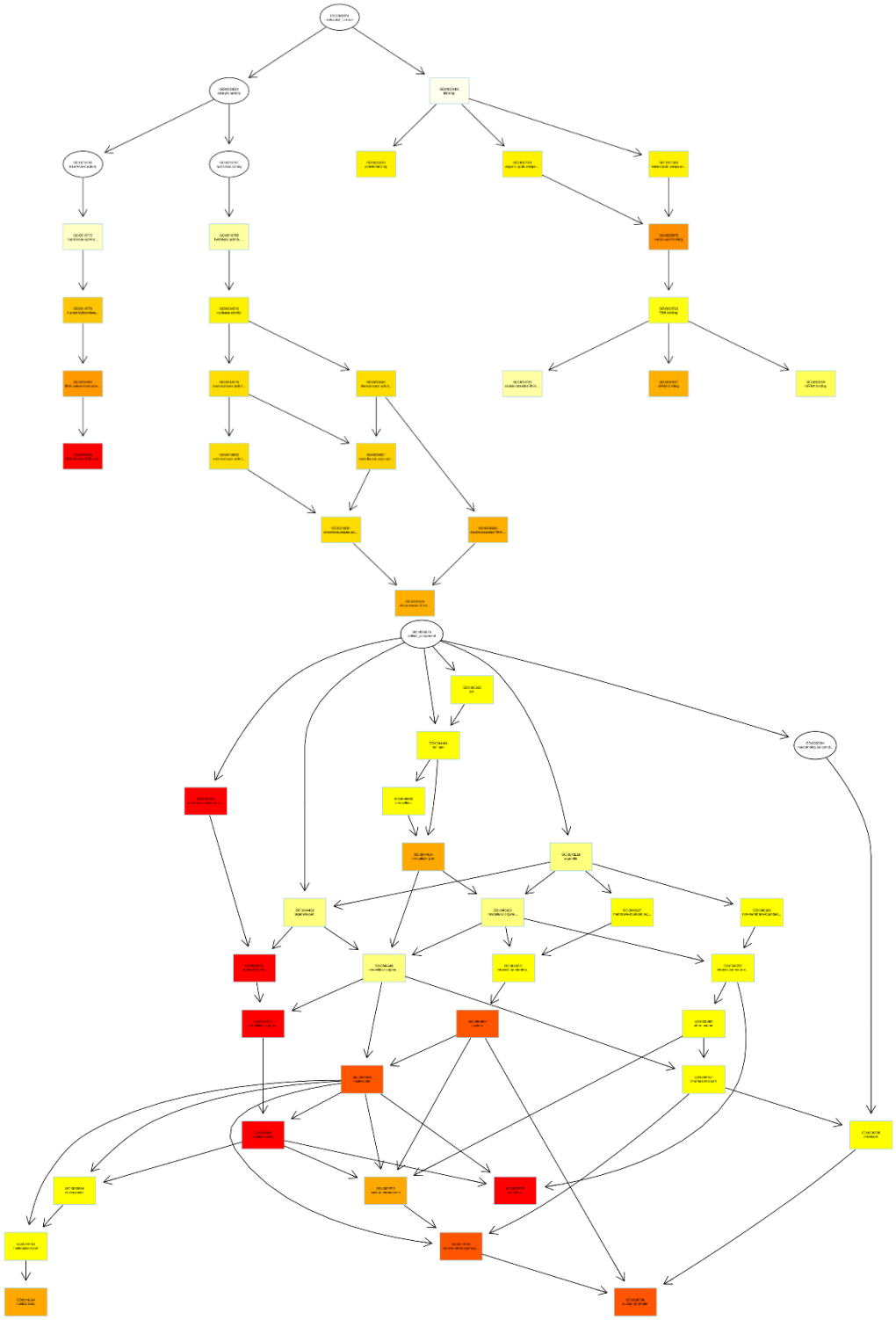


Figure 7. Gene enrichment pathway analysis map of BnRNAi genes

3.7 Analysis of subcellular locations of RNAi genes in *B. napus*

Cells are the basis for most biological and molecular functions in the eukaryotes. Identification of the cellular locations of the predicted proteins is a very crucial part of genomic research. Determination of the cellular compartment of the proteins would help to comprehend their functions at different cellular levels(Ehrlich et al., 2002; Glory and Murphy, 2007). The analysis results showed that BnDCL, BnAGO, and BnRDR proteins were located at various cellular segments including the nucleus, cytosol, chloroplast, vacuole, cytoskeletal, mitochondria, endoplasmic reticulum, plastid, peroxisome like

A.thaliana (**Figure 6(A-C)**). The heatmap showed that the maximum proteins were nucleus, chloroplast, and cytosol-centric with higher expression. The dark red color implied the level of intensity of the presence of the proteins in the respective CC. Among the BnDCL proteins, four proteins of the subfamily BnDCL3 (BnDCL3A and BnDCL3C) and BnDCL4 (BnDCL4A and BnDCL4C) were highly expressed in the nucleus (**Figure 6A**). In the case of BnAGO proteins, four members of the subfamily BnAGO4 (BnAGOA1, A2, C1, C2), two members of BnAGO6, three members of BnGAO7, BnAGO8, and five members of the BnAGO9 group were highly detected in the nucleus (**Figure 6B**). Four members of the BnAGO1 subgroup (BnAGO1A1, A2, C1, C2) and two members of the BnAGO5 subgroup (BnAGO5A and BnAGO) were chloroplast-centric. In the cytoplasm, two members of BnAGO10 (BnAGO10A, C), two members of BnAGO3(BnAGO3A,3C), BnAGO7C1, BnAGO6A were presented with higher level. For the BnRDR proteins, we observed that 7 proteins out of 17 (BnRDR1C1, BnRDR3A-C, BnRDR4, BnRDR5A1, BnRDR5C1, BnRDR5C3) were predicted to function in the nucleus (**Figure 6C**). Seven BnRDR proteins (BnRD5A1, A2, C2, C4, BnRDR6A, C1, C2) out of 17 were cytoplasmic oriented. Four proteins including BnRDR1A, BnRDR2A, 2C, and BnRDR5A2 were highly present in chloroplast. Reduction of particular mRNAs to lessen certain gene expression in plants typically happens in the cytoplasmic cellular positions that imply that the identified RNAi genes/proteins are intimately engaged in the PTGS(Agrawal et al., 2003). Cytosol is one of the vital CC in plants where the highest level of metabolism occurs(Lingel and Izaurralde, 2004) and we observed from the analysis that most of the proteins were also located in this component. Hence it may be assumed that the predicted proteins in *B.napus* located in cytoplasmic organelles are responsible for performing different chemical actions and energy transformations linked to rapeseed plant growth and development(Akond et al., 2022b). The proteins placed in mitochondria might have the role of engagement in developmental stages and different stress-responsive pathways(Liberatore et al., 2016). The CC chloroplast has a key role in oxygenic photosynthesis that generates energy and initial metabolism in plants. The chloroplast also has a vast role in the regulator of plant responses to abiotic and biotic stress conditions(Song et al., 2021).

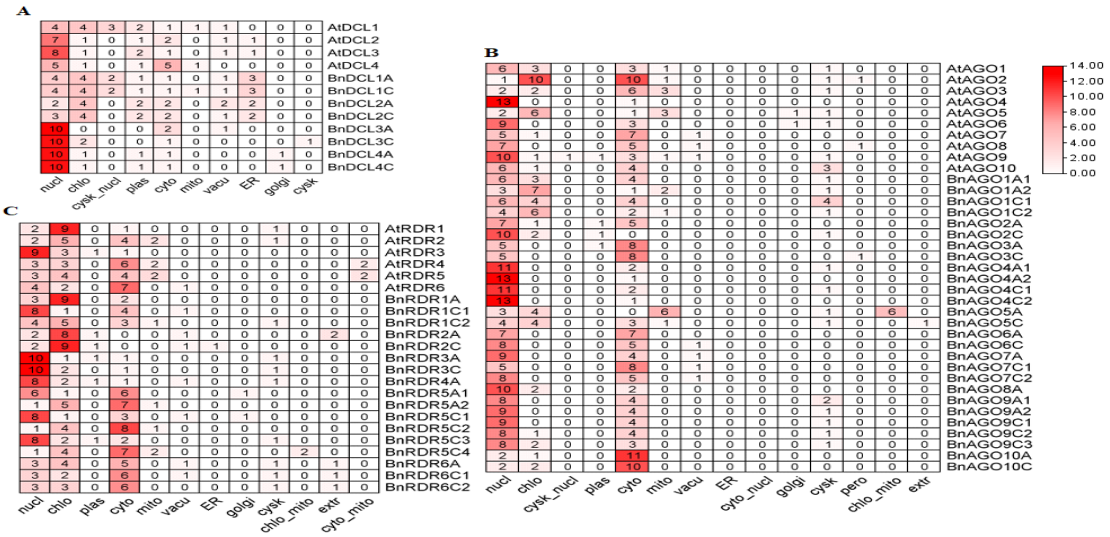


Figure 6. Sub-cellular localization analysis of BnRNAi and AtRNAi proteins. The dark red color represents the higher intensity of the presence of the genes in the respective cellular section. Protein percentages are found in several cellular compartments. In this analysis, nucl – nucleus, cyto –cytosol, chlo – chloroplast, vacu – vacuole, cysk – cytoskeletal, mito – mitochondria, E.R.–endoplasmic reticulum, plas – plastid, pero – peroxisome

4.0 Conclusion

Rapeseed-mustard crops are a great source of edible oil in Bangladesh. However, these crops frequently experience several pathogenic and environmental stressors that hamper the sufficient loss in yield production. Like other plants, in *B.napus*, some RNAi genes *BnDCL*, *BnAGO*, and *BnRDR* play important roles by controlling gene expression against stresses to help crops for healthy and timely growth and development. In this study, several bioinformatic techniques were applied to identify, characterize, and reveal the genomic diversity of the RNAi genes in *B.napus*. Our analysis finally suggests that the *B.napus* genome comprises 8 DCL, 27 AGO, and 17 RDR genes. Further, the phylogeny analysis exhibited that all members of these three gene groups keep their evolutionary relationships similar to their *Arabidopsis* orthologs. Functional domain analysis showed that all members of each *BnDCL*, *BnAGO*, and *BnRDR* group hold some definite functional domains consistent with the *AtDCL*, *AtAGO*, and *AtRDR* protein groups. Variations in the amino acid levels in the PIWI domain of the *BnAGO* proteins implied their involvement in physiological and biochemical activities and potential link to stress responses that are helpful for the higher yield production in *B.napus*. Gene structures also possess almost similar patterns to AtRNAi. Some potential BPs and MFs are associated with BnRNAi genes/proteins. Most of the proteins tend to be located in the nucleus, chloroplast, and cytosol. Hence, it can be concluded that this comprehensive bioinformatics analysis will further help biotechnologists/genetic engineers/ molecular plant breeders to explore the functional diversity of these genes/proteins in *B.napus* in regulating gene expression against different pathogenic and abiotic factors to improve mustard-rapeseed production in Bangladesh.

References

Agrawal, N., Dasaradhi, P. V. N., Mohmmed, A., Malhotra, P., Bhatnagar, R. K., and Mukherjee, S. K. (2003). RNA Interference: Biology, Mechanism, and Applications. *Microbiology and Molecular Biology Reviews*. doi: 10.1128/mmbr.67.4.657-685.2003

Ahmadovich Bozorov, T., Prakash Pandey, S., Dinh, S. T., Kim, S. G., Heinrich, M., Gase, K., et al. (2012). DICER-like Proteins and Their Role in Plant-herbivore Interactions in Nicotiana attenuata. *J Integr Plant Biol*. doi: 10.1111/j.1744-7909.2012. 01104.x

Akond, Z., Rahman, H., Ahsan, M. A., Mosharaf, M. P., Alam, M., and Mollah, M. N. H. (2022a). Comprehensive in Silico Analysis of RNA Silencing-Related Genes and Their Regulatory Elements in Wheat (Triticum aestivum L.). *Biomed Res Int* 2022. doi: 10.1155/2022/4955209

Alvarez-Diaz, J. C., Richard, M. M. S., Thareau, V., Teano, G., Paysant-Le-roux, C., Rigaill, G., et al. (2022). Genome-wide identification of key components of rna silencing in two phaseolus vulgaris genotypes of contrasting origin and their expression analyses in response to fungal infection. *Genes (Basel)* 13. doi: 10.3390/genes13010064

Arribas-Hernández, L., Marchais, A., Poulsen, C., Haase, B., Hauptmann, J., Benes, V., et al. (2016). The slicer activity of ARGONAUTE1 is required specifically for the phasing, not production, of trans-acting short interfering RNAs in arabidopsis. *Plant Cell* 28. doi: 10.1105/tpc.16.00121

Bai, M., Yang, G. S., Chen, W. T., Mao, Z. C., Kang, H. X., Chen, G. H., et al. (2012). Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in Solanum lycopersicum. *Gene* 501, 52–62. doi: 10.1016/j.gene.2012.02.009

Baumberger, N., and Baulcombe, D. C. (2005a). Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0505461102

Blevins, T., Rajeswaran, R., Shivaprasad, P. V., Beknazariants, D., Si-Ammour, A., Park, H. S., et al. (2006). Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res*. doi: 10.1093/nar/gkl886

Cao, J.-Y., Xu, Y.-P., Li, W., Li, S.-S., Rahman, H., and Cai, X.-Z. (2016). Genome-Wide Identification of Dicer-Like, Argonaute, and RNA-Dependent RNA Polymerase Gene Families in Brassica Species and Functional Analyses of Their Arabidopsis Homologs in Resistance to Sclerotinia sclerotiorum. *Front Plant Sci* 7, 1–17. doi: 10.3389/fpls.2016.01614

Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Gilbert, K. B., Montgomery, T. A., Nguyen, T., et al. (2012a). Functional analysis of three Arabidopsis argonautes using slicer-defective mutants. *Plant Cell*. doi: 10.1105/tpc.112.099945

Cenik, E. S., and Zamore, P. D. (2011). Argonaute proteins. *Current Biology*. doi: 10.1016/j.cub.2011.05.020

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-neolithic Brassica napus oilseed genome. *Science (1979)* 345. doi: 10.1126/science.1253435

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant* 13. doi: 10.1016/j.molp.2020.06.009

Cui, D. L., Meng, J. Y., Ren, X. Y., Yue, J. J., Fu, H. Y., Huang, M. T., et al. (2020a). Genome-wide identification and characterization of DCL, AGO and RDR gene families in Saccharum spontaneum. *Sci Rep*. doi: 10.1038/s41598-020-70061-7

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D. C. (2000a). An RNA-dependent RNA polymerase gene in arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell*. doi: 10.1016/S0092-8674(00)80864-8

Deleris, A., Gallago-Bartolome, J., Bao, J., Kasschau, K. D., Carrington, J. C., and Voinnet, O. (2006a). Hierarchical action and inhibition of plant dicer-like proteins in antiviral defense. *Science (1979)*. doi: 10.1126/science.1128214

Dubey, H., Kiran, K., Jaswal, R., Bhardwaj, S. C., Mondal, T. K., Jain, N., et al. (2020). Identification and characterization of Dicer-like genes in leaf rust pathogen (Puccinia triticina) of wheat. *Funct Integr Genomics*. doi: 10.1007/s10142-020-00745-w

Duvaud, S., Gabella, C., Lisacek, F., Stockinger, H., Ioannidis, V., and Durinx, C. (2021). Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. *Nucleic Acids Res* 49. doi: 10.1093/nar/gkab225

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform* 23, 205–211. doi: 10.1142/9781848165632_0019

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput Biol* 7. doi: 10.1371/journal.pcbi.1002195

Ehrlich, J. S., Hansen, M. D. H., and Nelson, W. J. (2002). Spatio-temporal regulation of Rac1 localization and lamellipodia dynamics during epithelial cell-cell adhesion. *Dev Cell* 3. doi: 10.1016/S1534-5807(02)00216-2

Fagard, M., Boutet, S., Morel, J.-B., Bellini, C., and Vaucheret, H. (2000). AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.200217597

Fang, Y., Wei, J., Huang, X., Li, Y., and Pan, X. (2023). Identification and Analysis of Protein Family Associated with RNA Interference Pathway in Juglandaceae. *Frontiers in Bioscience - Landmark* 28. doi: 10.31083/j.fbl2809218

Faysal, F., Id, A., Hossen, I., Sarkar, A. R., Konak, N., Zohra, F. T., et al. (2021). *Genome-wide identification of DCL , AGO and RDR gene families and their associated functional regulatory elements analyses in banana (Musa acuminata)*. doi: 10.1371/journal.pone.0256873

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution (N Y)*. doi: 10.2307/2408678

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in caenorhabditis elegans. *Nature*. doi: 10.1038/35888

Fusaro, A. F., Matthew, L., Smith, N. A., Curtin, S. J., Dedic-Hagan, J., Ellacott, G. A., et al. (2006). RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. *EMBO Rep*. doi: 10.1038/sj.embor.7400837

Galili, G. (2011). The aspartate-family pathway of plants. *Plant Signal Behav*. doi: 10.4161/psb.6.2.14425

Gan, D., Liang, D., Wu, J., Zhan, M., Yang, F., Xu, W., et al. (2016). Genome-Wide Identification of the Dicer-Like, Argonaute, and RNA-Dependent RNA Polymerase Gene Families in Cucumber (Cucumis sativus L.). *J Plant Growth Regul*. doi: 10.1007/s00344-015-9514-9

Gan, D., Zhan, M., Yang, F., Zhang, Q., Hu, K., Xu, W., et al. (2017). Expression analysis of argonaute, Dicer-like, and RNA-dependent RNA polymerase genes in cucumber (Cucumis sativus L.) in response to abiotic stress. *J Genet* 96. doi: 10.1007/s12041-017-0758-y

Gibbon, B. C., and Larkins, B. A. (2005). Molecular genetic approaches to developing quality protein maize. *Trends in Genetics*. doi: 10.1016/j.tig.2005.02.009

Glory, E., and Murphy, R. F. (2007). Automated Subcellular Location Determination and High-Throughput Microscopy. *Dev Cell* 12. doi: 10.1016/j.devcel.2006.12.007

Gupta, S. K. (2015). *Breeding Oilseed Crops for Sustainable Production: Opportunities and Constraints*. doi: 10.1016/C2013-0-19479-2

Hamar, E., Szaker, H. M., Kis, A., Dalmadi, A., Miloro, F., Szittya, G., et al. (2020). Genome-wide identification of RNA silencing-related genes and their expressional analysis in response to heat stress in barley (Hordeum vulgare l.). *Biomolecules*. doi: 10.3390/biom10060929

Henderson, I. R., Zhang, X., Lu, C., Johnson, L., Meyers, B. C., Green, P. J., et al. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet*. doi: 10.1038/ng1804

Höck, J., and Meister, G. (2008b). The Argonaute protein family. *Genome Biol* 9. doi: 10.1186/gb-2008-9-2-210

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: Protein localization predictor. *Nucleic Acids Res* 35. doi: 10.1093/nar/gkm259

Hunter, C., Sun, H., and Poethig, R. S. (2003). The Arabidopsis Heterochronic Gene ZIPPY Is an ARGONAUTE Family Member. *Current Biology*. doi: 10.1016/j.cub.2003.09.004

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., et al. (2010). Hsc70/Hsp90 Chaperone Machinery Mediates ATP-Dependent RISC Loading of Small RNA Duplexes. *Mol Cell* 39, 292–299. doi: 10.1016/J.MOLCEL.2010.05.015

Ji, L., Liu, X., Yan, J., Wang, W., Yumul, R. E., Kim, Y. J., et al. (2011). ARGONAUTE10 and ARGONAUTE1 regulate the termination of floral stem cells through two microRNAs in Arabidopsis. *PLoS Genet*. doi: 10.1371/journal.pgen.1001358

Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res*. doi: 10.1093/nar/gkw982

Jing, X., Xu, L., Huai, X., Zhang, H., Zhao, F., and Qiao, Y. (2023). Genome-Wide Identification and Characterization of Argonaute, Dicer-like and RNA-Dependent RNA Polymerase Gene Families and Their Expression Analyses in Fragaria spp. *Genes (Basel)* 14. doi: 10.3390/genes14010121

Johnson, L. S., Eddy, S. R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* 11. doi: 10.1186/1471-2105-11-431

Jovel, J., Walker, M., and Sanfaçon, H. (2007). Recovery of Nicotiana benthamiana Plants from a Necrotic Response Induced by a Nepovirus Is Associated with RNA Silencing but Not with Reduced Virus Titer. *J Virol*. doi: 10.1128/jvi.01192-07

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., et al. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat Genet*. doi: 10.1038/ng.548

Kapoor, M., Arora, R., Lama, T., Nijhawan, A., Khurana, J. P., Tyagi, A. K., et al. (2008a). Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. *BMC Genomics* 9, 1–17. doi: 10.1186/1471-2164-9-451

Kishor, P. B. K., Suravajhala, R., Rajasheker, G., Marka, N., Shridhar, K. K., Dhulala, D., et al. (2020). Lysine, Lysine-Rich, Serine, and Serine-Rich Proteins: Link Between Metabolism, Development,

and Abiotic Stress Tolerance and the Role of ncRNAs in Their Regulation. *Front Plant Sci* 11. doi: 10.3389/fpls.2020.546213

Krishnatreya, D. B., Baruah, P. M., Dowarah, B., Chowrasia, S., Mondal, T. K., and Agarwala, N. (2021). Genome-wide identification, evolutionary relationship and expression analysis of AGO, DCL and RDR family genes in tea. *Sci Rep* 11. doi: 10.1038/s41598-021-87991-5

Kumar, S., Stecher, G., and Tamura, K. (2016a). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33, 1870–1874. doi: 10.1093/molbev/msw054

Liberatore, K. L., Dukowic-Schulze, S., Miller, M. E., Chen, C., and Kianian, S. F. (2016). The role of mitochondria in plant development and stress tolerance. *Free Radic Biol Med*. doi: 10.1016/j.freeradbiomed.2016.03.033

Lingel, A., and Izaurralde, E. (2004). RNAi: Finding the elusive endonuclease. *RNA*. doi: 10.1261/rna.7175704

Liu, X., Lu, T., Dou, Y., Yu, B., and Zhang, C. (2014a). Identification of RNA silencing components in soybean and sorghum. *BMC Bioinformatics* 15, 1–13. doi: 10.1186/1471-2105-15-4

Llave, C. (2010). Virus-derived small interfering RNAs at the core of plant-virus interactions. *Trends Plant Sci*. doi: 10.1016/j.tplants.2010.09.001

MacRae, I. J., and Doudna, J. A. (2007). Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Curr Opin Struct Biol*. doi: 10.1016/j.sbi.2006.12.002

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., et al. (2017). CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*. doi: 10.1093/nar/gkw1129

Margis, R., Fusaro, A. F., Smith, N. A., Curtin, S. J., Watson, J. M., Finnegan, E. J., et al. (2006a). The evolution and diversification of Dicers in plants. *FEBS Lett*. doi: 10.1016/j.febslet.2006.03.072

Meng, F., Jia, H., Ling, N., Xue, Y., Liu, H., Wang, K., et al. (2013a). Cloning and characterization of two Argonaute genes in wheat (Triticum aestivum L.). *BMC Plant Biol*. doi: 10.1186/1471-2229-13-18

Mirzaei, K., Bahramnejad, B., Shamsifard, M. H., and Zamani, W. (2014). In silico identification, phylogenetic and bioinformatic analysis of argonaute genes in plants. *Int J Genomics*. doi: 10.1155/2014/967461

Mohanta, T. K., Khan, A., Hashem, A., Abdallah, E. F., and Al-Harrasi, A. (2019). The molecular mass and isoelectric point of plant proteomes. *BMC Genomics* 20. doi: 10.1186/s12864-019-5983-8

Mosharaf, M. P., Rahman, H., Ahsan, M. A., Akond, Z., Ahmed, F. F., Islam, M. M., et al. (2021a). *In silico identification and characterization of AGO, DCL and RDR gene families and their associated regulatory elements in sweet orange (Citrus sinensis L.)*. doi: 10.1371/journal.pone.0228233

Moura, M. O., Fausto, A. K. S., Fanelli, A., Guedes, F. A. D. F., Silva, T. D. F., Romanel, E., et al. (2019). Genome-wide identification of the Dicer-like family in cotton and analysis of the DCL expression modulation in response to biotic stress in two contrasting commercial cultivars. *BMC Plant Biol*. doi: 10.1186/s12870-019-2112-4

Mourrain, P., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., et al. (2000a). Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*. doi: 10.1016/S0092-8674(00)80863-6

Moussian, B., Schoof, H., Haecker, A., Jürgens, G., and Laux, T. (1998). Role of the ZWILLE gene in the regulation of central shoot meristem cell fate during Arabidopsis embryogenesis. *EMBO Journal*. doi: 10.1093/emboj/17.6.1799

Naim, D., Ahsan, A., Imtiaj, A., and Mollah, N. H. (2024). Genome-wide identification and in silico characterization of major RNAi gene families in date palm (Phoenix dactylifera). *BMC Genom Data* 25. doi: 10.1186/s12863-024-01217-x

Podder, A., Ahmed, F. F., Suman, M. Z. H., Mim, A. Y., and Hasan, K. (2023). *Genome-wide identification of DCL, AGO and RDR gene families and their associated functional regulatory element analyses in sunflower (Helianthus annuus)*. doi: 10.1371/journal.pone.0286994

Prakash, V., Devendran, R., and Chakraborty, S. (2017). Overview of plant RNA dependent RNA polymerases in antiviral defense and gene silencing. *Indian J Plant Physiol*. doi: 10.1007/s40502-017-0339-3

Qian, Y., Cheng, Y., Cheng, X., Jiang, H., Zhu, S., and Cheng, B. (2011a). Identification and characterization of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families in maize. *Plant Cell Rep* 30, 1347–1363. doi: 10.1007/s00299-011-1046-6

Qin, C., Li, B., Fan, Y., Zhang, X., Yu, Z., Ryabov, E., et al. (2017a). Roles of dicer-like proteins 2 and 4 in intra- and intercellular antiviral silencing. *Plant Physiol*. doi: 10.1104/pp.17.00475

Qin, L., Mo, N., Muhammad, T., and Liang, Y. (2018a). Genome-wide analysis of DCL, AGO, and RDR gene families in pepper (Capsicum Annuum L.). *Int J Mol Sci*. doi: 10.3390/ijms19041038

Qin, L., Mo, N., Zhang, Y., Muhammad, T., Zhao, G., Zhang, Y., et al. (2017b). CaRDR1, an RNA-dependent RNA polymerase plays a positive role in pepper resistance against TMV. *Front Plant Sci*. doi: 10.3389/fpls.2017.01068

Rédei, G. P. (2008). "CLUSTAL W (improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice)," in *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*. doi: 10.1007/978-1-4020-6754-9_3188

Rivas, F. V., Tolia, N. H., Song, J. J., Aragon, J. P., Liu, J., Hannon, G. J., et al. (2005a). Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol*. doi: 10.1038/nsmb918

Ros, R., Muñoz-Bertomeu, J., and Krueger, S. (2014). Serine in plants: Biosynthesis, metabolism, and functions. *Trends Plant Sci* 19. doi: 10.1016/j.tplants.2014.06.003

Sabbione, A., Daurelio, L., Vegetti, A., Talón, M., Tadeo, F., and Dotto, M. (2019). Genome-wide analysis of AGO, DCL and RDR gene families reveals RNA-directed DNA methylation is involved in fruit abscission in Citrus sinensis. *BMC Plant Biol*. doi: 10.1186/s12870-019-1998-1

Schmitz, R. J., Hong, L., Fitzpatrick, K. E., and Amasino, R. M. (2007). DICER-LIKE 1 and DICER-LIKE 3 redundantly act to promote flowering via repression of FLOWERING LOCUS C in Arabidopsis thaliana. *Genetics*. doi: 10.1534/genetics.107.070649

Schuurmans Stekhoven, F. M. A. H., Gorissen, M. H. A. G., and Flik, G. (2008). The isoelectric point, a key to understanding a variety of biochemical problems: A minireview. *Fish Physiol Biochem* 34. doi: 10.1007/s10695-007-9145-6

Song, Y., Feng, L., Alyafei, M. A. M., Jaleel, A., and Ren, M. (2021). Function of chloroplasts in plant stress responses. *Int J Mol Sci* 22. doi: 10.3390/ijms222413464

Stephen F. Altschul, Thomas L. Madden, A. A., Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and D. J., and (1997), L. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*.

Styles, B. T. (2003). Encyclopedia of Food Sciences and Nutrition. *Encyclopedia of Food Sciences and Nutrition*.

Wang, T., You, L., Li, R., Fu, D. Q., Zhu, B. Z., Luo, Y. B., et al. (2016). Cloning, identification, and expression analysis of a Dicer-Like gene family from Solanum lycopersicum. *Biol Plant*. doi: 10.1007/s10535-016-0620-8

Wang, W., Xu, M., Wang, G., and Galili, G. (2018). New insights into the metabolism of aspartate-family amino acids in plant seeds. *Plant Reprod*. doi: 10.1007/s00497-018-0322-9

Yadav, C. B., Muthamilarasan, M., Pandey, G., and Prasad, M. (2015). Identification, characterization and expression profiling of dicer-like, argonaute and rna-dependent RNA polymerase gene families in foxtail millet. *Plant Mol Biol Report*. doi: 10.1007/s11105-014-0736-y

Yang, H. Q., and Gao, H. J. (2007). Physiological function of arginine and its metabolites in plants. *Journal of Plant Physiology and Molecular Biology* 33.

Yun, S., and Zhang, X. (2023). Genome-wide identification, characterization and expression analysis of AGO, DCL, and RDR families in Chenopodium quinoa. *Sci Rep* 13, 3647. doi: 10.1038/s41598-023-30827-1

Zhao, X., Zheng, W., Zhong, Z., Chen, X., Wang, A., and Wang, Z. (2016). Genome-wide analysis of RNA-interference pathway in Brassica napus, and the expression profile of BnAGOs in response to Sclerotinia sclerotiorum infection. *Eur J Plant Pathol* 146, 565–579. doi: 10.1007/s10658-016-0942-6

Zhu, K., Zhao, J., Lubman, D. M., Miller, F. R., and Barder, T. J. (2005). Protein pI shifts due to posttranslational modifications in the separation and characterization of proteins. *Anal Chem* 77, 2745–55. doi: 10.1021/ac048494w

Zilberman, D., Cao, X., and Jacobsen, S. E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science (1979)*. doi: 10.1126/science.1079695

COMBINATION OF SOIL SAMPLING AND DRONE MAPPING EFFECTS ON FERTILIZER RECOMMENDATION FOR CROP PRODUCTION

ISTIAK AHMED[1], MOHAMMAD MASUDUZZAMAN MASUD[2], MOHAMMAD MUKHLESUR RAHMAN[1], NUR MOHAMMAD[1], KOUSHIK KUMAR SAHA[1] AND HABIB MOHAMMAD NASER[2]

[1]ASICT Division, BARI, Gazipur and [2]Soil Science Division, BARI, Gazipur

Abstract

This study explores the use of advanced machine learning techniques to map and analyze the spatial variability of soil nutrient content in agricultural fields using multispectral imagery captured by a DJI Phantom 4 Multispectral UAV. A total of 13 soil nutrient parameters were measured across 30 samples, with the distribution and variability assessed using boxplots and Completely Randomized Design (CRD) tests. Approximately 50% of the soil nutrients were found to deviate from a normal distribution, with significant variability observed between plots for several key nutrients. To capture and map this variability, an unsupervised deep learning approach was employed. An autoencoder was used to extract latent features from the multispectral image data, which were then clustered using the K-Means algorithm. The integration of autoencoder-based feature extraction with traditional clustering methods significantly improved the classification accuracy, allowing for the identification of 20 distinct classes within the study area. The spatial distribution of these classes was further analyzed, revealing clusters of plots with specific nutrient enrichments. The findings underscore the limitations of traditional parametric and unsupervised learning methods in capturing complex spatial patterns in soil nutrient content, particularly when using multispectral data during the bare soil period. The study demonstrates the potential of deep learning techniques, such as autoencoders, to enhance the analysis of soil properties and support more precise soil management practices. The results have important implications for the development of targeted agricultural interventions aimed at optimizing soil fertility and improving crop yields.

Keywords: Soil Nutrient Mapping, Multispectral Imagery, K-Means Clustering, Deep Learning, Autoencoder, UAV

Introduction

Agriculture has always been a cornerstone of human civilization, providing the essential resources needed for sustenance and economic growth. As the global population continues to rise, the demand for increased agricultural productivity and sustainability has never been more pressing. In response to these challenges, the agriculture sector is undergoing significant transformations through the adoption of precision agriculture technologies. Precision agriculture, a concept that has been evolving over the past few decades, aims to optimize resource use and maximize crop yields by utilizing advanced technologies for monitoring and managing agricultural practices. The advent of the Fourth Industrial Revolution (4IR) technologies has further accelerated these advancements, offering innovative solutions for enhancing agricultural efficiency, sustainability, and profitability. The Fourth Industrial Revolution (4IR) is characterized by the convergence of digital, physical, and biological systems, fundamentally altering how industries operate (Schwab, 2017). In the context of agriculture, 4IR technologies encompass a wide range of innovations, including the Internet of Things (IoT), artificial intelligence (AI), big data analytics, robotics, and unmanned aerial vehicles (UAVs) (Klerkx et al., 2019). These technologies have revolutionized precision agriculture by enabling farmers to collect and analyze data at an unprecedented scale and precision.

IoT devices, such as sensors and connected machinery, allow for real-time monitoring of soil moisture, temperature, humidity, and crop health. This data-driven approach enables farmers to make informed decisions regarding irrigation, fertilization, and pest control, ultimately optimizing resource use and reducing waste (Wolfert et al., 2017). AI and machine learning algorithms further enhance this process by analyzing complex datasets and providing predictive insights into crop performance and disease outbreaks (Liakos et al., 2018). Robotics and automation play a pivotal role in precision agriculture by streamlining labor-intensive tasks such as planting, harvesting, and weeding. Autonomous machinery equipped with AI-powered systems can operate with high precision and efficiency, reducing the need for manual labor and minimizing human error (Bechar & Vigneault, 2017).

One of the most transformative technologies in precision agriculture is the use of UAVs, or drones, which offer a unique vantage point for monitoring crops and soil conditions. Equipped with high-resolution cameras and multispectral sensors, UAVs can capture detailed imagery of fields, enabling the assessment of crop health, nutrient levels, and pest infestations. This information can be processed using advanced algorithms to create accurate maps that guide targeted interventions, such as variable-rate fertilizer application (Zhang & Kovacs, 2012). As a predominantly agrarian economy, Bangladesh faces numerous challenges in modernizing its agricultural practices. Despite these challenges, the country has made notable strides in adopting precision agriculture technologies. Government initiatives, international collaborations, and research projects have laid the foundation for the integration of 4IR technologies in agriculture (Kabir et al., 2020). The Bangladesh Agricultural Research Institute (BARI) and the Department of Agricultural Extension (DAE) have been actively involved in promoting precision agriculture through pilot projects and training programs. These initiatives aim to raise awareness among farmers about the benefits of precision agriculture and provide them with the necessary skills and knowledge to adopt these technologies (Roy & Chan, 2020). However, widespread adoption remains limited due to factors such as high initial costs, lack of infrastructure, and limited access to technology in rural areas (Alauddin & Quiggin, 2008). To overcome these barriers, partnerships between the government, private sector, and international organizations are crucial. Investments in infrastructure, capacity building, and technology transfer can facilitate the dissemination of precision agriculture technologies, empowering farmers to enhance productivity and sustainability (Hossain et al., 2020).

In Bangladesh, conventional blanket fertilizer application methods have been the norm for decades. This approach involves applying a uniform rate of fertilizers across an entire field, regardless of variations in soil fertility and nutrient requirements (Miah et al., 2016). Typically, farmers rely on generalized recommendations provided by agricultural extension services, which may not accurately reflect the specific needs of their fields (Roy & Chan, 2020). The demerits of blanket fertilizer application are significant. Firstly, it often leads to over-application or under-application of nutrients, resulting in imbalanced nutrient levels that can harm crop growth and yield (Rahman & Hasan, 2011). Over-application of fertilizers not only increases production costs but also poses environmental risks, such as soil degradation, water pollution, and greenhouse gas emissions (Islam et al., 2019). Conversely, under-application can lead to nutrient deficiencies, reducing crop productivity and quality (Haque et al., 2015). Inefficient nutrient management practices stemming from blanket fertilizer application contribute to the depletion of soil health and fertility over time. The lack of site-specific nutrient management hinders farmers' ability to optimize inputs, ultimately affecting profitability and sustainability (Sarker et al., 2017).

UAV-based nutrient mapping presents a promising solution to the challenges posed by conventional blanket fertilizer application methods. By utilizing UAVs equipped with multispectral sensors, farmers can obtain high-resolution images of their fields that capture variations in soil and crop conditions. These images can be processed using specialized software to generate nutrient maps, highlighting areas with different nutrient levels (Zhang & Kovacs, 2012). The advantages of UAV-based nutrient mapping are manifold. It enables farmers to implement site-specific nutrient management practices, ensuring that fertilizers are applied precisely where and when they are needed. This targeted approach minimizes resource wastage, reduces environmental impact, and enhances crop yields. Furthermore, UAVs provide rapid data collection, allowing for timely interventions and adaptive management strategies (Bendig et al., 2015). The feasibility of UAV-based nutrient mapping depends on several factors, including technological infrastructure, cost considerations, and user accessibility. In recent years, advancements in UAV technology have led to the development of more affordable and user-friendly drones equipped with multispectral sensors. These advancements have made UAV-based nutrient mapping increasingly accessible to farmers, even in resource-constrained settings (Torres-Sánchez et al., 2013). The cost-effectiveness of UAV-based nutrient mapping is evident in its potential to reduce input costs and increase crop productivity. By optimizing fertilizer use, farmers can achieve higher yields while minimizing production expenses. Additionally, the environmental benefits of reduced chemical inputs align with sustainable agriculture practices, contributing to long-term soil health and ecosystem preservation (Hunt et al., 2014). Multispectral sensors mounted on UAVs are capable of capturing variations in nutrient levels within fields. These sensors measure reflected light across different wavelengths, providing valuable information about plant health and soil properties. For instance, specific

spectral bands can indicate chlorophyll content, which is closely linked to nitrogen levels in plants. Similarly, other bands can detect moisture content and stress indicators, aiding in the assessment of nutrient deficiencies (Houborg et al., 2015). By analyzing multispectral data, farmers can gain insights into the spatial distribution of nutrients and identify areas that require targeted interventions. This capability enables more accurate and efficient nutrient management practices, ultimately improving crop health and productivity (Adao et al., 2017). UAV-based nutrient mapping offers numerous benefits for both farmers and policymakers. For farmers, the adoption of precision nutrient management practices leads to increased crop yields, reduced input costs, and enhanced profitability. By minimizing nutrient imbalances and optimizing resource use, farmers can achieve sustainable agricultural practices that promote long-term soil fertility and environmental stewardship (Bora et al., 2016). Policymakers can leverage UAV-based nutrient mapping to inform agricultural policies and development strategies. Accurate data on nutrient distribution and crop health can guide decisions related to fertilizer subsidies, resource allocation, and agricultural extension services. Additionally, UAV-based mapping can contribute to national food security by enhancing productivity and resilience in the agriculture sector (Schimmelpfennig, 2016).

This study utilized the DJI (Da-Jiang Innovations) Phantom 4 multispectral drone for mapping soil nutrient variability in crop fields. The drone's multispectral imaging capabilities enable the capture of data related to bare soil, which serve as indicators of soil nutrient content. By establishing correlations between these images and laboratory analysis of soil samples, a detailed map of soil nutrient variability was generated. This map, in turn, can guide the implementation of site-specific fertilizer application strategies, thereby enhancing nutrient-use efficiency and crop productivity.

Materials and Methods

This study was conducted with the collaboration between the Laboratory of the ASICT Division and the Soil Science Division at BARI, Gazipur. The Soil Science Division provided analytical support, while the ASICT Division offered technical support. The research was conducted in the experimental field of the Soil Science Division. The study area had the same soil type (chernozem), last season crop type (corn), tillage method (machine operation), and crop ripening system (one crop a year). The research was conducted during the bare soil period following crop harvest, the tillage layer had been plowed and the soil surface was left uncovered, as illustrated in Figure1. Furthermore, there were no occurrences of floods or droughts before the study period, and soil water content showed minimal variation.

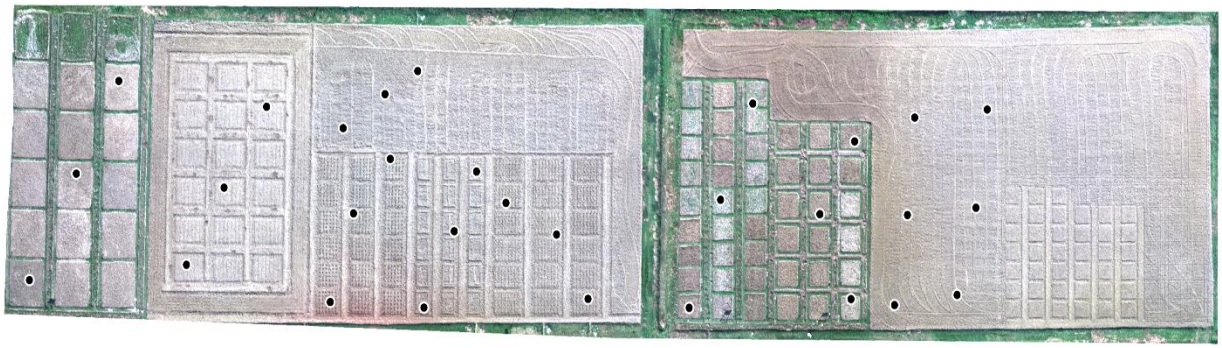


Figure 1. Sample distribution map of the study area

There were 10 plots and three soil samples were taken from each plot by using the three-point method within a range of 20 cm X 20 cm of each site. After removing plant roots, gravel, and other debris, the topsoil of 1 kg within the upper and lower 15 cm was mixed, bagged, and sealed. Finally, the soil was air-dried, ground, and screened at 2 mm, and the nutrient contents were determined by the potassium dichromate external heating method. Table1 shows the basic statistics of the considered nutrient contents of the soil samples.

Table 1. **Descriptive statistics of the considered nutrient contents of the soil samples**

Nutrient contents (unit)	max	mean	min	se
pH	6.99	5.8697	5.26	0.0617
Organic Matter (OM) (%)	1.90	1.5500	1.20	0.0386
Total Nitrogen (N) (%)	0.10	0.0817	0.06	0.0020
Calcium (Ca) (meq/100 ml)	14.90	12.2467	9.80	0.2857
Magnesium (Mg) (meq/100 ml)	3.90	3.0367	1.90	0.0809
Potassium (K) (meq/100 ml)	0.24	0.1373	0.06	0.0073
Phosphorus (P) (ppm)	21.10	11.8133	4.00	0.7991
Sulphur (S) (ppm)	30.90	22.5833	13.80	0.9234
Boron (B) (ppm)	0.99	0.6790	0.27	0.0361
Cupper (Cu) (ppm)	4.30	3.2733	2.60	0.0783
Iron (Fe) (ppm)	123.30	63.6967	22.30	4.1440
Manganese (Mn) (ppm)	29.60	17.7767	9.90	0.9326
Zinc (Zn) (ppm)	1.20	0.4217	0.00	0.0644

Multispectral remote sensing images with a resolution of centimeter-level were obtained during the bare soil period in the study area using a DJI Phantom 4 multispectral UAV (Figure2). The aerial photo was taken on 28 November 2023, with clear and breezy weather conditions. To ensure that the study's area was adequately covered, the route was reasonably planned using DJI GS Pro (Ground Station Pro), a DJI route planning professional software. The aerial parameters were set as follows: the altitude was 57 m, the spatial resolution of the image was 3 cm/px, and the course overlap rate and lateral overlap rates were both 80% (Figure2).

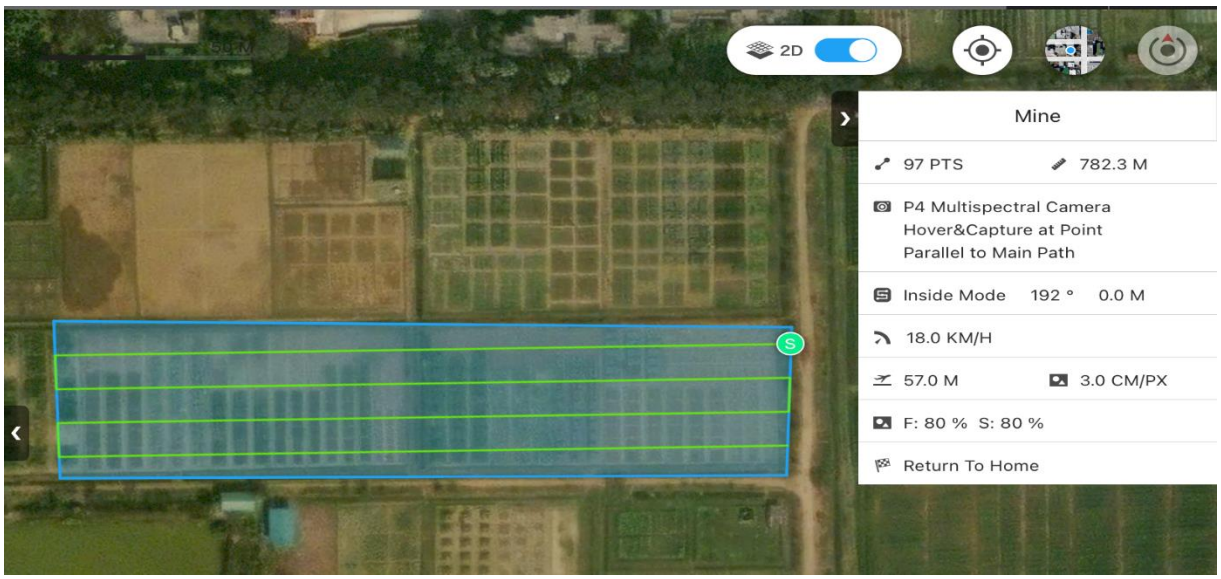


Figure 2. DJI Phantom 4 Multispectral and its flight plan

The multispectral images have five bands (Table 2). The obtained images underwent preprocessing, including the following steps: (1) checking photo quality to eliminate unnecessary photos such as take-off and landing images; (2) image photogrammetry and ortho-mosaicking to merge the photos; (3) generating a digital orthophoto image; (4) exporting multispectral images; (5) image band combination; (6) image clipping; (7) projection transformation; (8) resampling. Consequently, a multispectral image with a spatial resolution of 3 cm/px was generated. The coordinate system used was WGS84_UTM_Zone 45N. From image preprocessing to exporting the orthomosoic was done using Agisoft Metashape Professional software.

Table 2. Summary of DJI Phantom 4 multispectral UAV sensor characteristics.

Band #	Designation	Central Wavelength (nm)
1	Blue	450 ± 16
2	Green	560 ± 16
3	Red	650 ± 16
4	Red-edge	730 ± 16
5	NIR	840 ± 26

The collected soil samples were split into two categories (High and Low) using K-means clustering algorithm. Most of the commonly used clustering algorithms can be classified in two general categories namely; Hierarchical clustering and Non-Hierarchical clustering. Non-Hierarchical Clustering algorithms produce disjoint clusters and thus work well when a given set is composed of a number of distinct classes or when the data description is flat. means is one of the Non-Hierarchical Clustering algorithms. If we have n feature vectors $X_1, X_2, \ldots, X_n$ all belonging to the same class C and we know that they belong to K clusters such that $K < n$. If clusters are well separated, we can use a minimum distance classifier to separate them. We first initialize the means $\mu_1, \mu_2, \ldots, \mu_K$ of K clusters. One of the ways to do this is just to assign random number to them. We then determine the membership of each X by taking the $\|X - \mu_i\|$. The minimum distance determines X's membership in a respective cluster. This is done for all n feature vectors.

The study utilized multispectral imagery obtained from a DJI Phantom 4 Multispectral UAV. The UAV captured an orthomosaic image of the study area, which was subsequently used for classification. The orthomosaic image was loaded and processed using Python and the GDAL library, ensuring the preservation of geospatial information. To handle the large size of the orthomosaic image and prevent memory issues during processing, the image was divided into smaller, manageable patches. A patch size

of 64x64 pixels was selected, with no overlap between patches. Each patch was reshaped into a 1D vector for further analysis. An autoencoder is a type of artificial neural network used to learn efficient representations of data, typically for the purpose of dimensionality reduction. Unlike traditional supervised learning methods that require labeled data, autoencoders can operate in an unsupervised manner, making them suitable for tasks where labeled data is scarce or unavailable. In this study, the autoencoder was employed to extract latent features from the multispectral image patches, which were then used for unsupervised classification.

Architecture: The autoencoder consists of two main components: the encoder and the decoder. The encoder compresses the input data into a lower-dimensional latent space, effectively reducing the number of variables while retaining the most important features. This compressed representation is referred to as the "latent space" or "bottleneck" of the autoencoder. The encoder was designed as a fully connected neural network layer with a smaller number of neurons compared to the input layer, ensuring dimensionality reduction. The decoder attempts to reconstruct the original input data from the latent space representation. It mirrors the encoder in structure but reverses the process, aiming to expand the reduced-dimensionality data back to the original input size. The reconstructed output is compared to the original input to measure the accuracy of the reconstruction, guiding the training process. The autoencoder is trained to minimize the difference between the original input and the reconstructed output. This difference is quantified using a loss function, typically binary cross-entropy or mean squared error (MSE), depending on the data type and task. The loss function guides the optimization process, adjusting the weights and biases in both the encoder and decoder to improve the model's ability to reconstruct the input data from its compressed representation. During training, the autoencoder learns to capture the most salient features of the data in the latent space, effectively filtering out noise and redundancy. This process can be seen as a form of unsupervised feature learning, where the network autonomously identifies the most important aspects of the input data. Once the autoencoder is trained, the encoder part is used independently to transform new input data into the latent space. These latent features represent a compressed and refined version of the original data, capturing the underlying patterns and structures in a more compact form. These features are particularly useful for downstream tasks, such as clustering or classification. In this study, the latent features extracted from the multispectral image patches were used as input for the K-Means clustering algorithm. By working in the latent space rather than the original high-dimensional space, the clustering process becomes more efficient and often yields more meaningful results, as the latent features emphasize the most critical aspects of the data.

The encoded latent features extracted by the autoencoder were used as input for the K-Means clustering algorithm. The clustering algorithm was configured to classify the image into 20 distinct classes based on the similarity of the latent features. This approach allowed for unsupervised classification of the image without the need for labeled training data. The classified image, initially in array format, was saved as a GeoTIFF file to retain the georeferencing information of the original orthomosaic image. The GDAL library was used to write the classified data to a GeoTIFF file, preserving the spatial resolution and coordinate reference system of the original imagery.

The study will employ open-source software like QGIS, R, and Python for remote sensing image analysis and statistical analysis. Image georeferencing, processing, classification, and other related tasks were conducted at the Agro-Environmental Remote Sensing and Modeling (ARSAM) Laboratory of the ASICT Division at BARI.

Results and Discussion

The distribution of 13 soil nutrient contents was analyzed across 30 samples, and the results are illustrated using boxplots (Figure 3). The Shapiro-Wilk normality test was conducted to assess the distribution of each nutrient, with the p-values displayed above each boxplot's upper whisker. The test results indicated that approximately 50% of the nutrient parameters do not follow a normal distribution, suggesting a significant deviation from normality in half of the measured variables. Additionally, several outliers were detected, which could indicate anomalies or natural variability in the soil nutrient content across the study area.

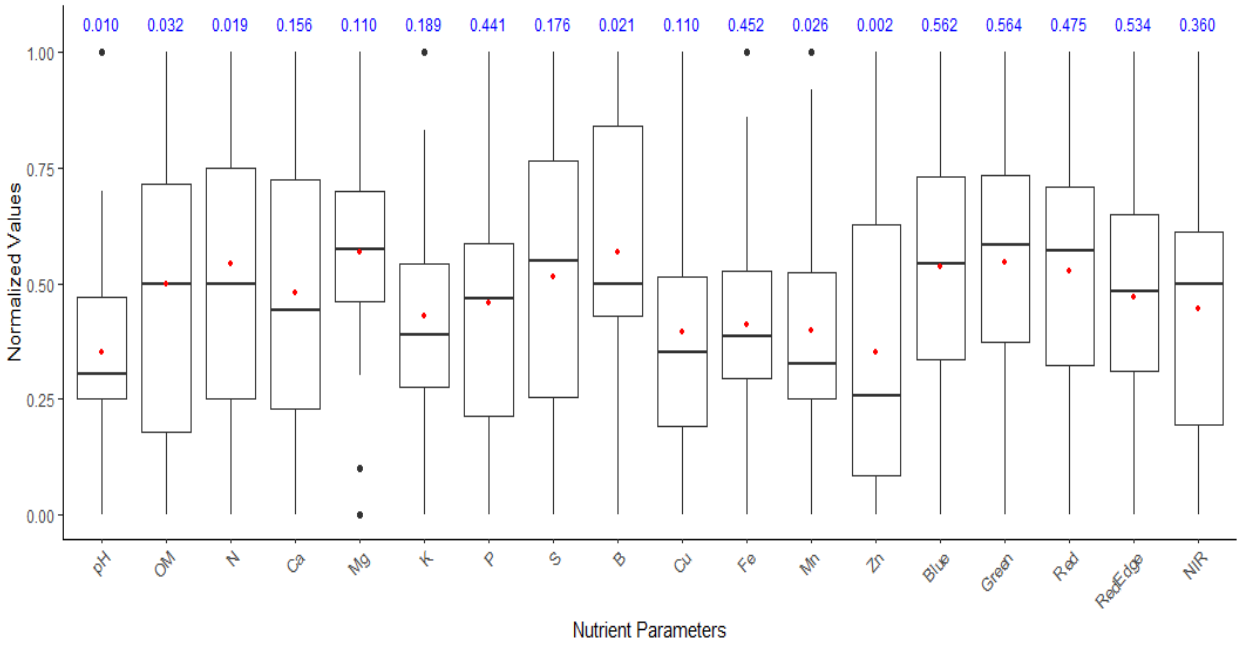


Figure 3. Boxplot of normalized soil nutrient content and band values

In contrast to the soil nutrients, the spectral band values from the multispectral imagery were found to be normally distributed, with no significant outliers observed. This difference in distribution between soil nutrients and band values highlights the complexity of soil properties and their potential impact on spectral reflectance data.

To evaluate the variability of soil nutrient content between different plots, a Completely Randomized Design (CRD) test was conducted. Since three samples were collected from each of the 10 plots, ANOVA was employed to determine whether there was significant variability in nutrient content between plots. The coefficient of variation (CV) for each nutrient and band value, along with the significance status derived from the ANOVA results, is presented in Figure 4.

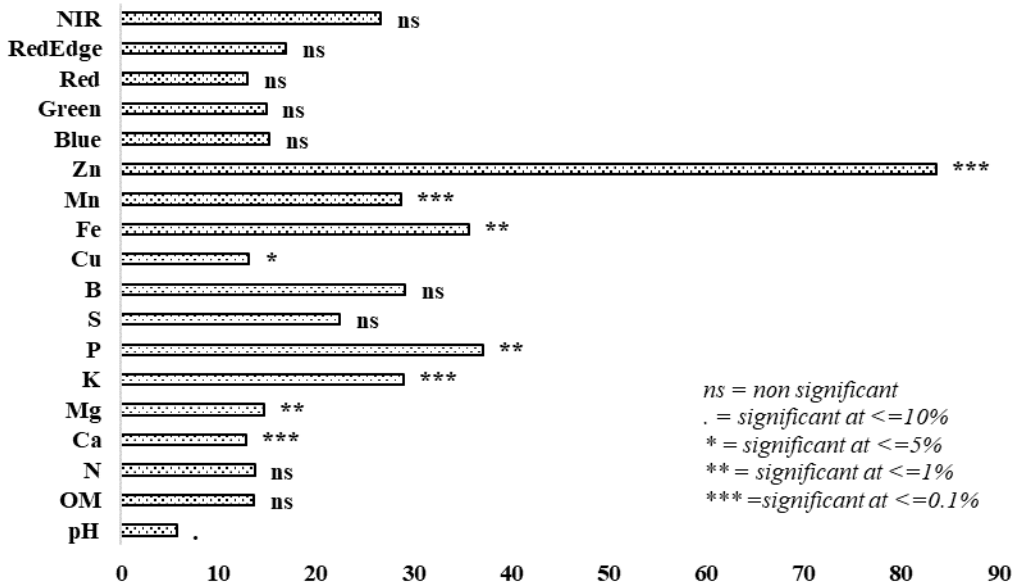


Figure 4. Coefficient of Variation (CV) with ANOVA output

The results indicate that several nutrient parameters, including potassium (K), phosphorus (P), sulfur (S), boron (B), iron (Fe), manganese (Mn), and zinc (Zn), exhibited high variability across the plots. This

suggests that these nutrients are more heterogeneous within the study area. In terms of spectral data, the Near-Infrared (NIR) band displayed the highest variability among the bands, which is consistent with its sensitivity to vegetation and soil moisture variations.

The ANOVA analysis further revealed that most of the nutrient parameters showed significant differences between plots, indicating spatial variability in soil fertility. However, organic matter (OM), nitrogen (N), sulfur (S), and boron (B) did not show significant differences, suggesting a more uniform distribution across the plots. For the spectral band values, none of the bands exhibited significant differences between plots, indicating that traditional parametric approaches may not effectively capture the spatial variability in soil properties during the bare soil period.

Given the significant variability observed in soil nutrient content, K-Means clustering was applied to classify the samples into two distinct groups based on their nutrient profiles. The clustering results, depicted in Figure 5, indicate that plots 1 to 5 formed one cluster, while the remaining plots (6 to 10) constituted a second cluster. This division suggests that roughly half of the plots exhibit similar nutrient characteristics, while the other half shows different nutrient profiles.

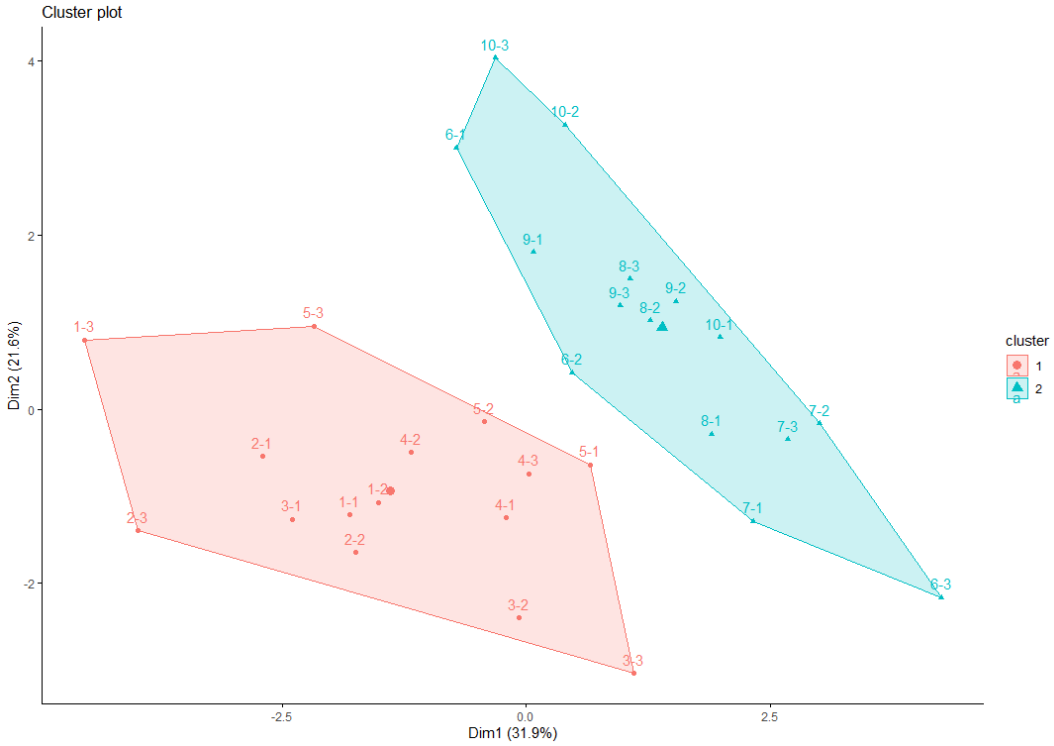


Figure 5. k-means clustering of plots based on nutrient parameters

Cluster means for each nutrient parameter were calculated and are presented in Table 3. Cluster 1 was characterized by higher levels of calcium (Ca), magnesium (Mg), sulfur (S), boron (B), copper (Cu), and iron (Fe), indicating nutrient enrichment in these elements. Conversely, Cluster 2 showed higher concentrations of organic matter (OM), nitrogen (N), potassium (K), phosphorus (P), manganese (Mn), and zinc (Zn). This clustering highlights the spatial heterogeneity in nutrient distribution and suggests that different management practices may be required for each cluster to optimize soil fertility.

Table 3. Cluster means of nutrient parameters

Clusters	pH	OM	N	Ca	Mg	K	P	S	B	Cu	Fe	Mn	Zn
1	6.04	1.52	0.08	13.6	3.35	0.12	10.37	22.9	0.74	3.47	70.97	15.08	0.17
2	5.70	1.57	0.08	10.89	2.73	0.15	13.26	22.27	0.62	3.08	56.43	20.47	0.68

To map the spatial variability of soil nutrient content at a finer scale, an autoencoder followed by K-Means clustering was employed. This approach allows for the integration of unsupervised deep learning techniques with traditional clustering methods to enhance the classification accuracy. The results, shown in Figure 6, reveal that the traditional unsupervised machine learning approach (such as K-Means alone) was unable to fully capture the spatial variations within the fields based on spectral band values. However, when combined with the autoencoder, the method successfully classified the multispectral image into 20 distinct classes, accurately reflecting the spatial variation in soil nutrient content.

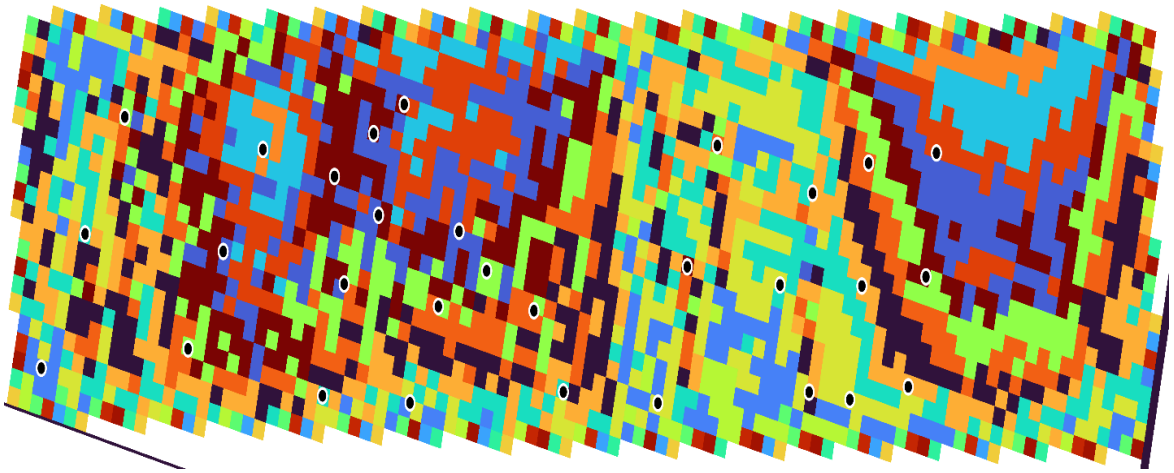


Figure 6. k-means clustering using autoencoder of plots based on spatial variation

The multispectral image of the plot is classified into 20 classes (0 to 19). This is done to accurately determine the classes and assess the classification matrix based on ground truth data. The ground truth data were then put on classified image and obtained the cluster-wise distribution of the classes of the ground truth point. The distribution was shown in the table 4.

Table 4. Distribution of the classes of clusters

Class / Clusters	**2**	**3**	**5**	**6**	**9**	**11**	**13**	**15**	**16**	**19**	**Total**
1						3	4	2	1	5	15
2	4	2	1	3	5						15

Class 11 to class 19 had comprised cluster 1 and class 2 to class 9 had comprised cluster 2. Thus, class 11 to class 19 could be grouped together and class 2 to class 9 could be grouped together to assess the spatial variation based on nutrient parameter (Figure 6).

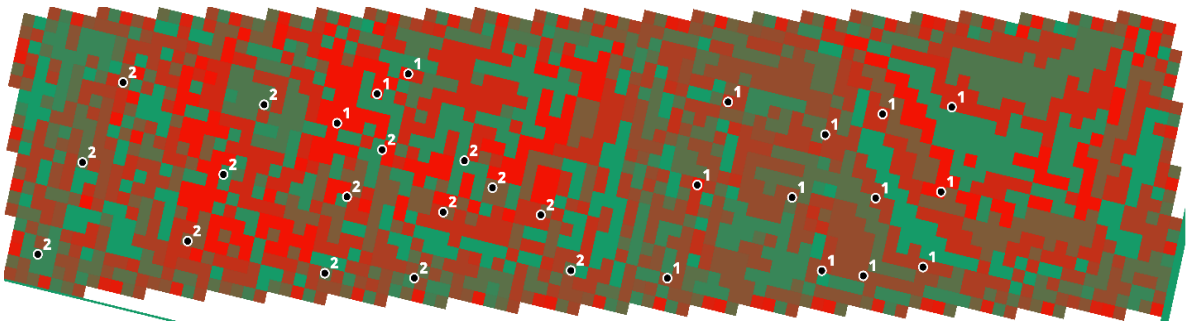


Figure 7. Grouping of the cluster based on reference data

Classes 11 to 19 were predominantly associated with Cluster 1, while classes 2 to 9 corresponded to Cluster 2. This classification enabled a more detailed spatial analysis of soil nutrient distribution, with

Cluster 1 (marked in red in Figure 6) indicating areas enriched in Ca, Mg, S, B, Cu, and Fe, and Cluster 2 (marked in green) highlighting regions enriched in OM, N, K, P, Mn, and Zn.

The spatial mapping and clustering results have important implications for precision agriculture and soil management within the study area. The identification of distinct nutrient clusters allows for targeted interventions, such as tailored fertilization strategies, that can address the specific nutrient needs of different plots. The use of deep learning techniques, such as autoencoders, in conjunction with traditional clustering methods, has proven to be effective in capturing the complex spatial variability in soil nutrient content, which might not be detectable using conventional methods alone.

Overall, the study demonstrates the potential of combining advanced machine learning techniques with multispectral imagery for enhancing soil nutrient management and optimizing agricultural productivity. Future research could explore the integration of additional data sources, such as soil moisture or topographic information, to further refine the classification and mapping of soil properties.

Conclusions

This study successfully demonstrated the integration of deep learning techniques, particularly autoencoders, with traditional clustering methods to map and analyze the spatial variability of soil nutrient content using multispectral imagery. The findings highlight the significant heterogeneity in soil nutrient distribution across the study area, with certain nutrients showing marked variability between plots. Traditional parametric approaches, ANOVA, revealed important insights into nutrient distribution but were insufficient in capturing the full extent of spatial variation, especially when applied to multispectral band values during the bare soil period. The use of an autoencoder followed by K-Means clustering provided a more nuanced understanding of the spatial patterns in soil nutrients. By reducing the dimensionality of the multispectral data and extracting latent features, the autoencoder enhanced the clustering process, resulting in a more accurate classification of the study area into distinct nutrient clusters. These clusters were then mapped spatially, allowing for precise identification of areas with specific nutrient enrichment, which is crucial for implementing targeted soil management practices. The study underscores the limitations of traditional unsupervised learning methods in fully capturing the complexities of soil and spectral data, especially in heterogeneous environments. It also highlights the potential of deep learning approaches to overcome these limitations and provide more detailed and actionable insights. The integration of machine learning techniques in agricultural research offers significant promise for improving soil management strategies, ultimately contributing to more sustainable and productive agricultural practices.

References

Adao, T., Hruška, J., Padua, L., Bessa, J., Peres, E., Morais, R., & Sousa, A. (2017). Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sensing*, 9(11), 1110.

Alauddin, M., & Quiggin, J. (2008). Agricultural intensification, irrigation and the environment in South Asia: Issues and policy options. *Ecological Economics*, 65(1), 111-124.

Bechar, A., & Vigneault, C. (2017). Agricultural robots for field operations: Concepts and components. *Biosystems Engineering*, 153, 4-18.

Bendig, J., Bolten, A., Bareth, G., Broscheit, J., Gnyp, M. L., & Schirrmann, M. (2015). UAV-based imaging for multi-temporal, very high resolution crop surface models to monitor crop growth variability. *Photogrammetrie-Fernerkundung-Geoinformation*, 2015(1), 69-79.

Bora, G. C., Alchanatis, V., & Lang, J. (2016). Recent trends in precision agriculture in the USA. *Engineering in Agriculture, Environment and Food*, 9(4), 253-262.

Haque, M. E., Molla, M. M. A., Biswas, J. C., & Duxbury, J. M. (2015). Sustainable soil management options for rice in Bangladesh. *Bangladesh Rice Journal*, 19(1), 1-15.

Hossain, M. S., Ali, M. Y., Hossain, M. A., & Kabir, M. S. (2020). Adoption of precision agriculture technologies in Bangladesh: Current status and future prospects. *Journal of Agricultural Engineering and Food Technology*, 7(1), 34-41.

Houborg, R., Fisher, J. B., & Skidmore, A. K. (2015). Advances in remote sensing of vegetation function and traits. *International Journal of Applied Earth Observation and Geoinformation*, 43, 1-6.

Hunt Jr, E. R., Daughtry, C. S., & McMurtrey III, J. E. (2014). A photogrammetric technique for estimating crop residue cover. *Remote Sensing of Environment*, 151, 122-129.

Islam, M. S., Ali, M. Y., & Shamsuzzaman, A. M. M. (2019). Environmental impacts of chemical fertilizer use in Bangladesh: A review. *The Agriculturists*, 17(1-2), 116-129.

Kabir, M. H., Hossain, M. S., & Khan, M. A. (2020). Adoption of smart farming technologies in Bangladesh: Current scenario and future directions. *Journal of Agricultural Informatics*, 11(1), 23-30.

Klerkx, L., Jakku, E., & Labarthe, P. (2019). A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and a future research agenda. *NJAS-Wageningen Journal of Life Sciences*, 90, 100315.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674.

Miah, M. A. M., Rashid, M. A., & Shiblee, S. M. A. (2016). Fertilizer use efficiency for sustainable agriculture development in Bangladesh. *Bangladesh Journal of Agricultural Research*, 41(1), 157-165.

Rahman, M. H., & Hasan, M. M. (2011). Economic and environmental impacts of improper use of fertilizers in agriculture: An overview of Bangladesh. *Bangladesh Journal of Agricultural Research*, 36(4), 723-732.

Roy, R., & Chan, N. W. (2020). Precision agriculture adoption in Bangladesh: A review. *Journal of the Indian Society of Remote Sensing*, 48(2), 217-230.

Sarker, M. A. R., Alam, K., & Gow, J. (2017). Exploring the relationship between climate change and rice yield in Bangladesh: An analysis of time series data. *Agricultural Systems*, 155, 12-21.

Schimmelpfennig, D. (2016). Farm profits and adoption of precision agriculture. *USDA-ERS Economic Research Report*, (217).

Schwab, K. (2017). *The Fourth Industrial Revolution*. Currency.

Torres-Sánchez, J., Peña, J. M., de Castro, A. I., & López-Granados, F. (2013). Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. *Computers and Electronics in Agriculture*, 103, 104-113.

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big data in smart farming–a review. *Agricultural Systems*, 153, 69-80.

Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: a review. *Precision Agriculture*, 13(6), 693-712.

MONTHLY HEATWAVE PREDICTION AND ITS EFFECT ON THE YIELD OF SUMMER TOMATO IN BANGLADESH

MOHAMMAD RASEL, ISTIAK AHMED AND M.A. MONAYEM MIAH

Abstract

The global mean temperature is gradually increasing, which is a concerning issue. The prolonged high temperatures cause heatwaves (HWs), leading to extensive damage to crop production. Summer tomatoes, particularly the heat-tolerant BARI-released hybrid varieties (BARI Hybrid tomato-4 and BARI Hybrid tomato-8), are important due to their increasing demand and profitability. This study used the Zero-Inflated Poisson (ZIP) model to predict monthly Heat Wave Count (HWC). It examined the impact of HWs, using simple linear regression (LR), on the yield of summer tomatoes cultivated in Gazipur district. The study utilized agro climatological gridded data from the assimilation model (MERRA-2) of the POWER project, supported by NASA (1982- 2023), as well as summer tomato yield data (2004-2023) provided by the Olericulture Division of Horticulture Research Center (HRC), Bangladesh Agricultural Research Institute (BARI). The study revealed that heat waves occurred in Gazipur district from March to May, with the highest frequency observed in April. Additionally, the number of heat wave days (HWDs) peaked in April, and the duration of HWDs increased over the years. The performance of the ZIP model showed its reliability possessing low Mean Square Error (MSE) and Mean Absolute Error (MAE) during the validation period 2022-2023. The result revealed the potential of the ZIP model to be used for reliable prediction of heatwaves in the context of climate change. Furthermore, the study found a significant negative association between HWs and the yield of summer tomatoes at a 5% significance level with a p-value of 0.029.

Keywords: Agroclimatology, MERRA-2, Heatwaves, Prediction, ZIP model, Summer tomato, Heat tolerant, Yield

Introduction

The average global temperature is increasing significantly, leading to various issues and raising concerns about climate change (Scherm and Van Bruggen, 1994). Hot days are becoming more frequent and hotter worldwide due to climate change (AU, 2014). In Bangladesh, winter is getting shorter while the summer season is lengthening (Chowdhury et al., 2012; Siddik and Rahman, 2014). Extreme heat during the summer in Bangladesh is causing significant damage to agricultural production, which is highly sensitive to extreme heat (Roberts and Schlenker, 2011). The prolonged extreme temperature leads to Heat Waves (HW), severely impacting crop production.

In Bangladesh, tomatoes (*Lycopersicon esculentum* Mill.) are one of the most important and popular vegetables. They are grown throughout the winter due to the favorable weather for healthy development and productivity (Zaman et al., 2006). With increasing demand and profitability, tomatoes are also being cultivated in the summer season (Hajong et al., 2018; Zaman et al., 2006). Recognizing the growing demand and significance of tomatoes, the Bangladesh Agricultural Research Institute (BARI) has developed off-season summer tomatoes. To date, BARI has created and introduced three heat-tolerant hybrid tomato varieties: BARI Hybrid Tomato-3, BARI Hybrid Tomato-4, and BARI Hybrid Tomato-8, which can be grown during the summer and rainy seasons under polytunnels (Hajong et al., 2018; Karim et al., 2009). Hajong et al. (2018) found that the best time to plant summer tomatoes is from April to May, sometimes extending to July, with harvesting taking place from August to late October. Optimal fruit setting in tomatoes requires temperatures below 26°C during the day and 20°C at night (Stevens and Rudich, 1978; Kuo et al., 1979). However, the average temperature in Bangladesh during the summer tomato growing season, from April to July, exceeds 28°C (Islam, 2009). Despite the heat-tolerant nature of summer tomatoes (Das et al., 2023), yield reduction may occur due to high temperatures. That's why this study is concerned with the prediction of HWs and revealing their effects, if any, on the yield of summer tomatoes.

A heat wave (HW) is a period of excessively hot weather, which may be accompanied by high humidity. There is no universal definition of Heat Wave (HW). Most of the definitions are based on the persistence of maximum, minimum or mean surface air temperature above a threshold value that is based on the upper tail of the temperature distribution over a region (Perkins, 2015). Some of the definitions may even consider surface humidity levels as humidity can worsen a HW effect. Della-Martha et al. (2007) defined HW as the number of consecutive 3-day periods in summer that exceed the long-term daily 80th percentile of daily maximum temperature. Srivastava et al. (2009) defined the event if the maximum temperature at a grid point is 3°C or more than the normal temperature, consecutively for 3 days or more. Mishra et al. (2015) considered HW as the period during which the daily maximum temperature stayed above the empirical 99th percentile consecutively for six or more days. Domeisen et al. (2023) defined HW as a temperature threshold such as the 90th percentile or higher, and include a persistence of at least three consecutive days.

According to local climatological conditions, Geneva based World Meteorological Organization (WMO) defined HW in "Events Types of Hazards and Extreme Events" as unusually hot weather over a region that persists for at least two consecutive days during the hot period of the year, with thermal conditions recorded above given threshold. This study took into account the WMO's definition and the Bangladesh Meteorological Department's (BMD) maximum temperature threshold, which are listed in Table 1. According to the BMD threshold and WMO definition, mild HW can be defined as experiencing a temperature greater or equal to 36°C for two or more consecutive days. A similar definition is applicable for moderate, severe, and extreme HW.

Table 1: The classification and threshold of maximum temperature as per BMD

Types	Heatwaves	For calculation HW
	Maximum temperature (℃)	
Mild	36-38	>=36
Moderate	38-40	>= 40
Severe	40-42	>= 42
Extreme	>42	>42

There have been numerous studies on the effects of HWs and summer tomatoes individually, each with different objectives. However, there are few studies that specifically examine the impact of HWs on the yield of summer tomatoes.

Mannan et al. (2021) studied the analytical study of HWs occurred over the Bangladesh during pre-monsoon (March - May) of 2021. They depicted that the presence of thermal low over western side of Bangladesh, anticyclonic circulation over the Bay of Bengal, presence of high humidity at lower tropospheric level over Bangladesh and negative magnitude of lifted index are well recognized conditions for formation, intensification and persistence of HW in Bangladesh. Karmakar (2019) revealed the pattern of climate change and its impact in northwestern Bangladesh. They found that the annual mean heat stress over northwest Bangladesh indicates significant increasing trends at Rajshahi and Rangpur and decreasing trend at Dinajpur during the period 1981-2016; whereas the annual heat stress at Dinajpur has increasing trends at +0.022/year since 1989. Rajshahi has the highest mean frequency of maximum temperature >36℃ in the month of May, whereas Dinajpur and Rangpur have the maximum mean frequency of maximum temperature >36℃ in the month of April. Heat waves will be more long lasting in Rajshahi during April-July.

Another study conducted by Nissan et al. (2017) to define and predict the HWs in Bangladesh using BMD data from 1989-2011. They found that almost all heat waves occur during the hot pre-monsoon summer season, between April and June, with most in May. They recommended a HW index for

predicting the mortality in Bangladesh due to HWs using generalized additive regression model. Prediction and projection of HWs are also made in Nissan et al. (2017) where daily to weekly predictions are made based on the drivers of HWs. Li (2023) conducted a study for monthly HW prediction based on Sweden using different machine learning techniques and determined related 16 features by correlation heatmap out of 21 extracting from remote sensing data for HW prediction. This study found soil moisture, precipitation, wind speed, wind direction, soil wetness, monthly mean temperature, humidity, sea level pressure, evaporation, land cover, slope, latent heat flux, geopotential height, latitude, longitude are significantly correlated with the occurrence and intensification of monthly HWs. Ibrahim, (2016) depicted tomato breeding for heat stress condition in study. It described that high temperature caused a lot of damage in tomato yield in summer season in Egypt. Heat stress are main reasons for poor fruit sets in tomatoes and decreases yield. Heat stress also significantly damages fruit sets of tomatoes which leads to the damage of tomato yield (Alsamir et al., 2021; Abdalla et al., 2020; Panthee et al., 2018).

Monthly heatwave count data are zero-inflated because HWs are usually observed from March to June and skewed, various study suggests various models for analyzing and predicting such zero-inflated skewed data. Green (2021) used Poisson and Negative binomial regression with their zero inflated and hurdle extensions which allow even more zeros. Beckett et al. (2014) showed how to model zero-inflated count data of natural calamities using ZIP regression and estimate the model.
Literature illustrates that ZIP models performs better than classical Poisson, negative binomial models and impact of HWs on the yield of summer tomatoes are not studied previously. To bridge the research gap, this study aims to predict the monthly HWs and explore the effects of HWs on the yield of summer tomatoes. This study used agro-climatological (1982-2023) and summer tomato data (2004-2023) for Gazipur district to develop ZIP model, predict HWs, and determine the effects of HWs on the yield of tomato for Gazipur district and then generalized it for all regions of Bangladesh.

However, the study has the following specific objectives:

- To get an exploratory view of the current heatwave scenario/pattern in Bangladesh.
- To forecast the monthly heatwave in Bangladesh.
- Help the policymakers to take necessary steps to reduce the damages of heatwaves in the agriculture sector.

Materials and Methods

1.0 Study Area

In this study, the primary purpose is to derive a statistical model to predict the monthly HW as accurate as possible and to find the relationship (in any) between HW and yield of summer tomato. As the yield trial of summer tomato is going on in Olericulture Division, Horticulture Research Center (HRC), BARI for past few years, we selected Gazipur district (Lattitude 23.9999, Longitude 90.4203) for monthly HW prediction.

2.0 Data Source

2.1 Agro-climatological data

To investigate the HW scenario in Gazipur district and develop a statistical model for monthly HW prediction the agro-climatological data were collected from the Prediction Of Worldwide Energy Resources (POWER) project funded by the National Aeronautics and Space Administration (NASA) Earth Science Directorate Applied Science Program (https://power.larc.nasa.gov/data-access-viewer/). The POWER project provides satellite and model-based sufficiently accurate, reliable solar and meteorological resource data over regions where surface measurements are sparse or nonexistent. The meteorological data in POWER are based upon the Goddard's Global Modeling and Assimilation Office (GMAO) Modern Era Retrospective-Analysis for Research and Applications (MERRA-2) assimilation

model and GEOS 5.12.4 FP-IT. The data in POWER are provided on a global grid with spatial resolutions equal to the input data. That resolution is ½° latitude by ⅝° longitude (About 50 km × 62.5 km) for the meteorological data sets with the WGS84 (World Geodic System 1984) grid reference system. The parameters, related to the HW, we collected from the POWER project data are daily maximum temperature, minimum temperature, specific humidity, relative humidity, wind speed at 2-meter, wind speed at 10-meter, wind direction at 2 meter and precipitation, surface soil wetness, profile soil moisture, and surface pressure from 01/01/1982 to 12/31/2023. The worldwide daily temperature, obtained from the MERRA-2 assimilation model of POWER project, is depicted in the figure-1.

2.2 Summer tomato data

The Olericulture Division of HRC, BARI has been conducting various experiments on two summer tomato varieties, BARI Hybrid tomato-4 and BARI Hybrid tomato-8, for a number of years with the aim of improving these varieties, as stated in their Annual Reports. These experiments include recording several parameters, such as the yield of these varieties. The annual yield (t/ha) data for summer tomatoes from 2004 to 2023 was obtained from the annual report of the Olericulture Division. The records for BARI Hybrid tomato-4 cover the years 2004 to 2007, while the rest of the data is for BARI tomato-8.

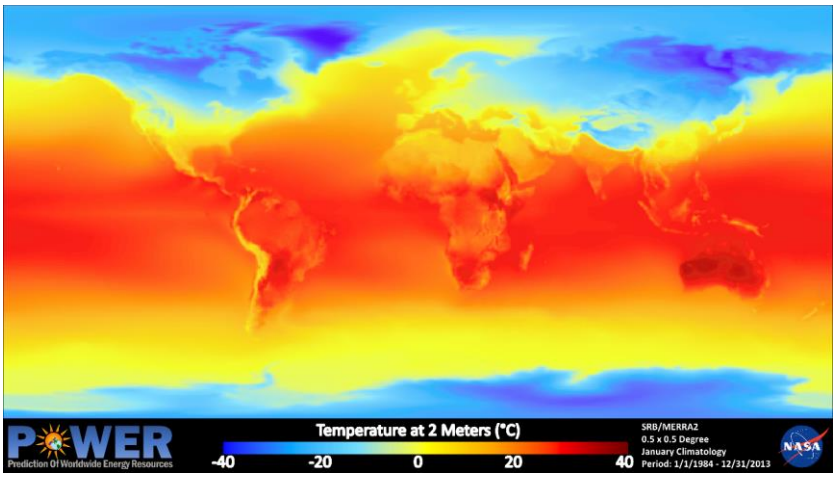


Figure-1: Daily worldwide temperature obtained by POWER project using MERRA-2 assimilation model from 01/01/1984 –12/31/ 2013

3.0 Data Arrangement

We obtained daily meteorological data from the above source and checked for outliers, missing values, other anomalies. Detected outliers were replace by the missing values in the data. Then, we searched for the Heat wave days (HWD) using the maximum daily temperature based on the definition of HWDs and count the number of monthly HWDs, HWs. Other parameters in the study were used as the monthly average.

4.0 Statistical Analysis

To investigate the HW scenerio and predict the monthly HW Spearman's Rank Correlation, Linear Regerssion (LR), Zero Inflated Poisson (ZIP) Regression model, Mean Sqaure Error (MSE), Mean Absolute Error (MAE) were used in this research program. All of the analyses were conducted with R software.

4.1 Spearman's rank correlation

The Pearson Correlation and Spearman Correlation are two different correlation measures that apply in specific situations. Pearson correlation detects linear relationships between quantitative variables with

data following a normal distribution (Schober et al., 2018). On the other hand, Spearman correlation uses data rank to measure monotonicity between ordinal or continuous variables (Hauke and Kossowski, 2011). The statistical distributions of the counts data are generally skewed, and hence not normally distributed (White and Bennetts, 1996). The monthly HW data are zero inflated and over dispersed, and right skewed, hence Spearman's rank correlation was used in this study to find the correlation between monthly number of HWs and other variables of interest. The formula for Spearman's rank correlation coefficient (Equation 1).

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} \quad \text{---------- (1)}$$

Where,

r_s = Correlation coefficients
n = Number of observations
d = The square of the difference in the ranks of the two coordinates for each point (x, y)

Spearman's rank correlation coefficient, denoted by r_s ,is a numerical value such that $-1 \leq r \leq 1$. The stronger the association, the closer r_s is to −1 or 1, and the weaker the association, the closer it is to 0 (Sedgwick, 2014).

4.2 Linear regression

We used linear regression to find the effects of HW on the yield of summer tomato. Linear regression analysis is used to predict the value of a variable based on the value of another variable (Weisberg, 2005; Seber and Lee, 2012). The variable we want to predict is called the dependent variable. The variable we are using to predict the other variable's value is called the independent variable.

This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a "least squares" method to discover the best-fit line for a set of paired data (Weisberg, 2005; Seber and Lee, 2012). We then estimate the value of Y (dependent variable) from X (independent variable). In this study, we used simple linear regression for analysis. It assumes that there is approximately a linear relationship between X and Y. Mathematically (Equation 2), we can write this linear relationship as follows (Seber and Lee, 2012; James et al., 2023):

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad \text{---------- (2)}$$

Here,
Y = Yearly yield of summer tomato in tons per hectare
X = Number of yearly heatwaves
β_0 = Intercept of the model
β_1 = Coefficent of X
ε = Random error term

4.3 Zero-inflated poission regression (ZIP)

ZIP regression is a model for count data with excess zeros (Lambert, 1992; Kim et al., 2019). Suppose a random variable, X follows the usual Poisson distribution with parameter λ, Poi(λ), with the following probability mass function (Equation 3).

$$P(X = k) = \exp(-\lambda)\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, 3, \ldots \quad \text{---------- (3)}$$

can be used to model many count data (for example: the number of yearly heatwaves). Note that *X* takes only nonnegative integer values. However, the Poi(λ) distribution may not be useful (or it gives a bad fit) when *X* takes the value 0 with a high probability (Beckett et al., 2014). In such a case a modified version of a regular Poi(λ) distribution known as the zero-inflated Poisson (ZIP) distribution becomes useful

(Lambert, 1992; Kim et al., 2019). The ZIP distribution with parameters π and λ, denoted by ZIP (π, λ), has the following probability mass function (Equation 4).

$$P(X = k) = \begin{cases} \pi + (1 - \pi)\, exp(-\lambda) & \text{if } k = 0 \\ (1 - \pi) exp(-\lambda) \frac{\lambda^k}{k!} & \text{if } k \in \{1, 2, \dots\}, \end{cases} \quad \text{---------- (4)}$$

where $0 \leq \pi \leq 1$ and $\lambda \geq 0$. The parameter π gives the extra probability thrust at the value 0; when it vanishes, ZIP (π, λ) reduces to Poi(λ).

4.4 Mean squared error (MSE)

The MSE is a measure of the amount of error in statistical models. It calculates the average squared difference between the observed and predicted values (Das, 2004). When a model has no error, the MSE equals zero. As model error increases, its value also increases. The mean squared error is also referred to as the mean squared deviation (MSD). In regression, the MSE represents the average squared residual (Hodson, 2022). The formula for MSE is provided below (Equation 5).

$$\text{MSE} = \frac{\Sigma(y_i - \hat{y}_i)^2}{n}, \quad \text{---------- (5)}$$

where,

y_i = i^{th} observed number of heatwaves
$\hat{y}_i$ = Corresponding predicted heat waves
n = Number of observations

A lower MSE indicates that the model's predictions are closure to the true values, reflecting better overall performance (Willmott and Matsuura, 2005).

4.5 Mean absolute error MAE)

The MAE is one of the fundamental metrics used for assessment of the regression model. It measures the average absolute difference between the predicted values and the actual target values. Unlike other metrics, MAE doesn't square the errors, which means it gives equal weight to all errors, regardless of their direction.

The formula for calculating MAE is as follows (Equation 6).

$$\text{MAE} = \frac{\Sigma_i^n |y_i - \bar{y}|}{n}, \quad \text{---------- (6)}$$

where, y_i is the i^{th} observed number of HWs, $\bar{y}$ is the average number of HWs and n is the number of counts of HWs.

The MAE value itself indicates the average absolute error between predicted and actual values. The smaller the MAE, the better the model's predictions align with the actual data. A MAE of 0 would mean a perfect prediction, but in most cases, achieving such perfection is unlikely (Karunasingha, 2022).

Results and Discussion

This section presents the compelling findings of our study, leading to a thorough discussion on the research objectives and a comprehensive comparative analysis.

1.0 Exploratory Data Analysis

The variables associated with the occurrences of HW are provided in Table 1 along with the descriptive analysis. Table 1 shows that the daily maximum temperature is 44.47 degrees Celsius, while the

minimum temperature is 3.43 ^{0}C. The maximum precipitation is 167.28 mm/day, and the average daily rainfall is 9.72 mm/day in Gazipur district. We found that the yearly maximum number of HW is 12, while no HW is also observed in a year with standard deviation 2.91. The average relative humidity, surface pressure, wind speed at 2-meter, and surface soil moisture is observed 70.56%, 100.67 Kpa, 2.04 m/s, 0.58 respectively. The later section revealed more findings about climate parameters and HWs obtained in our research.

The graph in figure 2 illustrates the trend of the average daily maximum temperature in Gazipur district from 2017 to 2023. The data shows a consistent increase in the mean daily maximum temperature over the years. This increasing pattern is causing an alarming situation for the people of Gazipur district, and hence Bangladesh. This situation has led us to conduct this research.

Table 2. Descriptive statistics of climate parameters (measured daily): 1982 - 2023

	Mean	Standard deviation	Minimum	Maximum
Variables of the interest	6.38	2.91	0	12
Maximum temperature (^{0}C)	31.24	3.93	16.38	44.47
Minimum temperature (^{0}C)	21.21	5.43	3.43	29.42
Specific humidity (g/kg)	14.93	5.77	2.2	23.99
Relative humidity (%)	70.56	17.76	12.56	95.75
Precipitation (mm/day)	4.89	9.72	0	167.28
Surface pressure (Kpa)	100.67	0.55	98.94	101.98
Wind speed at 2-meter (m/s)	2.04	1.01	0.4	9.16
Wind speed at 10-meter (m/s)	3.01	1.37	0.58	12.95
Wind direction at 2-meter (degree)	188.83	58.31	17.06	348.62
Surface soil wetness (0-1)	0.59	0.21	0.09	0.99
Profile soil moisture (0-1)	0.58	0.17	0.34	0.98

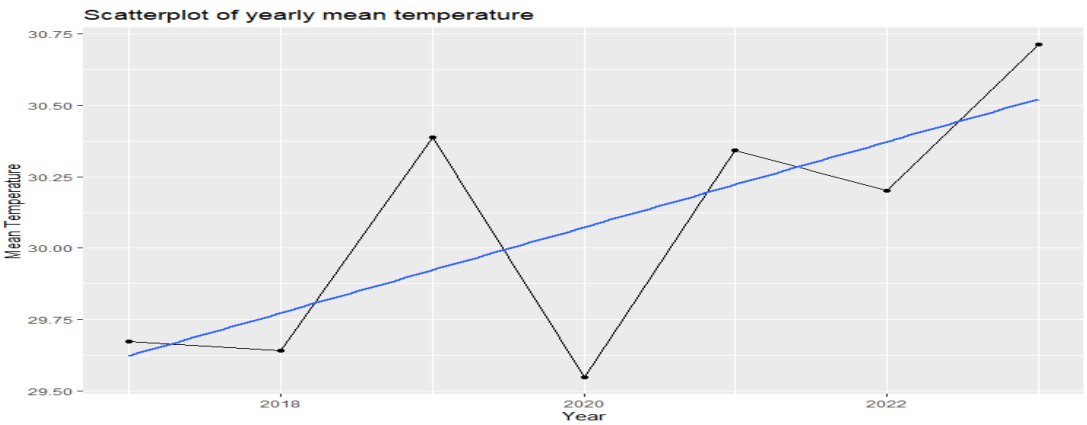


Figure 2: Scattter plot of yearly mean temperature with trend line

2.0 Scenerarios of Heatwaves

2.1 Yearly heatwave

According to the definition of heatwaves (HW) given in the introduction section (Table 1), we have calculated the occurrences of mild, moderate, severe, and extreme HWs. The analysis of heatwave occurrences is presented in Figure 3. The occurrences of mild and moderate heatwaves have increased over the years, but no severe and extreme heatwaves have been observed in Gazipur district from 2017 to 2023. This means no two or more consecutive days have a temperature greater than 40 degrees. However, it is possible for the temperature to be below 40 degrees one day and above 40 degrees the next day. On those days, we may feel extremely hot, but we do not count them as severe or extreme HW days. The graph in Figure 4(a) illustrates a gradual increase in the yearly count of mild heatwaves.

In our c
2018, in
not inve

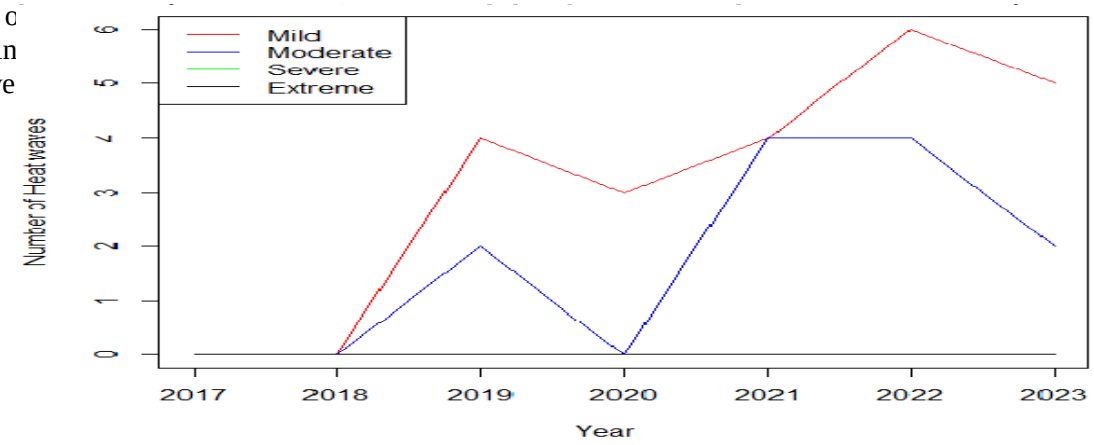
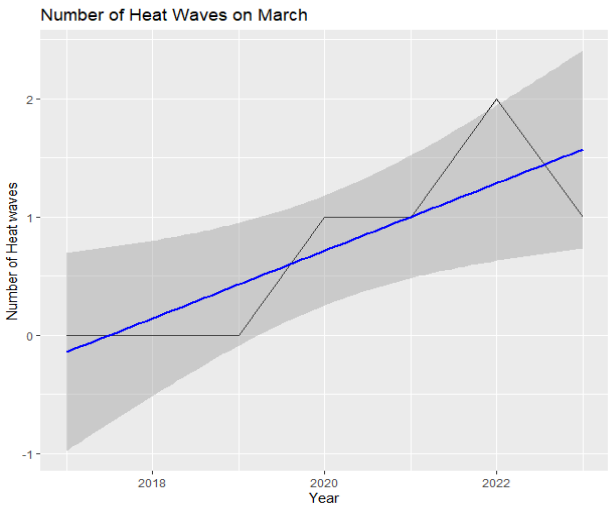


Figure 3: Yearly frequency of heatwave (Temperature ≥ 38℃ for ≥ two days)

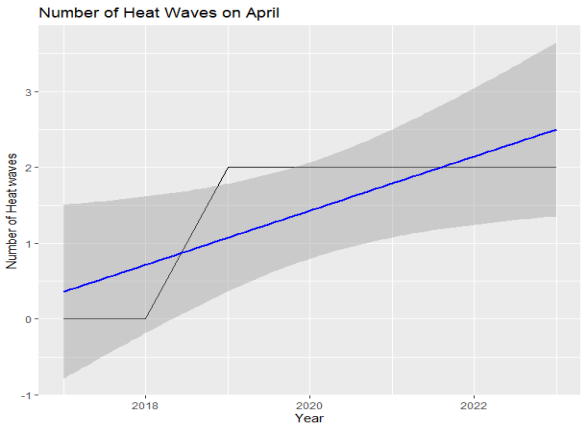
2.2 Monthly heatwave

In this research, one of our specific objectives was to explore the current scenario of HW in Bangladesh. The results for the HW scenarios in Gazipur district are presented in Sections 2.1 and 2.2, based on the POWER project data from NASA. We specifically chose to focus on Gazipur district because the POWER project provides gridded (50 km × 62.5 km) data (Duarte and Sentelhas, 2020). Using data for the entire Bangladesh would not accurately represent the country. Figure 4(b,c,d) illustrates the monthly heatwave scenario in Gazipur district for March, April, and May. No heatwaves are observed other than in these months, and we found that the number of heatwaves is gradually increasing. Additionally, we bserved no HW in March 2017–2019 and April–May 2017-2018 in the Figure 4(b,c,d).

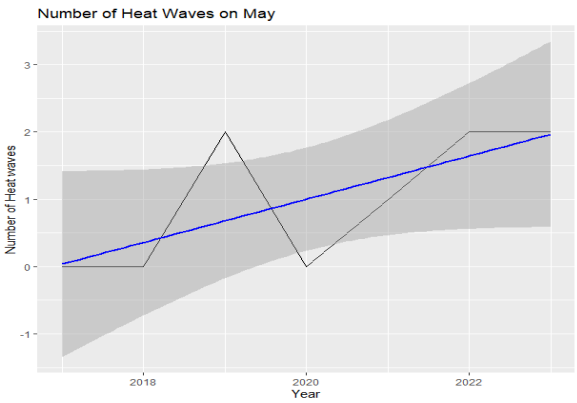
a Yearly mild HWs



b. HWC in March



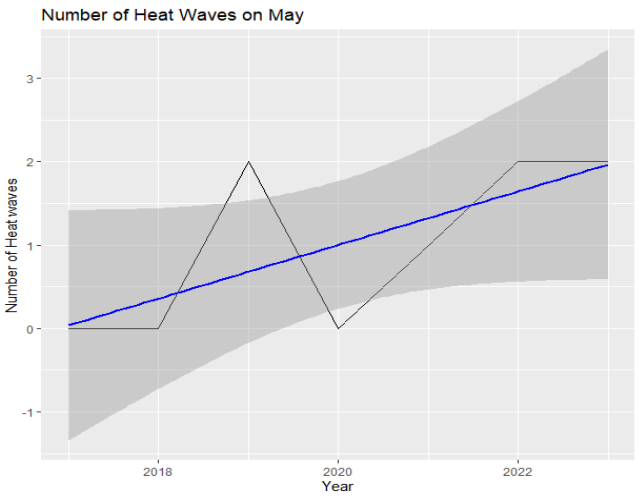
c. HWC in April



d. HWC in May

Fig 4. Monthly heatwave illustration of Gazipur district

The monthly duration of heat wave days (HWDs) is illustrated in Figure 5. The data shows that the size of HWDs is increasing from 2019 to 2023. More specifically, in 2019, March experienced a total of 8 days with temperatures exceeding 36°C, which was the maximum for that year. In 2020, the highest number of HWDs occurred in April, with 6 days exceeding 36°C. In 2021, April saw the maximum number of HWDs with 15 days exceeding 36°C. The year 2022 experienced the most HWDs, with 16 occurring in May. Lastly, in 2023, the maximum number of HWDs in April was 15 days with temperatures greater than 36°C.



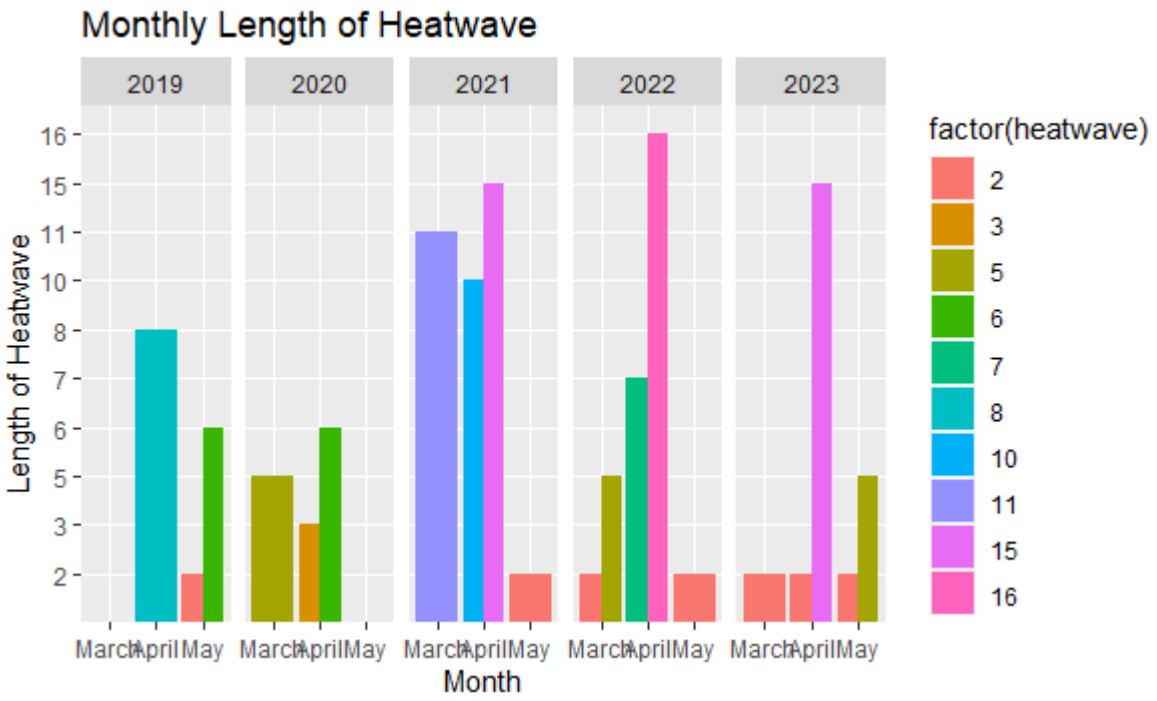


Fig 5. Monthly length of heatwaves: number of days having temperature greater than 36℃

3.0 Prediction of Monthly Heatwave Using ZIP Model

Before fitting the ZIP model in this section, we performed a Spearman's rank correlation analysis of the number of monthly heatwaves (HW) with other variables identified in the literature as potentially associated with heatwaves. We also conducted a correlation test to identify variables significantly correlated with the occurrence of heatwaves. The references for the literature are provided.

3.1 Correlation analysis

In Fig 6, after conducting a correlation analysis, we have decided to include the variables in the ZIP model. In figure M_HWC, the monthly heatwave count (MHWC) is moderately positively correlated (r = 0.62) with maximum temperature, and has a moderate negative relationship with relative humidity (RH), soil wetness (WET), and soil moisture (Moisture) with correlation coefficients of -0.46, -0.50, and -0.51 respectively. There is a weak association observed with specific humidity (SH), precipitation (Preci), surface pressure (SPS), wind speed at 2-meter & 10-meter (WS2M), and wind direction at 2-meter (WD2M).

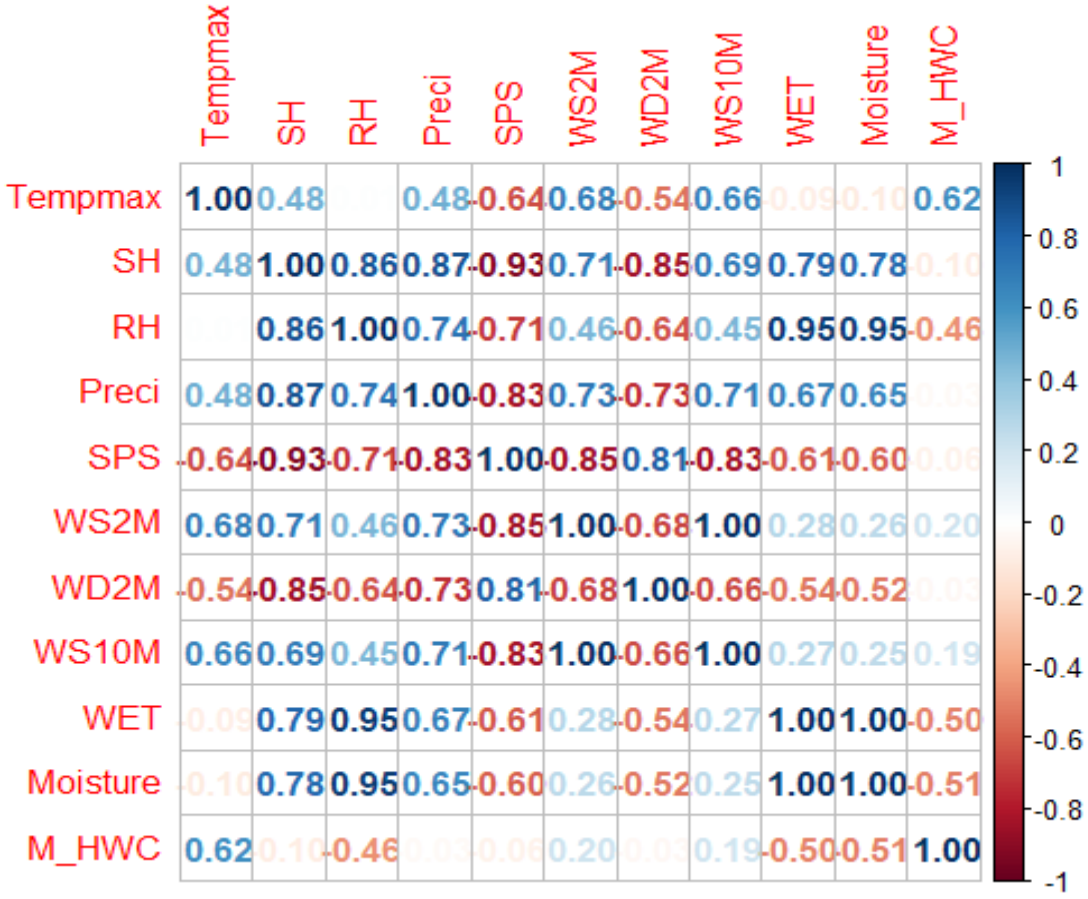


Figure 6: Correlation matrix of number of heatwave with climate parameters

In Table 3, we observed that the maximum daily temperature is highly significantly associated with MHWC, and relative humidity, soil wetness, and soil moisture are also highly significantly associated with MHWC with a p-value of 0.000. We found that as the maximum temperature increases, the MWHC will increase. Additionally, we found that an increase in relative humidity, soil wetness, and soil moisture leads to a decrease in MHWC. Unfortunately, we did not find a significant association between the amount of monthly precipitation and MHWC, but it can be inferred from the analysis that an increase in monthly precipitation results in a decrease in MHWC. Moreover, we found a significant positive correlation between wind speed at 2 meters and 10 meters, with p-values of 0.06 and 0.08, respectively. However, we did not find any significant association between surface pressure and wind direction.

3.2 Heatwave prediction

In the ZIP model, there are two parts: the count model and the zero-inflated model (binomial logit). Based on the correlation analysis, we included maximum temperature, relative humidity, soil wetness, soil moisture, and wind speed at 2 meters and 10 meters in the count model. These variables, with the exception of wind direction, are included in the zero-inflated model. The data was divided into two parts: a training set from 1982 to 2021 and a validation set from 2022-2023. The results of the ZIP model are presented in Table 4.

Table 3: Correlation test of MHWC with agroclimatogical parameters of POWER data

Climate Parameters with MHWC	Spearman's Rank Correlation Coefficient	P-value
Maximum temperature	0.62	0.00***
Specific humidity	0.10	0.37
Relative humidity	-0.46	0.00***
Precipitation	-0.03	0.80
Surface pressure	0.06	0.50
Wind speed at 2-meter	0.20	0.06 ▪
Wind speed at 10-metet	0.19	0.08 ▪
Wind direction at 2-meter	-0.033	0.76
Wetness	-0.50	0.00***
Soil moisture	-0.51	0.00***

Note: '***' Significant at 0.1%, '**'significant at 1%, "*" 5% level of significance

Our analysis in Table 3 shows that maximum temperature and soil moisture have a significant influence on the prediction of monthly heatwave count at a 0.1% significance level, with p-values of 0.004 and 0.003 respectively. Additionally, soil wetness is significantly associated with a 1% significance level. In the inflation model, we also found that temperature significantly contributes to the occurrence of zero heatwave counts.

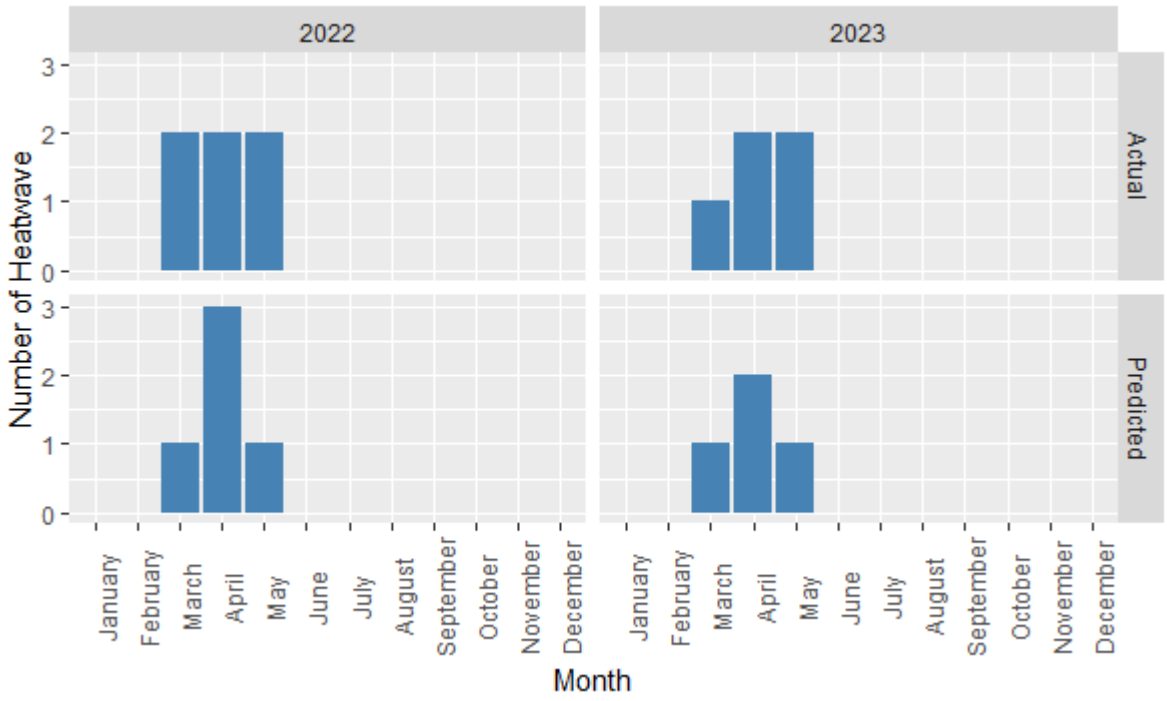


Fig 7. Monthly predicted heatwave counts with the actual count of heatwave

Table 4: Zero Inflated Poisson model estimations

	Coefficient	Std. Error	P-value
Count model coefficients with log link			
Maximum temperature	0.17	0.06	0.004 **
Relative humidity	0.002	0.023	0.927
Wind speed at 2-meter	1.29	2.71	0.635
Wind speed at 10-meter	-0.98	1.98	0.621
Soil wetness	7.28	2.92	0.012 *
Soil moisture	-11.84	4.09	0.003**
Intercept	-3.07	2.66	0.248
Zero-Inflation model coefficients with binomial logit link			
Maximum temperature	-4.8806	2.63	0.063 ▪
Relative humidity	0.2029	0.14	0.128
Wind speed at 2-meter	50.1241	46.55	0.281
Wind speed at 10-meter	-38.4949	35.37	0.276
Soil moisture	-20.19	15.94	0.205
Intercept	168.2504	15.94	0.077 ▪

Note: '*' Significant at 0.1%, '**' significant at 1%, "*" 5% level of significance**

Fig 7 presents the predicted and actual MWHC for the Gazipur district in 2022-2023 as obtained from the ZIP model. Heatwaves were observed only in March, April, and May. In 2022, the actual MWHC was 2 in March, April, and May, while the predicted MWHC was 1, 3, and 1 for March, April, and May respectively. In 2023, we observed actual MWHC values of 1, 2, and 2 in March, April, and May respectively, while the predicted values were 1, 2, and 1 for the same months. The accuracy of the ZIP model's predictions was assessed using MAE and MSE. A lower MAE and MSE indicate better predictive performance (Marmolin, 1986; Hodson, 2022). The obtained value for the predicted model is 0.1667, which is very close to zero, indicating a strong performance in predicting MWHC.

4.0 Effects of MHWC on the Yield of Summer Tomato

Two BARI-developed hybrid varieties namely BARI Hybrid tomato-4 and BARI Hybrid tomato-8 are heat-tolerant (Alam et al., 2010). However, we would like to investigate the potential impact of MHWCs and MHWDs on the yield of summer tomatoes in Bangladesh.

The average yield of summer tomatoes was 42.12 tons per hectare, with a standard deviation of 7.24. The minimum and maximum yields were 27.70 tons per hectare and 50.41 tons per hectare, respectively. In Fig 8, a line diagram illustrating the yearly yield of summer tomatoes, it is evident that there has been an increase in the yield of summer tomatoes.

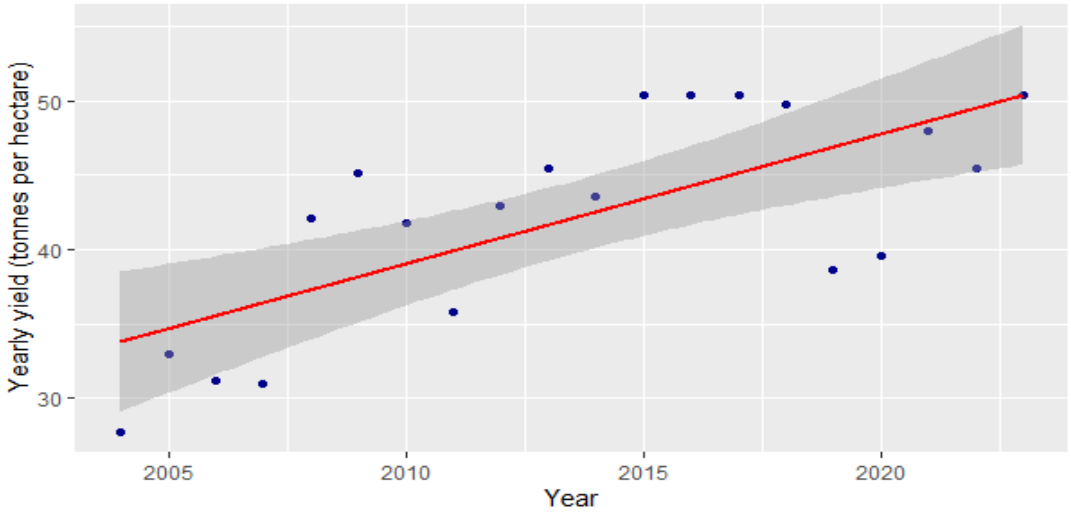


Fig 8. Line diagram of the yield of summer tomato

In Fig 9, a correlogram is presented. Scatter plots are visualized in the lower diagonal, density plots are in the diagonal, and the Pearson correlation coefficient with the significance level is displayed in the upper diagonal. We observed that as the yearly HWC and HWD increase, the yield decreases for each pair of scatterplots. We also found that there is a significant negative correlation between the yield and HWC, with a correlation coefficient of -0.488 at a significance level of 5%. Additionally, the correlation between HWD and yield is -0.272, indicating a negative linear relationship between them. These negative correlation coefficients indicate also that for the increase of HWCs and HWDs, the yield of summer tomato decreases.

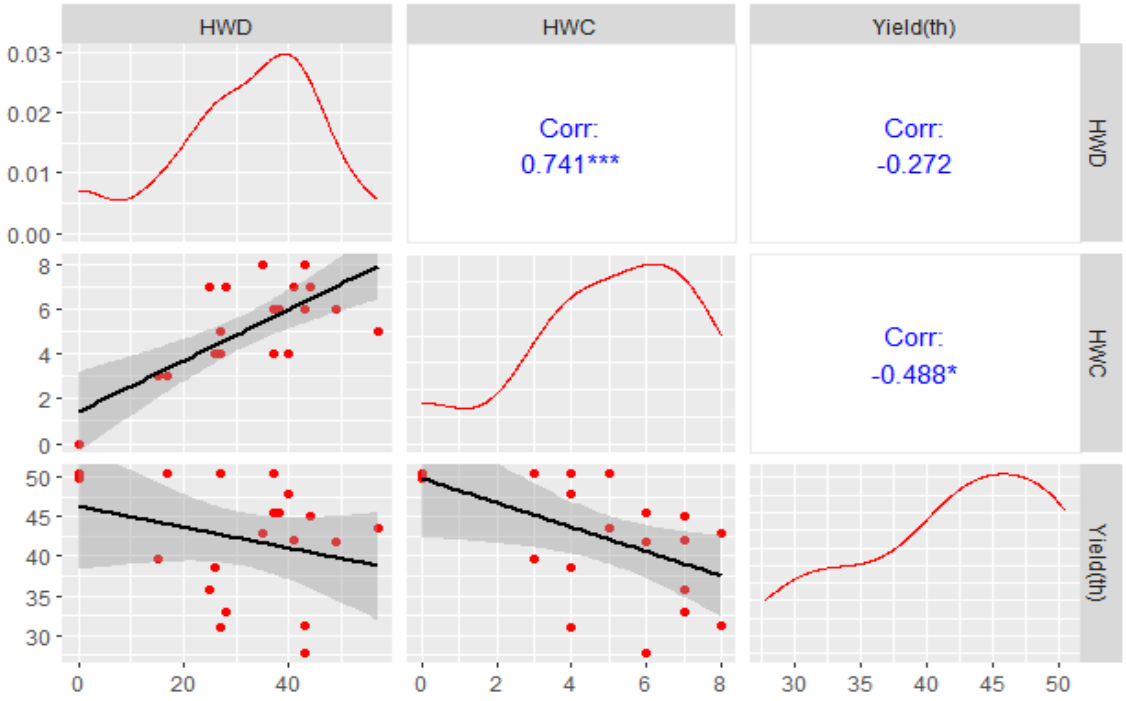


Fig 9. Correlogram of yearly HWCs, HWDs and yield of summer tomato

The correlation coefficients show the strength of the association between variables, but they do not establish causation (Poole and O'Farrell, 1971). To investigate the causation of yield and HWC, we conducted a linear regression analysis in this study. The LR estimates are presented in Table 5, and we observed that the overall model is significant with a p-value of 0.029. This indicates that the estimated

model is more effective than the null model in predicting and explaining the yield of summer tomatoes. The coefficient of determination, adjusted R-squared, is 20% for the model, meaning that only HWC explains 20% of the total variation in the yield of summer tomatoes. Other characteristics influencing yield, such as plant height, number of leaves per plant, number of branches per plant, virus infestation (%), fruit length and diameter, individual fruit weight, number of locules, and total soluble solid (TSS) content, as shown by Biswas et al, 2014, could increase the adjusted R-squared. However, this is beyond the scope of our study.

Table 5. Linear regression estimates of yield of summer tomatoes with HWC

Variables	Coefficients	Std. Error	t-value	p-value
HWC	-1.5414	0.6497	-2.373	0.029*
Intercept	49.8320	3.5584	14.004	0.000***

P-value for overall model significance: 0.02901

Adjusted R^2: 20%

Note: * Significant at 0.1%, ** significant at 1%, and * 5% level of significance**

Our analysis indicated that the coefficient of HWC is significantly different from 0, as its p-value is 0.029, which is less than 0.05. This provides evidence that HWC has a significant effect on the yield of summer tomatoes in Bangladesh, which was one of the specific objectives of this program. The intercept, 49.8320, is also highly significant with a p-value of 0.000, indicating that the average yield of summer tomatoes in Bangladesh is 49.8320 (t/ha) when no other variables are considered in the model. The coefficient of HWC is -1.5414, meaning that an increase in HWC by one day per year will result in a decrease of 1.5414 (t/ha) in the yield of summer tomatoes.

Conclusion

This study developed and evaluated the performance of a ZIP model for monthly heatwave (HW) prediction in Gazipur district, using assimilation model (MERRA-2) agro climatological data. The study also investigated the impacts of HWs on the yield of summer tomatoes. The predictive model performed well in predicting HWs during the validation period, 2022-2023, with very low (0.167 for both) Mean Squared Error (MSE) and Mean Absolute Error (MAE). This study also showed significant negative effects of HWs on the yield of summer tomatoes. The findings suggest that policymakers may utilize the predictive model to anticipate HWs two years in advance, helping to reduce damages to summer tomato production in Bangladesh. By adjusting the sowing and harvest time of summer tomatoes based on the early HW predictions, the overall yield can be increased, potentially preventing yield reduction. However, it is important to note that the monthly HW predictions are specific to the Gazipur district, and future research may focus on developing region-specific predictive ZIP models for short, mid, and long-term predictions, as well as creating spatial maps of HWs in Bangladesh. Additionally, forecasting the yield of summer tomatoes based on the findings of this study may contribute to a better understanding of growing demand.

Acknowledgement

The authors are thankful to the POWER project funded by NASA for their open-source gridded agro climatological data for the Gazipur district. I want to express my special gratitude to Dr. AKM Quamruzzaman, Chief Scientific Officer and Head of the Olericulture Division at HRC, BARI, and to Dr. Muhammad Nazim Uddin and Dr. Limu Akter, Senior Scientific Officers in the same department, for providing the summer tomato yield data for this study. Also, we acknowledge Abu Hanifa from the Institute of Statistical Research and Training (ISRT), University of Dhaka for his assistance in analyzing the data.

References

Abdalla, N., Taha, N., El-Ramady, H. and Bayoumi, Y. (2020). Management of heat stress in tomato seedlings under arid and semi-arid regions: a review. *Environment, Biodiversity and Soil Security*, 4:47-58.

Alam, M. S., Sultana, N., Ahmad, S., Hossain, M. M. and Islam, A. K. M. A. (2010). Performance of heat tolerant tomato hybrid lines under hot, humid conditions. *Bangladesh Journal of Agricultural Research*, 35(3).

Alsamir, M., Mahmood, T., Trethowan, R. and Ahmad, N. (2021). An overview of heat stress in tomato (*Solanum lycopersicum* L.). *Saudi journal of biological sciences*, 28(3):1654-1663.

AU, C. O. (2014). Heatwaves: Hotter, Longer, More Often. *Climate Council of Australia, 2014.*

Beckett, S., Jee, J., Ncube, T., Pompilus, S., Washington, Q., Singh, A. and Pal, N. (2014). Zero-Inflated Poisson (ZIP) distribution: Parameter estimation and applications to model data from natural calamities. *Journal of Mathematics*, 7(6): 751-767.

Biswas, M., Sarkar, D. R., Asif, M. I., Sikder, R. K., Mehraj, H.and Jamal Uddin, A. F. M. (2014). Comparison of growth and yield characteristics of BARI tomato varieties. *Journal of Bioscience and Agriculture Research*, 3(1): 1-7.

Chowdhury, A. F. M. K., Ahmmed, S.and Chowdhury, M. A. I. (2012). Trends of climatic variables (rainfall and temperature) at Sylhet, Bangladesh. *SUST Journal of Science and Technology*, 19(5): 87-93.

Das, K., Jiang, J.and Rao, J. N. K. (2004). Mean squared error of empirical predictor. *The Annals of Statistics, 32(2).*

Das, S. B., Islam, M. S., Uddin, M. N., Goswami, B. K., Abedin, S. S.and Baki, M. A. (2023). Evaluation of the performance of summer tomato lines in Bangladesh conditions. *Journal of Bio-Science*, 31(2): 13-23.

Della-Marta, P. M., Luterbacher, J., von Weissenfluh, H., Xoplaki, E., Brunet, M.and Wanner, H. (2007). Summer heat waves over western Europe 1880–2003, their relationship to large-scale forcings and predictability. *Climate Dynamics*, 29: 251-275.

Domeisen, D. I., Eltahir, E. A., Fischer, E. M., Knutti, R., Perkins-Kirkpatrick, S. E., Schär, C. and Wernli, H. (2023). Prediction and projection of heatwaves. *Nature Reviews Earth & Environment*, 4(1): 36-50.

Duarte, Y. C. and Sentelhas, P. C. (2020). NASA/POWER and daily gridded weather datasets—how good they are for estimating maize yields in Brazil? *International Journal of Biometeorology*, 64: 319-329.

Green, J. A. (2021). Too many zeros and/or highly skewed? A tutorial on modelling health behaviour as count data with Poisson and negative binomial regression. *Health Psychology and Behavioral Medicine*, 9(1): 436-455.

Hajong, P., Sikder, B., Mondal, S.and Islam, M. A. (2018). Adoption and profitability of summer tomato cultivation in Jashore district of Bangladesh. Bangladesh *Journal of Agricultural Research*, 43(4):575

Hauke, J. and Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae*, 30(2): 87-93.

Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, 1-10.

Ibrahim, E. S. Z. (2016). Tomato breeding for heat stress conditions. *European Journal of Academic Essays*, 3(2): 87-93.

Islam, A. S. (2009). Analyzing changes of temperature over Bangladesh due to global warming using historic data. Young Scientists of Asia Conclave, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), 15: 15-7.

James, G., Witten, D., Hastie, T., Tibshirani, R.and Taylor, J. (2023). Linear regression. In: An introduction to statistical learning: With applications in python . Cham: Springer International Publishing. pp. 69-134

Li, Z. (2023). Monthly heatwave prediction in Sweden based on machine learning techniques with remote sensing data. *Remote Sensing*, 15(16): 3921

Karim, M. R., Rahman, M. S. and Alam, M. S. (2009). Profitability of summer BARI hybrid tomato cultivation in Jessore district of Bangladesh. *Journal of Agriculture & Rural Development*, 7(1): 73-79.

Karmakar, S. (2019). Patterns of climate change and its impacts in northwestern. *Bangladesh. J. Eng. Sci*, 10: 33-48.

Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, 585: 609-629.

Kim, D. W., Deo, R. C., Park, S. J., Lee, J. S. and Lee, W. S. (2019). Weekly heat wave death prediction model using zero-inflated regression approach. *Theoretical and Applied Climatology*, 137: 823-838.

Kuo, C. G., Chen, B. W., Chou, M. H., Tsai, C. L. and Tsay, T. S. (1979). Tomato fruit-set at high temperatures. *HortScience*, 37(1): 172-175.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1): 1-14.

Mannan, M. A., Akter, F., Rashid, M. B., Hossain, M. M., Faruque, S. O., Ara, S. S., and Chowdhury, M. A. M. (2021). An analytical study of Heat Wave occurred over Bangladesh during pre-monsoon of 2021. *Dew-Drop*, 8(1): 152-160.

Marmolin, H. (1986). Subjective MSE measures. *IEEE transactions on systems, man, and cybernetics*, 16(3): 486-489.

Mishra, V., Ganguly, A. R., Nijssen, B. and Lettenmaie, D. P. (2015). Changes in observed climate extremes in global urban areas, *Environ. Res. Lett.*, 10: 024005. https://doi.org /10.1088/ 1748-9326/10/2/024005.

Nissan, H., Burkart, K., Coughlan de Perez, E., Van Aalst, M.and Mason, S. (2017). Defining and predicting heat waves in Bangladesh. *Journal of Applied Meteorology and Climatology*, 56(10): 2653-2670.

Panthee, D. R., Kressin, J. P. and Piotrowski, A. (2018). Heritability of flower number and fruit set under heat stress in tomato. *HortScience*, 53(9): 1294-1299.

Perkins, S. E. (2015). A review on the scientific understanding of heatwaves-Their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research*, 164: 242-267.

Poole, M. A.and O'Farrell, P. N. (1971). The assumptions of the linear regression model. *Transactions of the Institute of British Geographers*, 145-158.

Roberts, M. J.and Schlenker, W. (2011). Is agricultural production becoming more or less sensitive to extreme heat? Evidence from US corn and soybean yields. In: The Design and Implementation of US Climate Policy . University of Chicago Press. pp. 271-282.

Scherm, H. and Van Bruggen, A. H. C. (1994). Global warming and nonlinear growth: how important are changes in average temperature? *Phytopathology*, 84(12): 1380-1384.

Schober, P., Boer, C.and Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia*, 126(5): 1763-1768.

Seber, G. A.and Lee, A. J. (2012). Linear regression analysis. John Wiley & Sons.

Sedgwick, P. (2014). Spearman's rank correlation coefficient. *British Medical Journal*, 349: g7327

Siddik, M. A. Z. and Rahman, M. (2014). Trend analysis of maximum, minimum, and average temperatures in Bangladesh: 1961–2008. *Theoretical and applied climatology*, 116: 721-730.

Srivastava, A. K., Rajeevan, M. and Kshirsagar, S. R. (2009). Development of a high resolution daily gridded temperature data set (1969–2005) for the Indian region, *Atmos. Sci. Lett*., 10: 249–254, doi:10.1002/asl.232.

Stevens, M. A.and Rudich, J. (1978). Genetic potential for overcoming physiological limitations on adaptability, yield, and quality in the tomato. *HortScience*, 13(6): 673-678.

Weisberg, S. (2005). Applied Linear Regression (Vol. 528). John Wiley & Sons.

White, G. C.and Bennetts, R. E. (1996). Analysis of frequency count data using the negative binomial distribution. *Ecology*, 77(8): 2549-2557.

Willmott, C. J.and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1): 79-82.

Zaman, M. M., Huq, A. S. M. A.and Chowdhury, M. J. A. (2006). Production potentiality of summer tomato in Jamalpur region. *International Journal of Sustainable Crop Production*, 1(2): 12-15.

PREDICTIVE MODELING OF CLIMATE CHANGE IMPACTS ON POTATO YIELD: A STATISTICAL INVESTIGATION

JAMILA KHATUN PRIOTY, ISTIAK AHMED AND M.A. MONAYEM MIAH

Abstract

Amidst the shifting climate of Bangladesh, it is essential to study potato yield and climate change. This study has predicted the impacts of climate change on potato yield using machine learning models. We have used potato yield and agro-climatological data from the past 16 years (2006-07 to 2021-22) for major potato-growing districts like Bogura, Munshiganj, Rangpur, Joypurhat, and Rajshahi. The study analyzed climate data for long-term trends and changes, investigated the relationship between climate variables and potato yield, developed predictive models for future potato yield, and assessed the vulnerability of potato yield to climate change. Trend analysis has been used to identify significant shifts in historical climate data, while Pearson's correlation coefficient explored the link between climate variables and potato yields. Different predictive models such as multiple linear regression, Random Forest (RF), and Support Vector Machine (SVM) have been used to estimate the impacts of climate change on potato yield. The dataset is divided into an 80% training set (2006-07 to 2017-18) and a 20% testing set (2018-19 to 2021-22) for robust model training and validation. Results show that root zone soil wetness, surface soil wetness, profile soil moisture, surface pressure, and relative humidity at 2 meters exhibit an increasing trend with some fluctuations, while other climate variables display random patterns. Significant climate variables at the 10% level have included minimum temperature at 2 meters, root zone soil wetness, profile soil moisture, surface soil wetness, and maximum wind speed at 2 meters. SVM model is the best predictive model for Bogura, Munshiganj, Joypurhat, and Rajshahi. In contrast, the RF model performs best for Rangpur, based on root mean square error. Potato yields in the study areas showed a stable or increasing pattern despite fluctuations in significant climate parameters.

Keywords: Climate parameter, Potato, Random forest, Support vector machine

Introduction

Our world has undergone enormous changes in weather patterns and environmental conditions over the past few decades, disrupting the agriculture sector and posing serious problems for food security and sustainability. Noguer et al. (2001) predict that between 1990 and 2100, the average world temperature will rise between 1.4 and 5.8 C due to the ongoing global warming trend of 0.6 C + 0.2 since 1900. Studies by Mearns et al. (1984) and Hansen et al. (1988) revealed that even modest changes in mean temperature can lead to substantial variations in the frequency and intensity of temperature extremes. These extreme climatic events can have far-reaching impacts on both the environment and society. For instance, rising nighttime temperatures due to global warming have been linked to a decline in crop yields (Toros et al., 2019). Furthermore, there is mounting evidence suggesting that climate change is altering the temporal dynamics and overall levels of precipitation. Previous research has highlighted the critical role of daily rainfall timing in crop production (Shortridge, 2019). This type of climatic change would likely reduce crop productivity, however, the extent to which this occurs may vary greatly by region (McCarthy et al., 2001). Potato cultivation is diverse, thriving in various environments, yet it adapts optimistically to temperate climates (Haverkort, 1990). When temperatures rise above 17°C, as noted by Stol et al. (1991), the process of tuberization in potato diminishes (Reynolds and Ewing, 1989a). Additionally, potatoes are sensitive to frost, and when temperatures fall below 0°C, they risk severe damage (Hijmans et al., 2003). Several researchers have employed simulation models to investigate the impact of global warming on potato production. According to the SUBSTOR-Potato model, if we extrapolate future climate change scenarios to our current potato cropping systems, we can expect a moderate loss (-2% to -6%) in worldwide tuber production by 2055 and a bigger decline ranged from -2% to -26% by 2085 (Raymundo et al., 2018). Using the GLAM crop model and the ISI-MIP ensemble of global climate models, it is projected that global potato yields could increase by 9 to 20% by 2050 with adaptation measures, aligning with a 10 to 17% boost suggested by various climate models (Jennings et al., 2020).

Bangladesh is a South Asian disaster-prone country. The rain-fed agriculture of the country depends on seasonal rainfall, and the South Asian monsoon is the most important climatic phenomenon that is directly related to the intensity and frequency of rainfall and drought over the country (Dastagir, 2015). So, Bangladesh faces significant challenges in adapting to the impacts of climate change. It has often been cited as one of the most vulnerable countries to climate change. Bangladesh ranked seventh out of ten on the Global Climate Risk Index 2021 of the countries most affected by climate change since 2000 (Eckstein et al., 2021). According to the ND-GAIN index (Chen et al., 2015), Bangladesh is as vulnerable to climate change as 164 of the world's 182 countries. Many scientists have looked at global warming trends in surface temperatures, and they've found that different regions are experiencing it at different intensities and rates (IPCC, 2007; Kerr, 2009; Abbasnia et al., 2016; Abbasnia and Toros 2018a,b). There has been a strong warming in the warm period compared to the temperatures of the annual and cold periods (Toros, 2012). However, there has been limited research on climate extremes at the regional scale, which calls for ongoing observation. Several research projects have been done in Bangladesh on this topic. Recent years have seen an increase in the minimum temperature and a warming trend in the maximum temperature (Nishat and Mukherjee, 2013) Islam et al. (2004) found an increasing trend for the monthly maximum temperature, but the monthly minimum temperature shows both increasing and decreasing trends. Plant growth, pollination and reproductive systems are negatively impacted by higher temperatures (Klein Tank et al., 2006; Sacks and Kucharik, 2011). The growth and final output of crops can be significantly harmed by a brief period of extremely hot or low temperatures (Mearns et al., 1984). So it is important to investigate temperature and precipitation-related disasters and assess global warming impacts on crop yields. There has been a dearth of previous potato modeling studies compared to other crops, especially at the national scale, despite the importance of understanding the potential of potatoes to contribute to a climate-smart future. Salan et al. (2022) found that the ANOVA model in the Spatial-temporal framework was the most effective in assessing spatial and temporal variations in potato yield, revealing significant variability influenced by temperature, humidity, rainfall, and sunshine. Notably, wind speed showed no impact on Bangladeshi potato yields. Additionally, Rahman et al. (2018) projected a temperature increase of 5.32ºC by 2100 in Bangladesh, which is expected to reduce potato yields by 38.6%.

Potato is a popular food and vegetable crop in Bangladesh. It is used as a food crop as well as a vegetable by both poor and rich people. Besides, potatoes may not have a lot of protein, but the presence of important amino acids like lysine and metabolites boost their biological value and usage (Gorska et al., 2021). In addition to minerals like magnesium, iron, potassium, and phosphorus, potatoes include 0.5-2% of dietary fiber and are rich in vitamins like B_6 and C as well as trace quantities of folate, thiamin, niacin, and riboflavin (Beals et al., 2019). Bangladesh was chosen as the study country because it is located in a temperate zone, ranks seventh in terms of global potato production, and is third in Asia, behind only China and India, in terms of production (FAOSTAT, 2022). However, it is also important to keep in mind that pests and diseases hampered potato production due to the prevailing climatic conditions, which favor the development of specific diseases in Bangladesh (Chakraborty and Roy, 2016). According to Jannat et al. (2021), low rainfall and low sun radiation during the potato's maturation stage make it difficult to cultivate, while temperatures above or below the ideal range (18-22 °C) reduce yields. So, it is crucial for sustainable farming to understand the impacts of climate change on potato yield so that Bangladesh can meet its own needs and export good-quality potatoes to other countries. The specific objectives of this study are:

1. To analyze climate data to identify long-term trends and changes in climate variables;
2. To investigate the relationship between climate variables and potato yield;
3. To develop predictive models to estimate the impact of climate change on future potato yield and
4. To assess the vulnerability of potato yield to climate change and identify potential adaptation strategies for future improvement of this crop.

Materials and Methods

Data and Its Sources

In this study, we selected five potato-growing districts in Bangladesh- Bogura, Munshiganj, Rangpur, Joypurhat, and Rajshahi due to their collective contribution of about half (53.22%) to the country's total potato production.

The study utilized two types of data: potato yield data and agro-climatological data. The annual potato yield data, spanning the last 16 years (from 2006-07 to 2021-22), were sourced from various editions of the Yearbook of Agricultural Statistics of Bangladesh published by BBS. The agro-climatological data (2006-07 to 2021-22) were obtained from the Prediction of Worldwide Energy Resources (POWER) project, funded by the National Aeronautics and Space Administration (NASA) Earth Science Directorate Applied Science Program. The data can be accessed through the POWER Data Access Viewer. The POWER project aims to provide solar and meteorological data to assist in the development of renewable energy systems and climate research. The POWER project provides satellite and model-based solar and meteorological resource data, which are particularly valuable in regions where surface measurements are sparse or nonexistent. The meteorological data/parameters in POWER are based upon Goddard's Global Modeling and Assimilation Office (GMAO) Modern Era Retrospective-Analysis for Research and Applications (MERRA-2) assimilation model and GEOS 5.12.4 FP-IT. The data/parameters in POWER are provided on a global grid with spatial resolutions equal to the input data. That resolution is ½° latitude by ⅝° longitude (About 50 km × 62.5 km) for the meteorological data sets with the WGS84 (World Geodetic System 1984) grid reference system. The climate parameters, related to our research, we collected from the POWER project data are:

Table 1. Climate parameters and their unit

Climate parameters	Unit
Surface pressure	k Pa
Temperature at 2 meters	C
Relative humidity at 2 meters	%
Surface soil wetness	1
Temperature at 2 meters maximum	C
Temperature at 2 meters minimum	C
Profile soil moisture	1
Root zone soil wetness	1
Wind speed at 2 meters maximum	m/s
Wind speed at 2 meters minimum	m/s
Precipitation corrected sum	mm

Note: A value 1 indicates a completely saturated soil and a value 0 indicates a completely water-free soil.

For each year, data for the *Rabi* season (November to March) were processed as follows: We calculated the average values for temperature at 2 meters, relative humidity at 2 meters surface soil wetness, profile soil moisture and root zone soil wetness; summarized the data for surface pressure and precipitation

corrected sum; and recorded the maximum and minimum values for temperature at 2 meters maximum, wind speed at 2 meters maximum, temperature at 2 meters minimum, wind speed at 2 meters minimum.

Analytical procedures

We conducted a comprehensive analysis to examine the long-term trends in potato yield and climate parameters. Initially, we conducted trend analysis to discern patterns over time. Next, Pearson's correlation coefficient analysis was employed to elucidate the relationship between climate parameters and annual potato yield. Following this, we implemented various machine learning algorithms, including multiple linear regression, Random Forest, and Support Vector Machine models. The dataset was partitioned into an 80% training set (spanning from 2006-07 to 2017-18) and a 20% testing set (encompassing 2018-19 to 2021-22). Subsequently, we assessed and compared the predictive performance of these models by analyzing metrics such as Root Mean Square Error (RMSE).

Multiple Linear Regression Model

Multiple linear regression is a statistical technique used to model the relationship between one dependent variable and two or more independent variables. It extends simple linear regression by allowing multiple predictors, thus providing a more comprehensive analysis of how various factors influence the dependent variable. The model estimates the coefficients for each independent variable, which represent their respective contributions to the dependent variable. This technique is useful for predicting outcomes and understanding the relative impact of different predictors. It relies on assumptions such as linearity, independence, homoscedasticity, and normality of residuals.

In this context, the dependent variable, Y, represents the yearly yield of potatoes in tons per hectare, which is influenced by a range of climate parameters (e.g., Surface pressure, Temperature at 2 meters,…, Precipitation corrected sum). The regression model estimates the coefficients (β_1, β_2,…,βn) corresponding to each independent variable (X_1,X_2,…,Xn), quantifying the extent to which each climate parameter affects potato yield. The intercept (β_0) represents the expected yield when all climate parameters are at zero, although in practical terms, this scenario may be hypothetical. The empirical model can be represented as:

$$Y = \beta_0+\beta_1X_1+\beta_2X_2+\ldots\ldots\ldots\ldots\ldots\ldots+\beta nXn+ \varepsilon$$

Where,

Y = Yearly yield of potato in tons per hectare

β_0 = Intercept of the model

X_1, X_2,…,X_n=Climate parameters (e.g., Surface pressure, Temperature at 2 meters,…,Precipitation corrected sum)

β_1, β_2,….., β_n = Coefficients representing the contribution of each climate parameter

ϵ = Random error term, accounting for the variability in yield not explained by the model

Random Forest Model

Random forest is a powerful machine learning algorithm that constructs multiple decision trees, each trained on random subsets of data and features like surface pressure, temperature at 2 meters, relative humidity, soil moisture, wind speed, and precipitation data from regions like Bogura, Munshiganj, Rangpur, Rajshahi, and Joypurhat. By aggregating the predictions from all trees, Random Forest provides a more accurate and generalized prediction, minimizing the risk of overfitting. For regression tasks like yield prediction, the final output is the average of the predictions from all the trees. Additionally, Random Forest offers feature importance scores, which help identify the most influential climate parameters affecting potato yield.

Support Vector Machine (SVM) Model

The Support Vector Machine (SVM) is a powerful supervised learning algorithm that can be effectively applied to predict potato yield in tons per hectare using climate parameters such as surface pressure, temperature at 2 meters, relative humidity, soil wetness, and wind speed across regions like Bogura, Munshiganj, Rangpur, Rajshahi, and Joypurhat. SVM works by finding the optimal hyperplane that minimizes prediction errors while preventing overfitting, making it particularly effective in high-dimensional spaces. By using kernel tricks, SVM can handle both linear and non-linear relationships between climate factors and potato yield, ensuring accurate and robust predictions even with complex data.

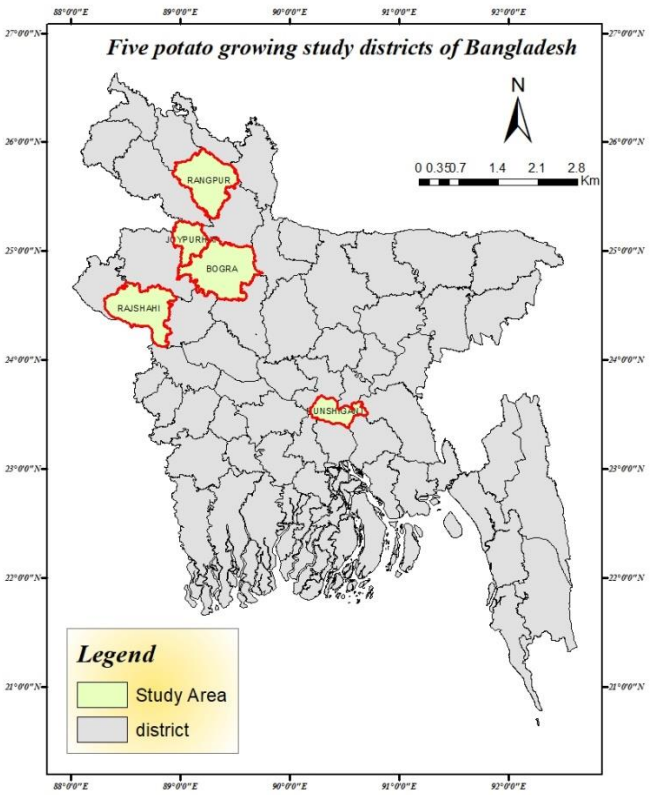


Figure 1. Map of Bangladesh showing the study areas

Results and Discussion

Long term trends and changes in climate variables

We have identified that the climate parameters: temperature at 2 meters minimum, root zone soil wetness, profile soil moisture, surface soil wetness, and wind speed at 2 meters maximum are significant (Table 2) for our selected districts. Therefore, we will create individual graphs for each significant climate variable, incorporating data from all five districts in each graph to identify long-term trends and changes in these variables.

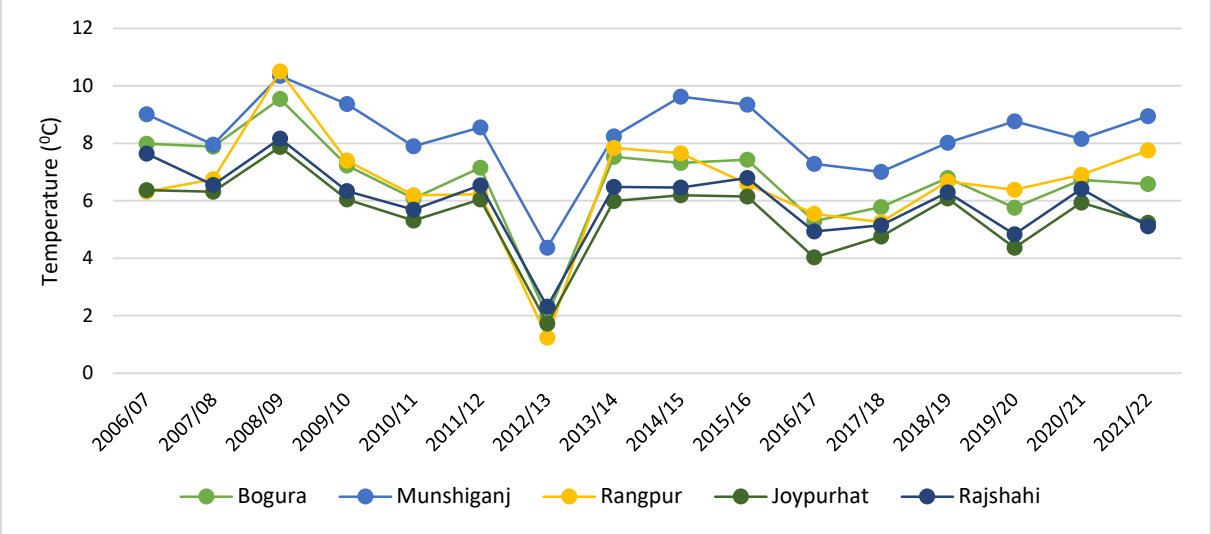


Fig 2. Trends of temperature at 2 meters minimum for the study areas

The above graph (Fig 2) indicates that the year 2012/13 is notable across all districts for having significant drops in minimum temperatures. Overall, each district shows variability in minimum temperatures without a clear increasing or decreasing trend over the years. We also find that Munshiganj stands out due to its consistently higher minimum temperatures, lower variability, and less severe temperature drops compared to the other districts. This indicates a more stable climate in terms of minimum temperatures for Munshiganj.

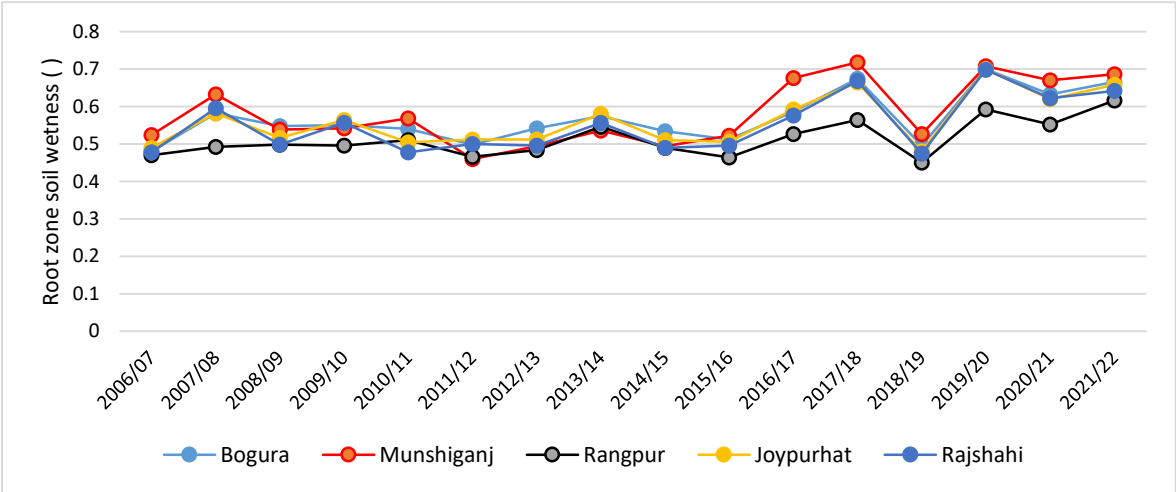


Fig 3. Trends of root zone soil wetness for the study areas

This table reflects the average soil wetness in the root zone for each location over the specified years. The values indicate the relative wetness, with higher values suggesting wetter conditions and lower values indicating drier conditions. 2019/20, with all districts recording high wetness values. 2011/12 and 2018/19, where multiple districts had some of their lowest wetness values (Fig 3).

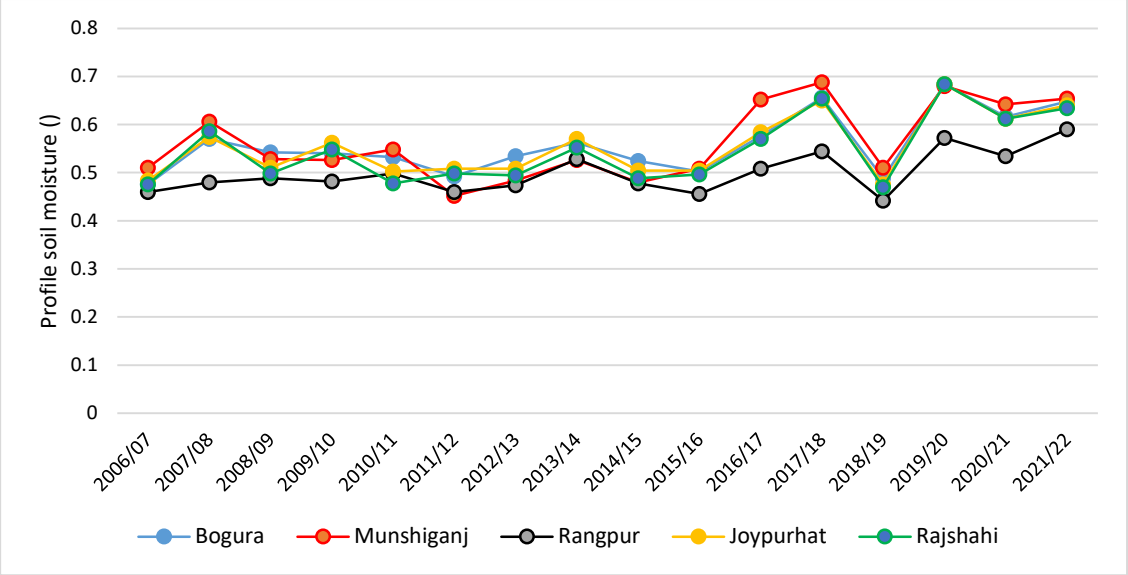


Fig 4. Trends of profile soil moisture for the study areas

Here the values indicate relative soil moisture 2019/20, with all districts recording high moisture values 2011/12 and 2018/19, where multiple districts had some of their lowest moisture values. Most districts show fluctuations in soil moisture with noticeable peaks in the later years of the dataset (Fig 4).

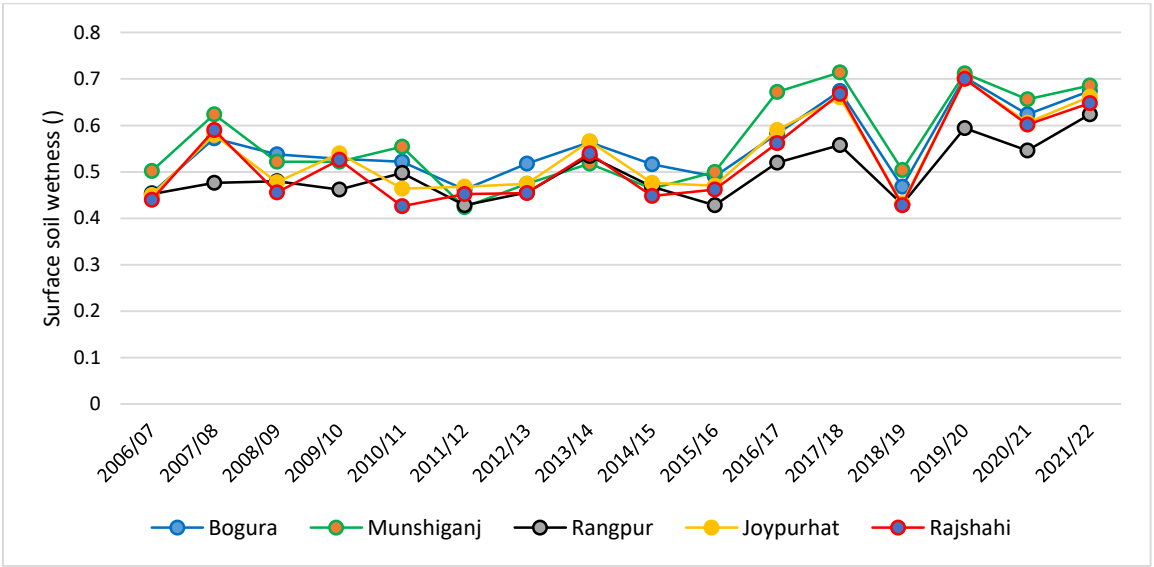


Fig 5. Trends of surface soil wetness for the study areas

Here is relative surface soil wetness 2019/20, with all districts recording high surface soil wetness values. 2011/12 and 2018/19, where multiple districts had some of their lowest surface soil wetness values (Fig 5).

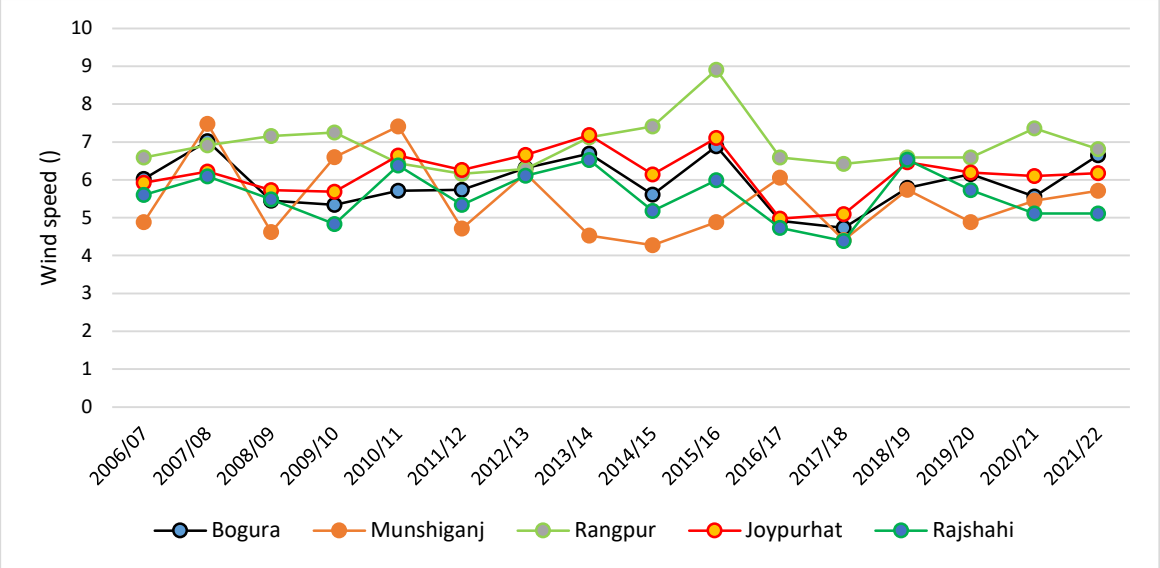


Fig 6. Trends of wind speed at 2 meters maximum for the study areas

The above graph (Fig 6) indicates that Bogura and Munshiganj exhibit more pronounced fluctuations in wind speed. Rangpur has relatively higher and more stable wind speeds with fewer drastic changes. Joypurhat shows moderate variability but remains relatively stable. Rajshahi has less variability and relatively stable wind speeds.

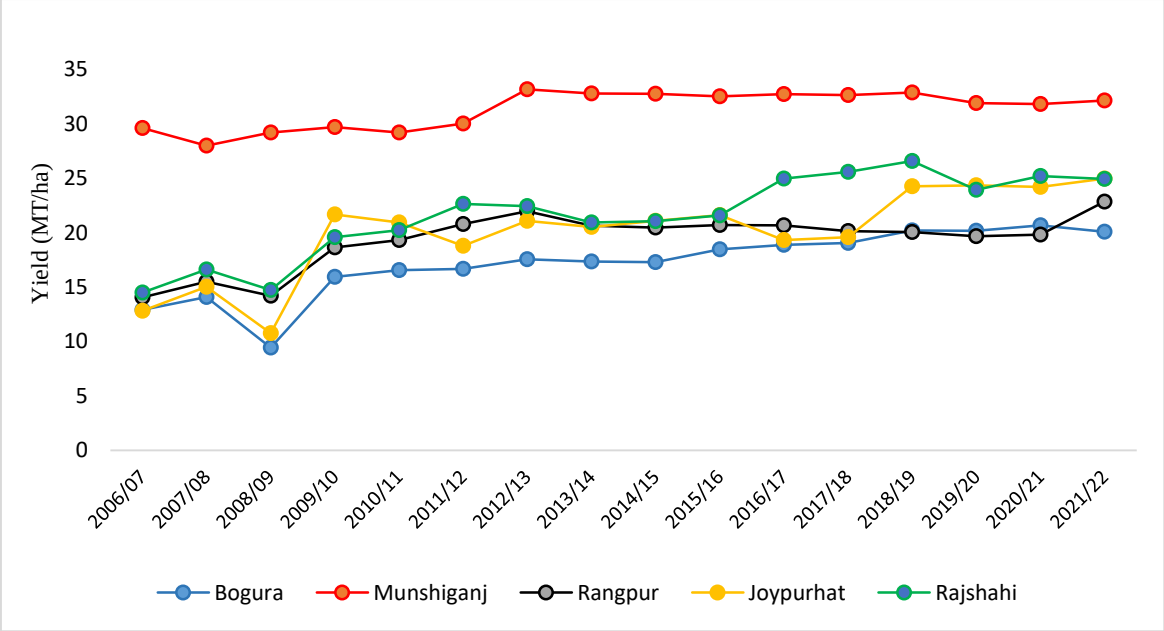


Fig 7. Comparative trend scenario of potato yields across the study areas

Overall, all five districts demonstrate an increasing trend in potato yield over the 16 years, with Munshiganj consistently yielding the highest and Joypurhat standing out as the most variable, with Bogura exhibiting moderate variability (Fig 7).

Climatic Factors Affecting Potato Yield

Bogura

For Bogura, we considered temperature at 2 meters minimum, root zone soil wetness, profile soil moisture, surface soil wetness, surface pressure, temperature at 2 meters, wind speed at 2 meters minimum, and relative humidity at 2 meters. We selected these 8 climate parameters because, when we developed a multiple linear regression model for Bogura, it resulted in an adjusted R^2 of 79.66% and ensured that the residuals were normally distributed. From our correlation analysis (as shown in Table 2), we have identified significant relationships between yield and several variables: temperature at 2 meters minimum, root zone soil wetness, profile soil moisture, and surface soil wetness, all at a 10% level of significance.

Upon assessment, we determined the following relationships:

- For every one-unit increase in temperature at 2 meters minimum, profile soil moisture, and surface soil wetness, the yield is expected to decrease by 1.78 units, 892.24 units, and 122.6 units, respectively, assuming all other factors remain constant.
- Conversely, for every one-unit increase in root zone soil wetness, the yield is expected to increase by 968.62 units, with all other factors held constant.

Also mention that, in the multiple regression model, temperature at 2 meters minimum, root zone soil wetness, profile soil moisture, and surface soil wetness contribute weights of 29.53%, 19.39%, 19.37%, and 1.48%, respectively.

Munshiganj

For Munshiganj, we considered wind speed at 2 meters maximum, surface pressure, temperature at 2 meters minimum, precipitation corrected sum, wind speed at 2 meters minimum, temperature at 2 meters, and relative humidity at 2 meters. We selected these 7 parameters because, when we developed a multiple linear regression model for Munshiganj, it resulted in an adjusted R^2 of 58.43% and ensured that the residuals were normally distributed. From correlation analysis, we find only a "wind speed at 2 meters maximum" significant relationship with yield at a 10% level of significance (Table 2).

Upon assessment, we determined the following relationships:

- For every one-unit increase in wind speed at 2 meters' maximum, the yield is expected to decrease by approximately 0.62 units assuming all other factors remain constant.

Also mention that, In the multiple regression model, wind speed at 2 meters' maximum contribute weights of 8.27%.

Rangpur

For Rangpur, we considered temperature at 2 meters minimum, root zone soil wetness, surface soil wetness and relative humidity at 2 meters. We selected these 4 parameters because, when we developed a multiple linear regression model for Rangpur, it resulted in an adjusted R^2 of 38.5% and ensured that the residuals were normally distributed. Based on our correlation analysis, we did not identify a significant relationship between the climate parameters and yield at a 10% level of significance (Table 2).

Joypurhat

For Joypurhat, we considered temperature at 2 meters minimum, profile soil moisture, surface soil wetness, wind speed at 2 meters maximum, temperature at 2 meters, and temperature at 2 meters maximum. We selected these 6 parameters because, when we developed a multiple linear regression model for Joypurhat it resulted in an adjusted R^2 of 38.24% and ensured that the residuals were normally

distributed. Based on our correlation analysis, we did not identify a significant relationship between the climate parameters and yield at a 10% level of significance (Table 2).

Rajshahi

For Rajshahi, we considered temperature at 2 meters minimum, root zone soil wetness, relative humidity at 2 meters, temperature at 2 meters, and profile soil moisture. We selected these 5 parameters because, when we developed a multiple linear regression model for Rajshahi, it resulted in an adjusted R^2 of 45.79% and ensured that the residuals were normally distributed. From correlation analysis, we find only a "temperature at 2 meters minimum" significant relationship with yield at a 10% level of significance (Table 2).

Upon assessment, we determined the following relationships:

- For every one-unit increase in temperature at 2 meters minimum, the yield is expected to decrease by approximately 2.53 units assuming all other factors remain constant.

Also mention that, In the multiple regression model, temperature at 2 meters minimum contributes weights of 32.7%.

Table 2. Correlation between yield and climate parameters at 10% level of significance

Rajshahi					Joypurhat					Rangpur					Munshiganj					Bogura				
Climate parameter	Yield	10%	Climate parameter	Yield	10%	Climate parameter	Yield	10%	Climate parameter	Yield	10%	Climate parameter	Yield	10%	Climate parameter	Yield	10%	Climate parameter	Yield	10%	Climate parameter	Yield	10%
T2M_MIN	0.02	Significant	T2M_MIN	0.11	NS	T2M_MIN	0.11	NS	T2M_MIN	0.11	NS	WS2M_MAX	0.09	Significant	T2M_MIN	0.03	Significant			
GWETROOT	0.11	NS	GWETPROF	0.15	NS	GWETROOT	0.21	NS	GWETPROF	0.21	NS	PS	0.11	NS	GWETROOT	0.07	Significant			
GWETPROF	0.11	NS	GWETROOT	0.16	NS	GWETPROF	0.21	NS	T2M_MIN	0.17	NS	PRECTOTCORR_SUM	0.24	NS	GWETPROF	0.09	Significant			
GWETTOP	0.13	NS	GWETTOP	0.19	NS	WS2M_MIN	0.26	NS	PRECTOTCORR_SUM	0.24	NS	WS2M_MIN	0.38	NS	GWETTOP	0.09	Significant			
PRECTOTCORR_SUM	0.27	NS	WS2M_MAX	0.31	NS	GWETTOP	0.27	NS	WS2M_MAX	0.38	NS	T2M_MAX	0.63	NS	PRECTOTCORR_SUM	0.39	NS			
RH2M	0.36	NS	PS	0.41	NS	PS	0.29	NS	T2M_MAX	0.63	NS	RH2M	0.41	NS						

NS	NS	NS	NS	NS
0.45	0.54	0.67	0.77	0.82
WS2M_MAX	WS2M_MIN	T2M	PS	T2M_MAX
NS	NS	NS	NS	NS
0.48	0.61	0.70	0.85	0.91
RH2M	T2M	WS2M_MIN	T2M_MAX	PRECTOTC
NS	NS	NS	NS	NS
0.56	0.78	0.82	0.96	0.97
T2M_MAX	RH2M	T2M	PRECTOTCO	WS2M_MAX
NS	NS	NS	NS	NS
0.63	0.64	0.65	0.88	0.96
GWETPROF	GWETROOT	GWETTOP	T2M	RH2M
NS	NS	NS	NS	NS
0.45	0.64	0.73	0.81	0.95
PS	T2M_MAX	T2M	WS2M_MIN	WS2M_MAX

Identifying best predictive model for forecasting in selected district

Root Mean Square Error(RMSE) measures the average magnitude of the prediction errors, with a lower RMSE indicating better model performance. It significantly penalizes larger errors because of the squaring of differences. For each district, we determined which model has the lowest RMSE.

Table 3. Model comparison based on Root Mean Square Error (RMSE)

District	Linear Regression	Random Forest (RF)	Support Vector Machine (SVM)	Lowest Root Mean Square Error (RMSE)
Bogura	3.95	2.72	1.30	SVM
Munshiganj	1.62	1.70	1.33	SVM
Rangpur	2.90	**1.54**	2.03	**RF**
Joypurhat	2.53	2.82	1.42	SVM
Rajshahi	3.42	4.00	1.74	SVM

Table 3 shows that the support vector machine model is better suited for predicting the climate-related outcomes in Bogura, Munshiganj, Joypurhat, and Rajshahi as it results in lower prediction errors. Conversely, the Random Forest model is more appropriate for the current predictive task in the Rangpur district. This is because the nonlinear machine learning techniques can capture the heterogeneous trend in the data set, leading to better performance compared to linear regression models.

Conclusions

The presented study aimed to investigate the relationship between climate variables and potato yield, and to develop predictive models to estimate the impact of climate change on future potato yield. From our research, we have determined that the climate parameters such as minimum temperature at 2 meters, root zone soil wetness, profile soil moisture, surface soil wetness, and maximum wind speed at 2 meters are

significant at the 10% level of significance for our selected districts. We also found that the best predictive model for Bogura, Munshiganj, Joypurhat, and Rajshahi is the Support Vector Machine (SVM) model. In contrast, for Rangpur, the Random Forest (RF) model performs best, based on the root mean square error (RMSE). From our analysis, potato yield in the districts shows resilience to varying climate conditions, with stable or increasing yields despite fluctuations in temperature, soil wetness, soil moisture, and wind speed. This resilience suggests that the existing potato varieties and agricultural practices are effective in mitigating the adverse effects of climate change. However, the significance of certain climate parameters indicates that ongoing monitoring and adaptive management are necessary to safeguard future production. To further enhance resilience, it is recommended that adaptation strategies focus on the development of more climate-tolerant potato varieties and the optimization of farming practices. Future research should explore the long-term effects of climate change on potato production and the potential for integrating advanced technologies, such as precision agriculture, to support adaptation efforts.

References

Abbasnia, M.; T. Tavousi and M. Khosravi (2016). Assessment of future changes in the maximum temperature at selected stations in Iran based on HADCM3 and CGCM3 models. *Asia Pac J Atmos Sci*, 52(4): 371–377.

Abbasnia, M. and H. Toros (2018a). Analysis of long-term changes in extreme climatic indices: a case study of the Mediterranean climate, Marmara Region, Turkey. *Pure Appl Geophys*, 175: 3861-3873. https://doi.org/10.1007/s0002 4-018-1888-8.

Abbasnia, M. and H. Toros (2018b). Monitoring observed changes in warm day's extremes over Turkey. *Nat Resour Conserv Res*, 1:1-4. https://doi.org/10.24294 /nrcr.v1i2.869.

Beals, K. A. (2019). Potatoes, nutrition and health. *American journal of potato research*, *96*(2): 102-110.

Chakraborty, R. and T. S. Roy (2016). Threats faced by brown rot of potato in Bangladesh. *Microbiology Research*, 7(1): 6258.

Chen, C.; I. Noble; J. Hellmann; J. Coffee; M. Murillo and N. Chawla (2015). Country ND-GAIN Index, Technical report. University of Notre Dame global adaptation index country index. https://gain.nd.edu/our-work/country-index/rankings/

Dastagir, M. R. (2015). Modeling recent climate change induced extreme events in Bangladesh: A review. *Weather and Climate Extremes*, 7: 49-60.

Eckstein, D.; V. Künzel and L. Schäfer (2021). *The global climate risk index 2021*. Bonn: German watch.

FAO STAT (2022). Food and Agriculture Data. Food and Agriculture Organization (FAO), Rome, Italy.

GAIN index summarizes a country's vulnerability to climate change and other global challenges in combination with readiness to improve resilience. https://gain.nd.edu/our-work/country-index/rankings /

Górska-Warsewicz, H.; K. Rejman; J. Kaczorowska and W. Laskowski (2021). Vegetables, potatoes, and their products as sources of energy and nutrients to the average diet in Poland. *International Journal of Environmental Research and Public Health*, 18(6): 3217.

Hansen, J. I.; I. Fung; A. Lacis; D. Rind; S. Lebedeff; et al. (1988). Global climate changes as forecast by Goddard Institute for Space Studies three-dimensional model. *J Geophys Res*, 93: 9341-9364. doi:10.1029/88JD00231.

Haverkort AJ. 1990. Ecology of potato cropping systems in relation to latitude and altitude. *Agr Syst* 32: 251-272.

Hijmans R. J.; B. Condori; R. Carillo and M. J. Kropff (2003). A quantitative and constraint-specific method to assess the potential impact of new agricultural technology: the ease of frost resistant potato for the Altiplano (Peru and Bolivia). *Agr Syst*. 76: 8g5-g11.

IPCC (2007). Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, 32 Avenue of the Americas, New York, NY10013-2473, USA. https://www.ipcc.ch/site/assets/uploads/2018/05/ar4_wg1_full_report-1.pdf.

Islam, A. K. M. R.; M. H. Ali and M.G.M. Amin (2004). Long-term variability of rainfall at different agro-ecological regions of Bangladesh. *J. Soc. Agric. Sci. Tech.*, 1(1&2):19–24.

Jannat, A.; Y. Ishikawa-Ishiwata and J. Furuya (2021). Assessing the impacts of climate variations on the potato production in Bangladesh: A supply and demand model approach. *Sustainability*, 13(9): 5011.

Jennings, S. A.; A. J. Challinor; P. Smith; J.I. Macdiarmid; E. Pope; S. Chapman, ... and T. Benton, (2022). A new integrated assessment framework for climate-smart nutrition security in sub-Saharan Africa: the integrated Future Estimator for Emissions and Diets (iFEED). *Frontiers in Sustainable Food Systems*, 6: 868189.

Kerr, R. A. (2009). What happened to global warming? Scientists say just wait a bit. *Science* 326 (5949): 28-29. https://doi.org/10.1126/science.326_28a.

Klein Tank, A. M. G.; T. C. Peterson; D. A. Quadir; S. Dorji; X. Zou; H. Tang; K. Santhosh; U. R. Joshi; A. K. Jaswal; et al. (2006). Changes in daily temperature and precipitation extremes in central and south Asia. *J. Geophys. Res.*, 111, D16105, doi: 10.1029/2005JD006316.

McCarthy, J. J. (Ed.). (2001). Climate change 2001: impacts, adaptation, and vulnerability: contribution of Working Group II to the third assessment report of the Intergovernmental Panel on Climate Change (Vol. 2). Cambridge University Press.

Mearns, L. O.; R. W. Katz and S. H. Schneider (1984). Extreme high-temperature events: changes in their probabilities with changes in mean temperature. *J ClimApplMeteorol* 23:1601-1608.

Nishat, A. and Mukherjee, N. 2013. Climate change impacts, scenario, and vulnerability of Bangladesh, In R. Shaw et al. (eds), *Climate Change Adaptation Actions in Bangladesh*, doi: 10.1007/978-4-431-54249-0_2.

Noguer, M.; P.J. van der Linden; X. Dai; K. Maskell and C. A. Johnson (2001). *Climate change 2001: the scientific basis* (Vol. 881, No. 9). J. T. Houghton, Y. D. J. G. Ding, & D. J. Griggs (Eds.). Cambridge: Cambridge University Press.

Rahman, A.; M. A. Mojid and S. Banu (2018). Climate change impact assessment on three major crops in the north-central region of Bangladesh using DSSAT. *International Journal of Agricultural and Biological Engineering*, 11(4): 135-143.

Raymundo, R.; S. Asseng; R. Robertson; A. Petsakos; G. Hoogenboom; R. Quiroz and J. Wolf (2018). Climate change impact on global potato production. *European Journal of Agronomy*, 100: 87-98.

Reynolds, M. P. and E. E. Ewing (989a). Effects of high air and soil temperature stress on growth and tuberization in *Solanum tuberosum*. *Ann Bot* 64(3): 241-247.

Sacks, W. J. and C. J. Kucharik (2011). Crop management and Phenology trends in the U.S. corn belt: Impacts on yields. *Agric. For. Meteorol.*, 151: 882–894.

Salan, M. S. A.; M. M. Hossain; I. H. Sumon; M. M. Rahman; M. A. Kabir and A. K. Majumder (2022). Measuring the impact of climate change on potato production in Bangladesh using Bayesian Hierarchical Spatial-temporal modeling. *Plos One*, 17(11): e0277933.

Shortridge, J. (2019). Observed trends in daily rainfall variability result in more severe climate change impacts on agriculture. *Climatic Change*, 157(3): 429-444.

Stol, W.; G.H.J. de Koning; A.J. Haverkort P.L. Kooman; H. van Keulen and F.W.T. Penning de Vries (1991). Agro-ecological characterization for potato production. A simulation study at the request of the International Potato Center (CIP), Lima, Peru. CABO-DLO, Report 155.

Toros, H. (2012). Spatio-temporal variation of daily extreme temperatures over Turkey. *Int J Climatol* 32(7):1047–1055.

Toros, H.; M. Mokari and M. Abbasnia (2019). Regional variability of temperature extremes in the maritime climate of Turkey: a case study to develop agricultural adaption strategies under climate change, *Modeling Earth Systems and Environment*. Doi.org/10.1007/s40808-019-00572-4.

GGE BIPLOT ANALYSIS FOR YIELD PERFORMANCE AND STABILITY ASSESMENT OF BARI RELEASED BT BRINJAL VARIETIES

MD. SHAKIL HOSSAIN, ISTIAK AHMED, TASLIMA ZAHAN, ZOBAER AKOND AND M.A. MONAYEM MIAH

Abstract

The stability and high yield of Bt brinjal variety is an important factor for long-term development and food security. The study investigated the effects of genotype (G) and environment (E) interaction on yield stability in four Bt brinjal varieties in seven different environments of Bangladesh. The study considered G and E as treatments, year as replication, and used a randomized complete block design (RCBD) with to construct the genotype plus genotype-vs-environment interaction (GGE) model. The joint analysis of variance revealed significant differences among the genotypes and environments (GE). The GGE biplot graphically showed the interrelationships between the tested environment and genotypes. The scores of PC1 (principal component 1) and PC2 (principal component 2) cumulatively explained approximately 88.09% of the total variation in GE interactions and were used to construct the GGE biplot. Bt brinjal-4 was the best genotype, with high average yields and high stability across the multi-locations. Sherpur and Satkhira were identified as the desired locations among the tested environments for growing all the genotypes. This study will help Bt brinjal growers select highly stable and high-performance varieties for a particular environment to achieve maximum production.

Keywords: Stability, Bt brinjal, GGE biplot, Multi-location, Principal component analysis

Introduction

Brinjal (*Solanum melongena*) is one of the most important and popular vegetables in Bangladesh that grown widely round the year. The crop is damaged severely by the notorious insect called brinjal shoot and fruit borer (BSFB) and the damage ranges from 30-70% depending upon the locality and edaphic conditions. To address this problem, farmers in major brinjal-producing areas in the country spray chemical insecticides every other day, 60-180 times per growing season (Shelton et al., 2020). The practice is unacceptable and unhealthy to consumers, farmers, and the environment. Bt brinjal is a transgenic brinjal developed by inserting a crystal protein gene (*Cry1Ac*) from the soil bacterium, *Bacillus thuringiensis* into various brinjal cultivars and these plants are found to be resistant against BSFB (Shelton et al., 2019).

Researchers often use yield and its contributed performance and a phenotypic expression for sorting and selection of crop cultivars under mega environment tests. Pathogenic infections, humidity, soil texture, and fertility, precipitation, and temperature may all play a role in the yield fluctuation caused by genotypes' responses to changing environments (Oladosu et al., 2016). This yield instability or fluctuation is named genotype by environment interaction (GEI) reported in different crops (De Vita et al., 2010; Karimizadeh et al., 2013). To determine the presence of GEI in a multi-environmental yield trial, an analysis of variance (ANOVA) is executed. The ANOVA procedures highlighted differentiation in fixed and random effects such as genotype, replication, and environment. Nevertheless, the major bottleneck of ANOVA is the failure to distinguish genotype variances in a non-additive manner as G × E interaction. A group of stability statistical measures was settled to analyze genotype stability that divulges several G × E interaction features, resulting in detecting stable genotypes across environments. To get a better understanding of genotypic stability patterns, two separate techniques such as univariate and multivariate of stability statistical analysis are used (Oladosu et al., 2017). Multivariate methods like as cluster analysis, pattern analysis, and principal component analysis or biplots are effective tools for uncovering patterns of GE interactions (Myint et al., 2019). The pattern analysis (combination of classification and ordination technique) will be efficient in uncovering and elucidating the GEI structure of intrinsic data under examination. Biplots are widely used to graphically show the interrelationships between genotypes (G), environments (E), and GEIs, as well as to demonstrate interaction patterns and identify comparably stable genotypes across environments (Oladosu et al., 2016).

Principal component analysis (PCA), also known as biplots, is a popular graphical representation of interaction outlines used to reveal inter-relationships between genotypes, environments, and GEI to identify genotypes that are well suited to specific environments or genotypes that are stable across environments. The two-way measures such as GEI data to singular value decomposition (SVD) and its graphical display are subjected to generate the biplots. A widely used biplot model is GGE biplot = genotype + genotype × environment. The GGE biplots are based on the singular value decomposition (SVD) of environment-centered provided by Yan et al. (2006) and graphically represent both genotype and genotype-by-environment based on primary sources of variation associated with genotype assessment. GGE biplots-based multi-environment trial (MET), genotype evaluation environmental valuation has been successfully executed for varietal stability analysis experimented by several researchers such as Mohammadi et al. (2009) and Kumar et al. (2012). The GGE biplot approach for decomposing genotype plus genotype-by-environment (G + G × E) is censured by Gauch et al. (2008). GGE biplot analysis has been used by several researchers to classify the mega environment, assess genotype rankings, and decide the discriminative and representative among the tested environments (Koutis et al., 2012). It is important to understand how GE interacts with cultivars to determine adaptation and stability (Yan and Tinker, 2006). The GGE biplot will aid researchers in better understanding complicated GE interactions in multi-environment breeding line trials and agronomic investigations (Yan and Kang, 2002). The GGE biplot was used to determine the performance of crop cultivars in a variety of stress conditions, ideal cultivars, mega-environments, and core testing sites. GGE biplots analysis is regarded as a useful statistical technique for producing phenotypically stable and superior cultivars, identifying stable genotypes across several environments, and achieving crop yield stability across multiple locations. As a consequence, the current study aims to identify superior genotypes with stable yield performance over a wide range of environments by evaluating the efficacy of various stability analysis methodologies. Another intention of this study was to examine how GEI influenced the yield and yield components of Bt brinjal varieties.

Limited research has focused on the stability and adaptability of BARI-released Bt brinjal varieties in Bangladesh, despite various studies conducted on Bt brinjal. Therefore, this study aims to fill this knowledge gap by assessing the performance and stability of BARI-released Bt brinjal varieties across diverse locations in Bangladesh. This study presents an important contribution to the understanding of the stability and adaptability of BARI-released Bt brinjal varieties across diverse locations in Bangladesh.

Materials and Methods

Experimental sites

The scientists of OFRD of BARI carried out some experiments on the farmers' fields of Bhola, Bogura, Kishoregonj, Manikgonj, Norsingdhi, Satkhira and Sherpur districts during 2019-2022 (Fig.1). These districts span a variety of agro-ecological zones, providing a comprehensive evaluation of the yield performance, pest resistance, and stability of Bt brinjal under different environmental conditions. The optimum temperature for its better growth is 15–21 °C. High to medium-high land is good for Bt brinjal. It can be grown on all types of soils from light to heavy soil. Higher yield may be obtained when soil becomes clay loam or silt. It does not grow well on highly acidic soil. The plants grow well when soil pH ranges from 5.5 to 6.5 (Crop production technology, BARC).

1. **Bhola**: Situated in the coastal zone, Bhola is influenced by tidal fluctuations, leading to periodic flooding and soil salinity issues. The district experiences high humidity and significant monsoonal rainfall, which can impact crop growth. This region is critical for evaluating Bt brinjal's performance under saline and waterlogged conditions.
2. **Bogura**: Located in the northern part of Bangladesh, Bogura has fertile alluvial soils from the Jamuna River and a moderate climate, making it a key agricultural hub (Karim et al., 2015). The district's agro-environment is ideal for assessing Bt brinjal under more traditional, less stressed growing conditions.
3. **Kishoregonj**: As part of the Haor region, Kishoregonj is prone to seasonal flooding, creating challenges for cultivation during the monsoon season. The waterlogged conditions make it a

significant location for testing Bt brinjal's flood tolerance and fruit quality under high water retention scenarios.

4. **Manikgonj**: This district is located near Dhaka and is characterized by a subtropical monsoon climate. The combination of fertile soils and adequate rainfall creates favorable conditions for agricultural productivity, making Manikgonj an appropriate location for testing yield consistency and pest resistance in Bt brinjal (Bangladesh Bureau of Statistics, 2021).
5. **Norsingdhi**: With loamy soils and moderate rainfall, Norsingdhi presents an ideal controlled environment for assessing the performance of Bt brinjal. The region's stable agricultural conditions allow for consistent monitoring of yield performance and resistance to the brinjal fruit and shoot borer (Ahmed et al., 2020).
6. **Satkhira**: Located in the southwestern coastal belt, Satkhira is affected by saline water intrusion from the Bay of Bengal. This district is particularly vulnerable to salinity and soil degradation, making it crucial for evaluating Bt brinjal's tolerance to saline conditions (Roy et al., 2011). The crop's ability to thrive in such environments is critical for sustainable cultivation in coastal areas.
7. **Sherpur**: Sherpur, in the northern part of Bangladesh, has relatively cooler temperatures during winter and fertile soils, making it an important district for assessing Bt brinjal's cold tolerance and performance during different seasonal phases. The district's agricultural landscape provides a unique opportunity to test adaptability under variable climatic conditions.

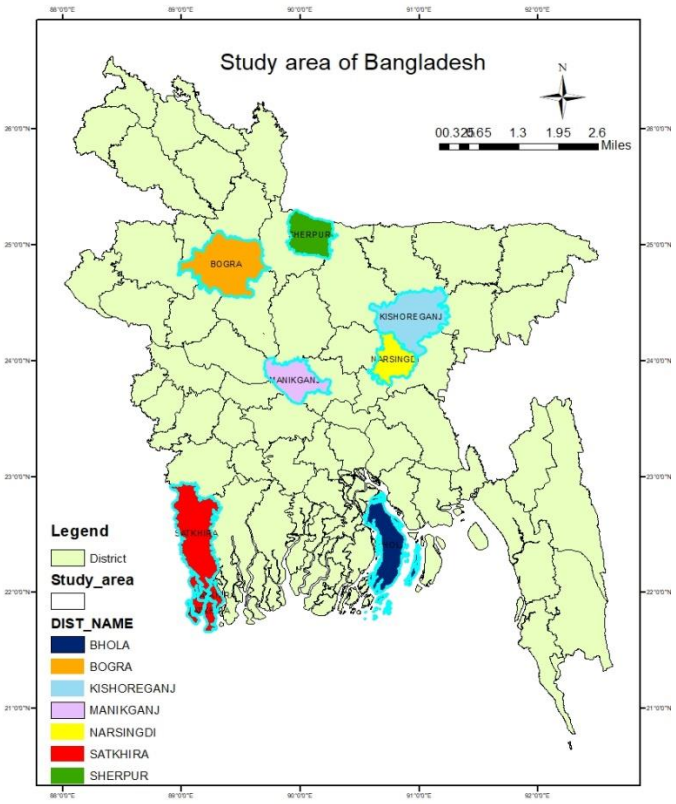


Figure 1: The map of Bangladesh showing study areas

Details of the experimental management

The showing time varied between the 15th of November and the 5th of December. Unit plot size was 400 m^2 with spacing of 100 cm × 80 cm. Stable bleaching powder was applied 20 days before transplanting @ 21 kg ha^{-1} for preventive measures against bacterial wilt. The fertilizers were used @ 138-40-100-18-1.7-3.6 kg ha^{-1} N-P-K-S-B-Zn and Cow dung 10 tha^{-1}. One-third of MP and the rest fertilizers except urea were applied during final land preparation. The remaining two-thirds of MP was divided into three splits

and applied at 20 DAP, at the flowering and fruiting stage. Urea was applied in four equal installments at 20 DAP, at the flowering, fruiting, and continuous fruiting stage. Plant protection measures for other pests were taken in some severely infested plots; Furadan was applied @ 33 kg ha^{-1} during transplanting, and Imidacloprid was sprayed @ 0.5 m L^{-1} water to control sucking pests. Similarly, Bavistin @ 2ml L^{-1} water was sprayed in 2-3 times to control fusarium wilt and Tilt @ 0.5ml L^{-1} was sprayed in 1-2 times to control phomopsis blight-infested plots. Epilachna beetle was controlled by destroying egg mass and larva, pupa & adult by hand picking. Other intercultural operations were done as and when necessary. The fruits were harvested between the months of February and May each year. from 7 February to 24 May. Ten plants were selected randomly from each plot to record plant height, individual fruit weight, length & breadth of fruit, and no. of total & infested fruit $plant^{-1}$. Yield data were calculated based on total plots.

Statistical analysis

The quantitative traits were subjected to analysis of variance (ANOVA) to estimate the existence of variations among the genotypes, locations, and genotype by location using the 'doe-bioresearch' package of RStudio. GGE biplot was created by using the 'metan' package of RStudio, which is an integrated development environment for R software. To explain the G × E interaction, the multivariate stability analysis was performed graphically based on the GGE biplot. The biplots were based on singular-value partitioning = 2, transformed (transform = 0), environment-centered (centering = 2), and standard deviation-standardized (scaling = 0).

Genotype and environmental modeling

The GGE biplot provides a graphical representation of variety-by-environment interactions, allowing for the identification of stable varieties and discriminating environments. The GGE model was used to evaluate genotype and mega-environment identification in multi-environment trials (METs). This model considers a GGE biplot, which was constructed by the first two symmetrically scaled principal components (PC1 and PC2) derived from singular value decomposition of environment-centered MET data. The GGE biplot graphically displays the G plus GE of a MET in a way that facilitates visual genotype evaluation and mega-environment identification (Yan et al., 2007). The mean yield of genotype i in environment j was described as follows:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varphi_{ij} \qquad (i)$$

where y_{ij} is the mean yield of genotype *i* in environment *j*, *i = 1, ... g; j = 1, ... e* are *g* and *e* are the numbers of genotypes and environments, respectively; μ is the grand mean; αi is the main effect of genotype *i*; βj is the main effect of environment *j*; and φ_{ij} is the interaction effect between genotype *i* and environment *j*. Subjecting $\varphi_{ij}$to singular value decomposition (SVD) results in the GGE model. The deletion of α allows the variation explained by this term to be absorbed into the φ_{ij}term. In the genotype plus genotype-vs-environment interaction (GGE) model, the α term was deleted from the above model, and then the environment-centered data matrix, φ_{ij}, was subjected to SVD (Yan and Kang, 2002). Specifically, we have

$$\varphi_{ij} = y_{ij} - \mu - \beta_j = \sum_{k=1}^{p} \xi_{ik}^{*} \eta_{jk}^{*} \qquad (ii)$$

Where, $\xi_{ik}^{*} = \lambda_k^{\alpha} \xi_{ik}^{*}$; $\eta_{jk}^{*} = \lambda_k^{1-\alpha} \eta_{jk} \lambda_k$ is the *kth* eigenvalue from the SVD *(k = 1 ... p)*, with p ≤ min (eg); α is the singular value partition factor for the principal component (PC) *k*; and ξ_{ik}^{*} and η_{jk}^{*} are the PC scores for genotype *i* and environment *j*, respectively.

Results and Discussion

The yields of brinjal

The exploratory data analysis revealed that the yields of Bt brinjals varied greatly from 18.26 to 41.32 t ha^{-1} depending on the variety (Fig.2.) and location. As all the data were from the farmers' fields under the supervision of the scientists of OFRD, BARI, the mean yield was found to be 28.56 t $ha^{-1,}$ which was above the country's mean yield of 21.86 t ha^{-1} (BBS, 2022). The analysis also revealed that the overall yield was approximately normally distributed (Fig. 3). The location, genotype, and year-wise box plots also supported the normality assumption. Although there were some outliers, expert opinion suggested that these were not outliers but rather extreme observations. Therefore, these extreme observations were retained in the statistical analysis.

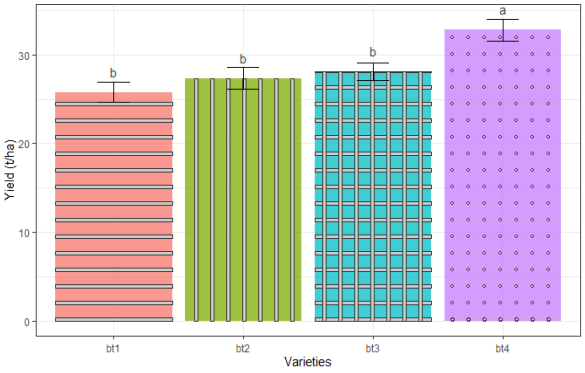


Fig.2. Variation in brinjal yields

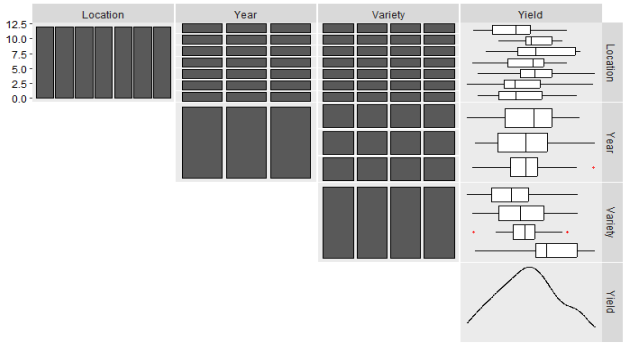


Fig.3. Distribution of yields approximately normal

Interaction effect of brinjal yields

In order to describe the main effect and quantify the interactions among and within the source of variations combined analysis of variance (ANOVA) was performed. The ANOVA is displayed in Table 1. The mean square of locations, genotypes, and genotypes by locations showed significant differences ($p \leq 0.01$, $p \leq 0.05$) for yield per hectare. Highly significant differences in locations and genotypes may be attributed to changes in environmental conditions and genetic makeup that differ from one environment to the next. A greater difference among locations for genotype means resulting in most of the variation presence in genotype performance. Oppositely lower variation was found in genotype by location. In this current research, a significant level divergence among G × E interaction and genotype effect indicated the certainty of the presence of diverse multi-environments with different genotype as well as high yield potential. Nevertheless, the variance component analysis is not enough to clarify the details of the genotype by environment interaction. Henceforth, additional statistical techniques such as multivariate analysis can be more fruitful in unfolding and understanding the GEI. The genotype-by-environment interaction effect primarily highlights the fact that genotypes responded inversely to various locations, emphasizing the need of genotype assessment in diverse environments.

Table 1. **Analysis of variance (**ANOVA) across environments

Source of variance	df	Sum Sq.	Mean Sq.	F-value	P-value
ENV	6	219.39	36.57	1.5421	0.0236

				0.8665	
REP(ENV)	14	331.94	23.71		0.5979
GEN	3	574.91	191.64	7.0038	0.0006
GEN:ENV	18	544.81	30.27	1.1062	0.0098
Residuals	42	1149.19	27.36	--	--
Total	101	3365.05	33.32	--	--

Yield variability in genotype across the environments

The mean performance results of genotype across the locations over the years showed that Bt brinjal-4 exhibited high yields across all locations except Bhola (Fig. 4). Bt brinjal-1 had a low yield in the Bhola and Satkhira and Sherpur districts. Bt brinjal-2 had a poor yield in the Satkhira district. Bt brinjal-3 exhibited moderate yields in Bogura and Manikgonj locations. High yields of Bt brinja-3 and Bt brinjal-4 were detected throughout the study period.

Fig.4 Mean performance of genotype across locations

Effects of genotype plus genotype-environment (GGEs) across environments

The GGE model was performed using standard deviation as a scaling parameter, global centering, and symmetrical singular value partitioning. In standard deviation scaling parameter, each value is divided by the standard deviation of its corresponding environment (column). Global centering means the inclusion of the combine effect (E + G + GE) in the model and symmetrical singular value partitioning means the singular value is symmetrically partitioned into the genotype and the environment eigenvectors. The GGE biplot based on PC1 and PC2 scores (Fig. 5) revealed that the type of variation was due to genotype and that the GEI was present in the multi-location trial (MLT) data. The biplot explained 88.09% of the total variation that was due to GGE (Fig. 5). The distance of the genotypes from the biplot origin indicates yield performance across the tested locations. Genotypes presented above-average yielders on the right side of the biplot origin, i.e., the high-yielding genotype, Bt brinjal-4, and the remaining genotypes had moderate yields (Fig. 5).

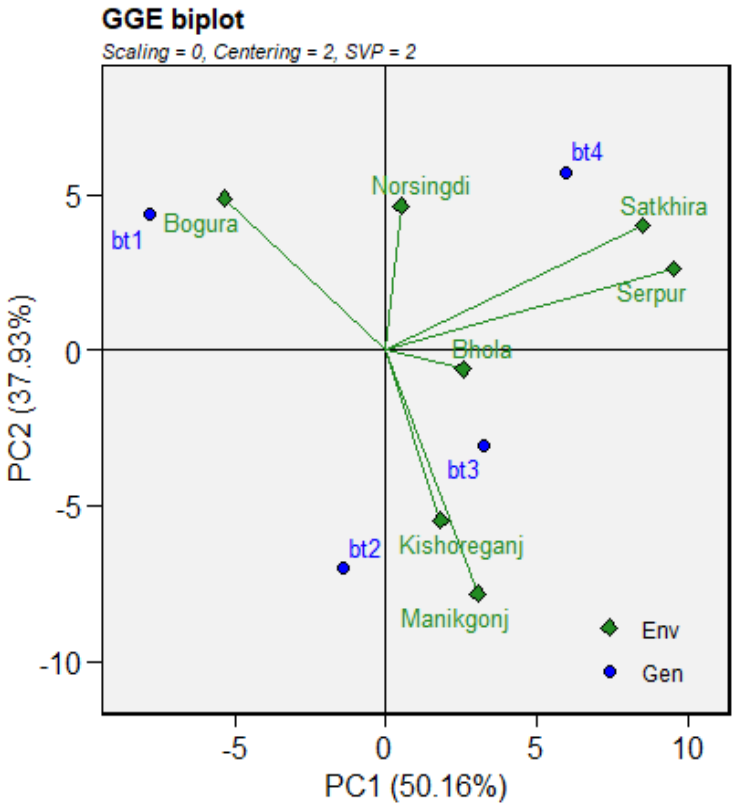


Fig.5 GGE biplot of genotypes across location

Which-won-where pattern

The polygon view of the GGE biplot graphic analysis is presented (Figure 6) for the identification of winning genotypes by visualizing the interaction patterns between genotypes and environments. It helps identify crossover and non-crossover genotype-by-environment interactions as well as the possible existence of different mega-environments in multi-location yield trials (Yan and Tinker, 2006). As displayed in (Figure 6) genotypes, Bt1, Bt2, and Bt4 were the vertex genotypes. These genotypes perform best or worse in some or all environments because they are the furthest from the biplot's commencement, and they are regarded as specifically suited genotypes since they are more responsive to environmental change. They thrive in environments that are part of their specific sector in the GGE's polygon view-biplot. In Bogura Bt-1 was the most successful genotype. In Kishoregonj Bt 2 is the winning genotype. Bt brinjal-3 is the winning genotype in Bhola. Bt brinjal-4 is the winning genotype in Sherpur and Satkhira. Similar results were reported by Oladosu et al. (2016), and Akhter et al. (2015), who characterized genotypes' which-won-where patterns. They found that some genotypes performed better in a specific environment than others and that some genotypes performed moderate in some environments.

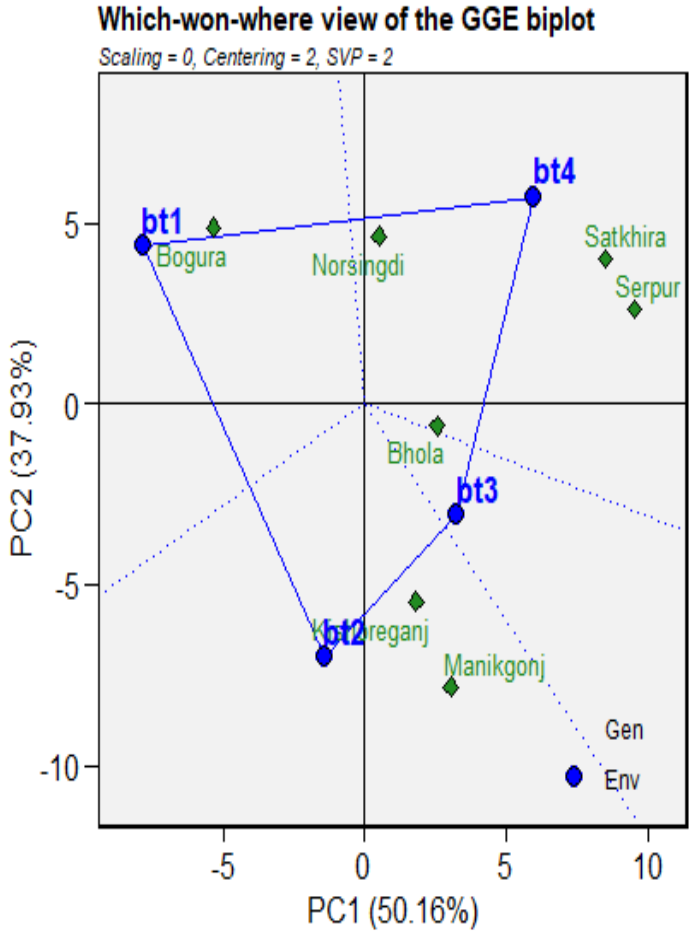


Fig.6 The polygon view displaying the genotype main effect plus G × E interaction effect across location

Evaluation of Environments Relative to Ideal Environments

According to Yan and Tinker (Yan and Tinker., 2006), an ideal environment has the highest discriminating capability and representativeness, which are important properties of a test environment. Yan and Kang (2002) defined an ideal environment as one that is highly differentiating for the tested genotypes while at the same time representative of the target environments. In this regard, Sherpur had a smaller angle with the average environment axis (AEA), while Kishoregonj, Norsingdi, and Bogura had a large angle with the average environment axis (Figure 7). Satkhira and Bhola were close to the center with very short vectors (Figure 7) and provided less helpful discriminating information about the genotypes. As a result, Sherpur was the most representative environment, whereas Kishoregonj and Norsingdhi were the least representative. Also, Sherpur had the highest discriminating capability of the genotypes (Figure 7). As a result, Sherpur is the most favorable environment for the selection of superior genotypes.

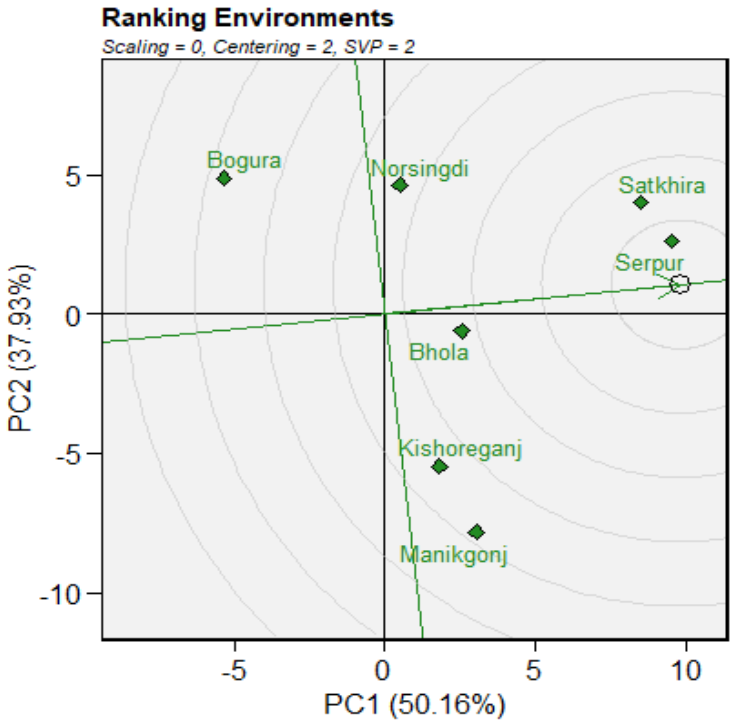


Fig.7 Ranking of genotypes tested across different locations

In the environment-focused GGE biplot, the ideal environment is positioned in the first concentric circle, and the desired environments are defined as those that are closest to the ideal environment. In this regard, Sherpur is in the first concentric circle and has been in the ideal environment (Figure 7).

Discussion

Our utilization of multi locational trials, a cornerstone of breeding programs, not only generated valuable data for assessing GEI but also facilitated the comprehensive evaluation of Bt brinjal varieties under various conditions. The joint analysis of variance revealed highly significant GEIs, both in terms of crossover and non-crossover interactions. This emphasizes the need to consider the performance of Bt-brinjal genotypes across multiple environments, whether for the development of new varieties or for the selection of suitable varieties for targeted environments. This finding aligns with similar observations in various crops, including wheat, rice, garlic, chickpea, lentil, barley, and maize (Balestre et al., 2009; Munawar et al., 2013; Karimizadeh et al., 2013).

GGE biplot analysis emerged as a powerful tool for exploring and analyzing the multivariate data obtained from field level study. This approach allowed visualizing two-way tables, effectively highlighting the similarities and differences among the tested Bt brinjal varieties. The study revealed significant variations within the variety across the locations, holding immense potential for the development of new, high-yielding, and stable Bt brinjal cultivars. One particularly notable finding was the identification of a mega-environment consists of Sherpur, Satkhira, Manikgonj, and Norsingdi for Bt brinjal-4 in Bangladesh. While such findings have been documented for other crops, such as maize and rice (Balestre et al., 2009; Munawar et al., 2013), this is one of the first instances where mega-environment has been recognized for a particular variety of Bt brinjal in Bangladesh. This information is invaluable for tailoring cultivation strategies to specific regions, ensuring that Bt brinjal varieties are optimally matched with their respective environments for maximum yield and stability.

According to previous studies, the vertex genotypes are corner genotypes possessing maximum distance from the biplot origin, and these are the most responsive genotypes to either high-yielding or poor-yield plants (Bocianowski et al., 2019). The "which-won-where" tool of the GGE biplot played a pivotal role in the selection of suitable locations for Bt-brinjal genotypes. By identifying four vertex genotypes and designating one as the winning genotype across all locations, we discerned that some genotypes were highly responsive in terms of yield, while others were less responsive. Bt brinjal-4 and Bt brinjal-2 emerged as high-yielding genotypes, while Bt brinjal-1 exhibited poorer yields. These findings are in alignment with studies in barley, garlic, and other crops that identified vertex genotypes as high-yielding varieties (Batth et al., 2013; Koutis et al., 2012, Munawar et al., 2013, Bocianowski et al., 2019).

Assessing mean performance and stability is a crucial step in genotype selection. In this study, Bt brinjal-4 emerged as a desirable genotypes based on both their mean performance and stability. These genotypes exhibit high yields and greater stability across the studied locations, making them promising choices for cultivation. The use of an ideal genotype as a reference further reinforced these findings. Bt brinjal-4, positioned close to the ideal genotype, stands out as a high-yielding and stable choice for brinjal cultivation in Bangladesh. Furthermore, the suitability of environments for Bt brinjal cultivation revealed that Sherpur and Satkhira, environments with large PC1 scores and small PC2 scores, are more informative and representative locations among the tested areas. This environment excels in accurately differentiating the performance of brinjal genotypes, making it an ideal choice for cultivation. Sherpur and Satkhira closely resemble the ideal environment, indicating that they are well suited for cultivating Bt brinjal-4.

Ultimately, this study highlighted the importance of GEI in brinjal farming and provides a robust methodology for future endeavors in increasing Bt brinjal production. Through the utilization of GGE biplot analysis, this study provided valuable insight into the performance of several Bt brinjal varieties and successfully identified highly promising varieties suitable for cultivation in different locations in Bangladesh.

5. Conclusions

This study rigorously evaluated four Bt brinjal genotypes based on their mean yield performances across seven diverse locations in Bangladesh by employing GGE biplot analysis. The analysis of variance underscored the substantial and highly significant effects of genotype (G), environment (E), and genotype-environment interaction (GEI) on yield. This robust interaction effect indicates the dynamic nature of yield responses among genotypes under varying environmental conditions. The application of GGE biplot analysis enabled a detailed examination of genotype-by-environment interactions, aiding in the identification of genotypes most suited to specific environment(s). Notably, Bt brinjal-4 emerged as standout, exhibiting both high mean yields and remarkable stability across all locations for the mega environment. From this investigation, a clear imperative emerges as the development of stable, high-yielding Bt brinjal varieties tailored to diverse environments is a priority. These findings not only contribute to our understanding of Bt brinjal cultivation in Bangladesh, but also provide valuable insights for breeding programs, agricultural policymakers, and farmers. In the future, further research endeavors can build upon these foundations, exploring avenues to enhance brinjal breeding and cultivation strategies that can better withstand the challenges posed by varying environmental conditions under a changing

climate. In doing so, we can take steps toward ensuring food security through Bt brinjal and other major crops for sustaining agricultural productivity in Bangladesh.

References

Ahmed, F., Rahman, M. M., & Sultana, N. (2020). Effect of Bt Brinjal on Yield Performance and Farmer's Pest Management in Bangladesh. *International Journal of Agricultural Research*, 15(2), 100-110.

Akter, A., Hasan, M. J., Kulsum, U., Rahman, M. H., Khatun, M. and Islam, M. R. (2015). GGE biplot analysis for yield stability in multi-environment trials of promising hybrid rice (*Oryza sativa* L.). *Bangladesh Rice Journal*, *19*(1): 1-8.

Batth, G. S., Kumar, H., Gupta, V. and Brar, P. S. (2013). GGE biplot analysis for characterization of garlic (Allium sativum L.) germplasm based on agro-morphological traits. *International Journal of Plant Breeding*, *7*(2): 106-110.

Balestre, M., Von Pinho, R. G., Souza, J. C. and Oliveira, R. L. (2009). Genotypic stability and adaptability in tropical maize based on AMMI and GGE biplot analysis. *Genetics and Molecular Research*, *8*(4): 1311-1322.

Brdar-Jokanović, M., Girek, Z., Ugrinović, M., Sikora, V., Đalović, I. and Zdravković, J. (2016). AMMI model in the analysis of genotype by environment interaction of conventionally and organically grown onion. *Genetika*, *48*(3): 1027-1038.

Bocianowski, J., Księżak, J. and Nowosad, K. (2019). Genotype by environment interaction for seeds yield in pea (Pisum sativum L.) using additive main effects and multiplicative interaction model. *Euphytica*, *215*(11): 191.

De Vita, P., Mastrangelo, A. M., Matteu, L., Mazzucotelli, E., Virzi, N., Palumbo, M. ... and Cattivelli, L. (2010). Genetic improvement effects on yield stability in durum wheat genotypes grown in Italy. *Field crops research*, *119*(1): 68-77.

Gauch Jr, H. G., Piepho, H. P. and Annicchiarico, P. (2008). Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop science*, *48*(3): 866-889.

Karimizadeh, R., Mohammadi, M., Sabaghni, N., Mahamadi, A. A., Roustami, B., Seyydi, F. and Akbari, F. (2013). GGE biplot analysis of yield stability in multi-environment trials of lentil genotypes under rainfed condition. *Notulae Scientia Biologicae*, *5*(2): 256-262.

Koutis, K., Mavromatis, A. G., Baxevanos, D. and Koutsika-Sotiriou, M. (2012). Multi-environmental evaluation of wheat landraces by GGE biplot analysis for organic breeding.

Kumar, A., Verulkar, S. B., Mandal, N. P., Variar, M., Shukla, V. D., Dwivedi, J. L. and Raman, A. (2012). High-yielding, drought-tolerant, stable rice genotypes for the shallow rainfed lowland drought-prone ecosystem. *Field crops research*, 133: 37-47.

Luo, J., Pan, Y. B., Que, Y., Zhang, H., Grisham, M. P. and Xu, L. (2015). Biplot evaluation of test environments and identification of mega-environment for sugarcane cultivars in China. *Scientific reports*, 5(1): 15505.

Mohammadi, R., Haghparast, R., Amri, A. and Ceccarelli, S. (2009). Yield stability of rainfed durum wheat and GGE biplot analysis of multi-environment trials. *Crop and Pasture Science*, *61*(1): 92-101.

Munawar, M., Hammad, G. and Shahbaz, M. (2013). Evaluation of maize (Zea mays L.) hybrids under different environments by GGE biplot analysis. *American-Eurasian Journal of Agriculture & Environmental Science*, *13*: 1252-1257.

Myint, K. A., Amiruddin, M. D., Rafii, M. Y., Abd Samad, M. Y., Ramlee, S. I., Yaakub, Z. and Oladosu, Y. (2019). Genetic diversity and selection criteria of MPOB-Senegal oil palm (Elaeis guineensis Jacq.) germplasm by quantitative traits. *Industrial Crops and Products*, *139*: 111558.

Oladosu, Y., Rafii, M. Y., Abdullah, N., Hussin, G., Ramli, A., Rahim, H. A. ... and Usman, M. (2016). Principle and application of plant mutagenesis in crop improvement: a review. *Biotechnology & Biotechnological Equipment*, *30*(1): 1-16.

Oladosu, Y., Rafii, M. Y., Abdullah, N., Magaji, U., Miah, G., Hussin, G. and Ramli, A. (2017). Genotype× Environment interaction and stability analyses of yield and yield components of established and mutant rice genotypes tested in multiple locations in Malaysia. Acta Agriculturae Scandinavica, Section B,*Soil & Plant Science*, 67(7): 590-606.

Roy, A., Islam, M. S., & Zaman, M. (2011). Coastal Agriculture in Bangladesh:Challenges and Opportunities. *Journal of Agriculture and Rural Development*, 9(1), 33-40.

Shelton, A. M., Hossain, M. J., Paranjape, V., Prodhan, M. Z., Azad, A. K., Majumder, R. ... and Hossain, M. A. (2019). Bt brinjal in Bangladesh: the first genetically engineered food crop in a developing country. *Cold Spring Harbor perspectives in biology*, *11*(10), a034678.

Shelton, A. M., Sarwer, S. H., Hossain, M. J., Brookes, G. and Paranjape, V. (2020). Impact of Bt brinjal cultivation in the market value chain in five districts of Bangladesh. *Frontiers in Bioengineering and Biotechnology*, *8*: 541115.

Williams, W. T. (1971). Principles of clustering. *Annual Review of Ecology and Systematics*, 303-326.

Yan, W. and Kang, M. S. (2002). GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC press.

Yan, W. and Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *Canadian journal of plant science*, *86*(3): 623-645.

Yan, W., Kang, M. S., Ma, B., Woods, S. and Cornelius, P. L. (2007). GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Science*, *47*(2): 643-653.

UAV-SFM-BASED 3D RECONSTRUCTION OF FRUIT TREES TO ESTIMATE TREE PARAMETERS

KOWSHIK KUMAR SAHA[1], ISTIAK AHMED[1], MOHAMMAD MUKHLESUR RAHMAN[1], NUR MOHAMMAD[1], MONIRUL ISLAM[2]

[1]ASICT DIVISION, BARI, GAZIPUR AND [2]POMOLOGY DIVISION, HRC, BARI, GAZIPUR

Abstract

Plant characteristics such as height, width, volume, and leaf area are crucial for assessing plant growth, health, and yield predictions. In orchard conditions, these parameters guide the application of targeted treatments for individual trees. Traditionally, these measurements have been taken through direct observation or manual measurement, often relying on visual inspection. This study utilized UAV-SFM (Unmanned Aerial Vehicle - Structure from Motion) technology to generate a detailed point cloud of a mango orchard, capturing the spatial information of mango trees with high resolution. The reconstructed point cloud of the orchard displayed a detailed and precise representation, capturing both the canopy structure and the topography of the orchard with high density and accuracy. Application of Cloth Simulation Filter (CSF) classifier was utilized to differentiate between ground and non-ground points, resulting in clear segmentation essential for precise analysis. The segmented point clouds were used for height estimation and measurement of tree canopies. The linear regression analysis between the measured tree heights and the tree heights estimated from the UAV-SfM 3D point cloud yielded an R^2 value of 0.75. Collectively, these results showed the effectiveness of UAV-SFM and advanced filtering techniques in providing detailed, usable information for precision horticulture and orchard management.

Keywords: 3D point cloud, unmanned aerial vehicle (UAV), structure from motion (SfM), orchard, fruit trees

Introduction

Plant parameters e.g., height, width, volume, leaf area etc. are the important parameters which are often observed to understand plant growth, vigor and predict yield. In orchard conditions, specific treatments required to individual trees are implemented based on these parameters (Zhang et al., 2021). In an orchard condition, due to various factors tree growth and fruit production are affected. Long-term factors, including soil conditions, topography, and microclimate, significantly impact the growth and productivity of a tree (Charrier et al., 2015). In addition, short-term influences such as insect pests and orchard management practices also play a crucial role in determining the tree's development and fruit yield. The integration of spatio-temporal data into orchard management practices enables the implementation of precision agriculture techniques (Zude-Sasse, et al., 2016). Precision agriculture combines data from various sources, including remote sensing, soil sensors, and climate monitoring systems, to make informed decisions. For example, using satellite imagery and on-ground sensors, managers can identify areas with varying crop health, moisture levels, and pest activity (Postolache et al., 2022). This information allows for the application of site-specific treatments, ensuring that each part of the orchard receives the appropriate care according to its unique conditions. Traditionally, in-situ measurement or human eye observation is done to estimate these parameters. However, these techniques are time-consuming, labor-intensive and often biased because lack of skilled manpower.

Several remote sensing techniques have been invented in the last few decades to understand vegetation status and spatio-temporal heterogeneity. Aerial and satellite-based imaging techniques are increasingly used in modern orchard tree management, offering detailed and timely information about crop health and productivity. Aerial platform such as Unmanned Aerial Vehicles (UAVs), equipped with high-resolution cameras and multispectral sensors, provide high-frequency, close-range imaging that allows orchard managers to monitor tree health, detect pest infestations, and assess water stress with precision (Hafeez, et al., 2023). For example, UAVs can capture detailed two-dimensional (2D) images of the orchard canopy, enabling the calculation of vegetation indices like normalized difference vegetation index (NDVI) or enhanced vegetation index (EVI) at the plant level, which helps in identifying areas required specific management (Gallardo-Salazar & Pompa-García, 2020). This real-time data can be used to make informed decisions about irrigation, fertilization, and pest control, optimizing resource use and improving

overall tree health. On the other hand, satellite imaging offers broader spatial coverage and temporal consistency, which is beneficial for tracking larger-scale trends in orchard management. Satellites provide periodic imagery that can be used to monitor changes in orchard conditions over time, evaluate seasonal variations, and assess the impact of weather events on tree health. The integration of satellite data with ground-based observations allows for a comprehensive analysis of orchard conditions, facilitating strategic planning and management decisions.

In comparison to 2D remote sensing approaches, the utilization of three-dimensional (3D) methods is relatively novel and emerging (Farahani, et al., 2017). 3D point cloud technology transforms the non-invasive measurement of plant phenotypic parameters, providing crucial data for both agriculture and research. 3D plant monitoring, often facilitated by advanced technologies such as LiDAR (Light Detection and Ranging) and other imaging techniques, offers several benefits in the context of agriculture and plant management (Lin, 2015). 3D plant monitoring provides a detailed and accurate representation of the plant's structure, including the canopy, branches, and individual leaves. This level of detail is valuable for understanding the growth patterns and structural characteristics of plants, which can be critical for optimizing cultivation practices. The ability to accurately model and measure fruit trees in three dimensions is crucial for various aspects of agricultural management (Rosell, et al., 2009). Traditional methods often involve manual measurements, which are labor-intensive, time-consuming, and prone to errors. UAV-SfM-based 3D reconstruction offers a non-invasive, cost-effective alternative that provides detailed structural information without disturbing the environment or the crops themselves (Hobart, et al., 2020).

UAV (Unmanned Aerial Vehicle) is also an emerging remote sensing platform which can be facilitated to acquire data from a large area in a short period of time and also possible to acquire multi-temporal dataset for the specific location. UAVs have emerged as versatile platforms for capturing high-resolution aerial imagery, offering significant advantages over traditional methods, including reduced operational costs, high temporal and spatial resolution, and the ability to access hard-to-reach areas (Zhang et al., 2021). When equipped with cameras, UAVs can capture overlapping images from multiple angles around the target, which are then processed using SfM algorithms to generate dense 3D point clouds and detailed orthophotos (Nex & Remondino, 2014).

Structure from Motion is a photogrammetric technique that reconstructs 3D structures from 2D image sequences by identifying and matching feature points across multiple images (Westoby et al., 2012). The SfM process involves several steps: image acquisition, feature detection and matching, camera pose estimation, and point cloud generation (Snavely et al., 2008). This method does not require pre-calibrated cameras or known camera positions, making it flexible and widely applicable.

In this proposed study, multi-temporal 3D point cloud of fruit trees will be generated and analyzed to estimate the plant parameters. It is expected that the finding of this study can be beneficial for fruit tree management and production system in Bangladesh by providing timely and precise plant information. The objectives of the study were- a) To reconstruct 3D point cloud of fruit trees in orchard condition from multi-view UAV images (RGB) and b) To estimate plant parameters analyzing reconstructed 3D point cloud

Materials and methods

Test site

On 3 July 2024, measurement campaigns were carried out at the mango orchard of Horticultural Center, BARI in Joydebpur, Gazipur located at coordinates 23°59'0.35"N, 90°24'25.79"E (Fig. 1). The orchard, planted in 2014, consists of eleven rows with 55 mango trees of the BARI Mango 3 and 4 varieties. For this study, six rows, covering about 550 m², were selected which contained 30 trees. This orchard was chosen because of its significant variation in tree height, and the gaps between trees made site-specific management practical. The mango trees were pruned at regular intervals, spaced about 4 meters apart in rows, with approximately 4 meters between each tree. Reference tree data such as tree height, area etc. were collected through manual measurements from the orchard during the UAV campaign.

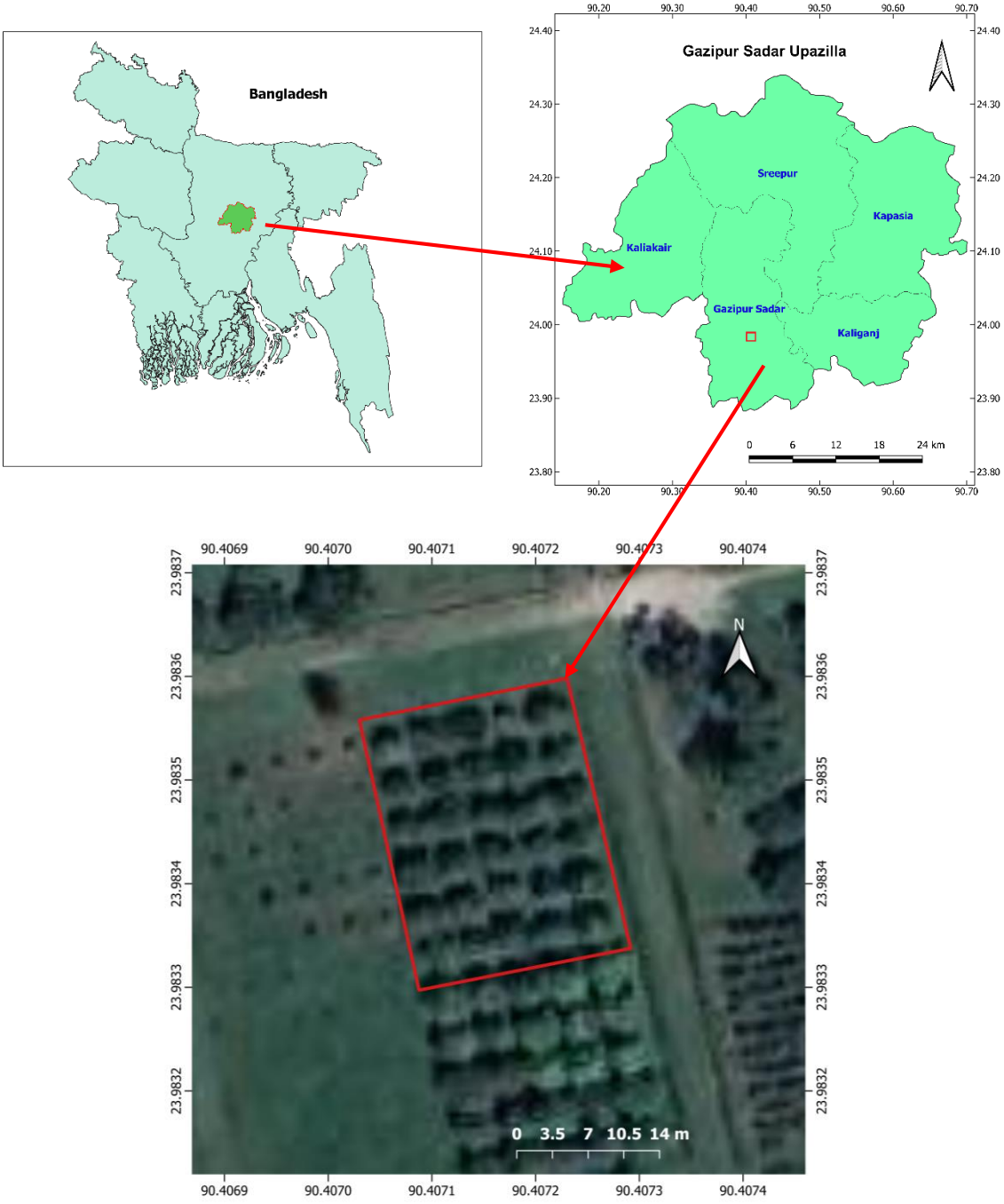


Fig. 1. Study area location was depicted through GIS maps.

UAV data acquisition

The UAV measurement campaigns were conducted well before the harvest, utilizing a P4M drone (DJI, Shenzhen, Guangdong, China) also known as the Phantom 4 Multispectral equipped with a high-resolution RGB camera (P4 Multispectral, 2024). The UAV system, with a takeoff weight of approximately 1.5 kg, could operate for up to 27 minutes per flight. To ensure image stability and minimize blurring, a three-axis gimbal was employed, keeping the camera angle fixed during flights. The drone is equipped with six 1/2.9 inch Complementary Metal Oxide Semiconductor (CMOS) sensors. These include one RGB sensor for capturing visible light images and five monochromatic sensors for multispectral imaging. Each sensor has a resolution of 2.08 megapixels, contributing to a total resolution of 2.12 megapixels. The filters for the five monochromatic sensors used in multispectral imaging are as

follows: Blue (B) at 450 nm ± 16 nm; Green (G) at 560 nm ± 16 nm; Red (R) at 650 nm ± 16 nm; Red edge (RE) at 730 nm ± 16 nm; and Near-infrared (NIR) at 840 nm ± 26 nm. The image resolution is 1600 × 1300 pixels (aspect ratio 4:3.25), with a controllable tilt range from −90° to +30°. The ground sample distance (GSD) is H/18.9 cm per pixel. The drone supports two image formats: JPEG for visible light images and TIFF for multispectral images.

The P4M drone was operated using DJI GS Pro (version 2.0.17, DJI, China) flight control software on an iPad mini (Apple Inc., Elk Grove, Calif., USA). At first an overview flight was conducted to capture the entire orchard from a broader perspective. In this flight, the UAV captured five spectral images of the orchard implementing all sensors at an altitude of 30 meters, capturing images at nadir. These images were ortho mosaiced and georeferenced using Agisoft Metashape (version 2.1.2, Agisoft LLC, Russia). These images were used for region of interest (ROI) selection and identify the orchard's boundary and layout.

After the initial overview flight, a detailed flight was conducted at a lower altitude of 12 meters. The flight plan was shown in Fig. 2. During this flight, the UAV flew along the tree rows, capturing oblique images at a 45° angle from nadir to ensure a clear view of the trees. The flight speed was set to 5 m/s, providing an image overlap of 80 % in both the forward and side flight paths. This setup ensured high-quality image coverage. These UAV flights produced detailed, high-resolution imagery, crucial for accurate 3D reconstruction and analysis of the orchard. The combination of overview and detailed flights enabled comprehensive data collection, capturing both the overall orchard layout and the finer details of individual tree rows.

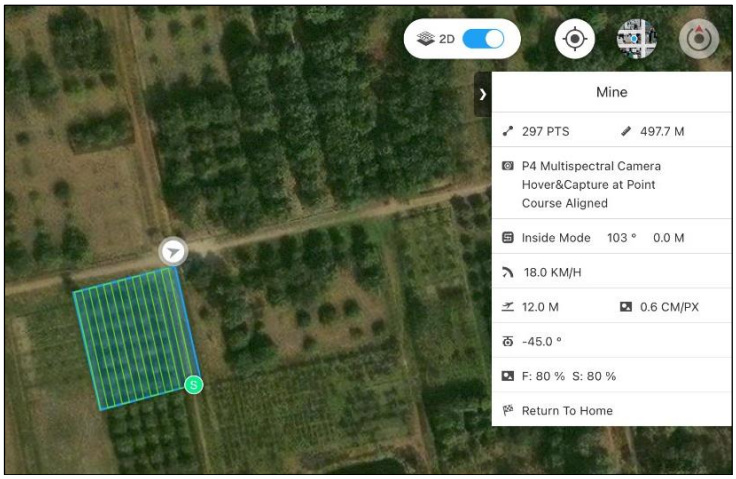


Fig. 2. Flight plan of UAV data acquisition for the study area.

3D reconstruction of tree point clouds

The images from UAV flight campaign were processed using photogrammetry software (Metashape, Agisoft LLC, Russia) to generate 3D point clouds. This process employed Structure from Motion (SfM) and multi-view stereo reconstruction techniques. High-quality settings were used for determining image positions, orientations, and matching overlapping images. To enhance the quality of the dense point cloud reconstruction, the sparse point cloud was manually refined. Points were eliminated by setting maximum thresholds for reprojection error, reconstruction uncertainty, and projection accuracy to 0.1, 50, and 10 pixels, respectively. The image positions, orientations, and matches were updated based on the remaining tie points. The dense 3D point cloud reconstruction was performed with "high" quality and "mild" depth filtering settings. The time required to compute the point clouds from 340 photographs was approximately 120 minutes, respectively.

Tree parameters estimation

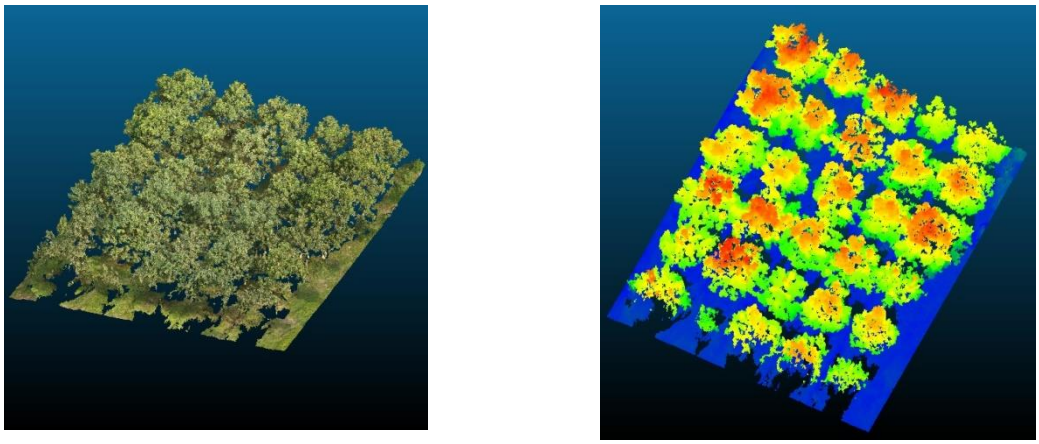
After the initial 3D reconstruction of the point clouds from flight campaign images, the next steps involve several key processes to refine the data and extract tree parameters. These steps were ground points removal, tree height estimation, and ground-projected canopy area estimation.

The 3D point clouds often contain noise and artifacts that can interfere with accurate analysis. Cloth simulation filter (CSF) classifier algorithm was applied using CloudCompare software (GPL Software, 2018) to separate ground points and non-ground points . This method adjusts the points based on their local geometry to minimize inconsistencies and noise. The cloth simulation algorithm iteratively adjusts the point positions to fit a smoother surface, effectively filtering out small, irrelevant deviations and finally using a threshold (0.2 m) to separate the point cloud above the created surface. Once the ground points were removed through CSF filtering, the next task was tree height estimation. This involves determining the elevation of various points in the 3D tree model relative to the lowest point of the point cloud. Interpolation approach was employed to estimate the height of points across the point cloud. By analyzing the ground control points, precise height measurement was achieved.

Tree canopy area is an important tree parameter often used to determine leaf area index (LAI) or tree growth. It was estimated as ground-projected area from 3D point cloud for individual mango trees. To determine the ground-projected area from the point cloud, the point data was projected onto the horizontal plane (x-y). This involved calculating the area covered by the point cloud when viewed from directly above. The ground-projected area was estimated by generating a 2D representation of the point cloud, removing elevation data, and then calculating the area based on the spatial distribution of the points (Saha et al., 2022). These later data analysis steps were performed in developed Python scripts.

Results

Fig. 3a presented a detailed 3D view of UAV-SfM point cloud of the mango orchard selected for this study. From the visual observation it can be stated that the point cloud achieved high resolution and accuracy through this imaging technique. The point cloud effectively captures the complex spatial arrangement of the mango trees within the orchard, illustrating a dense network of data points that delineate the canopy structure, tree trunks, and the interstitial spaces between the trees. The high density of points contributes to a rich three-dimensional representation, with minimal gaps or noticeable distortions, indicating a high-quality reconstruction of the orchard. The precision of the point cloud is further evidenced by the clear delineation of individual tree canopies and the accurate depiction of the orchard's overall topography.



(a) (b)

Fig. 3. (a) RGB 3D point cloud of the mango orchard generated by UAV-SfM method and (b) classified point cloud as ground points and non-ground point using CSF classifer.

The data reveals detailed variations in canopy size and tree distribution, which are critical for subsequent analyses such as tree growth monitoring and yield estimation. Furthermore, the point cloud exhibits minimal noise, with a consistent level of detail throughout the dataset, reflecting the efficacy of the UAV-SfM method in capturing both the macro and micro features of the orchard environment.

In Fig. 3b, classification of ground and non-ground points was illustrated in the point cloud of the mango orchard, utilizing the CSF classifier. This figure displayed a clear distinction between ground and non-ground points achieved through advanced filtering techniques. The ground points were shown in blue color and non-ground points depicted in height ramp color (green, yellow and red). The CSF classifier effectively separated the ground surface from the vegetation and other objects by simulating a cloth-like draping over the point cloud. This method allowed for the identification of ground points based on their elevation and surface continuity, while the non-ground points, which include the mango tree canopies and other structures, were isolated due to their differing height and spatial characteristics. The classification process demonstrated here enhances the interpretability of the point cloud data by clearly differentiating between ground and vegetation, which is crucial for applications such as terrain analysis, canopy volume estimation, and overall orchard management.

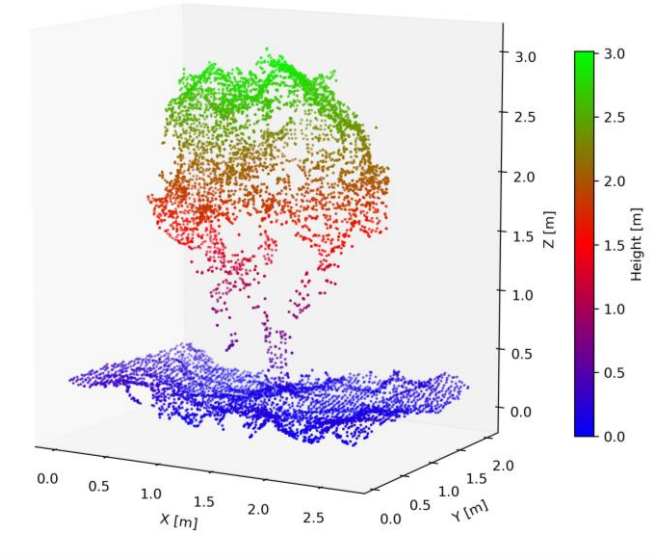


Fig. 4. 3D point cloud of a segmented mango tree.

In Fig. 4, a manually segmented mango tree point cloud was shown for tree height estimation in the mango orchard. This figure highlighted the segmenting point cloud to isolate different height levels, facilitating accurate measurement of tree canopies and overall vegetation structure. By categorizing points based on their elevation, the segmentation allowed for detailed analysis of vertical dimensions, which was essential for assessing tree growth, canopy volume, and spatial distribution. All the mango trees (n = 30) were estimated by analyzing segmented point clouds.

The relationship between the measured tree height and the tree height estimated using UAV-SfM 3D point cloud data is presented in Fig. 5. The scatterplot illustrates a positive correlation between the two variables, with most data points clustered around the line of best fit. The x-axis represents the measured tree height, ranging from 3 to 7 m, while the y-axis represents the estimated tree height derived from the UAV-SfM 3D point cloud method, covering the same range. The linear regression analysis yielded an R^2 value of 0.75, indicating that 75 % of the variance in the estimated tree height can be explained by the measured tree height. This suggested a strong positive linear relationship between the two variables, affirming that the UAV-SfM 3D point cloud method is a reliable tool for estimating tree height. The positive slope of the best fit line further supports this conclusion, showing that increases in measured tree height correspond to increases in estimated tree height.

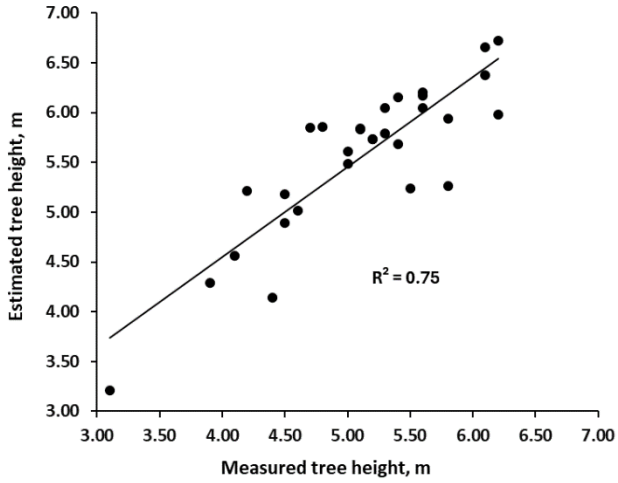


Fig. 5. Scatterplot between measured tree height and UAV-Sfm 3D point cloud estimated tree height.

Despite the strong correlation, some variability was observed in the scatter of the data points around the best fit line. This variability suggests that while the UAV-SfM method generally provides accurate height estimates, there are instances where the estimated values deviate from the measured heights. Notably, this deviation appears more pronounced at the lower end of the height spectrum, where some data points lie further from the regression line. These discrepancies may be due to factors such as tree species differences, environmental conditions during data acquisition, or limitations in the UAV-SfM method under certain conditions.

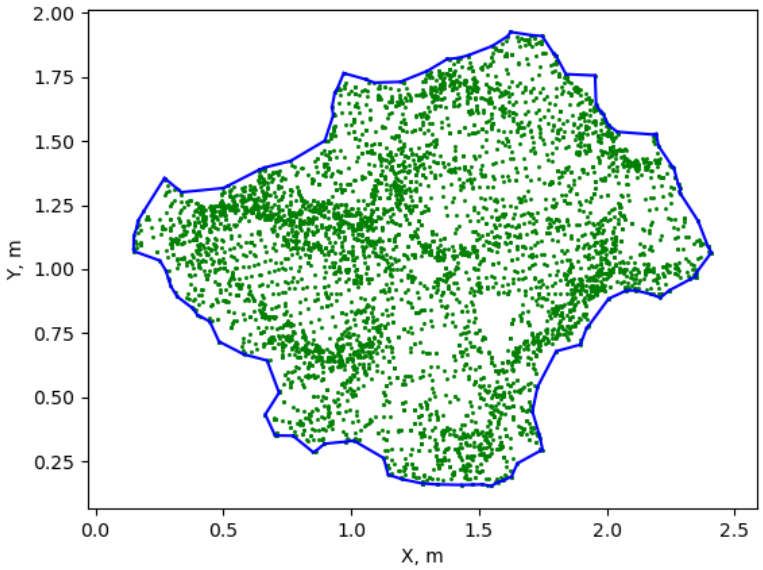
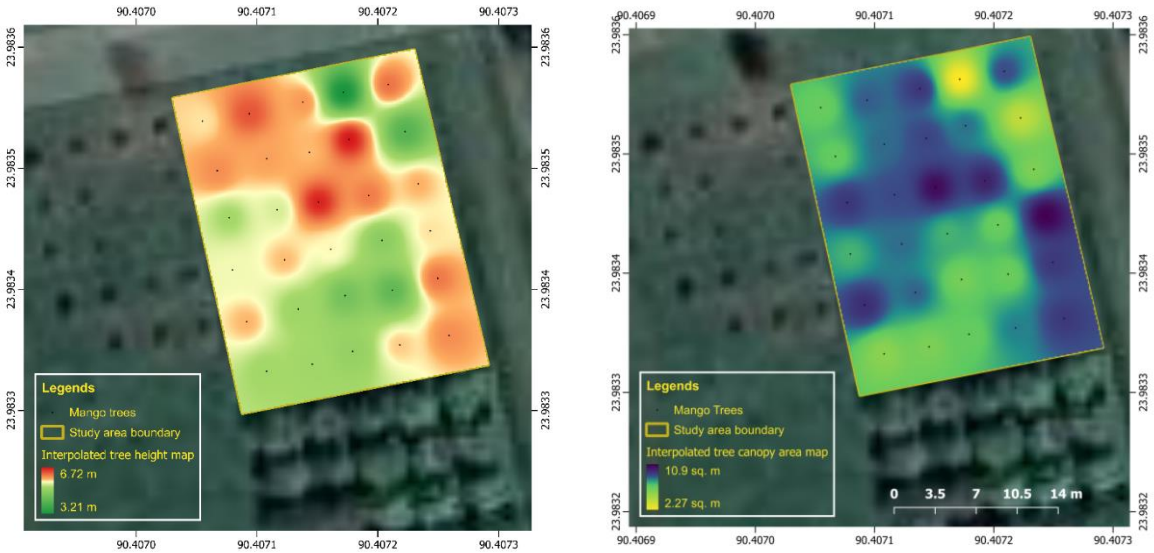


Fig. 6. Estimation process of ground projected tree canopy area (in blue polygon, estimating area of 3.14 m^2) of a sapmled mango tree.

Fig. 6 illustrated the process of estimating the ground-projected canopy area of a sampled mango tree, which calculated about 3.14 m². Similarlly all the segmented (n = 30) mango tree point clouds were utilized for the canopy area estimation. Manually, the point cloud was segmented to isolate the specific mango tree of interest, and its canopy was defined by projecting the three-dimensional canopy outline onto the x-y horizontal plane representing the ground. The figure showed visual overlays showing the canopy's footprint on the ground. Finally, by connecting all the outer points created a polygon and polygon area was determined.

Fig. 7a and b showed the spatial maps for interpolated data of mango tree characteristics (tree height and tree canopy area) estimated from 3D point cloud within the study area. Fig. 7a illustrated the interpolated tree height distribution for the mango trees within the study boundary. Tree heights are represented using a color gradient from green (lower heights, around 3.21 m) to red (higher heights, up to 6.72 m) and the tree locations are marked by black dots. Similarlly, Fig. 7b showed the distribution of mango tree canopy areas within the study boundary. The canopy areas were represented using a color gradient from yellow (indicating smaller canopy areas, around 2.27 sq. m) to purple (indicating larger canopy areas, up to 10.9 sq. m). Both maps provided a spatial visualization of the tree characteristics, with clear variations across the study area, which could be useful for future investigations of the factors influencing these tree attributes. The spatial maps of mango tree canopy area and height can be used as valuable tools for precision orchard management. Various precise application can be achieved such as optimized resource allocation, tree health monitoring, yield estimation, and targeted interventions such as precision pruning and irrigation utilizing these information.



(a) (b)

Fig. 7. Spatial distribution map of mango (a) tree height and (b) tree canopy area in the study area.

Conclusion

This study demonstrated the efficacy of UAV-SFM technology combined with advanced filtering techniques for precise monitoring and management of fruit trees in horticultural production system. The high-resolution dense point cloud was generated that provided a detailed representation of the mango orchard, with effective differentiation of ground and non-ground points achieved through the CSF classifier. The ability to segment and estimate tree canopy height and ground-projected area revealed the potential of these methods for detailed analysis of orchard conditions. The accurate measurement of the ground-projected canopy area of individual trees depicted the practical applications of these technologies in optimizing orchard management practices. By providing detailed spatial information, these maps can help orchard managers to make data-driven decisions that enhance productivity, sustainability, and profitability, ensuring that each tree receives specific treatment based on its specific needs and growth patterns.

References

Charrier, G., Ngao, J., Saudreau, M., & Améglio, T. (2015). Effects of environmental factors and management practices on microclimate, winter physiology, and frost resistance in trees. Frontiers in plant science, 6, 259. https://doi.org/10.3389/fpls.2015.00259.

Farahani, N., Braun, A., Jutt, D., Huffman, T., Reder, N., Liu, Z., ... & Pantanowitz, L. (2017). Three-dimensional imaging and scanning: current and future applications for pathology. Journal of pathology informatics, 8(1), 36. https://doi.org/10.4103/jpi.jpi_32_17.

Gallardo-Salazar, J. L., & Pompa-García, M. (2020). Detecting individual tree attributes and multispectral indices using unmanned aerial vehicles: Applications in a pine clonal orchard. Remote Sensing, 12(24), 4144. https://doi.org/10.3390/rs12244144.

GPL Software. CloudCompare; France. Available online: https://www.danielgm.net/cc/ (accessed on 20 May 2024).

Hafeez, A., Husain, M. A., Singh, S. P., Chauhan, A., Khan, M. T., Kumar, N., ... & Soni, S. K. (2023). Implementation of drone technology for farm monitoring & pesticide spraying: A review. Information processing in Agriculture, 10(2), 192-203. https://doi.org/10.1016/j.inpa.2022.02.002.

Hobart, M., Pflanz, M., Weltzien, C., & Schirrmann, M. (2020). Growth height determination of tree walls for precise monitoring in apple fruit production using UAV photogrammetry. Remote Sensing, 12(10), 1656. https://doi.org/10.3390/rs12101656.

Lin, Y. (2015). LiDAR: An important tool for next-generation phenotyping technology of high potential for plant phenomics?. Computers and electronics in Agriculture, 119, 61-73. https://doi.org/10.1016/j.compag.2015.10.011.

Nex, F., & Remondino, F. (2014). UAV for 3D mapping applications: a review. Applied geomatics, 6, 1-15. https://doi.org/10.1007/s12518-013-0120-x.

P4 Multispectral. Available online: https://www.dji.com/bg/p4-multispectral (accessed on 7 July 2024)

Postolache, S., Sebastião, P., Viegas, V., Postolache, O., & Cercas, F. (2022). IoT-based systems for soil nutrients assessment in horticulture. Sensors, 23(1), 403. https://doi.org/10.3390/s23010403

Rosell, J. R., Llorens, J., Sanz, R., Arnó, J., Ribes-Dasi, M., Masip, J., ... & Palacín, J. (2009). Obtaining the three-dimensional structure of tree orchards from remote 2D terrestrial LIDAR scanning. Agricultural and Forest Meteorology, 149(9), 1505-1515. https://doi.org/10.1016/j.agrformet.2009.04.008.

Saha, K. K., Tsoulias, N., & Zude-Sasse, M. (2019). Removing ground points from 3D point cloud in apple orchard. (https://www.cabidigitallibrary.org/doi/full/10.5555/20193344031).

Saha, K. K., Tsoulias, N., Weltzien, C., & Zude-Sasse, M. (2022). Estimation of vegetative growth in strawberry plants using mobile LiDAR laser scanner. Horticulturae, 8(2), 90. https://doi.org/10.3390/horticulturae8020090.

Snavely, N., Seitz, S. M., & Szeliski, R. (2008, June). Skeletal graphs for efficient structure from motion. In 2008 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-8). IEEE. https://doi.org/10.1109/CVPR.2008.4587678.

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., & Reynolds, J. M. (2012). 'Structure-from-Motion'photogrammetry: A low-cost, effective tool for geoscience applications. Geomorphology, 179, 300-314. https://doi.org/10.1016/j.geomorph.2012.08.021.

Zhang, C., Valente, J., Kooistra, L., Guo, L., & Wang, W. (2021). Orchard management with small unmanned aerial vehicles: A survey of sensing and analysis approaches. Precision agriculture, 22(6), 2007-2052. https://doi.org/10.1007/s11119-021-09813-y.

Zude-Sasse, M., Fountas, S., Gemtos, T. A., & Abu-Khalaf, N. (2016). Applications of precision agriculture in horticultural crops. European Journal of Horticultural Science, 81 (2), 78-90. https://doi.org/10.17660/eJHS.2016/81.2.2.

SERVICES OF THE ASICT DIVISION 2023-2024

SYED MD. ABDULLAH SHIBLEE
CHIEF SCIENTIFIC OFFICER, ASICT DIVISION, BARI, GAZIPUR

Introduction

Agricultural Statistics and Information & Communication Technology (ASICT) Division comprises two parts: Agricultural Statistics and ICT. Concerned scientists of both parts conduct different types of research relevant to Agricultural Statistics and ICT. In addition, the scientists of both parts are providing agricultural statistics and ICT-related support services to BARI personnel and other stakeholders of the agriculture system to facilitate agricultural research and development. Besides, this division has been working to implement the *National ICT Policy 2018* and *Digital Bangladesh 2021*. However, the ASICT division is constantly keeping its vital role for BARI as a whole.

Web Information

BARI-developed technologies and related information are being published regularly through its website (www.bari.gov.bd) and mobile apps. In addition to technologies, some important issues like tender circulars, job circulars, journal publications, annual reports, etc. are also hosted as and when necessary. The uploaded information along with their number has been shown in Table 1. The number of upload documents is 23.03% lower in 2023-24 compared to 2022-23.

Table 1. List of information uploaded in the website and mobile apps during 2023-2024

S.N	Type of information	Uploaded number	Type of document
1	Notices	232	PDF
2	Official Leave Related Order	502	PDF
3	Order/Publicity	944	PDF
4	Question/Answer	31	PDF
5	e-Tender	46	PDF
6	Higher Education	24	PDF
7	Seminar and Workshop	61	PDF
8	Training-domestic	110	PDF
9	NOC of Officers	265	PDF
10	Loan Related Order	633	PDF
11	Retirement Related Order	249	PDF
12	Photogallery	4028	Images
13	Officers	04	PDF
14	Labourer Office Order	124	PDF
15	Labourer Promotion/Transfer	28	PDF
16	BARI in News	85	PDF
17	Produced Agricultural Technology	04	PDF
18	Promotion and Transfer	230	PDF
19	Forms	02	PDF
20	Common Documents	01	PDF
21	Monthly/Quarterly/Annual Inspection/Assessment Report	36	PDF
22	Recruitment Notice/Examination Schedule	14	PDF
23	Training Abroad	26	PDF
24	Quarterly/Annual Assessment Report	14	PDF
25	Director General and Directors	19	PDF

26	Yearly Salary Increment	137	PDF
27	Citizens' Charter	55	PDF
28	On Farm Research Division	12	PDF
29	Crop Research Centers/subcenters	08	PDF
30	Sim List of Office Incharges	06	PDF
31	Resignation/Termination	12	PDF
32	Result of Practical Examination	01	PDF
33	Focalpoint Officer and Appeal Officer	03	PDF
34	CDO and Appeal Officer	04	PDF
35	Integrity strategy and action plan	05	PDF
36	Voluntarily releasable information	10	PDF
37	BARI Success	04	PDF
38	Inspection and Assessment Report	03	PDF
39	Scientists/Officers/Stuffs	01	PDF
40	Quarterly/Half-yearly Inspection/Assessment Report	02	PDF
41	ADP Allotment	02	PDF
42	SIM Related Office Order	01	PDF
43	List of SIMs of Scientists	01	PDF
44	SIM Transfer and Receive Form	02	PDF
45	Result of Written Examination	07	PDF
Total Upload		**7988**	-

Note: The total number of uploads in 2022-23 was 10,378

Web-based Query and Answer

BARI has started online e-agriculture services for its beneficiaries. Any stakeholder can ask agriculture-related questions with the help of online facilities of the BARI website and Mobile apps. On the other hand, it has been providing different feedback to the end users through the BARI website (www.bari.gov.bd) and Mobile apps (baritechnology.org/m). The number of queries/answers addressed by different centers/divisions during 2023-24 is 27.21% higher compared to the year 2022-23 (Table 2).

Table 2. Number of queries &answers addressed by different centers/divisions during 2023-24

Sl. No.	Name of Centre/Division	Number of queries & answers
1.	Vegetable Division, HRC	200
2.	Pomology Division, HRC	160
3.	Floriculture Division, HRC	20
4.	Tuber Crops Research Centre	50
5.	Oil Seed Research Centre	20
6.	Pulses Research Sub-station	15
7.	Regional Spices Research Centre	65
8.	Soil Science Division	25
9.	FMPE Division	20
10.	IWM Division	5
11.	Vertebrate Division	5
12.	Seed Technology Division	15
13.	Pathology Division	10
14.	Biotechnology Division	2
15.	On Farm Research Division	3
16.	Entomology Division	15
17.	Plant Breeding Division	5
18.	ASICT Division	100
Total		**735**

Note: The total number of queries & answers in 2022-23 was 535.

Citizen Information Service Centre

BARI has started disseminating technological information to the citizen. Any citizen can come and obtain his/her required agricultural technology-related information/services directly from Citizen Information Service Centre and/or from the respective crop Centre/research divisions at BARI. During 2022-2023, a total of **137** citizens of different professions from **13** districts visited the BARI citizen services information Centre and got services in **17** different areas.

Ensuring access to information for citizens through the Right to Information Act

BARI has appointed personnel to ensure citizens access to the information in the easiest way. During 2022-2023, the following information has been provided to the citizens (Table 3).

Table 3. Citizen's access to information in 2023-2024

SN	Name of the authority	No. of application received regarding RTI Act, 2009	No. of queries solved through providing information	No. of decisions for not providing requested information with reasons	No. of appeals against the decision of the officer in charge	No. of settlement appeals	No. of disciplinary action taken by the authorities	The amount received as the value of information as per rule 8 of the RTI Act, 2009	Details of activities taken by the authorities
1	BARI	5	5	-	-	-	-	-	-

Online-based Services to the Citizen

BARI has started providing online-based different services to the citizen since 2016-17 (Table 4). ASICT Division develops at least one online-based service each year. During 2023-2024, one service entitled '***Vromonsuchi Byabosthapona Software'*** has been developed. Scientists from BARI Headquarters or outer station of BARI can get tour through the permission from higher authority by using the online Vromonsuchi Byabosthapona Software through the link (https://baripmis.org/monitoring/login). A list of online-based services of BARI has been shown in Table 4.

Table 4. List of online-based services in BARI

Sl.No.	Service	Link	Year
1.	Krishi Projukti Bhandar	http://m.baritechnology.org/	2015-16
2.	Job application	https://bari.taletalk.com.bd	2016-17
3.	District-wise cropping pattern	https://baritechnology.org/crpt	2017-18
4.	Online APA reporting	https://baripmis.org/apa	2018-19
5.	Online loan management	baripmis.org/form-app	2018-19
6.	BARI telephone directory	https://play.google.com/store/apps/details?id=com.baridirectorey.iamsh.baridirectorey	2019-20
7.	Gene bank	http://pgrcbari.org/	2020-21
8.	Matri Kalam supply form	https://baripmis.org/sapling/	2021-22

9.	Transportation management system	https://baripmis.org/transport/	2022-23
10	Online Vromonsuchi Byabosthapona Software	https://baripmis.org/monitoring/login	2023-24

Service Simplification to the Citizen

BARI started web-based service simplification using ICT from 2017-2018. The ASICT Division of BARI has to simplify at least one web-based service for its stakeholders each year. During 2023-2024, one service titled **'*Online Vromonsuchi Byabosthapona Software*'** has been simplified for providing instant and easiest tour facilities to the scientists, officers, and staffs of BARI (Table 5).

Table 5. List of web-based service simplification at BARI

SN	Type	Service	Year
1.	Information centre	Citizen information services	2017-18
2.	Collection of seed and seedling	Limited collection of BARI develop variety seeds and sapling.	2018-19
3.	Collection and preservation of personal information	Collection & preservation of the information of BARI scientists/officers/staff	2019-20
4.	Online-based Query-Answer	Providing quick responses to the citizen's queries regarding the BARI developed variety and technology	2020-21
5.	Determination of pesticide residue in agri-products, foods, soil, and water	Providing the results of pesticide residue in agri-products, foods, soil, and water to the stakeholders in the quickest possible time.	2021-22
6	BARI Vehicle Management System	Providing instant and easiest transport facilities to the scientist, officers, and staffs of BARI.	2022-2023
7.	Online Vromonsuchi Byabosthapona Software	https://baripmis.org/monitoring/login	2023-24

ICT-based Innovation to the Citizen

BARI started developing ICT-based innovation services for the citizen from 2016-2017. The ASICT Division of BARI has to develop at least one ICT-based innovation for its stakeholders each year. In the year 2023-2024, ASICT Division developed an online "*Online Vromonsuchi Byabosthapona Software*" for providing on line tour registration system for scientists (Table 6).

Table 6. List of ICT-based innovations to the citizen services at BARI

Sl. No.	Type	Innovation	Year
1.	Apps	Krishi Projukti Vander (Mobile apps)	2016-17
2.	Online service	Faster analysis of soil samples	2017-18

3.	Apps	Mobile apps for mango yield estimation	2018-19
4.	Web Apps	GeoMango: Satellite-based mango orchard mapping	2019-20
5.	Web Apps	Provision for Advisory on Necessary Irrigation	2020-21
6.	Pension System	Providing pension directly to the users bank account	2021-22
7.	Matri Kalam Form Online Registration	Providing information on the availability of fruit saplings and registering the form for purchasing those fruit saplings.	2022-2023
8.	Web Apps	Online Vromonsuchi Byabosthapona Software	2023-2024

Social Networking (Facebook)

BARI has been publicizing its regular activities among the public through social networking. Any person can ask and interact with BARI authority regarding BARI-mandated crop varieties, improved technologies, and other agricultural information through its official Facebook page (www.facebook.com/BD.GOV.BARI). At present BARI Facebook page is followed by **23,884** Facebook users (it was **21,235** in 2022-23). During 2023-24, a total of 604 posts (in 2022-23, it was 500) and 256 public Quarry/Answer were addressed through the BARI Facebook page.

Social Networking (YouTube)

BARI disseminates its regular activities among the public through its **YouTube** channel. Any person can easily know about BARI technologies through its official YouTube channel titled 'Bangladesh Agricultural Research Institute' (https://www.youtube.com/@BD.GOV.BARI.). At present BARI YouTube channel has **848** subscribers and 25 videos. During 2023-24, **185** subscribers were added to the channel with **6100** views (161.3 hours watch time).

Web-based Mail Services

BARI has its domain of email connectivity under the name "bari.gov.bd". At present, a total of 824 web mails have been assigned under the BARI domain till 2023-2024. It has been decided to assign email addresses to all scientists under the BARI domain gradually.

Networking and Antivirus Maintenance

ASICT Division is giving services on LAN & Antivirus maintenance, especially for the scientists and officers of BARI headquarters. Rendering 24-hour internet services at BARI headquarters. At present more than **600** computers are connected to the network and provided with a corporate version of antivirus. The number of networking and antivirus maintenance done by the ASICT Division during 2023-2024 has been stated in Table 7.

Table 7. Networking and antivirus maintenance during 2023-24

SN	Subject	Number
1.	Trouble shooting for LAN & internet connectivity	889
2.	Trouble shooting for Antivirus	73
3.	New internet connectivity/ Wi-Fi	161
4.	Network connection repair	124
5.	Network switch replace	29
	Total	**1276**

Wi-Fi Connectivity

The ASICT division has established Wi-Fi connectivity in different places at BARI. This division is providing services on Wi-Fi & maintenance, especially for the scientists and officers of BARI headquarters. At present **1093** devices are connected to the Wi-Fi network.

Statistical Analysis Service

ASICT division has been giving services on statistical analysis through computer package software such as R, Python, SPSS and QGIS. Some important analyses have been given to the scientists of different Centers/Divisions which is shown in the following Table 8.

Table 8. Statistical analysis services rendered to different centers/divisions during 2023-2024

Name of Centre/ Division/RARS	Number of Experiment	Required Analysis
Plant Breeding Division	3	CRD, RCBD, Factorial, Split plot, Combined, Correlation & Regression, MVA, Error Bar, Time Series, GGE, AMMI, GIS Map, T-test, Drone Image Processing, etc.
Plant Physiology Division	3	
Plant Pathology Division	4	
Seed Technology Division	1	
Agricultural Economics Division	4	
T & C	1	
SRC, Bogura	4	
HRC	11	
PRC	3	
PGRC	4	
IWM	8	
TCRC	7	
ORC	17	
OFRD	13	
OFRD, Satkhira	4	
OFRD, Khulna	3	
OFRD, Rangpur	1	
BINA	2	
BSRI	3	
Cimmyt	5	
IRRI	2	
Supreme Seed Ltd	11	
Total	**114**	

E-Governance

(a) E-filing System

Access to Information (a2i) has been established e-filing system in BARI Headquarters (HQ), Crops Center, and RARS. ASICT division assists the e-filing system especially for the Headquarters, Crops Center, and RARS personnel. At present **187** personnel are connected with the e-filing system. The number of letters issued in BARI headquarters through the e-filing system decreased 3.86% from 2022-23 to 2023-24 (Table 9).

Table 9. List of file and letter issued in HQ through e-filing system during 2023-24

SN.	Name of Wing/Centre/Division/Section	No. of file **(2023-2024)**	No. of letter issued		
			2022-23	2023-24	% increase
1.	Support and Services wing	5	60	15	-75.0
2.	Research Wing	11	67	96	43.3
3.	Training & Communication Wing	49	878	798	-9.1
4.	Planning and Evaluation Wing	9	122	238	95.1
5.	Horticulture Research Center	39	509	530	4.1
6.	Oilseed Research Center	9	171	107	-37.4
7.	Tuber Crops Research Center	16	350	109	-68.9
8.	Plant Genetic Research Center	13	341	139	-59.2
9.	Agronomy Division	7	509	270	-47.0
10.	Soil Science Division	2	208	307	47.6
11.	Entomology Division	44	110	48	-56.4
12.	On-Farm Research Division	4	163	149	-8.6
13.	Agricultural Economics Division	9	475	151	-68.2
14.	Plant Physiology Division	4	80	47	-41.3
15.	Plant Pathology Division	10	95	150	57.9
16.	Plant Breeding Division	3	434	83	-80.9
17.	IWM Division	0	284	267	-6.0

SN.	Name of Wing/Centre/Division/Section	No. of file	No. of letter issued		
18.	ASICT Division	22	143	168	17.5
19.	Postharvest Technology Division	36	104	235	126.0
20.	Biotechnology Division	35	54	133	146.3
21.	Seed Technology Division	1	173	54	-68.8
22.	Vertebrate Pest Management Division	9	96	129	34.4
23.	FMPE Division	9	257	147	-42.8
24.	Farm Division	9	307	367	19.5
25.	Machinery Repair and Maintenance	13	102	63	-38.2
26.	Regional Spice Research Center	3	131	58	-55.7
27.	Pulse Research Sub-station	1	90	73	-18.9
28.	Administration Section	29	2254	3266	44.9
29.	Finance and Accounts	243	584	575	-1.5
30.	Transport Section	1	111	52	-53.2
31.	Building and Ground Section	4	4	2	-50.0
32.	Procurement and Store Section	9	24	67	179.2
33.	Common Service	0	110	144	30.9
Total BARI		**658**	**9400**	**9037**	**-3.86**

(b) E-tendering System

Central Procurement Technical Unit (CPTU) has established an e-tendering system at BARI. ASICT division assists on the e-tendering system, especially for the headquarters through two Procurement Entity (PE) offices like Building & Ground section and Procurement & Store section. During 2023-2024, a total of **194** e-tendering have been implemented through two PE offices (Table 10).

Table 10. List of E-tendering during 2023-2024

Sl. No	Name of office	Number of e-tender
1.	Procurement & Store section	81
2.	Building & Ground section	113
	Total	**194**

(c) Labour Management Software

BARI labour management automation process is going on in full swing. At present, 53 Centre/divisions/sections and **3235** labour information are included in this automation. This automation software is divided into three parts viz, labour information, labour salary, and labour report. Each part can be operated solely and has the option to be integrated as per requirements.

(d) Salary System Software

BARI payroll management system has been developed for the BARI personnel. This system is user-friendly. Any person can use his/her own BARI ID number and prepare his/her salary statement. At present, a total of **584** scientists/officers/staff are using this system.

(e) Management Information System (MIS) Software

MIS software is divided into three separate modules viz. Personnel Management Information System (PMIS), Training Management Information System (TMIS), and Publications Management Information System. Each module can be operated solely and has the option to be integrated as per requirements.

Table 11. Present status of the three MIS software modules

MIS Software Modules	Achievement (data entry)
Personnel Management Information System (PMIS)	100 %
Training Management Information System (TMIS)	100 %
Publications Management Information System	100%

Human Resource Development

ASICT division has engaged in developing human resources through various types of training programs related to ICT. In 2023-2024, ASICT Division successfully conducted seven training programs on e-governance and the Right to Information Act 2009, and a workshop on Innovation Showcasing.

Table 12. List of training/workshops conducted at ASICT division during 2023-2024

SN	Course Title	No. of training	Duration (day)	No. of participant	Venue
1.	Training on implementation of e-Governance Action Plan	2	01	40	SRC, Shibgonj, Bogura
2.	Workshop on Innovation Showcasing	1	01	80	Seminar Room, FMPE Division
3.	Training on Right to Information Act 2009	2	01	40	Biometrical Lab, ASICT Division

বার্ষিক কর্মসম্পাদন চুক্তির **ফসলের উৎপাদন ও উৎপাদনশীলতা বৃদ্ধি** সংক্রান্ত কার্যক্রমের অগ্রগতি প্রতিবেদন

দপ্তর বা সংস্থার নাম: **বাংলাদেশ কৃষি গবেষণা ইন্সটিটিউট, গাজীপুর-১৭০১**

অর্থবছর: ২০২৩-২৪ (জুলাই, ২০২৩ – জুন, ২০২৪)

ফসলের উৎপাদন ও উৎপাদনশীলতা বৃদ্ধি

ক্রম	কার্যক্রম	কর্মসম্পাদন সূচক	কর্মসম্পাদন সূচকের মান	লক্ষ্যমাত্রা (অসাধারণ)	একক	১ম ত্রৈমাসিক অগ্রগতি	২য় ত্রৈমাসিক অগ্রগতি	অর্ধবার্ষিক অগ্রগতি (১ম+২য়)	৩য় ত্রৈমাসিক অগ্রগতি	৪র্থ ত্রৈমাসিক অগ্রগতি	বার্ষিক চূড়ান্ত অগ্রগতি (১ম+২য়+৩য়+৪র্থ)	দাবীকৃত নম্বর
০১	০২	০৩	০৪	০৫	০৬	০৭	০৮	০৯	১০	১১	১২	১৩
০১	১.২ উদ্ভাবিত জাত এবং প্রযুক্তির সম্প্রসারণ	১.২.৮ বার্ষিক গবেষণা রিপোর্ট প্রকাশিত	১৫	১	সংখ্যা	-	৯/১০/২৩	৯/১০/২৩	-	-	৯/১০/২৩	১৫
		১.২.৯ লিফলেট, নিউজলেটার, বুকলেট, জার্নাল ইত্যাদি প্রকাশিত	৫	১	সংখ্যা	-	১	১	-	১	২	৫
		১.২.১১ আয়োজিত মেলায় অংশগ্রহণ	৫	১	সংখ্যা	-	১	১	-	-	১	৫
											সর্বমোট দাবীকৃত নম্বর =	**২৫**

বার্ষিক কর্মসম্পাদন চুক্তির **ই-গভর্ন্যান্স ও উদ্ভাবন** সংক্রান্ত কার্যক্রমের অগ্রগতি প্রতিবেদন

দপ্তর বা সংস্থার নাম: **বাংলাদেশ কৃষি গবেষণা ইন্সটিটিউট, গাজীপুর-১৭০১**

অর্থবছর: ২০২৩-২৪ (জুলাই, ২০২৩ – জুন, ২০২৪)

ই-গভর্ন্যান্স/উদ্ভাবন পরিকল্পনা ও বাস্তবায়ন												
ক্রম	কার্যক্রম	কর্মসম্পাদন সূচক	কর্মসম্পাদন সূচকের মান	লক্ষ্যমাত্রা (অসাধারণ)	একক	১ম ত্রৈমাসিক অগ্রগতি	২য় ত্রৈমাসিক অগ্রগতি	অর্ধবার্ষিক অগ্রগতি (১ম+২য়)	৩য় ত্রৈমাসিক অগ্রগতি	৪র্থ ত্রৈমাসিক অগ্রগতি	বার্ষিক চূড়ান্ত অগ্রগতি (১ম+২য়+৩য়+৪র্থ)	দাবীকৃত নম্বর
০১	০২	০৩	০৪	০৫	০৬	০৭	০৮	০৯	১০	১১	১২	১৩
০১	[১.১] সেবা সহজিকরন / ডিজিটাইজেশনের মাধ্যমে উদ্ভাবনী ধারণা বাস্তবায়ন	[১.১.১] সেবা/অফিস ব্যবস্থাপনা সহজিকরন/ডিজিটাইজেশনের মাধ্যমে ন্যূনতম একটি উদ্ভাবনী ধারণা বাস্তবায়িত	১০	১৬/০৩/২০২৪	তারিখ	-	-	-	৬/৩/২৪	-	৬/৩/২৪	১০
০২	[২.১] ইতঃপূর্বে বাস্তবায়িত সহজিকৃত ও ডিজিটাইজকৃত সেবা চালু অব্যাহত রাখা	[২.১.১] ইতঃপূর্বে উদ্ভাবিত/সহজিকৃত/ডিজিটাইজকৃত সেবাসমূহের ডেটাবেজ হালনাগাদকরন ও ডেটাবেজের সেবাসমূহ অব্যাহত রাখা	৫	৪	সংখ্যা	১	১	১	১	১	৪	৫
০৩	[৩.১] ইনোভেশন শোকেজিং	[৩.১.১] আওতাধীন অফিসসমূহের অংশগ্রহণে ন্যূনতম একটি ইনোভেশন প্রদর্শনী (শোকেসিং) আয়োজিত এবং শ্রেষ্ঠ উদ্ভাবনী উদ্যোগ নির্বাচিত	৮	০৯/০৫/২০২৪	তারিখ	-	-	-	-	২৮/০৪/২০২৪	২৮/০৪/২০২৪	৮
০৪	[৪.১] ই-নথির ব্যবহার বৃদ্ধি	[৪.১.১] ই-ফাইলে নোট নিষ্পত্তিকৃত	৮	৮০%	%	৯৮.৫০%	৯৭.৬৭%	৯৮.০৮৫%	৯৭.৫৫%	৯৭.০০%	৯৭.৮৫%	৮
০৫	[৫.১] তথ্য বাতায়ন হালনাগাদকরন	[৫.১.১] তথ্য বাতায়ন হালনাগাদকৃত	৭	৪	সংখ্যা	১	১	২	১	১	৪	৭
		[৫.১.২] আওতাধীন অফিসসমূহের তথ্য বাতায়ন হালনাগাদকরন নিশ্চিতকরন	৩	৪	সংখ্যা	১	১	২	১	১	৪	৩
০৬	[৬.১] স্মার্ট বাংলাদেশ বিনির্মাণে কর্মশালা আয়োজন	[৬.১.১] স্মার্ট বাংলাদেশ বিনির্মাণ বিষয়ক কর্মশালা/সভা/সেমিনার আয়োজিত	৫	২	সংখ্যা	-	-	-	-	২	২	৫
		[৬.১.২] স্মার্ট বাংলাদেশ বিনির্মাণ বিষয়ক কর্মপরিকল্পনা প্রণয়নকৃত	৪	২৫/০৩/২০২৪	তারিখ	১২/০৭/২০২৩	-	১২/০৭/২০২৩	-	-	১২/০৭/২০২৩	৪
											সর্বমোট দাবীকৃত নম্বর=	**৫০**

বার্ষিক কর্মসম্পাদন চুক্তির **তথ্য অধিকার** সংক্রান্ত কার্যক্রমের অগ্রগতি প্রতিবেদন

দপ্তর বা সংস্থার নাম: **বাংলাদেশ কৃষি গবেষণা ইন্সটিটিউট, গাজীপুর-১৭০১**

অর্থবছর: ২০২৩-২৪ (জুলাই, ২০২৩ – জুন, ২০২৪)

তথ্য অধিকার সংক্রান্ত কর্মপরিকল্পনা

কর্মসম্পাদনের ক্ষেত্র	মান	কার্যক্রম	কর্মসম্পাদন সূচক	একক	কর্ম সম্পাদন সূচকের মান	লক্ষ্যমাত্রা অসাধারণ	১ম ত্রৈমাসিক অগ্রপতি	২য় ত্রৈমাসিক অগ্রপতি	অর্ধবার্ষিক অগ্রগতি (১ম+২য়)	৩য় ত্রৈমাসিক অগ্রপতি	৪র্থ ত্রৈমাসিক অগ্রপতি	বার্ষিক চূড়ান্ত অগ্রগতি (১ম+২য়+৩য় +৪র্থ)	দাবীকৃত নম্বর
১	২	৩	৪	৫	৬	৯	১০	১১	১২	১৩	১৪	১৫	১৬
প্রাতিষ্ঠানিক	০৬	[১.১] তথ্য অধিকার আইন অনুযায়ী নির্ধারিত সময়ের মধ্যে তথ্য প্রাপ্তির আবেদন	[১.১.১] নির্ধারিত সময়ের মধ্যে তথ্য প্রাপ্তির	%	০৬	১০০%	১০০%	১০০%	১০০%	১০০%	১০০%	১০০%	৬
সক্ষমতাবৃদ্ধি	১৯	[২.১] স্বতঃ প্রণোদিতভাবে প্রকাশযোগ্য সকল তথ্য হালনাগাদ করে ওয়েবসাইটে প্রকাশ	[২.১.১] হালনাগাদকৃত তথ্য ওয়েবসাইটে প্রকাশিত	তারিখ	০৪	০১/১২/২০২৩ থেকে ৩১/১২/২০		১৭/১২/২৩	১৭/১২/২৩	-	-	১৭/১২/২৩	২
						০১/০৫/২০২৪ থেকে ৩১/০৫/২০২৪	-	-	-	-	২৯/৫/২০২৪	২৯/৫/২০২৪	২
		[২.২] বার্ষিক প্রতিবেদন প্রকাশ	[২.২.১] নির্ধারিত সময়ে বার্ষিক প্রতিবেদন	তারিখ	০৫	১৫/১০/২০২৩	-	৯/১০/২৩	৯/১০/২৩	-	-	৯/১০/২৩	৫
		[২.৩] তথ্য অধিকার আইন ও এর বিধিবিধান সম্পর্কে জন সচেতনতা বৃদ্ধিকরণ	[২.৩.১] প্রচার কার্যক্রম সম্পন্ন	সংখ্যা	০৬	৩	১	২	৩	-	-	৩	৬
		[২.৪] তথ্য অধিকার আইন, ২০০৯ ও এর বিধিমালা, প্রবিধানমালা, স্বতঃপ্রণোদিত তথ্য প্রকাশ নির্দেশিকাসহ সংশ্লিষ্ট বিষয়ে কর্মকর্তা/কর্মচারীদের প্রশিক্ষণ	[২.৪.১] প্রশিক্ষণ আয়োজিত	সংখ্যা	০৪	২	-	-	-	১	১	২	৪
											সর্বমোট দাবীকৃত নাম্বার =		**২৫**

APPENDIX I

SCIENTISTS WORKING AT ASICT DIVISION, 2023-2024

Sl. No.	Name	Designation	Discipline	Remarks
1.	Dr. Syed Md. Abdullah Shiblee	CSO	Agricultural Economics	In-charge
2.	Kowshik Kumar Saha	Agril. Eng.	Agril. Engineering	--
3.	Moh. Mukhlesur Rahman	SSO	Statistics	--
4.	Nur Mohammad	SSO	Statistics	--
5.	Kazi Saidur Rahman	SSO	Statistics	--
6.	Istiak Ahmed	SO	Statistics	--
7.	Mohammad Rasel	SO	Statistics	--
8.	Jamila Khatun Prioty	SO	Statistics	--
9.	Md. Shakil Hossain	SO	Statistics	--
10.	Jahangir Hossain,	Asst. Progr	Computer Science	--

APPENDIX II

SUPPORTING STAFF WORKING AT ASICT DIVISION, 2023-2024

Sl. No.	Name	Designation	Remarks
1.	Nazma Islam	Computer Operator	--
2.	Md. Shajahan Morol	Computer Operator	Cashier
3.	Jebun Nahar	Computer Operator	--
4.	Mohammad Abdur Rahim	Office Assistant cum Computer Typist	Storekeeper
5.	Kawser Mahmud	Office Assistant cum Computer Typist	--
6.	Md. Reduanul Islam	Lab Technician	--
7.	Md. Sayed Khan	Statistical Assistant	--
8.	Monir Hossen	Chainman	PA to CSO
9.	Md. Delwar Hossain	Office Assistant	--

APPENDIX III

LABOUR WORKING AT ASICT DIVISION, 2023-2024

Sl. No.	Name	Designation	Remarks
1.	Md. Topan Miah	Data Entry Operator	Regular
2.	Shamima Akhter	Photocopy Operator	Irregular
3.	Md. Parvez Mosharof	Office Assistant	Irregular
4.	Md. Ismail Hossain	Cash Assistant	Irregular
5.	Md. Azizur Rahman	Internet Technician	Irregular
6.	Sohela Akhter Happy	Data Entry Operator	Master roll
7.	Rekha Akhter	Office Assistant	Master roll
8.	Md. Al-Amin	Electrician	Master roll

APPENDIX IV

LIST OF PUBLICATIONS 2023-24

Scientific Journal

Tsoulias, N., **Saha, K. K.**, & Zude-Sasse, M. (2023). In-situ fruit analysis by means of LiDAR 3D point cloud of normalized difference vegetation index (NDVI). *Computers and Electronics in Agriculture*, *205*, 107611.

Singh, N., **Saha, K. K.**, Makaram, P., & Zude-Sasse, M. (2023). Leaf area estimation of strawberry plants using commercial low-cost LiDAR. In *XXXI International Horticultural Congress (IHC2022): III International Symposium on Mechanization, Precision Horticulture, and 1360* (pp. 23-28).

Penzel, M., Tsoulias, N., **Saha, K. K.**, Handtke, N., Gubin, E., & Zude-Sasse, M. (2023). Cherry tree growth in response to varying soil apparent electrical conductivity. In *XXXI International Horticultural Congress (IHC2022): International Symposium on Innovative Perennial Crops Management 1366* (pp. 289-296).

Saha, K. K., & Zude-Sasse, M. (2023). Estimation of chlorophyll content in banana during shelf life using LiDAR laser scanner. *Postharvest Biology and Technology*, *192*, 112011.

Mottalib, M. A., Hossain, M. A., Amin, M. N., **Saha, K. K.**, ... & Alam, M. M. (2024). Agro-Geoinformatics-based approach for predicting requirements of BARI seed planter machine in the southern region of Bangladesh. *International Journal of Engineering Inventions, 13 (2),* 73-89.

Tapia-Zapata, N., **Saha, K. K.**, Tsoulias, N., & Zude-Sasse, M. (2024). A geometric modelling approach to estimate apple fruit size by means of LiDAR 3D point clouds. *International Journal of Food Properties*, *27*(1), 566-583.

Saha, K. K., Weltzien, C., Bookhagen, B., & Zude-Sasse, M. (2024). Chlorophyll content estimation and ripeness detection in tomato fruit based on NDVI from dual wavelength LiDAR point cloud data. *Journal of Food Engineering*, *383*, 112218.

Rahman, M.M., Islam, M.A., Mahboob, M.G., Mohammad, N., **Ahmed, I.** (2024). Forecasting of Potato Yield Estimation by Satellite Based Remote Sensing Technique, *Acta Informatica Malaysia (AIM)*, 8 (2), 45-51.DOI: 10.26480/aim.02.2024.45.51

Mohammad, N., M. A. Islam, M. M. Rahman, I. Ahmed, and M. G. Mahboob. 2023. A model of estimation maize yield based on weather, agronomical and satellite data. Bangladesh Journal of Agricultural Research, 48(4):433-449

Khatun, M. F., Islam, M. S., **Ahmed, I**., Jahan, N., Islam, M. R., Ahmed, Q. M., & Jahan, N. (2023). Assessing Genetic Variability of Snake Gourd (Trichasanthes Cucumerina Var. Anguina L.) Germplasm Through Morphological Characterization and Multivariate Analysis. *International Journal of Smart Agriculture*, 1(1), 40-49.DOI: https://doi.org/10.54536/ijsa.v1i1.2271

Saif, H. B., Ruma, K. F., **Ahmed, I**., Islam, M. R., Akhi, A. H., Nasim, F. A., ... & Karim, M. R. (2024). Generation Change of Cropping Intensity in Bangladesh: A Systematic Review. *Archives of Current Research International*, 24(5), 33-43. DOI: 10.9734/acri/2024/v24i5676

Ahmed, I., Biswas, J. C., Jahan, M. A. H. S., Ishtiaque, S., Zahan, T., Islam, M. A., ... & Hossain, A. (2024). Yield stability among potato varieties suitable for different agroecological regions of Bangladesh. *Heliyon*, 10(11).

Prioty, J. K., Rahman, K. S., & Miah, M. A. M. (2023). Growth and instability analysis of black gram (Vigna mungo L.) in Bangladesh. *Bangladesh Journal of Agriculture*, *48*(2), 30-38. DOI: 10.3329/bjagri.v48i2.70156

Prioty, J. K., Rahman, K. S., & Miah, M. A. M. (2024). Growth and Instability Analysis of Minor Pulses in Bangladesh. *Food Science and Engineering*, 200-207.DOI: 10.37256/fse.5120243708

Prioty, J.K., Rahman, K.S., & Miah, M.A.M. (2023). Growth and Instability Analysis of Lentil in Bangladesh. *Bangladesh Journal of Agricultural Research, 48(2)*, 249-257.

Prioty, J. K., Rahman, K. S., & Miah, M. A. M. (2024). Assessment Of Trend And Instability Of Area And Production Of Mung Bean In Bangladesh. *SAARC Journal of Agriculture*, *22*(1), 181-189.DOI: 10.3329/sja.v22i1.70168

Rapporteur's Report

Internal Research Review and Program Planning Workshop 2024

Agricultural Statistics and Information & Communication Technology Division

Date: 10-09-2024

Comments and Suggestions on Research Report 2023-24

Serial	Comments and Suggestions	Action Taken
1.	Justification for the study **"**Detection of differences in vegetation and chlorophyll content in agricultural field using unmanned aerial vehicles in abstract and introduction section" should be include.	✓ Justification added
	In results and discussion create sub-headings according to the study's objectives and discuss the results incorporating Tables, figures, etc.	It has done accordingly
	Follow the style of the report on the Mustard yield prediction study	It has done accordingly
	Write a conclusion section based on your study findings.	It has done accordingly
	Maps need to be much more visual	It has done accordingly
	Why the negative relationship for NDVI?	Corrected
	More indices (NDVI, SAVI etc.) should be included and best one should be chosen.	It has done accordingly
	Add justification of the study "Forecasting of onion production in Bangladesh using different models and machine learning algorithm" and key findings in abstract.	Justification added
	Avoid excessive theoretical discussion of different models.	It has done accordingly
	Compare your forecasted onion production with the findings shown in a BARC study.	It has done accordingly
	Write a conclusion section based on your study findings.	It has done accordingly
	More indices should be included	It will be followed in future research
	Recheck the maximum yield, NDVI maximum, write up	Checked
	Rewrite the abstract of "Forecasting onion yield by using satellite-based remote sensing technique in Bangladesh". Avoid unnecessary words and sentences. Add key results according to the objectives.	It has done accordingly
	Also run regression model with 2-year average data. Then the result will be compared or validated with 3rd year data. Check this statement "In most cases yields were maximum, but NDVI were not maximum".	It will be followed in next year & checked
	Write a conclusion section based on your study findings.	It has done accordingly
	Introduction of "Combination of Soil Sampling and Drone Mapping Effects on Fertilizer Recommendation for Crop production" is good but too large. Try to avoid unnecessary sentences.	Necessary steps will be taken.

	Illustrate or discuss the percentage of accuracy in Results and Discussion section.	Accuracy will be shown based on second year outcome.
	Write a conclusion section based on your study findings.	Necessary steps will be taken.
	To be done at multiple locations	Not possible as lab test involve.
	Prediction accuracy should be shown	Based on second year data.
	It is not enough to do with only high and low nutrients, if you get the value then it is better	It is not possible based only on multispectral data.
	Work with active remote sensing instead of passive remote sensing to get clear image. Better to work with RADAR data. (Abed sir)	Necessary steps will be taken.
	Add an objective in abstract and introduction section regarding the impact of heatwave on the yield of summer tomatoes.	It has done accordingly
	In Fig 5,of Results and Discussion write the no. of HW in the Y-axis. Create a para on the Limitations of the study. Discuss with the relevant scientists and change tomato variety names.	It has done accordingly
	BARI hybrid tomato-4 and BARI hybrid tomato-8,correct it	It has done accordingly
	Summer tomato planting time June to July correct it.	It has done accordingly
	Heatwave should be classified with the length of the heatwave of days	May be a separate research.
	Use historical data not only 5 years, make it for western region; Jashore district.	Historical data will be taken.
	Better to conduct your study on western region of Bangladesh	It will be done in future research
	Working in collaboration with TCRC for “Predictive modeling of climate change impacts on potato production: a statistical investigation” would be good	Will be done in the following year
	Title should be changed: Use Climatic parameters instead of climate change	Tittle has been changed accordingly
	Potato yield is closely related to disease, hence disease and vegetative parameters may be combined	Will be done in the following year
	Try to collect the responses of farmers or related scientists on the preference of Bt brinjal varieties and discuss it in the report.	This will be addressed in future studies
	Write a conclusion section based on your study findings for “GGE bi-plot analysis for yield performance and stability assessment of BARI released Bt Brinjal varieties”.	The conclusion section has been modified based on the findings
	It should be seen whether the weather parameters can be related according to the area.	Weather parameters such as temperature and rainfall have been included in the report.
	Add justification for the study UAV-SfM-based 3D reconstruction of fruit trees to estimate plant parameters	Justification added
	Please measure all the parameters of a single mango tree and show the comparative results in a Table, and discuss it along with their accuracy.	Will be done in the following year

	Write a conclusion section based on your study findings.	Conclusion has been added
	Density, Volume can be measured or not?	Can be measured
	Fruit yield can be estimated or not?	Will be tried
	Validate the parameters with the real tree to see the accuracy	Already validated for tree height
	Is it possible to linkup ASICT Division with the phenotype system of Spices Research Centre?	Yes, it is possible
	Low cost green house, shade house of BARC should be linked up with ASICT, BARI.	Necessary steps will be taken.
	BARC's sensors are linked up in ASICT Division and utilized	Necessary steps will be taken.
	Software has to be developed to run BARC's sensors	Necessary steps will be taken.
	Can we collaborate with any external universities or BUET who make sensors?	Will try to collaborate in future.
	It should be seen whether the parts can be bought from outside and assembled to make the sensor	Necessary research will be taken.
	Develop your own sensor/software and trial instead of trailing others sensor/software.	Necessary research will be taken.
	Start apply in high value crop/hydroponic/in-garden system	Action already taken.
51.	Work with biotechnologist and oilseed scientists for wet lab validation	Will try to collaborate in future.
52.	The gist finding of bioinformatics analysis of Brassica napus should be extracted and given to the oilseed scientists.	Work with collaboration.
53.	Capacity enhancement should be developed among relevant scientists at BARI to handle OMICS big data/Genome Data and services should be provided.	Action will be taken.
54.	Name of the station in Online "Vromon shuchi" software should be added below the applicant signature	Action should be taken.
55.	Application list in the interface of Online "Vromon shuchi" software should be rearranged as: Local -> Inter-district -> Final application list	Action should be taken.
56.	Technical support for Online "Vromon shuchi" software may be taken from BCC for smooth performance	Action should be taken.

Comments and Suggestions from the Expert Panel for ASICT Division

SN	Comments and Suggestions
1.	Password protected division wise common document/data storage system can be developed for data, research related materials sharing.
2.	A common power point present format for BARI may be prepared and implemented for all Divisions.
3.	To act collectively according to the needs of the nation and to ensure that the work of Remote sensing and GIS section is done with international impact.

4.	Collaboration between the ASICT division and the National Institute of Biotechnology (NIB) should be considered to take the initiative to research bioinformatics and crop genomics for sustainable crop improvement.
5.	Instead of working with passive remote sensing, try working with active remote sensing. Because the image is not clear
6.	Common documents management can be arranged. Information can be online based data, research related anything. May have division based passwords.

Comments and Suggestions on Research Program 2024-25

SN	Comments and Suggestions
1.	Tittle of the program "Crop-type mapping and acreage & production estimation of major crop in dry seasons using remote sensing and machine learning algorithms" can be changed as "Crop typo mapping and yield prediction".
2.	Not only yield can be predicted, area also has to be estimated for program "Crop typo mapping and yield prediction".
3.	The title must be changed, make it statistical basis for "Foreseeing the field emergence of Bt eggplant seed lots unifying with the vigor test**"**
4.	Selected or assessment must be added to the title "Efficiency of Agricultural Practices for Food Security in Barind Region of Bangladesh" and "Efficiency of Agricultural Practices for Food Security in Char Land Region of Bangladesh"