iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71198

Satellite Soil Moisture Data-Driven to Predict Paddy Production Using Machine Learning Approach

Rebina Ferdous[1], Md Mehedi Imam[2], Md. Hafizur Rahman[3]
[1]Basic & Applied Research on Jute Project, Bangladesh Jute Research Institute, Dhaka-1207, Bangladesh
[2]Department of Computer Science and Engineering, Jagannath University, Dhaka-1100 Bangladesh
[3]ICT cell, Jagannath University, Dhaka-1100, Bangladesh

Abstract: Measuring and mapping how wet the soil is accurately are very important for farming in a sustainable way, managing water resources, and studying the climate. On top of that, it also important to find out that, based on soil moisture of a land, what could be the potential paddy production. This study aims to explore the mapping of soil moisture of Rajshahi district in Bangladesh using multi-spectral remote-sensing satellite images and based on historical paddy production data, what could be the possible paddy production in the year 2025. To find out the possible production in the year of 2025, I have used the Machine Learning techniques to predict the paddy production based on the data generated from the soil moisture map and integrated the data with the paddy production statistics data of Bangladesh Bureau of Statistics (BBS). A total of 14 Landsat scenes covers the entire Rajshahi district. Thus, a set of Landsat imagery (a total of 14 scenes) for the year 2017, 2020 and 2023 was used in this study to map the soil moisture of Bangladesh through the application of Geographical Information System (GIS) and Remote Sensing. Satellite Image preprocessing, correction, and analysis were done with the ArcGIS software (version 10.8, developed by Environmental Systems Research Institute, USA) and prediction was found out through Google Colab. The map in this study indicates the level of soil wetness in Rajshahi, ranging from Very Dry, Dry, Wet to Very Wet soils. Based upon the study, it is found out that, it is possible to increase paddy production rate by 47.95% in the year 2025 due to improvement of soil moisture conditions.
Index Terms: Key words: Agriculture, GIS, Landsat, Remote Sensing, Satellite Image, Soil Moisture Dynamics ,Paddy Production, Machine Learning.

I. INTRODUCTION

This study investigates how soil moisture affects paddy production in Rajshahi, Bangladesh, using GIS and remote sensing techniques. It maps the soil wetness levels from Landsat images of 2017, 2020 and 2023, and predicts the paddy production for 2025 using machine learning and BBS data. The results show that soil moisture improvement can increase paddy production by 47.95% in 2025.

Using of GIS and remote sensing to map soil moisture in Bangladesh from multispectral Landsat images of 2022. The map aims to help the government improve agriculture and rural development. The study also analyzes the long-term and seasonal changes of soil moisture and their impact on climate modeling. The study provides accurate and timely spatial soil moisture data for environmental and socioeconomic purposes. The map classifies soil moisture into four levels: very wet, wet, dry, and very dry. The land cover and soil classifications have high accuracy and consistency [1]. From the above study, they did not demonstrate the recent information about soil moisture level of drought, non-aqueous surfaces, moderate drought, non-aqueous surfaces, humidity and water surface of the Rajshahi district land and also missing the main point of that the prediction of possible paddy production of the lands. To predict the paddy production rate of 2025, there is also missing of this massive part. There are several ways to find out paddy production rate of a season. Soil Moisture level runs a significant role in this study. Our proposed techniques help to determine the showcase soil moisture map of different times of different year and prediction of earlier paddy production rate.

II. PROBLEM STATEMENT

The previous research used LANDSAT 8 satellite images and GIS to map soil moisture in Bangladesh. Soil moisture data is important for disaster, environmental, and hydro-logical purposes. The study shows that multi-spectral remote sensing can map soil moisture accurately and reliably. The method can be used in other areas with similar conditions.

But in our research, it is found out that the soil moisture level of land changes of a time. In 12 months, there are significant changes of soil moisture level in different time of the year. For example, in the summer season (aus Season), there is a lowest reading of moisture level and in the raining season, the soil moisture level is quite high. When the soil moisture level is quite high, it is always better for to seed any crops. In the aus season, it is quite low of paddy production since the soil moisture level is quite low. Here we have also observed that in recent years, there is an improvement of soil moisture level. For example, in the year 2017, the moisture level was quite low, for this reason, paddy production was low. In the year 2020, soil moisture level was better than the soil moisture level of 2017. So, there is also an improvement in paddy production on that year. In the year of 2023, the soil moisture level was better than ever. According to BBS [2], it was that last highest rate of paddy production on this year. Our goal is, since the soil moisture level is improving day by day, it is also possible to predict the paddy production rate of the year of 2025.

III. RELATED WORKS

This study was conducted using data collected from agriculture farms anywhere in Bangladesh. We learned more about the data collection process and the analysis of the data of this study. Based on the review of the preliminary analysis of the data and what was previously reported data in the soil moisture is selected as the primary factor in training the ML algorithms. In the study, it was found the results of a neural network (NN) to avoid over-fitting on a small dataset to an ensemble of neural networks (ENN), which helps build a more stable algorithm at the expense of error rate by using machine learning method.

A. Related Works on Soil Moisture Map

The study used multi-spectral satellite images to map soil moisture in Bangladesh, which is important for sustainable farming, water management, and climate studies. Those study aimed to produce a soil moisture map that can assist the government in improving agriculture and rural development. Those study used 14 Landsat [3] scenes covering the whole country for the year 2022, and applies GIS and remote sensing techniques to process, correct, and analyze the satellite images. The study also examines the long-term and seasonal changes of soil moisture in Bangladesh and their implications for climate modeling. The study provides accurate and up-to-date spatial soil moisture data at low cost and time, which is useful for environmental modeling, risk assessment, decision-making for various government and development agencies, and socio-economic development. The study classifies soil moisture into four levels: very wet, wet, dry, and very dry [4]. The study achieved high accuracy and consistency for land cover and soil classifications using Random Forest and maximum likelihood methods. [1]

This paper presents a data-driven method for developing PA solutions for collecting and modeling data. It uses soil moisture, a crucial factor for crop growth, as an example to show the effectiveness of our method. On the collection side, we design a reactive wireless sensor node that uses MicaZ mote and VH400 soil moisture sensor to measure the changes of soil moisture. We test the device on field soil to show its performance and sensitivity. On the data analysis side, we create a unique, site-specific soil moisture prediction framework based on machine learning models using SVM and RVM. The framework forecasts soil moisture n days ahead using the same soil and environmental features that our sensor node can collect. Because of the large data size needed by the machine learning methods, we use the Illinois historical data, not the sensor data from the field, to evaluate our framework. It achieves low error rates (15%) and high correlations (95%) between the predicted and the actual values for 9 different sites when predicting soil moisture for about 2 weeks ahead. We also show that the prediction outputs can stay accurate for a long time (one year) when we feed reliable data to the model every 45 days. [5]

The other study aims to predict soil moisture accurately for water management and farm operations. Soil moisture depends on many soil, crop, and weather factors, which makes it hard to model mathematically. We tested different machine learning methods for predicting soil moisture in the RRVN. The methods we used were CART, RFR, BRT, MLR, SVR, and ANN. The goal of this study was to compare the performance of these methods and assess the importance of predictors. The best methods were RFR and BRT. [6]

Soil moisture affects agricultural production and hydro-logical cycles, and its accurate prediction is essential for water resource use and management. However, soil moisture has complex structural and meteorological factors, and it is hard to model mathematically. Current prediction models have issues such as accuracy, generalization, and multi-feature processing, and they need to improve. For this reason, we used the Beijing area as a case study and proposed a deep learning regression network with big data fitting ability to build a soil moisture prediction model. We integrated the dataset, analyzed the time series of the predictors, and used the Taylor diagram to show the relationship between features and predictors. We chose some meteorological parameters that can give effective weights for moisture prediction. The test results show that the deep learning model works well for soil moisture prediction. It can use more input features and still predict soil moisture data accurately. It also provides a useful theoretical foundation for water-saving irrigation and drought prevention. [7]

B. Related Works on Machine Learning

Agriculture is crucial for the economic development of any country. However, the growing population, changing climate, and limited resources make it hard to meet the food demand of the current population. Precision agriculture or smart farming is a new tool to tackle the current challenges in agricultural sustainability. This tool is powered by machine learning (ML), which enables the machine to learn without being programmed. ML and IoT (Internet of Things) enabled farm equipment are the main elements of the next agriculture revolution. In this article, authors review the ML applications in agriculture. The areas they focus on are soil parameter prediction such as organic carbon and moisture content, crop yield prediction, disease and weed detection in crops and species detection. They also review ML with computer vision for crop image classification to monitor crop quality and yield. This method can be used for better livestock production by predicting fertility patterns, diagnosing eating disorders, cattle behavior using ML models and collar sensor data, etc. They also review intelligent irrigation and harvesting techniques that reduce human labour. This article shows how knowledge-based agriculture can enhance the sustainable productivity and quality of the product. [8].

This study presents the second stage of a system that uses modern hardware and software tools to monitor and control soil conditions accurately. The system collects data from the sensor network in the soil of a strawberry farm to estimate the final physicochemical properties of the fruit at harvest time near the sensor locations. The paper explores neural networks and Gaussian process regression models together to predict the most important physicochemical attributes of strawberry. For example, color, alone or with soluble solids content (sweetness), can be predicted with only 9% and 14% error, respectively. This accuracy will allow the next stage of the system to control the soil conditions and optimize quality and resource use for sustainable and high-quality strawberry production. [9]

In another study, soil moisture is important for agriculture, ecosystems, and some natural disasters driven by water. However, monitoring data is affected by instrument noise, extreme values, and changing response to rainfall when ground conditions change. Also, current soil moisture models are not good at predicting for more than a few hours. To make better predictions, we propose two datadriven models, the NAR and the AEAR. These models are based on deterministic, physical hydrology, and we test their ability to predict soil moisture for longer than a few hours. The model parameters reflect the physical processes of gravity and suction that redistribute water in unsaturated soil. We test our models with soil moisture and rainfall data from a steep, post-fire site in southern California. Data analysis is hard because of fast landscape change in steep, burned slopes after small or moderate rain events. The NAR and AEAR models are competitive with other existing and new models in prediction experiments. The AEAR model works well for three different soil textures at different depths (5, 15, and 30 cm). We also show similar strong results in controlled, lab experiments. Our AEAR model has easyto-understand hydro-logic parameters and makes more accurate predictions than other models for 10–24 h. These longer warnings for natural disasters like floods and landslides can help save lives and property. [10]

Another work demonstrates how to handle diverse information and data from real datasets that gather physical, biological, and sensory values. As productive companies public or private, large or small want to increase profitability and reduce costs, finding suitable ways to use data that are constantly collected and available can be the right option to achieve these goals. The agricultural field is not really resistant to the digital technology and the "smart farm" model is more and more common by using the Internet of Things (IoT) paradigm applied to environmental and historical information through time-series. The aim of this study is the design and implementation of practical tasks, from crop harvest prediction to missing or incorrect sensors data recovery, using and comparing different machine learning techniques to suggest which direction to put efforts and investments. The results show how there are many opportunities for innovation while meeting the demands and needs of companies that want to have a sustainable and optimized agriculture industrial business, investing in both technology and the knowledge and the skilled workforce needed to get the best out of it. [11]

IV. DATA SET, ENVIRONMENTAL SETUP, MATERIALS AND METHOD

This chapter presents the main contribution of the project: the selection of and collect of remote sensing data from United States Geological Survey (USGS). Satellite data has 14 different types of bands and among the bands, we need band 5 and band 6 for our proposed work. In this stage, satellite images are needed to be corrected. After selection and download of the raw data, environmental setup is needed for process the data. The process method is done through a systematic way to determine the materials we need and the initial result we want. In the Materials and Method section, it has been discussed about the preprocessing, creation of subset of bands, Mosaic Path and Rows, Spatial Analysis and Soil Moisture Calculations.

A. Remote Sensing Data

In the study, we used 14 in total scenes which were collected from the Landsat 8 [3] Operational Land Imager (OLI) sensor with the help of the Landsat 8 satellite.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

This sensor captures the imagery of the Earth's surface including bi-dimensional bands. OLI mainly captures the reflectance and radiance of the object. After that, this imagery has been digitally analyzed.

Our dataset scenes cover paths 135-139 and rows 42-46, covering the whole of Bangladesh. Inside this covering, we have selected the Rajshahi district in polygon shape. This polygon shape covers the all-around area of Rajshahi. The Landsat 8 image used in the study was downloaded from USGS Earth Explorer site http://earthexplorer.usgs.gov, with ¡10% cloud cover. The spatial resolution varies with wavelength: 15 m in the OLI Panchromatic band, 30m in the OLI Multi-spectral bands, and 100 m in the TIRS Thermal bands (resampled to 30 m to match multi-spectral bands).

B. Satellite Image Correction

We used ENVI 5.1 Radiation Correction tools to adjust the soil moisture data. We also used methods and formulas that make Landsat data better for radiometric calibration. It changes Digital Number to brightness using calibration numbers. We did this for the needed bands because the original images are not good for calculating indices. In atmospheric correction, the brightness will be changed to reflectance. This helps to reduce the effect of the atmosphere on Landsat 8 images, especially dust, water vapor, and to make the estimation and classification more precise. We only did this for the visible and near-infrared bands. We got all the data we needed for the calibration and correction process from the header file (metadata) that came with the satellite images.

C. Environmental Setup

It is used Microsoft Windows 11 Enterprise Edition and two software programs, ArcMap and ENVI 5.1, for our study. ArcMap is part of the ArcGIS Desktop suite, which is a commercial GIS software made by Esri. This software can do many things with geographic data, such as creating, managing, analyzing, and visualizing it. It also has many tools with different features to study different kinds of geographic data. This software can be used for many purposes, such as urban planning, natural resource management, and environmental analysis. We used it to study the soil moisture in Rajshahi district area.

We used a software program called ENVI 5.1 to study the Landsat 8 satellite image. This software is made for commercial use and can study remote sensing data, including Landsat images. Harries made this software, which has many features such as data import, visualization, image processing, feature extraction, classification, change detection, geospatial analysis, GIS integration, and customization automation. We needed a lot of computing power to study the satellite image with ArcMap and ENVI, so we used Microsoft Windows 11 Enterprise Edition desktop.

D. Materials and Method

We used two bands (Band-5 NIR and band-6 SWIR) from Landsat-8 to measure soil moisture. Fig. 1 shows the steps of our method. Fig. 2 shows where Rajshahi district is on the world map. It is between 24.10 and 24.40 north and 88.20 and 90.00 east. Fig. 1 shows the diagram of our model. We used the data from Landsat 8 OLI. This data has many bands, but not all of them are useful. We used PCA to choose the best bands and remove the rest. Then we used ENVI to process the data. We joined the bands together to make a smooth mosaic. Then we used spatial analysis to find the pattern and relationship in the geographic data.

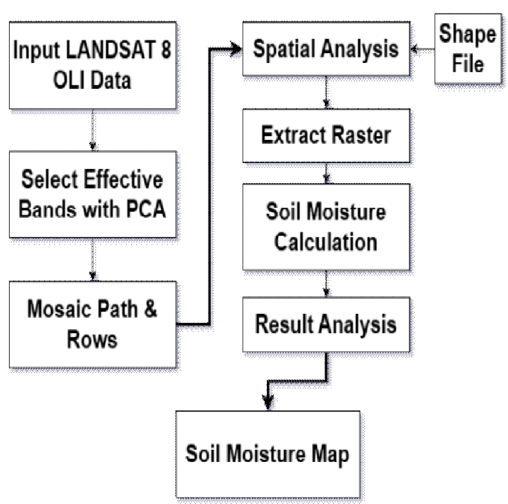


FIGURE 1. Working flow architecture.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

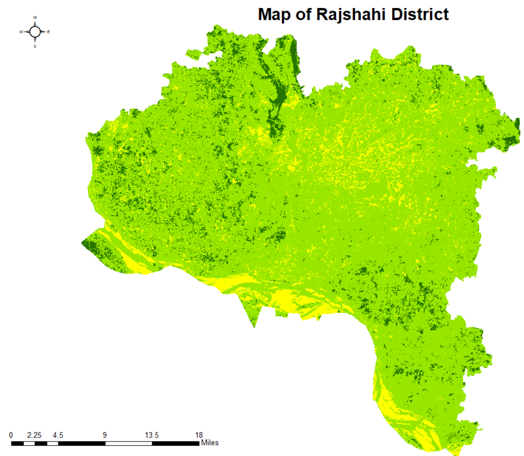


FIGURE 2. Rajshahi district is shown in RGB of Landsat 8 image of 2023.

1) Preprocessing

Satellite Raw images needed to be clipped to remove the background black color from the images. This can be done by the raster clip tool in Data Management of ArcToolbox in ArcMap. An example is shown in Fig. 3.

2) Creating a Subset of Bands

PCA is a powerful algorithm that can reduce the number of features and choose the most important ones from a large set of features. It can pick the best number of bands from the multi-band images with many dimensions by eliminating the dependence between the bands. In our situation, the output images of the data mostly focused on the first few bands of the image. We need to find the first few bands to select the fine details of the image. Also, we need to improve the first few bands and remove the other bands. In our research, we used PCA in the ArcMap to choose the best few bands by looking at the first few bands and getting rid of the other bands.

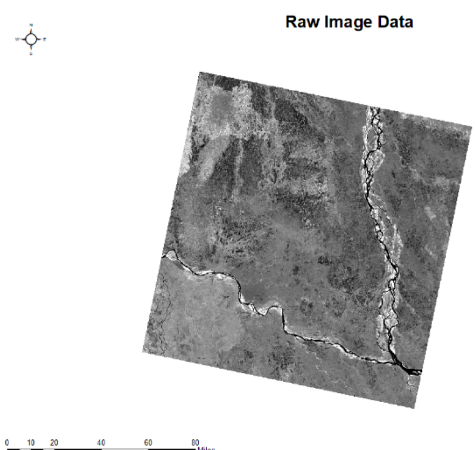


FIGURE 3. Raw Image Data.

3) Mosaic Path and Rows

We have to join all the cut images of each band separately to make one raster of our research area. Mosaics are mixtures or joins of two or more images. Fig. 4 is shown as mosaic form of Fig. 3. With ArcGIS, we can join many raster data sets to make one raster data set. We have taken out the Area of Interest (AOI) from the result of the mosaic image using the shape file of Bangladesh.

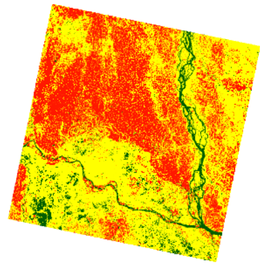


FIGURE 4. Colorful Mosaic of Raw Image Data.

4) Spatial Analysis

This has been done using the Extract by mask of Spatial Analysis Tools in ArcGIS. The shape file was downloaded from GeoDASH (https://geodash.gov.bd) [12] website. We need to use a multiband image layer for image classification. The Composite Bands function in ArcGIS allows us to combine raster to form a multi-band image. Therefore, we need to use the Combine Bands tool to load individual bands into a new multiple-band image.

5) Soil Moisture Calculation

To create a Soil Moisture Map, Near Infrared (NIR) and Short Wavelength Infrared (SWIR) were used which are Band-5 and 6 of Landsat-8 images, respectively. It is calculated as a ratio between the Near Infrared (NIR) and Short Wavelength Infrared (SWIR) values in traditional fashion as follows:

$$SoilMoisture = \frac{NIR - SWIR}{NIR + SWIR} = \frac{Band\ 5 - Band\ 6}{Band\ 5 + Band\ 6}$$

V. RESULT ANALYSIS

To assess the performance of the prediction model we utilized accuracy as evaluation metrics. The metrics mentioned are the standard evaluation measures utilized to assess the performance multi-class classification datasets. The outputs of Soil Moisture and Soil Classifications are shown in maps. This section also includes the analysis of the results including the integration of paddy production rate of Rajshahi district by the year of 2017, 2020 and 2023, trend analysis, feature selection, model training and the prediction of the paddy production of the year of 2025.

We used two bands (NIR Band 5 and SWIR Band 6) from Landsat-8 to make the Soil Moisture Map. We used a soil moisture equation to do the calculation. The calculation gives a fair accuracy for the Soil Moisture map. The map shows how wet or dry the soil is in Rajshahi district. It is harder to estimate soil moisture when the land has a lot of plants or snow and when the land has big changes in height. The best results are when there is little or no soil cover, especially when the land is flat. It is hard to estimate soil moisture when the land has plants or snow on it. Also, big changes in height make it harder to estimate. Soil moisture estimation works better when the land is flat and does not have snow, plants, or forests.From the study of Soil Moisture Level of Rajshahi district for the year of 2017 is quite low in comparison with value of other two years. The data is classified into 4 categories. These are: Very Dry, Dry, Wet and Very Wet. This level is classified with grid Code following from 1, 2, 3 and 4. Here we have found the Very Dry area is 0.0576 Hectares which is the 0.000024% of the total land of Rajshahi district. 31087.13 Hectares area is classified as Dry area which is 12.76498% of the total area. Best area for paddy production is found in the Wet level which is classified as the grid Code level 3. Here we found the 158465.1 Hectares area is suitable for paddy production which is 65.06888% of the total land area. The last one is the Very Wet area, classified as grid Code 4. 53982.12 Hectares area is found Very Wet which is around 22.16611% of total land area, additionally, which is not also suitable for paddy production. All of this numbers are visualized through the Soil Moisture map of Rajshahi district for the year of 2017 in the fig 5 down below.

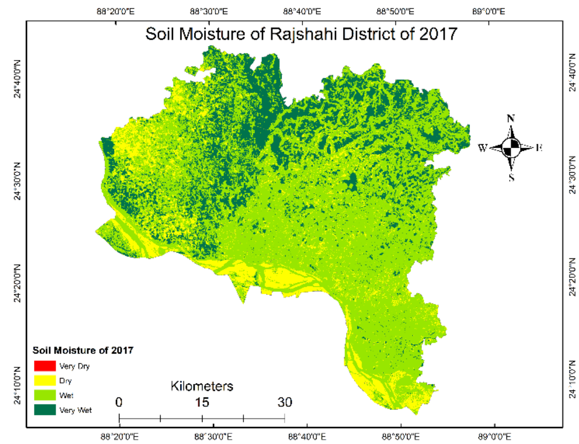


FIGURE 5. Soil Moisture of Rajshahi District of 2017

From the study of Soil Moisture Level of Rajshahi district for the year of 2020 is lower in comparison with value of year 2017. The data is also classified into 4 categories. These are: Very Dry, Dry, Wet and Very Wet. This level is classified with grid Code following from 1, 2, 3 and 4. Here we have found the Very Dry area is 2.042123 Hectares which is the 0.000839% of the total land of Rajshahi district. 43602.03 Hectares area is classified as Dry area which is 17.90404% of the total area which is also 5% increase of Dry area. Best area for paddy production is found in the Wet level which is classified as the grid Code level 3. Here we found the 148037.7 Hectares area is suitable for paddy production which is 60.78782% of the total land area which is 5% less than the value of the year of 2017. The last one is the Very Wet area, classified as grid Code 4. 51890.06 Hectares area is found Very Wet which is around 21.3073% of total land area, additionally, which is not also suitable for paddy production. Overall, the quality of soil moisture is decreased in comparing with soil moisture quality of the year of 2017. All of this numbers are visualized through the Soil Moisture map of Rajshahi district for the year of 2020 in the fig 6 down below.

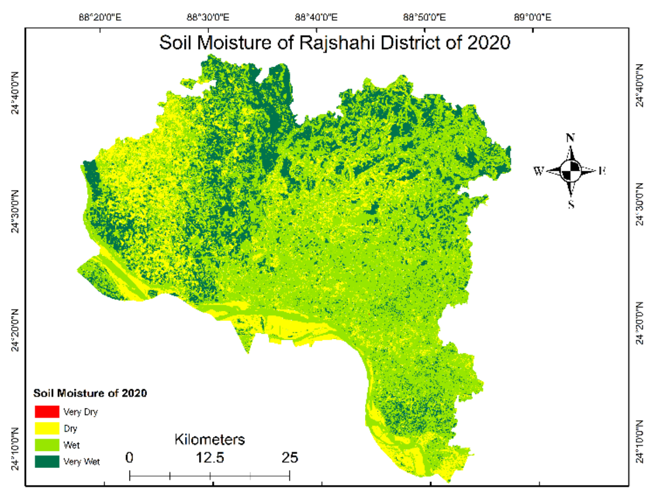


FIGURE 6. Soil Moisture of Rajshahi District of 2020

From the study of Soil Moisture Level of Rajshahi district for the year of 2023 is higher in comparison with value of year 2017 and 2020. The data is also classified into 4 categories. These are: Very Dry, Dry, Wet and Very Wet. This level is classified with grid Code following from 1, 2, 3 and 4. Here we have found the Very Dry area is 0.057634 Hectares which is the 0.000024% of the total land of Rajshahi district. 23296.27 Hectares area is classified as Dry area which is 9.565818% of the total area which is also 8% decreased of Dry area. Best area for paddy production is found in the Wet level which is classified as the grid Code level 3. Here we found the 200467 Hectares area is suitable for paddy production which is 82.31493% of the total land area which is 22% more than the value of the year of 2017. The last one is the Very Wet area, classified as grid Code 4. 19773.29 Hectares area is found Very Wet which is around 8.119227% of total land area, additionally, which is not also suitable for paddy production.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Overall, the quality of soil moisture is increased in comparing with soil moisture quality of the year of 2017 and 2020. All of this numbers are visualized through the Soil Moisture map of Rajshahi district for the year of 2023 in the fig 7 down below.

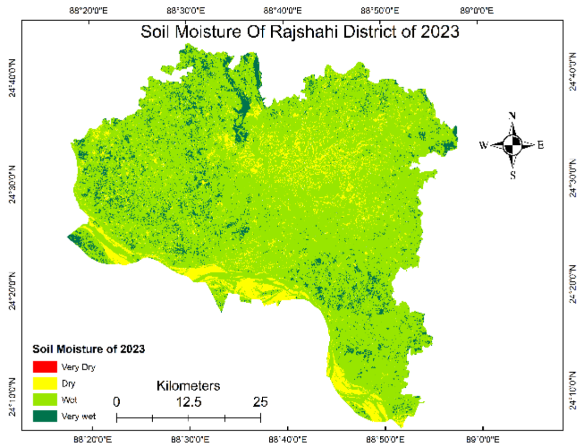


FIGURE 7. Soil Moisture of Rajshahi District of 2023

The study compares the soil moisture level of Rajshahi district in Bangladesh for the years 2017, 2020, and 2023. These maps focus on the Wet category, which is the best for paddy production, and has a grid Code of 3. The soil moisture map of Wet shows that the Wet area was the largest in 2023 (82.31%), followed by 2017 (65.07%), and then 2020 (60.79%). Here the study also indicates that the Dry and Very Wet areas changed over the years, while the Very Dry area remained almost the same. The text concludes that the soil moisture quality improved in 2023 compared to 2017 and 2020. Here is a map of Wet level of Rajshahi District for the year of 2017, 2020 and 2023 in the fig 8, 9 and 10.

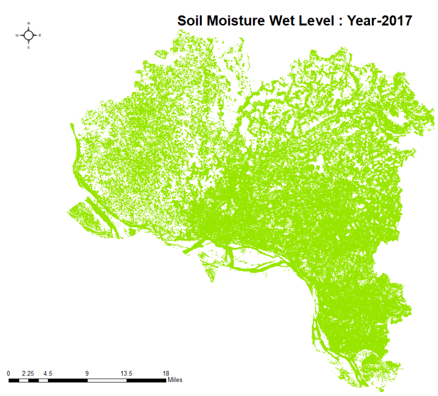


FIGURE 8. Soil Moisture Wet Level: Year 2017

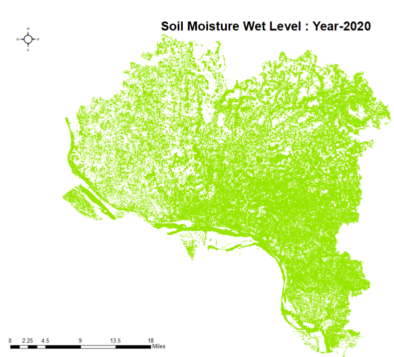


FIGURE 9. Soil Moisture Wet Level: Year 2020

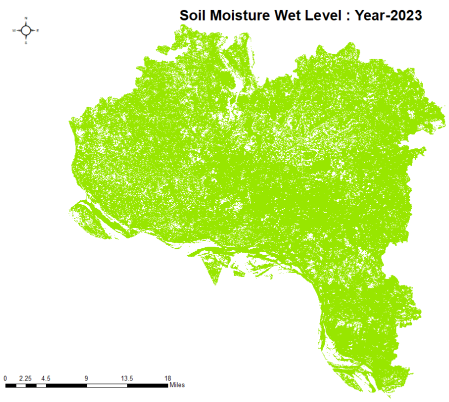


FIGURE 10. Soil Moisture Wet Level: Year 2023

Here the maps compare the Dry area of Rajshahi district in Bangladesh for the years 2017, 2020, and 2023. These maps shows that the Dry area was the smallest in 2023 (9.57%), followed by 2017 (12.76%), and then 2020 (17.90%). Here study also indicates that the Dry area, which is increased in 2017 and again decreased in 2023, while the Very Dry area remained almost the same. The result concludes that the soil moisture quality improved in 2023 compared to 2017 and 2020.

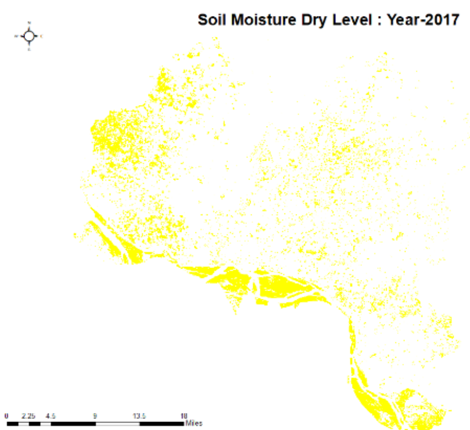


FIGURE 11. Soil Moisture Dry Level: Year 2017

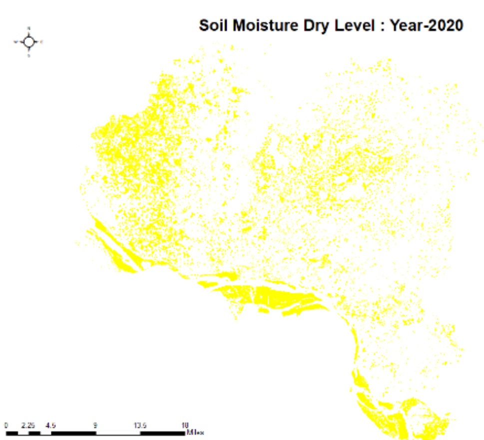


FIGURE 12. Soil Moisture of Rajshahi District of 2017

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

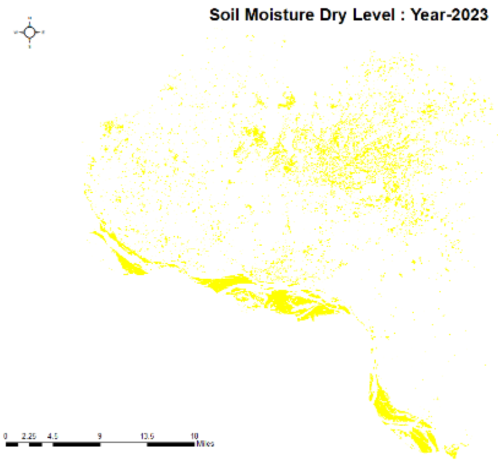


FIGURE 13. Soil Moisture of Rajshahi District of 2017

VI. PADDY PRODUCTION PREDICTION

For the prediction of paddy production for the year of 2025, production rate of paddy for the year of 2017, 2020 and 2023 is collected from Bangladesh Bureau of Statistics (BBS) and integrated the BBS data with the data of soil moisture map data. In this section, it has been analyzed the production trend and analysis, feature selection, model training and prediction of potential paddy production for the year of 2025.

1) Trend Analysis

Paddy production for 2017 was 103737.00 metric tons while land area was used 38824.00 hectares. In the year of 2020, production rate was improved by a significant margin. While 47295.00 hectares of land was used, paddy production was increased by 22.79% which is approximately 134357.00 metric tons. In recent year of 2023, 45384.00 hectares of land was used but the paddy production reduced by 0.48% due to a significant increase of dry land in the year of 2020 and the production rate is 133706.00 metric tons. The fig. 14 down below represents the situation of paddy production by year.

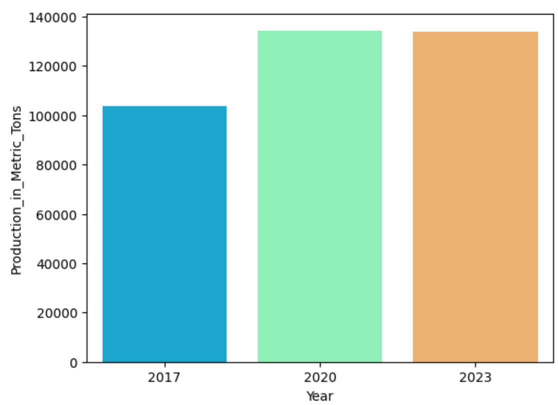


FIGURE 14. Paddy Production Trend over the Years.

2) Feature Selection

Feature selection is essential in machine learning, focusing on the most impactful features from many variables. It prevents over-fitting by removing non-contributing features and boosts accuracy by selecting those closely linked to the target variable. It also shortens training and prediction times by reducing data complexity. Wrapper methods aid in this selection, and Python's Pandas library simplifies data management, enhancing model clarity.

3) Analysis by Grid Code and Visualization

Analyzing the data of cultivated area and production in metric tons based on the grid Code value. The grid Code is a categorical variable that indicates the level of soil moisture, from 1 (Very Dry) to 4 (Very Wet). We filter the data into two subsets: one where grid Code is 3 (Wet) and one where grid Code is not 3. Then, we understand the descriptive statistics of the two subsets, such as the count, mean, standard deviation, minimum, maximum, and quarterlies.The result of this analysis shows that the Wet area (grid Code 3) has a higher mean and standard deviation of both cultivated area and production than the other areas (grid Code not 3). This suggests that the Wet area is more suitable for paddy production than the other areas. The result also shows that the Wet area has a smaller range of values than the other areas, indicating less variability in the data. The result is visualized using a box plot to compare the distributions of the two subsets in the fig. 15, 16 and 17

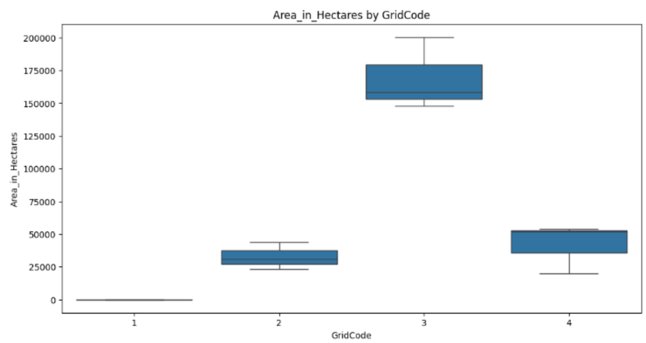


FIGURE 15. Total Area Analysis by grid Code.

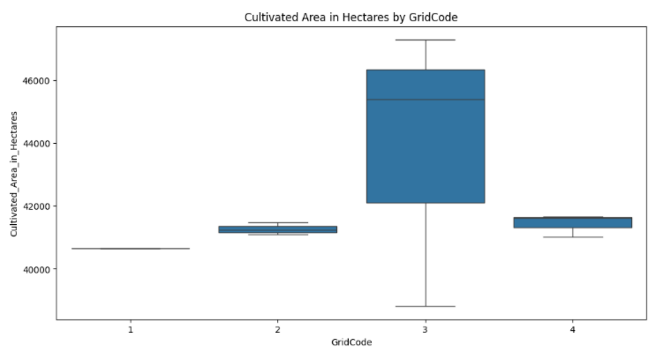


FIGURE 16. Cultivated Area Analysis by grid Code.

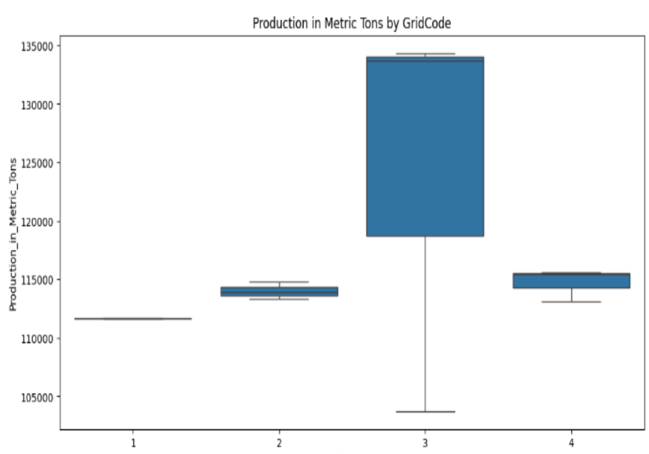


FIGURE 17. Production Rate Analysis by grid Code.

4) Soil Moisture Impact

Soil Moisture also impacts on the cultivation. When (Soil Moisture Wet) Wet, the analysis represents the most area was used for cultivation.

In some cases, the analysis represents the highest production rate while the area is Wet. Fig 18 and 19 is shown below.

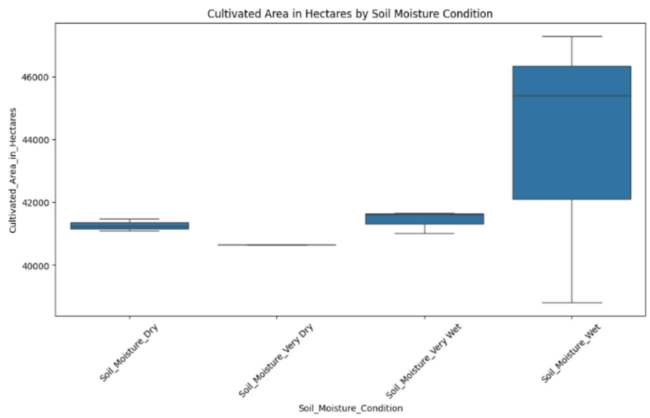


FIGURE 18. Area of Cultivation by Soil Moisture Level.

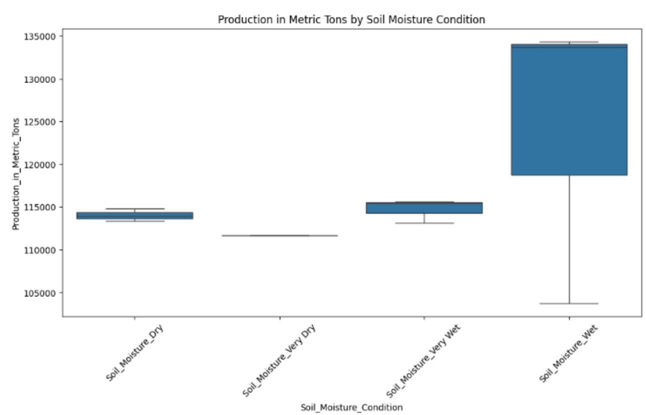


FIGURE 19. Production Rate by Soil Moisture Level.

5) Model Training

The method is used simple linear regression [13], a statistical method that estimates the relationship between a single predictor variable (X) and a response variable (y). It imports the necessary libraries and modules, such as sklearn, NumPy, matplotlib, and seaborn [14]. It assigns the predictor variable (X) and the response variable (y) from a DataFrame (df) that contains the data. The features and target are the names of the columns that store the predictor and response variables, respectively. It splits the data into training and test sets, using 80% of the data for training and 20% for testing. The random state parameter ensures that the same split is reproduced every time the code is run. It creates a linear regression model object (model) and fits it to the training data (X train and y train) using the fit method. This method calculates the coefficients of the linear equation that best fits the data:$y = \beta_0 + \beta_1 X$. The predictions on the test data (X test) using the predict method and stores them in y pred. This method applies the fitted linear equation to the new data and returns the predicted values of y. It calculates the root mean squared error (RMSE) of the predictions, which is a measure of how well the model fits the data. The RMSE is the square root of the average of the squared differences between the actual and predicted values of y. A lower RMSE indicates a better fit. The code prints the RMSE value to the console.

6) Prediction and Visualization

Multiple linear regressions, a statistical method that estimates the relationship between one or more predictor variables and a response variable. By creating a new DataFrame (df 2025) with the values for 2025, based on the predictor variables: Area in Hectares,

Cultivated Are in Hectares, grid Code, and Soil Moisture. It makes a prediction for 2025 using the model object (model) that was previously fitted to the original data. The model object contains the coefficients of the linear equation that best fits the data:
The predicted production in 2025 is 183354.0926 metric tons, which is 47.95% higher than the actual production in the cultivated area. The comparison of the actual and predicted production is shown in the following graph Fig. 20:

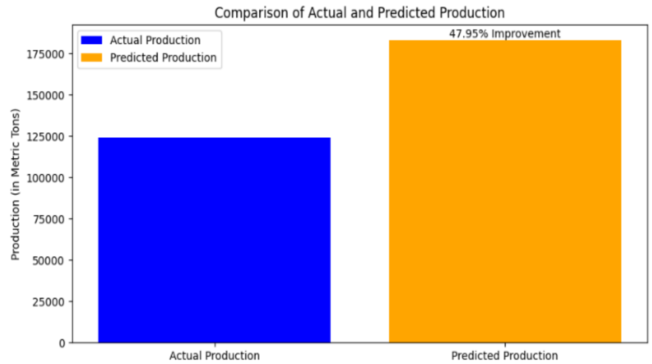


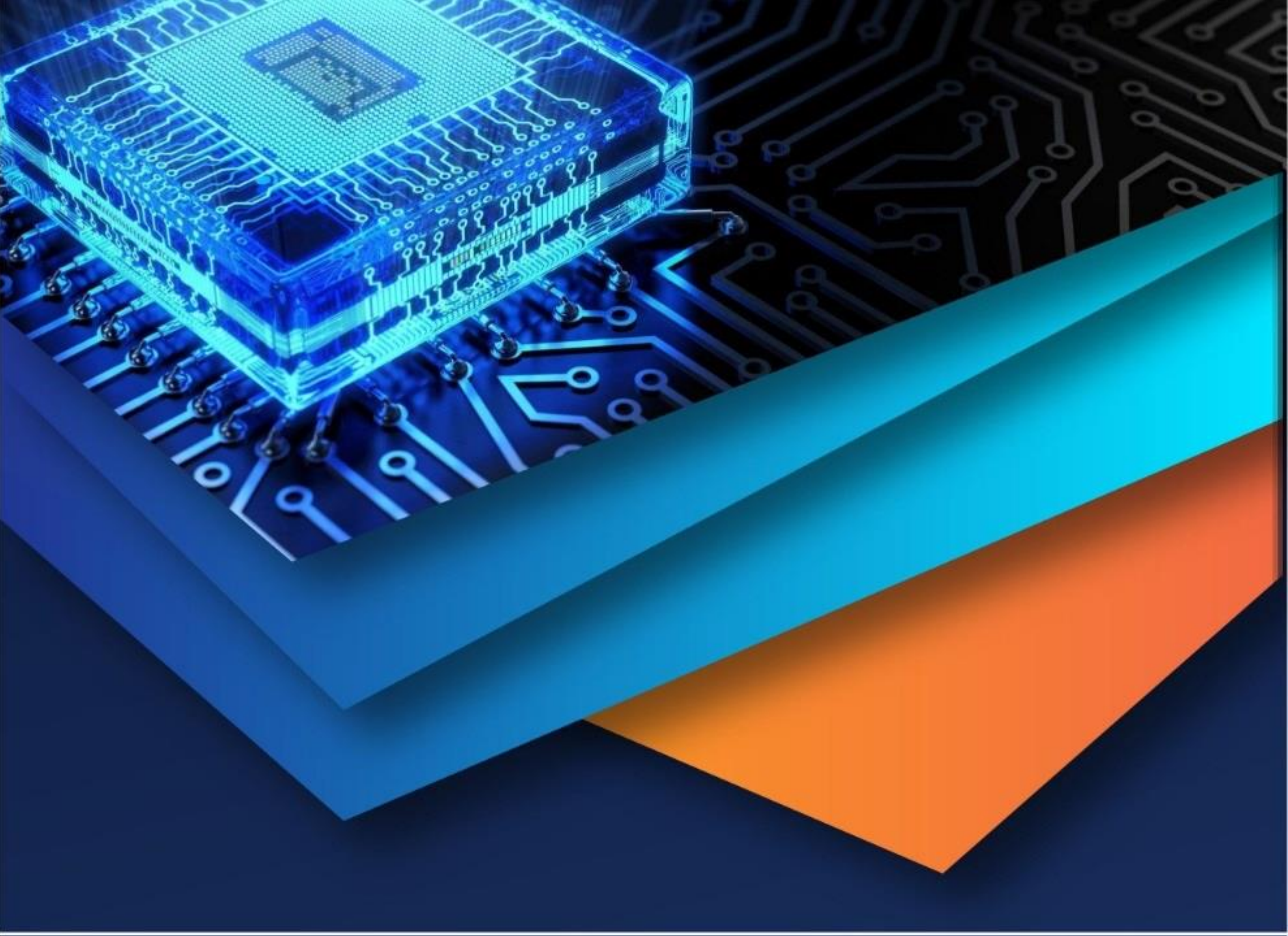
FIGURE 20. Actual vs Predicted Production for 2025.

VII. CONCLUSIONS

This work shows how LANDSAT 8 OLI/TIRS satellite image can be used to study soil moisture, which is very important for disaster prediction, environmental monitoring, and hydro-logical applications. Data from Remote Sensing (RS) have helped to improve electronic land modeling on different scales in recent years. This study also shows that Landsat 8 OLI/TIRS images can measure soil moisture of Bangladesh. The study uses the Geographic Information System (GIS) technique, a powerful tool that improves the accuracy of soil moisture and land use mapping by studying remote sensing data. This study also uses high spatial resolution satellite (Landsat 8 data) to map the soil moisture and give correct and detailed soil moisture information that is up to date. The results of this study show that multi-spectral remote sensing satellite imagery can map soil moisture well at a regional scale. The study also combines satellite data with ground measurements and advanced modeling techniques to make the soil moisture estimates more accurate and reliable. The method used in this study can be used in other places with similar environmental features to watch and control soil moisture levels well.
Finally, LANDSAT 8 satellite data and GIS technique is used to map Soil Moisture in Rajshahi district and evaluates the potential paddy production rate. We have also found out yearly increase of soil moisture quality level (Wet Level). There could be an increasing number of potential paddy production in 2025 by 47.95%.

REFERENCES

[1] M. M. B. W. A. W. M. M. H. Md. Mamun Hossain, Asswad Sarker Noman and A. S. M. Miah, “Exploring Bangladesh’s soil moisture dynamics via multispectral remote sensing satellite image,” 2023.
[2] “Bangladesh bureau of statistics,” https://bbs.gov.bd/site/page/3e838eb6-30a2-4709-be85-40484b0c16c6/Yearbook-of-Agricultural-Statistics.
[3] “Earth explorer,” Earth Explorer: https://earthexplorer.usgs.gov/.
[4] “Normalized difference water index,” https://eos.com/make-an-analysis/ndwi.
[5] Z. K. Zhihao Hong and R. Iyer, “A data-driven approach to soil moisture collection and prediction using a wireless sensor network and machine learning techniques,” 978-1-5090-0898-8/16/31.002016IEEE, 2016.
[6] A. L. M. D. Umesh Acharya and P. G. Oduor, “Machine learning for predicting field soil moisture using soil, crop and nearby weather station data in the red river valley of the north,” https://doi.org/10.3390/soilsystems5040057, 2021.
[7] X. Z. L. Z. X. X. Yu Cai, Wengang ZhengID, “Research on soil moisture prediction model based on deep learning,” https://doi.org/10.1371/journal.pone.0214508, 2019.
[8] P. G. ABHINAV SHARMA, ARPIT JAIN and V. CHOWDARY, “Machine learning applications for precision agriculture: A comprehensive review,” Digital Object Identifier 10.1109/ACCESS.2020.3048415, 2020.
[9] R. Elashmawy and I. Uysal, “Precision agriculture using soil sensor driven machine learning for smart strawberry production,” MDPI, 2023.
[10] O. J. M. Aniruddha Basak, Kevin M. Schmidt, “From data to interpretable models: machine learning for soil moisture forecasting,” International Journal of Data Science and Analytics, 2023.
[11] D. I. Fabrizio Balducci and G. Pirlo, “Machine learning applications on agricultural datasets for smart farm enhancement,” Machines 2018, 6, 38; doi:10.3390/machines6030038, 2018.
[12] “Geodesh,” http://data.gov.bd/dataset/geodash.
[13] “Python library,” https://docs.python.org/3/library/index.html.
[14] “ML library,” https://scikit-learn.org/stable/index.html.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY