

## Research Article

# Prediction Approaches for Smart Cultivation: A Comparative Study

**Amitabha Chakrabarty**<sup>1</sup>, **Nafees Mansoor**<sup>2</sup>, **Muhammad Irfan Uddin**<sup>3</sup>,  
**Mosleh Hmoud Al-adaileh**<sup>4</sup>, **Nizar Alsharif**<sup>5</sup>, and **Fawaz Waselallah Alsaade**<sup>6</sup>

<sup>1</sup>Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh

<sup>2</sup>Department of Computer Science and Engineering, University of Liberal Arts Bangladesh, Dhaka, Bangladesh

<sup>3</sup>Institute of Computing, Kohat University of Science and Technology, Kohat, Pakistan

<sup>4</sup>Deanship of E-Learning and Distance Education, King Faisal University, Hofuf, Saudi Arabia

<sup>5</sup>Department of Computer Engineering and Science, Al-Baha University, Al Bahah, Saudi Arabia

<sup>6</sup>College of Computer Sciences and Information Technology, King Faisal University, Hofuf, Saudi Arabia

Correspondence should be addressed to Nafees Mansoor; [nafees@nafees.info](mailto:nafees@nafees.info)

Received 27 January 2021; Revised 17 March 2021; Accepted 26 March 2021; Published 9 April 2021

Academic Editor: Furqan Aziz

Copyright © 2021 Amitabha Chakrabarty et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Crop cultivation is one of the oldest activities of civilization. For a long time, crop production was carried out based on knowledge passed from generation to generation. However, due to the rapid growth in the human population of the world, human knowledge-based cultivation is not enough to meet the demanding need. To address this issue, the usage of machine learning-based tools has been studied in this paper. An experiment has been carried out over 0.3 million data. This dataset identifies 46 prominent parameters for cultivation, which is collected from the Department of Agriculture Extension, Bangladesh. Comparison between neural networks and numbers of machine learning algorithms has been carried out in this research. It is observed that the neural network outperforms the other methods by maintaining an average prediction accuracy of 96.06% for six different crops. Other contemporary machine learning algorithms, namely, support vector machine, random forest, and logistic regression, have average prediction accuracy of around 68.9%, 91.2%, and 62.39%, respectively.

## 1. Introduction

To feed the rapidly growing global population, modern-day agriculture faces the demand for rising production of food. Hence, the latest technologies are transpiring in the agricultural sector to enhance net productivity by gathering and processing information. Besides, the distressing climate changes have also been hinted at the inevitable demand for modernizing the agriculture domain with the latest tools and technologies. Therefore, in the modern era, agricultural and farming domains are adapting and applying state-of-the-art technologies, namely, machine learning and the Internet of things (IoT), as agents for booming the net productivity and utilizing agricultural resources efficiently [1–4]. This has also

exposed a new direction of innovative research in the agricultural sector.

On the other hand, agriculture remains the single most important avenue for mankind, and therefore in most countries, the largest part of the workforce is in some way involved in this sector [5]. Being one of the most densely populated countries and one of the fastest-growing economies in the world, smart agriculture can have a profound impact on Bangladesh [6]. This in turn can contribute to as much as 17% of the country's GDP and almost half of the working population who are involved in the agriculture sector in Bangladesh [7]. Bangladesh has around 70% of the agricultural land among its total area where the major local crops are considered to be rice, jute, and wheat [8]. However, the age-old cultivation process is still in practice in

Bangladeshi cultivation, which operates on the information passed over generations. With the current demand for maximizing production, human experience-based models sometimes fail to meet the optimum requirement [9].

Hence, the objective of this research is to explore the possibilities of using a modern technology-driven prediction model to assist the farmers in efficiently selecting crops and maximizing the yield potentialities [10]. It is very important to explore alternatives to traditional agriculture systems so that the continuously increasing food demand of overly populated countries like Bangladesh can be addressed. Several machine learning techniques have been considered in this paper for 6 main crops' yield prediction, including 3 variations of rice and 3 of corn, jute, and potato. As efficient crop selection and yield prediction have been identified to be a multivariant problem; this paper has accumulated a total of 46 influential parameters that include several environmental parameters, various fertilizers, land type, soil structure, and essential soil types. Along with these parameters, roughly 0.3 million data are also acquired for the machine learning models to provide an effective solution.

To generate a diverse crop yield prediction system, the artificial neural network is adopted in this work [11, 12]. Furthermore, this work is an extension of the previous work presented in [11] to identify the effective crop yield selection method. The research has studied the prediction model using different machine learning techniques, namely, the deep neural network, support vector machine algorithm, random forest algorithm, and logistic regression. It is also found that environmental parameters along with soil types and composition have a remarkable influence on the overall crop production. Multiple hidden layer-enabled deep neural network has shown better accuracy compared with other explored methods.

The rest of the paper is organized as follows. Section 2 describes the existing prediction-based models for the agricultural domain. Prepared dataset and data analyses are presented in Section 3. A discussion on the methodology and experimental results is presented in Section 4, and the conclusion is presented in Section 5.

## 2. Background Study

Ji et al. in [13] created a model for an agricultural administration that wants accurate and simple estimation techniques to predict rice yields in the planning process. The goal of the research was to (1) see whether artificial neural system (ANN) models could successfully anticipate Fujian rice yield for ordinary atmosphere state of the mountain district, (2) assess ANN model performance relative to different parameters, and (3) compare the effectiveness of multiple linear regression models with ANN models.

Another research proposed by Dahikar and Rode in [14] used parameters related to soil and atmosphere to predict essential crops using an artificial neural network. Areal image detection-based agriculture using machine learning was proposed by Treboux and Genoud in [15]. It showed that the decision tree ensemble outperforms other image

identification-based solutions. A paper by Shakoor et al. [16] proposed a crop rank that suggests a list of cost-effective crops for cultivation in Bangladesh. This paper used data collected from Agricultural Statistics and Bangladesh Agricultural Research Council and used supervised machine learning algorithms to generate the crop list. Similar predictive approaches in agriculture domain have been demonstrated by Wang et al. in [17]. The paper proposed the usage of statistical models for yield prediction in response to the change in temperature and perception. Three different statistical models, namely, time series, panel, and cross-sectional models, were used to predict climate change based on the given data. Similar proposals have been presented in [18, 19].

Shastri and Sanjay in [12] presented that the customized artificial neural network (C-ANN) model performs better with a higher  $R^2$  statistic and the least percentage prediction error than the MLR and D-ANN models on the test dataset. Furthermore, the prediction of crop yield is very essential in the domain of agriculture. In this study, the wheat yield was predicted by considering its different parameters, and better wheat yield was predicted by applying the C-ANN model.

Jabjone and Jiamrum in [20] proposed the model for predicting rice yield in the Phimai district, Thailand. Specifically, a multilayer feed-forward neural networking model in combination with the back-propagation algorithm is applied to generate the model. Besides, data between 2002 and 2007 are used for training the model. This model was used to predict the rice yield between 2008 and 2012. Moreover, the input data from six meteorological factors, rainfall, water distribution, evaporation, transpiration, temperature, humidity, and wind speed, were used. Evapotranspiration (ET) was found by using the Penman-Monteith equation. Their consequence showed that ANN (8, 19, and 17) provided the lowest value of Root Mean Squared Error (RMSE) (10.57) and MAPE (2.3). The rice yield prediction of ANN (8, 19, and 17) and actual data have a linear relationship ( $R^2 = 0.99$ ). Their predicting model, therefore, was precise and appropriate to predict the rice yield.

Most of the work reported in the literature on predictive agriculture conducted analysis on limited data as well as small scale in terms of number algorithms used. In this work, we have collected extensive datasets and used those data for making a prediction using most of the common prediction algorithms. A dataset of almost 0.3 million entries and more than 45 attributes will definitely help to generate more accurate predictions compared to other reported works.

## 3. Dataset and Data Analysis

Similar to any other intelligent system, the dataset is a fundamental block to foster the development of an efficient crop yield prediction system. Since a proper dataset for the mentioned purpose was absent, a dataset has been prepared and presented in the latter part of the paper along with its analysis.

*3.1. Descriptions of the Dataset.* In any machine learning-based system, the system acquires knowledge from the dataset and predicts future trends. Considering the agricultural philosophy, cropping pattern, soil composition, and so on, 30 agroecological zones can be identified in Bangladesh. Agricultural zone 28 has been chosen as the target area for this research. This zone covers seven districts, which are Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi. As agricultural zone 28 has been considered for the dataset, the research first identifies the sources to get agricultural data. We have found various establishments to be the data source which include Soil Resource Development Institute (SRDI), Bangladesh Agricultural Research Council (BARC), Bangladesh Bureau of Statistics-Government (BBS), and Bangladesh Meteorological Department (BMD). Among these, the SRDI works on soil nutrients where BARC also works in the agricultural domain. The other two sources process and store data for various domains, along with the farming domain. However, all these sources are yet to prepare any agricultural data for employing machine learning algorithms. Regardless of the data acquisition method, data collection and processing are always diligent and time-consuming. Furthermore, we found that these four organizations are assigned with dissimilar concerns. Hence, the SRDI library, which is filled with thousands of books on different agricultural zones, has been accessed to collect data. Among these, we have identified 70 books that are focused on the agricultural data for the period ranging from 2008 to 2017. Hence, these data from the collected books are required to be translated and converted to digital form from the printed version. Similarly, nondigitized data are collected from BBS and BMD and digitized. Once the collected data were compiled, it has been noticed that there are features to be considered for prediction purposes. Hence, a total of 46 different features are identified that have a great impact on agricultural crop yield production. The volume of the collected data is 206126 instances (approximately 0.3 million). To process, we have scaled those data because the value was in a different unit. Moreover, we have cleaned the initial data, because there were many null values. Besides, many columns were in a categorical format so that we have converted those to unique numerical values. Finally, we compiled all the data into a single tabulated form to apply the machine learning algorithms and composed the tabulated Excel format into a comma separated value (CSV) format.

On the other hand, according to the BBS, the top six crops for Bangladesh are Aus rice, Aman rice, Boro rice, jute, wheat, and potato. Thus, the research focuses on these six for further exploration. It is also found that environmental parameters, namely, rainfall, maximum-minimum temperatures, and humidity, have exceedingly impacted the dataset [21]. Besides, we have considered four principal fertilizers in the dataset which are urea, triple superphosphate (TSP), diammonium phosphate (DAP), and Muriate of Potash (MOP) [22]. Moreover, we have categorized the inundation land types to be highland, medium highland, medium low land, low land, and very low land. These categorizations are done based on the elevation of these lands from the sea level. Additionally,

soil structure, a key influence in cultivation, is referred to as the shape of soil structural units. Hence, based on the structure, we may also get to know in what land crops yield most. The soil type is also a very dominant factor in farming to efficiently harvest the crop [5]. We have considered 19 essential soil types which are categorized into calcareous alluvium, noncalcareous alluvium, acid basin clay, floodplain calcareous brown, floodplain calcareous grey, floodplain calcareous dark grey, floodplain noncalcareous grey floodplain soil, noncalcareous dark grey floodplain soil, peat, made land, noncalcareous brown, shallow terrace red-brown, deep terrace red-brown, terrace mottled brown, terrace shallow grey, terrace deep grey, valley grey, and hill brown for the research [22]. The perfect soil should be considered to be a loam, which refers to the fact that particular soil has an even section of sand, silt, and clay. Thus, based on this proposition, we have included the soil moisture, texture, consistency, and reaction in the dataset for enhanced visibility of the individual soil type.

*3.2. Data Analyses.* The data analyses considering various features are presented in the following parts.

*3.2.1. Data Analysis of Maximum Temperature.* Figure 1 illustrates the maximum temperature in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi from 2008 to 2017, where the annual maximum temperature was recorded between 22 degrees Celsius and 36 degrees Celsius and has an impact on crop production. As an overall trend, it is seen that the highest peak of maximum temperature was in the year 2009 and the lowest peak of maximum temperature was in the year 2011. Moreover, in 2008, the maximum temperature stayed constant at approximately 34 degrees Celsius except for Gazipur and Mymensingh, where the temperature was around 30 degrees Celsius. As we move forward, the maximum temperature rose until the year 2011. However, the temperature plummeted during the year 2011 which was nearly 23 degrees Celsius except for Gazipur and Mymensingh. After that, the temperature climbed sharply in the following years.

*3.2.2. Data Analysis of Minimum Temperature.* Figure 2 presents the annual minimum temperature for seven districts including Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts between 2008 and 2017, where the annual minimum temperature was recorded between 10 degrees Celsius and 22 degrees Celsius. It is evident that during this period the minimum temperature of the Mymensingh district was at an apex and remained stable between 20 degrees Celsius and 22 degrees Celsius. However, it was opposite for Tangail district, which was approximately 10 degrees Celsius to 13 degrees Celsius. Moreover, in 2009, the annual minimum temperature increased to 3 units and rose to 15 degrees Celsius in Dhaka, Narayanganj, and Narsingdi districts. On the other hand, initially, the minimum temperature of Gazipur was 20 degrees Celsius and slumped to 11 degrees Celsius in the

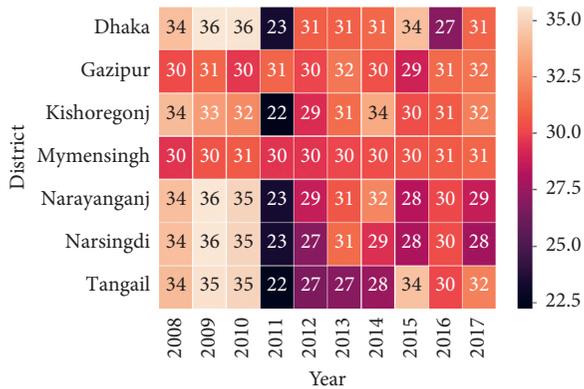


FIGURE 1: Analysis of maximum temperature from 2008 to 2017 against district.

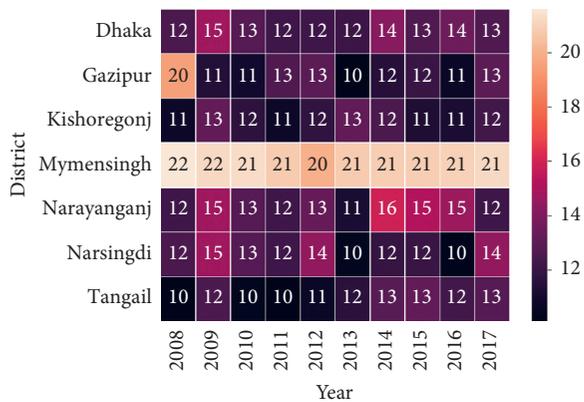


FIGURE 2: Analysis of minimum temperature from 2008 to 2017 against district.

following year and then stayed stable between the temperature of 11 degrees Celsius and 13 degrees Celsius. The average minimum temperature of Narayanganj was approximately at the same level till the year 2013 but increased suddenly in the following year, which was 16 degrees Celsius. Hence, we can conclusively say that the average minimum temperature stayed between 10 degrees Celsius and 16 degrees Celsius except for the Gazipur district.

**3.2.3. Data Analysis of Average Rainfall.** Figure 3 demonstrates average rainfall in the agricultural zone 28 including Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts between 2008 and 2017. Annual average rainfall was recorded from 12 mm to 24 mm. As we can visualize that Dhaka and Narsingdi have received the highest peak of average rainfall between 2008 and 2014, which was approximately 24 mm, and performed the least in 2012 and 2014, which was nearly between 12 mm and 14 mm. The most stable average rainfall occurred in the Kishoregonj district, where the average rainfall was roughly 16 mm to 22.3 mm. However, in the year 2010, average rainfall drastically fell for all the districts. This resulted because during that time the annual average maximum

temperature was rocketed in all the districts. Consequently, the precipitation was poor.

**3.2.4. Data Analysis of Humidity.** Figure 4 illustrates the average humidity between 2008 and 2017 for the agricultural zone 28, which includes Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts. As an overall trend, it is clearly seen that Kishoregonj, Tangail, and Mymensingh districts had an apex of average humidity throughout those years, but Dhaka, Gazipur, and Narsingdi districts had a lower amount of average humidity during the same period. In addition, the average humidity was at the lowest peak during the year 2010 in all the districts. It is apparently seen that the average humidity was recorded approximately between 66% and 80%, except for the district Dhaka, Gazipur, Narayanganj, and Narsingdi, where average humidity remained constant at nearly 65%. Moreover, in 2009, the percentage of average humidity rocketed around 80%. However, it dramatically fell until the year 2010. After that, the average humidity rises by approximately 5% every year. In summary, Kishoregonj, Tangail, and Mymensingh districts had an apex of average humidity, causing a lot of rain.

**3.2.5. Data Analysis of Aus Rice.** In Figure 5, data analysis of Aus rice has been depicted from the viewpoint of crop production and crop yield. Hence, Figure 5(a) shows the production of Aus rice between 2008 and 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts. It is pointedly comprehended that Mymensingh is the district where Aus rice is produced most, approximately between 110000 and 50000 metric tons in the studied term. It is also found that, among the seven locations, Kishoregonj is the second most producing district of Aus. On the other hand, Tangail, Narsingdi, Narayanganj, and, Dhaka have been producing the lowest. In 2012, Mymensingh has the highest crop which is nearly 110000 metric tons. The reason behind this heavy growth rate is the enormous amount of chemical fertilizer, which was found in this district over these years. Moreover, the amount of average humidity is also higher, which is nearly (78%), even though the precipitation was low compared to the other years. Furthermore, the maximum temperature was quite decent, even though the minimum temperature for Mymensingh was upward compared to the other districts. Moreover, in the Mymensingh district, there are a huge number of noncalcareous dark grey floodplain soils, non-calcareous grey floodplain soils, deep red-brown terrace soils, shallow red-brown terrace soils, and acid basin clay soils [10]. Furthermore, we have found numerous soil textures in the Mymensingh area. As a result, it shows the peak amount of production level of Aus rice that year. In addition, Kishoregonj shows a preferable amount of Aus rice production nearly 45000 metric tons, after the Mymensingh district. On the other hand, it is apparently evident that both Narsingdi and Dhaka districts had a small amount of Aus rice production. This is because, except for triple super-phosphate, most other chemical fertilizers are deployed

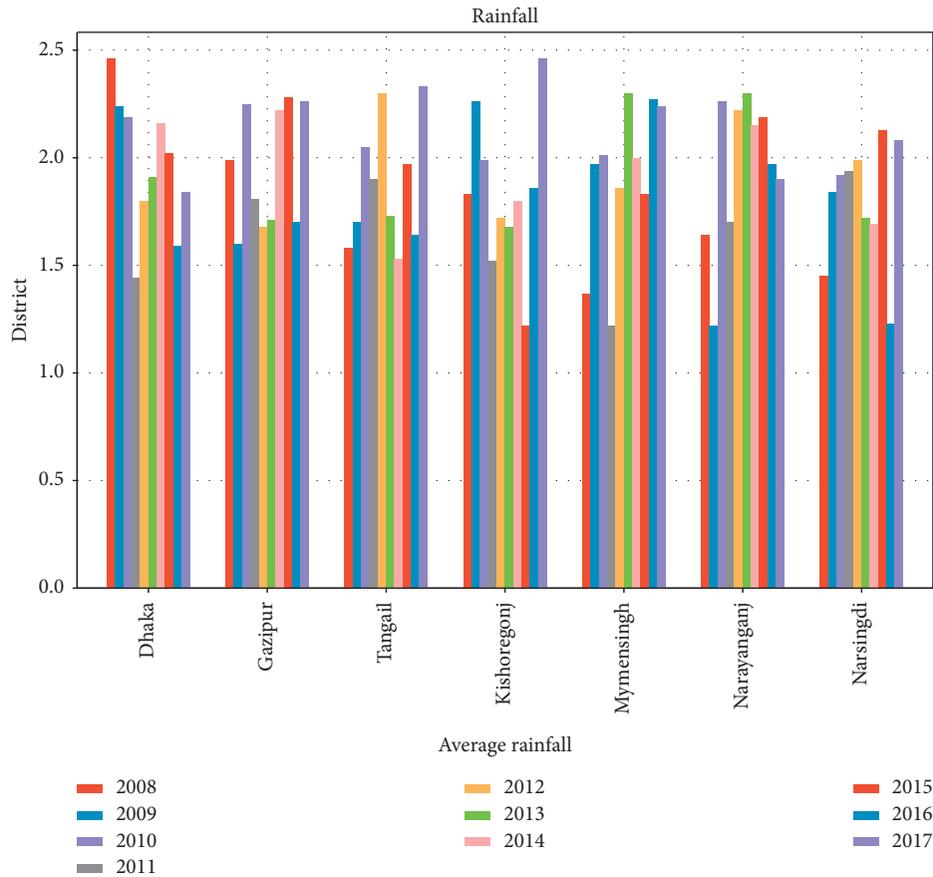


FIGURE 3: Analysis of average rainfall.

heavily in Mymensingh than in Dhaka. Furthermore, rainfall levels of both Narsingdi and Dhaka are still on the higher side.

In Figure 5(b), the crop yield of Aus rice between 2008 and 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts is described. This is the agricultural output of Aus rice. There is a difference between crop production and crop yield. We are analyzing the crop yield to understand how much production occurs in a particular area. For measuring the crop yield, we need to plug in the value of the production level and divide it with the area of the production level. It is noticeable that Mymensingh did the higher production level, but compared to its amount of area, it did not perform up to the level. Gazipur has an enormously small amount of land compared to the Mymensingh district, but it performed well between 2014 and 2017. Again, in the year 2010, Narayanganj had an apex amount of Aus crop yield rate just under a 3.0-hectare rate, among other districts. Hence, we can conclusively say that the crop yield rate of Aus rice was almost the same level except for Tangail and Narsingdi districts.

**3.2.6. Data Analysis of Aman Rice.** In Figure 6, data analysis of Aman rice has been depicted from the viewpoint of crop production and crop yield. Therefore, Figure 6(a) illustrates

the production of Aman rice between 2008 and 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts. The graph indicates that, similar to Aus rice, Mymensingh is still the leading district, where Aman rice produced the most. This is the same reason as Aus rice production. The pattern of Aus rice production and the production of Aman rice characteristics is quite similar. From the above figure, during 2011 to 2013, the Kishoregonj district seems to produce a high amount of Aman rice. However, if we observe the figure, the crop yield gradually increased from 2011 to 2017. During 2014 to 2017, the land area of that particular district fell slowly, but the production rate rose, which took the crop yield unit to the maximum level. These increasing productions are due to the changes in chemical fertilizer, soil, and land type. Kishoregonj has noncalcareous grey floodplain soil, noncalcareous dark grey floodplain soil, acid basin clay, noncalcareous alluvium, deep red-brown terrace soil, and shallow red-brown terrace soil which plays an important role to be capable of producing abundant crops [22]. Moreover, if we concentrate on the years between 2008 and 2010, we will notice that all the four chemical components were low in amount. However, it tends to increase gradually, which supports Aman rice to expand its production level with proper soil consistency, soil moisture, soil reaction, and soil texture. However, in the year 2008, an excessive amount of rainfall and an unstable atmosphere cause a decrease rate of the production level.

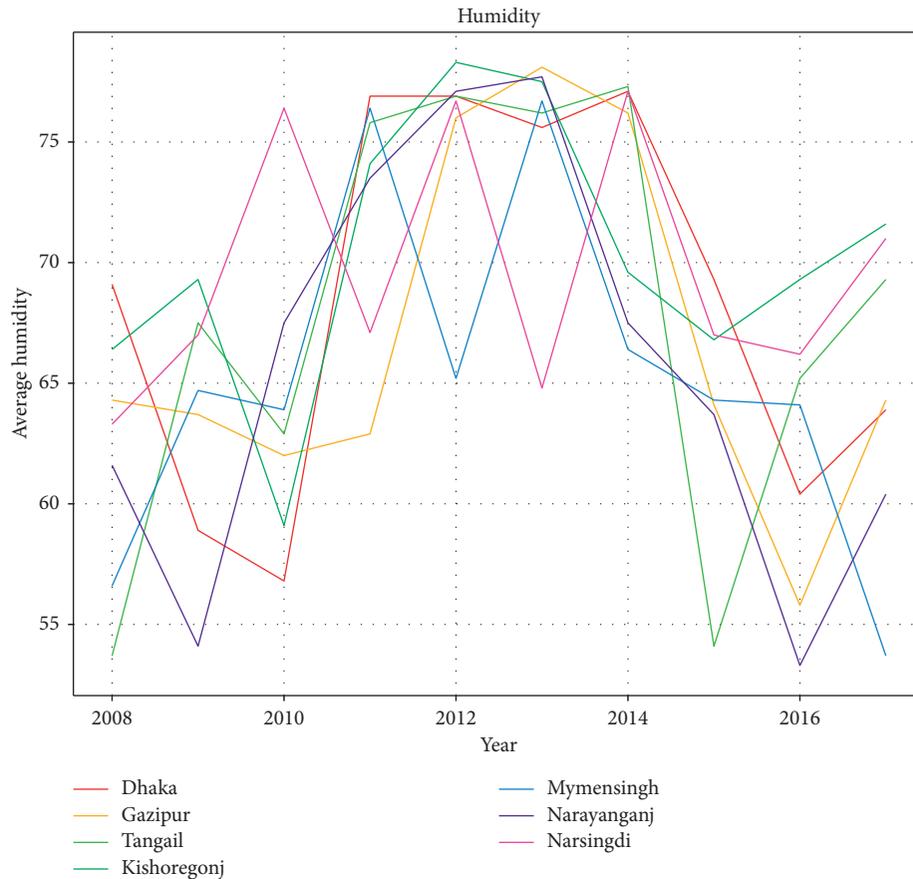


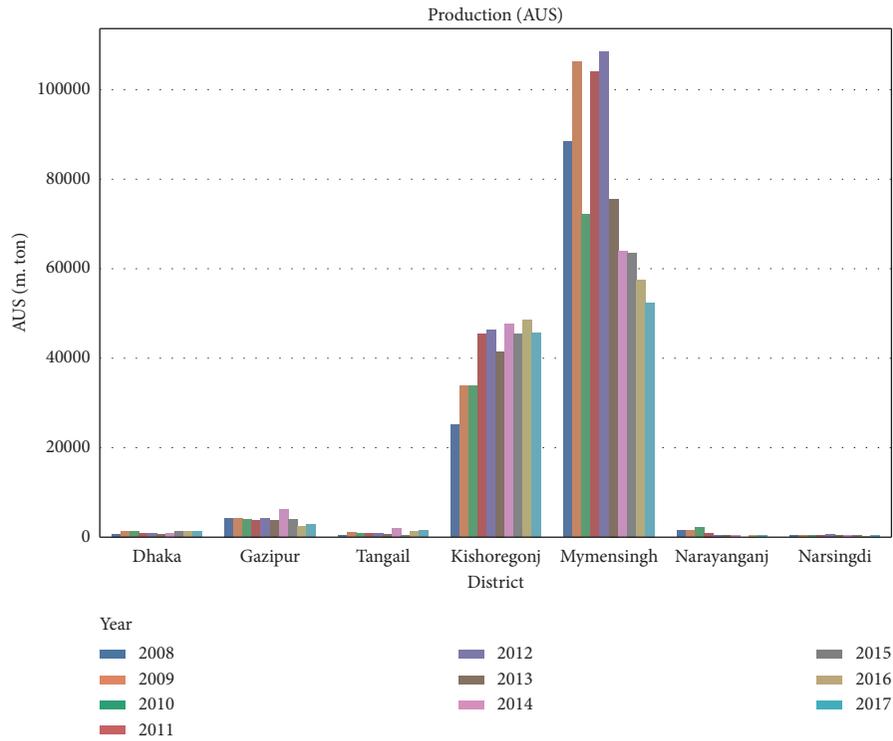
FIGURE 4: Analysis of average humidity.

Next, Figure 6(b) describes the crop yield of Aman rice between 2008 and 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts. Gazipur and Narsingdi yield Aman rice most, due to their land area, which plays an important role here [22]. Both Gazipur and Narsingdi districts consist of the same type of land, which is inundation land medium highland, inundation land highland, inundation, and low land, inundation land medium low land, and miscellaneous land. One major drawback is a small amount of land, which causes a lower-yielding rate. In summary, it can be said that Gazipur and Narsingdi would be suitable districts for both Aus and Aman rice production.

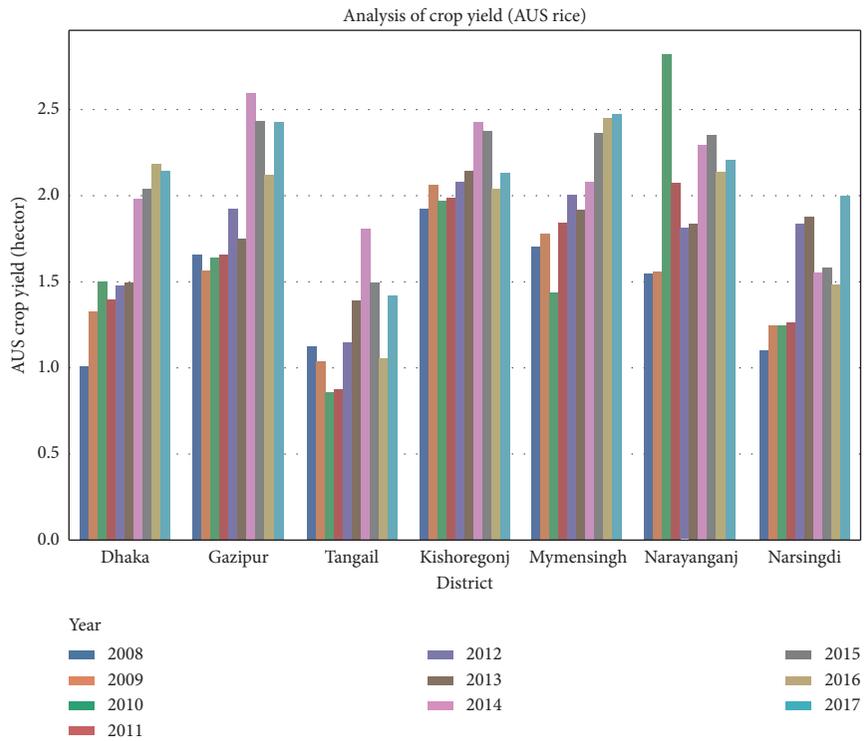
**3.2.7. Data Analysis of Boro Rice.** Figure 7(a) presents the production of Boro rice from 2008 to 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts. As an overall trend, it is seen that Kishoregonj produces the highest amount of Boro rice during these ten years, which varies between 1400000 metric tons and 700000 metric tons. Next, the Mymensingh district produces almost the same level during those periods, which varies between 1000000 metric tons and 800000 metric tons. The Boro production of Tangail remained constant in between 750000 metric tons. Besides, the Boro production of Dhaka, Gazipur, Narayanganj, and Narsingdi was at the same level of nearly 200000 metric tons.

On the other hand, Figure 7(b) describes the crop yield of Boro rice between 2008 and 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts. It is seen that Dhaka has the highest amount of Boro rice yield and it was climbing gradually. Dhaka has inundation land highland, inundation land medium low highland, inundation land low land, and miscellaneous land, which consists of noncalcareous alluvium soil, acid basin soil, and calcareous dark grey floodplain soil with high concentration, which improves Boro rice cultivation [22]. However, the Mymensingh district is not suitable for cultivating Boro rice, because of its atmosphere. Moreover, Mymensingh has a lower amount of yield rate during 2008 to 2013. After that, the crop yield rate increased slowly. To conclude, it is quite evident that the yield rate of Boro rice is approximately at the same level for all the districts.

**3.2.8. Data Analysis of Jute.** Figure 8(a) represents the jute production in the mentioned seven different localities for ten years starting from 2008. It is perceived from the figure that Tangail happens to be the most jute growing district with a maximum production rate, whereas Narayanganj has the most moderate yielding rate. Tangail becomes the most favorable place for growing jute due to the variance in textures and components in the soil. Consequently,



(a)



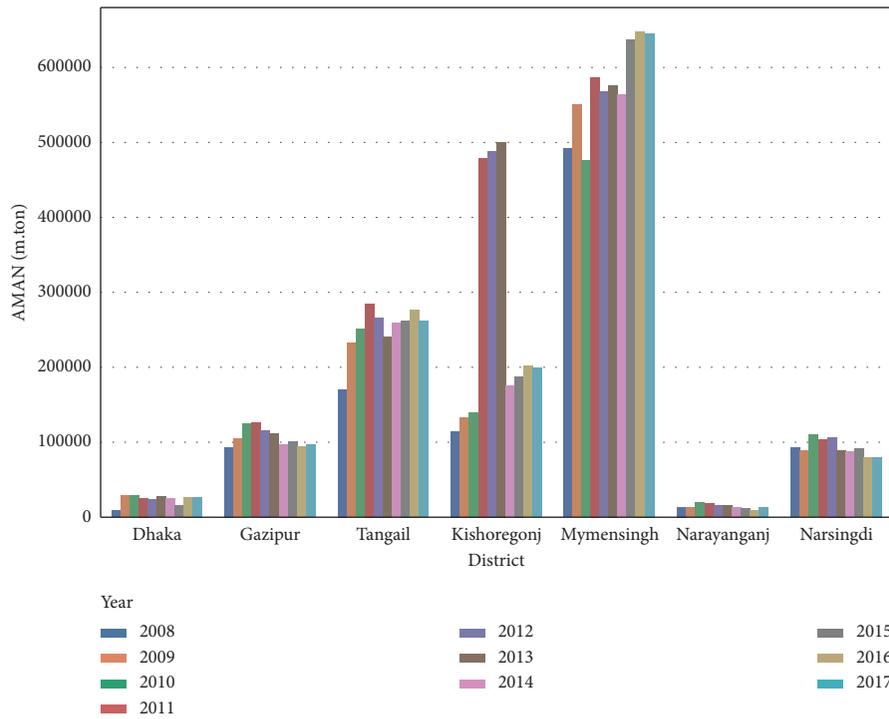
(b)

FIGURE 5: Data analysis of Aus rice. (a) Crop production for Aus rice. (b) Crop yield for Aus rice.

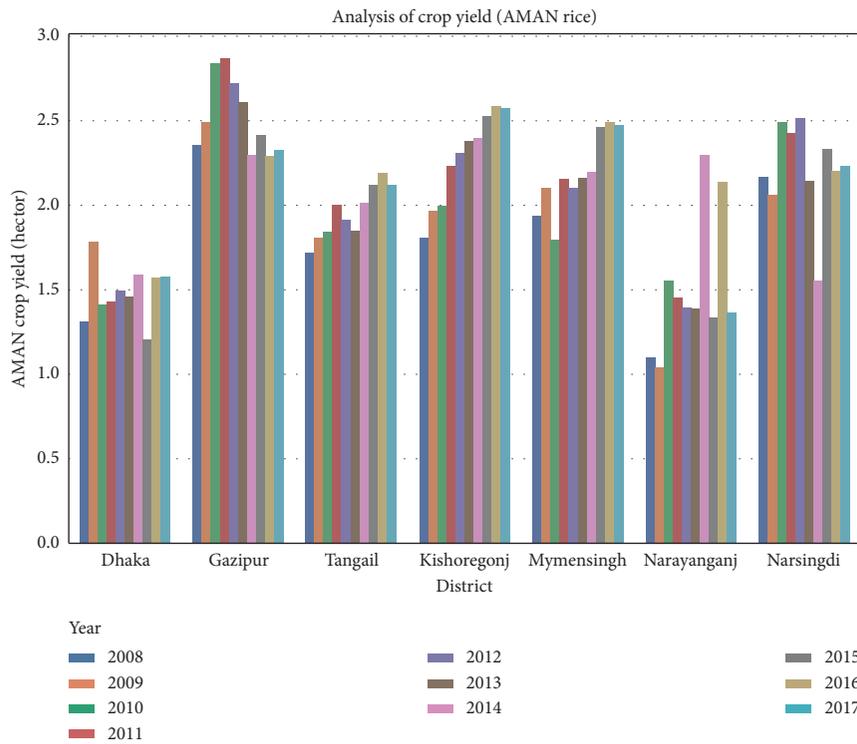
Tangail has the highest amount of production rate. Since jute plants are typically dry, a location that offers plentiful rain along with hot and humid weather becomes most suitable for plantation. Hence, the Tangail district offers

the more desirable choice for jute due to its soil and climate.

Furthermore, in Figure 8(b), a description of jute between 2008 and 2017 in agricultural zone 28 is depicted. This



(a)

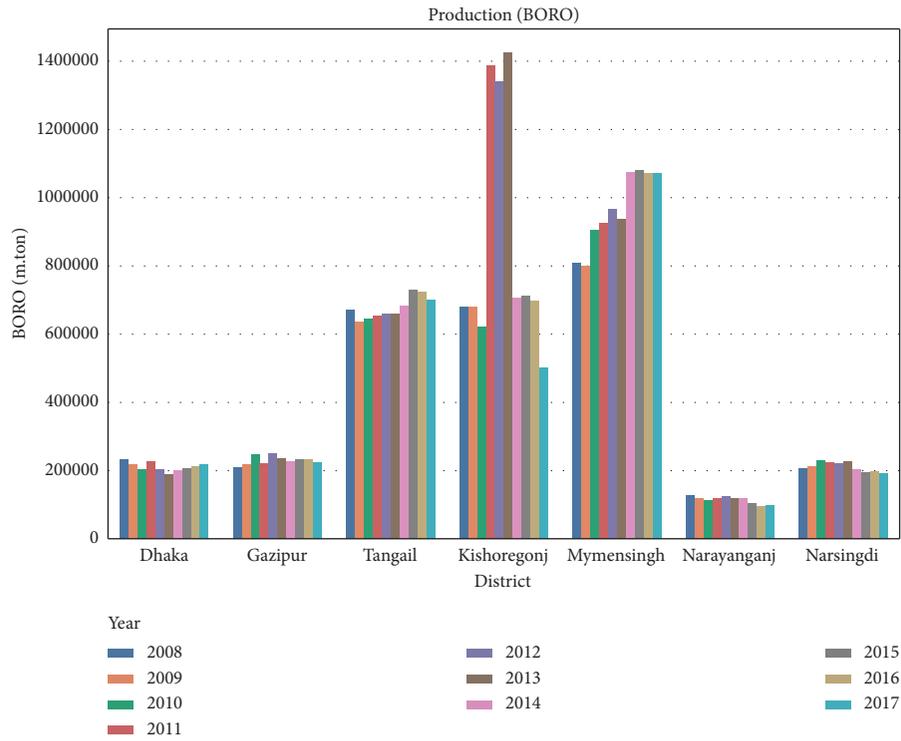


(b)

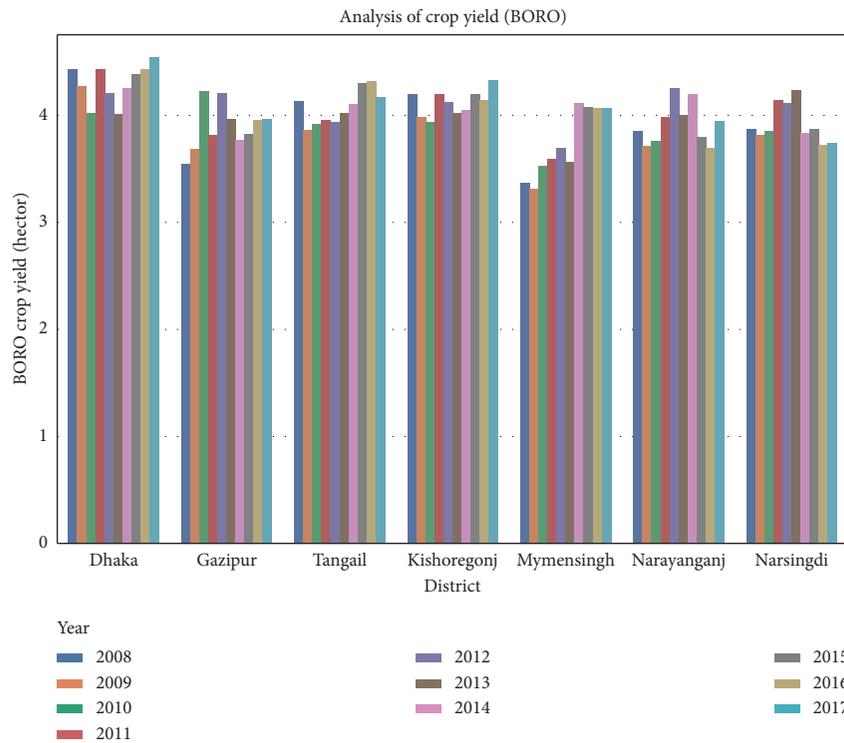
FIGURE 6: Data analysis of Aman rice. (a) Crop production for Aman rice. (b) Crop yield for Aman rice.

is the agricultural output of jute. As an overall trend, it is seen that Gazipur has the highest amount of jute yield rate of approximately 8-hectare rate to 12-hectare rate, though its overall production is lower. Besides, Kishoregonj and Narsingdi had a good amount of yielding rate because of the

stable required environment, atmosphere, soil, and fertilizer [22]. Furthermore, the yielding rate of Gazipur fluctuates during this period. In the year 2008, Gazipur has the highest amount of yielding a rate of approximately (12-hectare rate) but it decreased drastically in the year 2009. In 2009, the



(a)

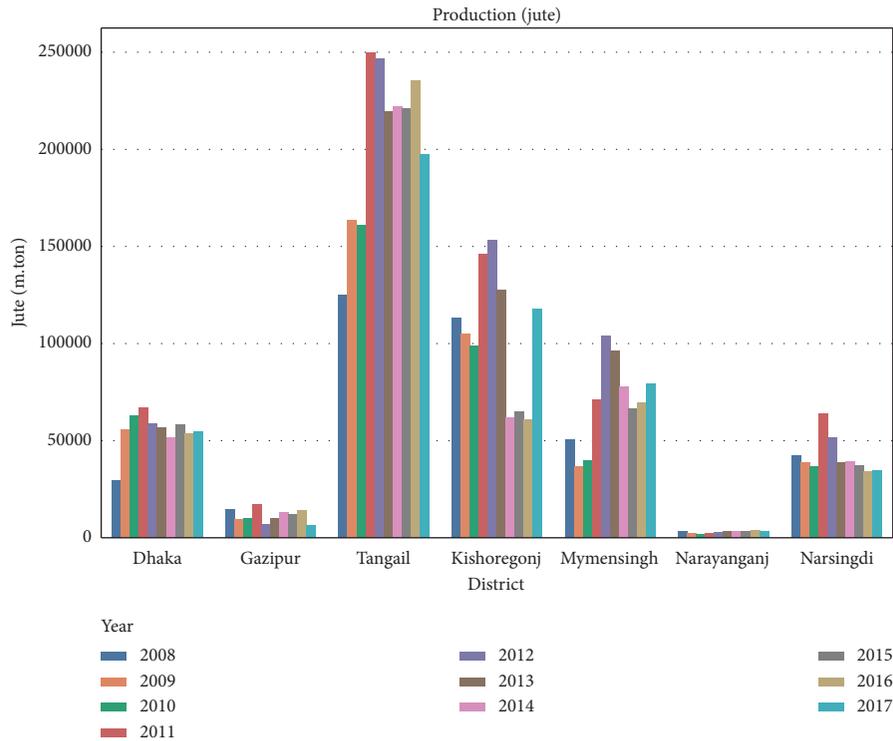


(b)

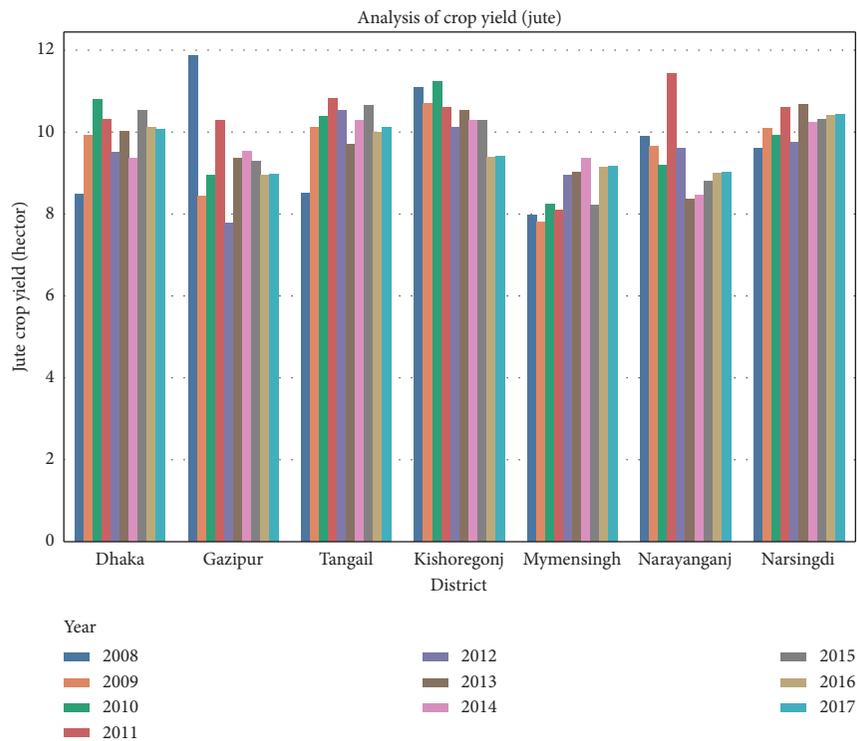
FIGURE 7: Data analysis of Boro rice. (a) Crop production for Boro rice. (b) Crop yield for Boro rice.

minimum temperature of Gazipur was nearly 11 degrees Celsius, which is not feasible for keeping the rate the same as in 2008. However, Tangail has the highest production rate but decreased slightly after the year 2015.

3.2.9. *Data Analysis of Wheat.* In Figure 9(a), the analysis of the production of wheat from 2008 to 2017 in agricultural zone 28 including Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts is



(a)

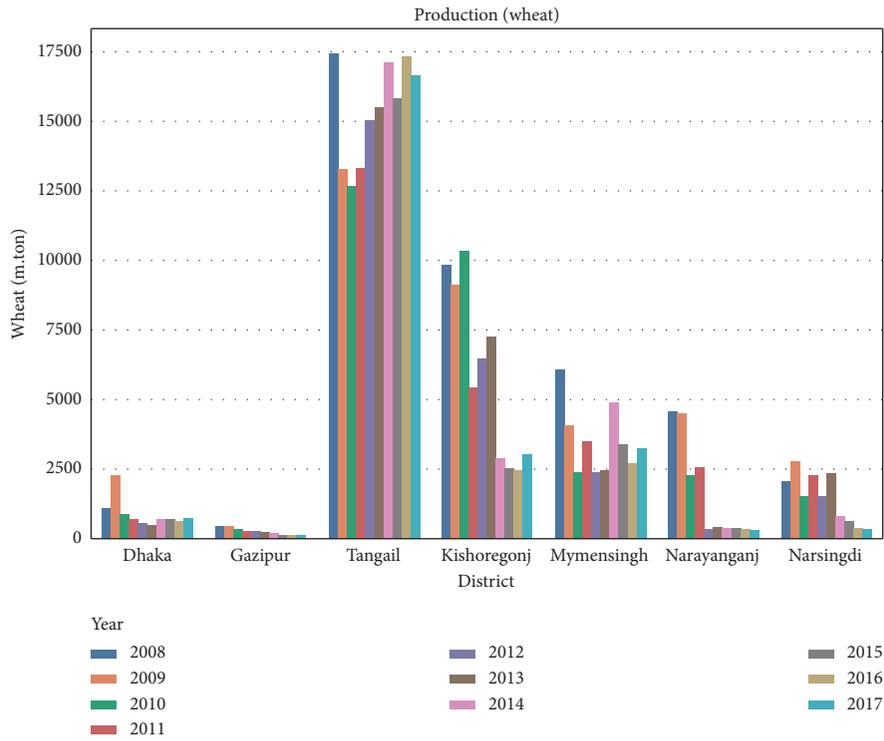


(b)

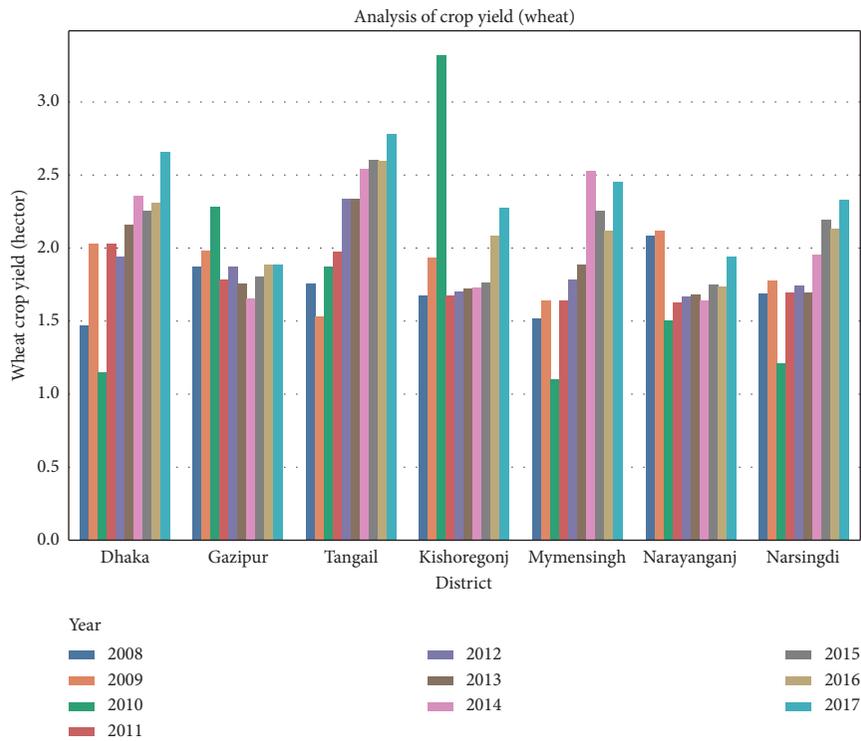
FIGURE 8: Data analysis of jute. (a) Crop production for jute. (b) Crop yield for jute.

highlighted. It is quite evident that Tangail produced the highest amount of wheat rather than other districts. Moreover, Tangail has average fertilizers with a little bit high amount of urea, lower average minimum temperature than

usual, high humidity, and an unstable rainfall which creates less effect in the production of wheat. As a result, Tangail always has a moderated amount of production among all districts. On the other hand, Gazipur failed to produce an



(a)



(b)

FIGURE 9: Data analysis of wheat. (a) Crop production for wheat. (b) Crop yield for wheat.

average production of wheat due to variation of soil type in Gazipur such as deep red-brown terrace soil, shallow red-brown terrace soil, acid basin clay, grey valley soil, shallow grey soil, and calcareous alluvium soil [22].

In Figure 9(b), the crop yield of wheat between 2008 and 2017 in agricultural zone 28 (Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts) is presented. It is seen that in most of the districts the wheat

yield was about 1.5 hectares (metric ton) to 2.5 hectares (metric ton). In the year 2010, the Kishoregonj district had the highest amount of wheat yield, which was more than 3.0 hectares (metric tons) but eventually it falls. Besides, during the year 2010, there was the lowest humidity and average temperature as well as average precipitation, which can be an effect on the highest growth on wheat production of the year 2010 from all the districts. In summary, during this period, the yield rate of wheat remains constant nearly 1.5 hectares and 2.5 hectares.

*3.2.10. Data Analysis of Potato.* Figure 10(a) represents the production of potato between 2008 and 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts. It is seen that Narayanganj has the highest growth rate of potato. Besides, the soil type of Narayanganj has inundation land medium low land, inundation land low land, and inundation land medium highland and miscellaneous land. These types of land contain noncalcareous dark grey floodplain soil, non-calcareous grey floodplain soil, deep red-brown terrace soil, shallow red-brown terrace soil, acid basin clay, and non-calcareous alluvium [22]. This land proportion and soil proportion play a vital role in producing potatoes because it is a starchy and tuberous crop. Due to the decreasing amount of maximum temperature, small amount of diammonium phosphate, and humidity, the increasing rate of minimum temperature from 2008 potato growth gets higher till 2017. However, Gazipur, Tangail, and Kishoregonj districts have the same level of production rate due to a similar type of environment, atmosphere, and soil features.

Moreover, Figure 10(b) shows the crop yield of potatoes between the years 2008 and 2017 in Dhaka, Gazipur, Narayanganj, Tangail, Kishoregonj, Mymensingh, and Narsingdi districts. This is the agricultural output of the potato. The graph indicates that the district of Narayanganj is the apex for potato yields where the Dhaka district remains in the 2nd position. As Dhaka has average rainfall, average humidity, average temperature, and average fertilizers, it helps to yield potatoes in this region. Moreover, both Dhaka and Narayanganj have the same proportion of soil consistency which leads to their potato production in a higher amount. During the same period, Gazipur, Tangail, Kishoregonj, and Mymensingh tend to have a similar amount of potato yield rate.

## 4. Experimental Results

In this research, we have used two loss functions to evaluate the prediction accuracy of the test scores. One of the loss functions is Mean Squared Error (MSE) as it gives the average of the squared errors [23]. In other words, MSE is applied to indicate the difference between the predicted and actual target variables. Besides, the RMSE is adopted to measure the differences between values predicted by the model as RMSE denotes the standard deviation of the predicted errors. Thus, a glimpse of the distributed predicted errors is obtained from this loss function.

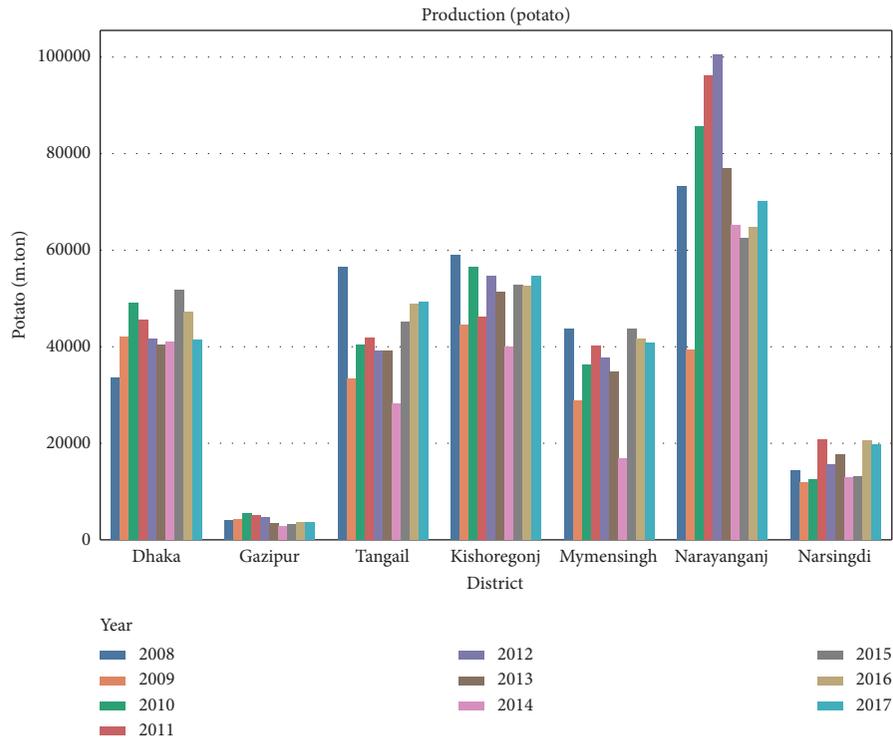
*4.1. Performance Evaluation in terms of MSE.* Here, to show the level of achieved accuracy, the research considers the “tanh” activation function with 20 percent of the validation set considered from the training set. Moreover, we have used the “sgd” optimizer with 100 epochs on the model for performance evaluation. Besides, such arrangements are also considered to ensure nonoverfitted data.

Loss and validation loss are measured to evaluate the performance of the model during the training and testing phases. Hence, Figure 11 represents the MSE of the model during these two phases. It is observed that the loss decreases gradually in the proposed model. This also qualifies for the increasing accuracy prediction of the system. Though the loss rate of the model is high at the opening epochs, and later, as observed in Figure 11, with each iteration, the model gradually learns itself and improves its accuracy. Eventually, the loss decreases sharply and after 100 epochs, the loss of the system is around 0.059, and the validation loss is about 0.063. Furthermore, the result analysis also shows that, for Dhaka district in 2014, the actual result was 1.988 where the proposed model is close to the actual with the predicted value of 2.01.

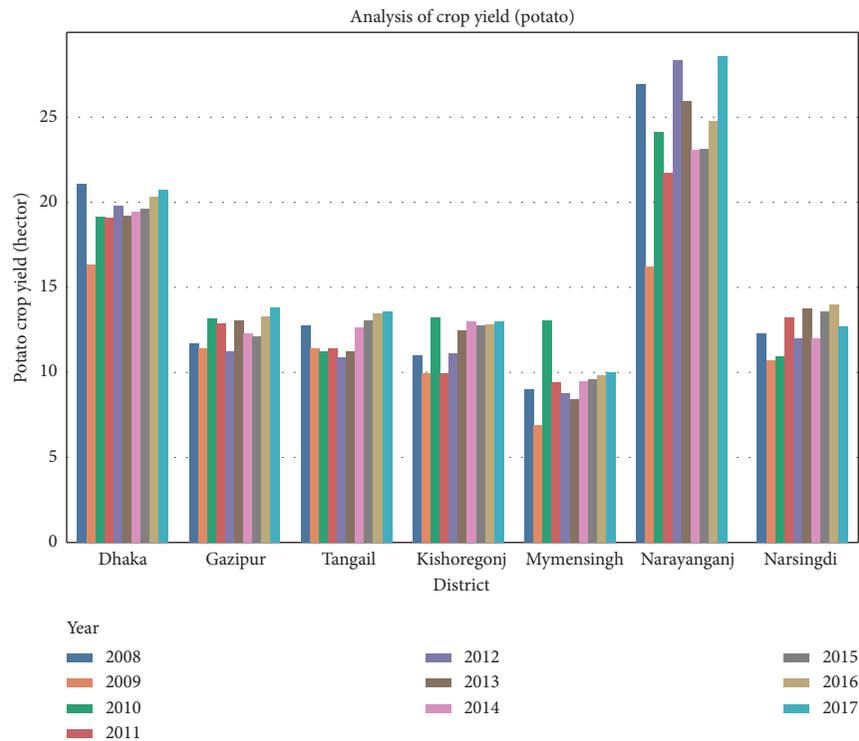
*4.2. Performance Evaluation in terms of RMSE.* This is the second loss function for the performance evaluation of the model. The density of the neurons in the hidden layers is unchanged where the activation function is set to be the same as the “tanh” while keeping the validation set to 20 percent. Similar to the MSE, Figure 12 shows that the RMSE of the model also decreases gradually. After, 100 epochs, we can maintain the loss nearing 0.267 and validation loss nearing 0.42. On a contrary, the measured result slightly hinted at the chance of overfitting. The prediction result for the Dhaka district in the year 2014 shows that the model predicted a value close to 1.47 where the actual value is 1.988 for the same.

Accuracy is considered one of the most important evaluation indicators to measure the performance of any technique. Here, in this section, we present different performances of the four different methods which are shown in Figure 13 for all six crops. It is found that the deep neural network is performing much better than the rest in all cases. As shown in the figure, for Aus rice, the random forest has an accuracy of just over 90 which is the second best result after the deep neural network. In the Aman rice dataset, we have considered 80% as a training set and 20% as a testing set, for all four algorithms. Hence, we obtained a result of 97.7% accuracy for the deep neural network referring to the MSE to be 2.3 percent. Furthermore, for Aus rice, the support vector machine shows an accuracy of 73.3 percent with 26.7% MSE where 52.57% accuracy is found for the logistic regression algorithm with 47.43% MSE. Hence, we are choosing a deep neural network approach for Aus rice selection and yield prediction purposes.

Besides, Figure 13 also displays that deep neural networks have a much better accuracy rate than the logistic regression, support vector machine, and random forest for Aman rice. As depicted in the figure, we managed to get an



(a)



(b)

FIGURE 10: Data analysis of potato. (a) Crop production for potato. (b) Crop yield for potato.

accuracy of 94.6% using the deep neural network for Aman rice. This also refers to the fact that the deep neural network has the least MSE which is 5.4% where the support vector machine has 30.3% MSE, the random forest has 7.9% MSE,

and with 49.3% accuracy, the logistic regression has 50.7% MSE. Hence, considering these results, for Aman rice selection and yield prediction purposes, the deep neural network approach can be recommended too.

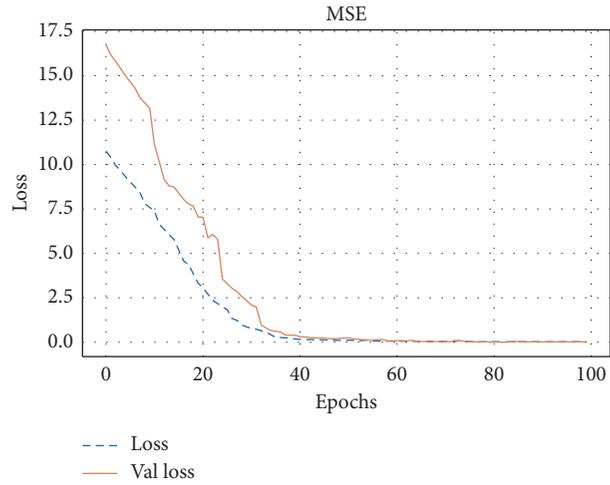


FIGURE 11: Validation of the model in terms of MSE.

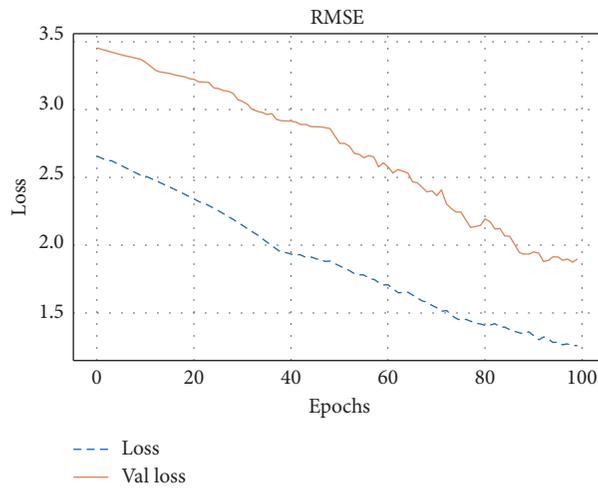


FIGURE 12: Validation of the model in terms of RMSE.

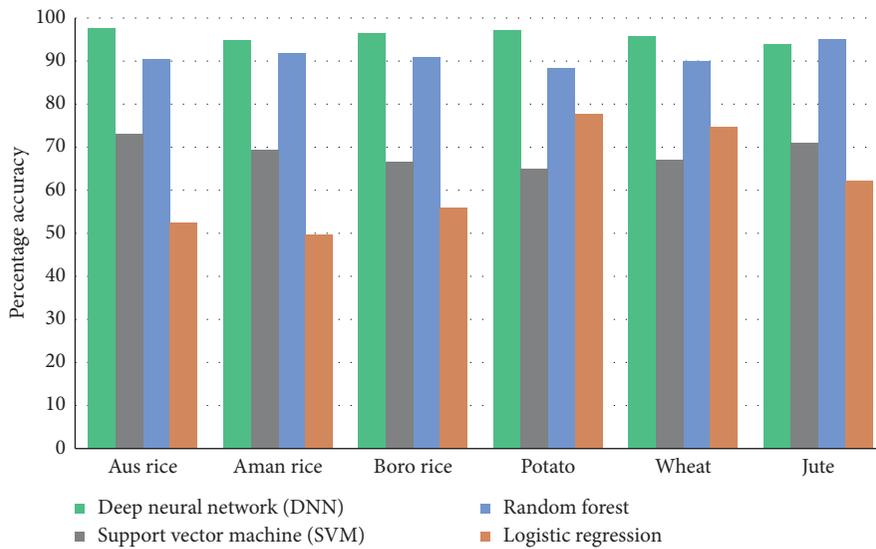


FIGURE 13: Accuracy percentage of various crops in different algorithms.

Furthermore, for Boro rice, the deep neural network is also found to be the most efficient one compared with the logistic regression, support vector machine, and random forest. In the Boro rice dataset, the deep neural network shows an accuracy of 96.7% with a 3.3% MSE. Besides, as illustrated in the figure, 67% accuracy is found for support vector machine, 91.2% accuracy for the random forest, and 56.11% accuracy for logistic regression. In addition, in the jute dataset, we got 94.1% accuracy for the deep neural network with 5.9% MSE, 71.4% accuracy for support vector machine with 28.63% MSE, 95.3% accuracy for the random forest with 4.7% MSE, and 62.56% accuracy for logistic regression with 37.44% MSE.

Moreover, as shown in Figure 13, the deep neural network-based model also has better accuracy for potatoes as well as for wheat. For potato, the deep neural network-based model maintained 97.3% accuracy. However, considering the same dataset, we have found 65% accuracy in the support vector machine, 88.3% accuracy for the random forest, and 78% accuracy for logistic regression-based models. Thus, considering the accuracy results, we can say that the neural network-based model outperforms the other three for potato selection and yield prediction purpose. It is worthwhile to mention that we have also found 96% accuracy of a deep neural network for wheat selection and yield prediction where the MSE is 4%. For the same crop, 67% accuracy is obtained in the support vector machine-based model, and 90% accuracy is obtained in the random forest-based model where the logistic regression-based model shows 75% accuracy with 25% MSE.

## 5. Conclusion

An extensive study has been carried out in this paper in relation to the use of machine learning tools in agriculture. In countries like Bangladesh, agriculture employing most of the workforce still follows traditional human prediction-based methods to produce crops. This research makes an attempt to include the prediction models in the old-fashioned agriculture system used in Bangladesh. This paper focuses on a novel attempt to collect huge data from various government organizations of Bangladesh that deal with agriculture. An attempt is made on using various machine learning algorithms to predict crop production based on the collected dataset. It has been shown that neural networks have outperformed other machine learning algorithms and justify the requirements of using machine learning tools in agriculture. There are mobile apps available for doing small-scale management and accounting for crop production, but these apps cannot make a prediction of crop production based on historical data. In the future, the researchers would like to extend this model for other regions of Bangladesh to see an overall picture of using predictive models in agriculture.

## Data Availability

Data used to support the findings of the study are available in the manuscript.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: a review," *Sensors*, vol. 18, no. 8, p. 2674, 2018.
- [2] K. S. Pratyush Reddy, Y. M. Roopa, and K. Rajeev L, N. S. Nandan, IoT based smart agriculture using machine learning," in *Proceedings of the Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 130–134, Coimbatore, India, July 2020.
- [3] Z. Ünal, "Smart farming becomes even smarter with deep learning—A bibliographical analysis," *IEEE Access*, vol. 8, pp. 105587–105609, 2020.
- [4] L. García, L. Parra, J. M. Jimenez, J. Lloret, and P. Lorenz, "IoT-based smart irrigation systems: an overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture," *Sensors*, vol. 20, no. 4, p. 1042, 2020.
- [5] D. Patil, M. S. Shirdhonkar, and M. Rice, "Rice crop yield prediction using data mining techniques: an overview," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 7, no. 5, pp. 427–431, 2017.
- [6] N. Mansoor, "Conceptualizing and realizing a smart city model for Bangladesh," 2012, <https://arxiv.org/abs/2012.03055>.
- [7] M. Mondal and H. Sanaul, "The implications of population growth and climate change on sustainable development in Bangladesh," *Jambá: Journal of Disaster Risk Studies*, vol. 11, no. 1, pp. 1–10, 2019.
- [8] H. Khushi, M. Moniruzzaman, and N. Tabassum, "Growth scenarios of area, production and yield of rice in Bangladesh," *Bangladesh Journal of Agricultural Economics*, vol. 41, no. 2, 2020.
- [9] X. Zhou, H. B. Zheng, X. Q. Xu et al., "Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 246–255, 2017.
- [10] M. S. G. Adnan, A. Y. M. Abdullah, A. Dewan, and J. W. Hall, "The effects of changing land use and flood hazard on poverty in coastal Bangladesh," *Land Use Policy*, vol. 99, Article ID 104868, 2020.
- [11] T. Islam, T. A. Chisty, and A. Chakrabarty, "A deep neural network approach for crop selection and yield prediction in Bangladesh," in *Proceedings of the IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, pp. 1–6, West Java, Indonesia, 2018 December.
- [12] K. A. Shastry and H. A. Sanjay, *Cloud-based Agricultural Framework for Soil Classification and Crop Yield Prediction as a Service in Emerging Research in Computing, Information, Communication and Applications*, Springer, Singapore, Singapore, 2019.
- [13] B. Ji, Y. Sun, S. Yang, and J. Wan, "Artificial neural networks for rice yield prediction in mountainous regions," *The Journal of Agricultural Science*, vol. 145, no. 3, p. 249, 2007.
- [14] S. S. Dahikar and S. V. Rode, "Agricultural crop yield prediction using artificial neural network approach," *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, vol. 2, no. 1, pp. 683–686, 2014.

- [15] J. Treboux and D. J. Genoud, "High precision agriculture: an application of improved machine-learning algorithms," in *Proceedings of the Swiss Conference on Data Science (SDS)*, pp. 103–108, Bern, Switzerland, April 2021.
- [16] M. T. Shakoor, K. Rahman, S. N. Rayta, and A. Chakrabarty, "Agricultural production output prediction using supervised machine learning techniques," in *Proceedings of the International Conference on Next Generation Computing Applications (NextComp)*, pp. 182–187, Balaclava, Mauritius, September 2019.
- [17] S. Wang, S. Di Tommaso, J. Faulkner et al., "Mapping crop types in southeast India with smartphone crowdsourcing and deep learning," *Remote Sensing*, vol. 12, no. 18, p. 2957, 2020.
- [18] R. N. Bhimanpallewar and M. R. Narasingarao, "Alternative approaches of machine learning for agriculture advisory system," in *Proceedings of the International Conference on Cloud Computing, Data Science & Engineering*, pp. 27–31, Noida, India, August 2018.
- [19] R. Medar, V. S. Rajpurohit, and S. Shweta, "Crop yield prediction using machine learning techniques," in *Proceedings of the International Conference for Convergence in Technology (I2CT)*, pp. 1–5, India Pune, August 2019.
- [20] S. Jabjone and C. Jiamrum, "Artificial neural networks for predicting the rice yield in Phimai district of Thailand," *International Journal of Electrical Energy*, vol. 1, no. 3, pp. 177–181, 2013.
- [21] M. M. Rahman, N. Haq, and R. M. Rahman, "Machine learning facilitated rice prediction in Bangladesh," in *Proceedings of the Annual Global Online Conference on Information and Computer Technology*, pp. 1–4, Louisville, KY, USA, December 2014.
- [22] Bangladesh Bureau of Statistics, "Bangladesh Bureau of statistics," 2008, <http://www.bbs.gov.bd>.
- [23] S. Khaki and L. Wang, "Crop yield prediction using deep neural networks," *Frontiers in Plant Science*, vol. 10, p. 621, 2019.