Pakistan Journal of Engineering and Technology (PakJET)
Multidisciplinary & Peer Reviewed
Volume: 5, Number: 1, Pages: 68- 77, Year: 2022

A Predictive Machine Learning and Deep Learning Approach on Agriculture Datasets for New Moringa Oleifera Varieties Prediction

Muhammad Ayoub[1], Shabir Hussain[2,*], Akmal Khan[4], Muhammad Zahid[3], Junaid Abdul Wahid[2], Liao Zhifang[1], Rukshanda Rehman[5]

[1]School of Computer Science and Engineering, Central South University, Changsha, China
[2]School of Information Engineering, Zhengzhou University, Zhengzhou, China
[3]Department of Agronomy, University of Agriculture Faisalabad, Faisalabad, Pakistan
[4]Department of Data Science, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan
[5]Department of Zoology, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan

*Corresponding author: Shabir Hussain (e-mail: shabir@gs.zzu.edu.cn).

Abstract- **Moringa oleifera, the best known of the thirteen species of the genus Moringacae, has achieved importance due to its multipurpose usage with high nutritional value. Very little work has been done in the advancement of moringa varieties in Pakistan. Thus, it needs to develop a new variety of Moringa with better nutritional value. The agrarian performs many experiments like interbreeding Moringa oleifera germplasm with exotic germplasm. Furthermore, they grow it in the nursery and then move it on the field, which almost took six months in a traditional approach. It consumes various resources and time to access the quality of newly developed varieties. This work aims to use machine learning and deep learning approaches to reduce the utilization of various resources and time used by the agrarian to develop a new moringa variety. We used machine learning and deep learning approaches to make predictions about new varieties before their proper plantation. This research work took two moringa parents' varieties with their required features like plant height, protein, potassium, etc. We trained machine learning and deep learning models on the feature values of parents' varieties. Our proposed machine learning model made the best predictions, using parents' plant features to determine these parameter values in their offspring varieties, which will help to choose the best interbreed variety of moringa oleifera.**

***Index Terms*-- Machine Learning, Deep Learning, Agronomy, Moringa Oleifera**

I. INTRODUCTION

Moringa oleifera is a tree with remarkable and significant inherent capacity as it is fast-growing and has drought tolerance. It is a widely cultivated species of the monotypic family, the Moringaceae, which is aboriginal to south Asia. It maturates in Himalayan foothills from northeastern Pakistan to northern West Bengal. All parts of Moringa oleifera (flowers, leaves, seeds, and roots) are valuable for humans and animals' utilization. It is called the world's most useful tree due to its outstanding properties. Such as medicinal use for the treatment of various diseases like endemic diseases. This tree is used for biomass production, ornamental plantings, bio-pesticide, water purification, and industrial and various agronomic uses [1]. Studies reported that dry leaves of Moringa oleifera carry ten times Vitamin A than carrot, 17 times calcium than milk, 15 times potassium than bananas, 25 times iron than spinach, nine times proteins yoghurt, and seven times vitamin C than Orange [2]. The Moringa seeds were found to show the property of normal coagulants/flocculants, which permits developing the tree with the end goal of water and sewage treatment plant to clear turbidity in drinking water ooze in sewage [3]. In Pakistan, there are two notable species of Moringa: Moringa concanensis and Moringa oleifera. Moringa concanensis is exceptionally uncommon and grew and found in the remote territory of Tharparkar, Sindh, Pakistan. In comparison, Moringa oleifera species are found in the Sindh and irrigated fields of Punjab [4]. Due to the multi-benefits of Moringa, it is necessary to develop new varieties. Although many experiments are performed to develop new moringa varieties. The difficulty is that Moringa almost took six months to grow up to 18 feet [5]. The most crucial thing is that we must wait four to six months to analyze the new variety. It depletes a variety of resources, including waster, fertilizer, and various other vital and valuable resources; if the newly offspring variety is not significantly better. It is a complete waste of time and resources. This is unsatisfactory because what are the advantages of the technological period if we still lack technological benefits in agriculture, which is the primary economic resource in Pakistan.

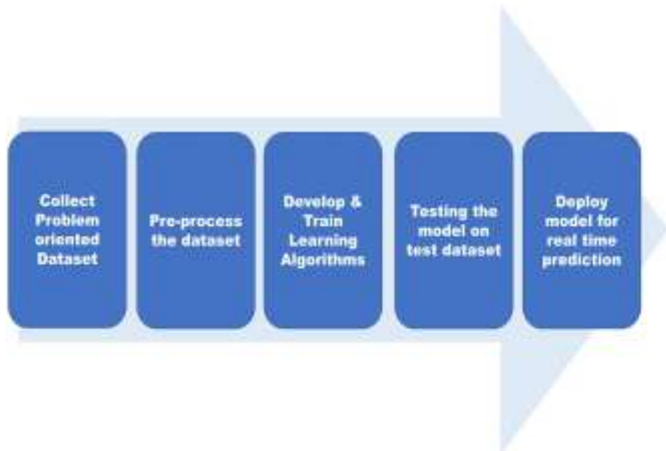

FIGURE 1: Training and deployment process of machine learning models

As a result, our research intends to employ machine learning and deep learning technologies to predict offspring reliably based on their parents and their different parameters before their proper plantation. Machine learning is the sub-field of Artificial Intelligence; it teaches a computer system how to make the best future forecast based on the facts about the provided data. As we know, machine learning models depend on training data in which it finds hidden patterns, and based on these hidden patterns from data, and the model predicts the output for new data points. In short, the more and good training data, the model prediction will be more accurate. The workflow process of how the machine learning algorithm is trained and deployed is shown in Fig. 1.

In general, we first need many data to predict the future using machine learning. After it, we need to check the type of data as it is labelled data or not. For this, machine learning has different types of an algorithm like supervised algorithms for labelled data and unsupervised for non-labelled data. Pre-processing is also performed on the dataset to normalize it and remove unwanted data. After this, we have an essential step: the training of our algorithm in which we split our dataset as training and testing part and feed training data to the model for training. Finally, we have a trained model after all of this process. When we give any observation to that model without its output, our model makes the best prediction and tells us its output on behalf of training. Machine learning models rely on training data in which hidden patterns are discovered, and the model predicts the output of subsequent observations based on these hidden patterns. In other words, the more and better training data there is, the more accurate the output will be.

A) PROBLEM STATEMENT

As mentioned above, no newly developed moringa variety has ever been reported in Pakistan. So agrarians conducted numerous experiments to develop a new moringa variety, such as interbreeding of Moringa oleifera germplasm with foreign germplasm, growing it in a nursery, and transplanting it to the field. This process takes about six months and consumes a significant amount of resources. We will be able to fully access the quality of the newly developed variety after this period and usage of numerous resources. If this newly generated variety is insufficient or inferior to existing types, efforts and resources will have been wasted.

- Agrarian spends a lot of time and resources on interbreeding and planting.
- Agrarians have to wait for six months to analyze the developed variety properly.
- All efforts and resources will be wasted if the newly developed variety is inferior to the existing variety.

B) CONTRIBUTIONS

This research intends to leverage machine learning and deep learning technologies to save time and resources by predicting novel varieties before appropriate interbreeding and planting. In this research work, we developed machine learning and deep learning models, which predict all parameters for Moringa oleifera offspring varieties before their planting. In other words, we employed machine learning and deep learning technologies to reduce the time and resources required by agrarians to generate a new moringa variety. These technologies make very accurate predictions about newly developed varieties based on their parents' varieties.

II. RELATED WORK

Due to the multi-benefits of Moringa, it needs time to develop new varieties. However, many experiments are performed to develop new moringa varieties. There are only a few named varieties of Moringa. Jaffna, grown in South India, produces fruits $60 - 90$cm in length and with the soft flesh of good taste. The other types are Chavakacheri murunga, a Jaffna type, which bears fruits as long as $90 - 120$cm.[6] developed a new variety of Moringa, i.e., Christened PKM-2 annual Moringa, by crossing MP 31 (Eppothum vendran local) and MP 28 (Arasaradi local) varieties. This hybrid exhibits a higher yield by 48% than PKM-1 as far as pods weight has recorded a production of 71.58% greater than PKM-1. On average, each pod weighs around 280g. [7] worked on two moringa varieties, PKM-1 and PKM-2, and observed that fruit set percentage was 16.00, 32.00, and 64.00 in PKM-1 and 47.20, 24.00, 36.00, and 68.00 in PKM-2 under natural conditions but under controlled conditions of crossing, fruit set percentage increased to 68.00 and 78.00%. As the above literature shows, many minor works have been done to develop a new moringa variety in this domain. This process took much time and effort to interbreed Moringa oleifera germplasm with exotic germplasm to develop and evaluate the newly developed variety. Hence, we used machine learning and deep learning approaches to avoid the wastage of time, experimental costs, and other resource used by the agrarian to develop a new moringa variety. Machine Learning is now used in medicine, agriculture, and many more.

Moring is a remarkable tree and is known as a source of income for the local nation by making many valuable products like medicine and an influential role in water and soil conservation. The study [8] used various machine learning and deep learning approaches to predict the regions susceptible to moringa peregrine recovery. The experimental results reported that rainfall is the key indicator that determines the success of the plant establishment. To analyze the potential use of hyperspectral data in predicting biomass yield of different cultivars of Moringa oleifera, the authors of [9] gathered data of five moringa oleifera cultivars. The authors used random forest regression and classification techniques to analyze the data. Experimental results reported that moringa oleifera cultivars could be discriminated from each other using their first-order derivative of reflectance and random forest classifier.

Similarly, the authors [11] used the Plant Village dataset of tomato leaf diseases and rained Convolutional Neural Networks to classify leaf disease of tomato plant. Their experimental results using Convolutional Neural Networks (CNN) reported good accuracy over pre-trained models like VGG-16, Inception v3, etc. They achieved 91.2% accuracy using CNN for the nine diseases and one healthy class. To early detect and classify the leaf disease of rice, the authors [12] obtained a dataset from the UCI machine learning repository of rice leaf disease and trained different machine learning algorithms. Their experimental results using the decision tree algorithm reported a 97% accuracy.

The early identification of the disease of plants will be helpful for farmers to avoid further losses. Hence, the authors [13] used different supervised machine learning algorithms for maize plant disease detection using the image dataset available on the plant village dataset. Using a random forest algorithm, their experimental results achieved the highest accuracy of 79.23% for plant disease detection. The authors [14] used pre-trained deep learning models for face mask detection and classification in three classes (face with a proper mask, face with an improper mask, and face without mask). They achieved 99.81% accuracy by using VGG-16. Another research work for the crop monitoring system is done by [15].

They use different sensors to measure humidity, wind speed, temperature, etc., from the agriculture sector, which affects the crop's health, collects and stores them using IoT applications. These data further feed Artificial Neural Network (ANN) to predict various factors that affect crop growth and yield. On behalf of these predictions, a decision is taken which is helpful for the former. In the same way, [16] used various machine learning techniques for accurate yield prediction and nitrogen state prediction. Their overall accuracy is 81% by using the Supervised Kohonen Networks algorithm. Weed management is essential for crop productivity. Hence, the authors of [17] used a drone of Mavic Pro with the Parrot Sequoia multispectral camera to capture a commercial lettuce crop. They used a support vector machine and YOLO V2 for weed estimation in lettuce crops. Their results achieved F1-scores of 88% 94%, respectively.

Crop productivity is an essential factor for any country; when policymakers know the amount of crop production, this will help them make decisions. [18] use data mining and machine learning algorithms named adaSVM and adaNaive Bayes to predict crop production. They also compared their results with simple SVM and Naive Bayes algorithms in which they achieved higher accuracy by using adaSVM and adaNaive Bayes. Early identification of the stresses in the paddy crops is the key to preventing qualitative and quantitative loss of agricultural yield. Hence, the authors of [19] used a deep convolution neural network to automatically identify and classify different biotic and abiotic paddy crop stresses using the field images, and they achieved 92.89% accuracy.

III. MATERIALS & METHOD

A. DATA STATISTICS

The motivation behind this work is; research conducted on Moringa for creating its new varieties by exploiting indigenous and exotic germplasm at the University of Agriculture Faisalabad, Pakistan. They used three moringa varieties, made six combinations, and initiated pods formation. They used the following moringa varieties for the crossing:

- White
- Black
- PKM-1

Crossing of following germplasm fulfilled by making their combinations as:

I. White × black
II. White × PKM-1
III. Black × White
IV. Black × PKM-1
V. PKM-1 × White
VI. PKM-1 × Black

From these combinations, 3 and 4 were not achieved due to climatic fluctuations and alteration of seasonal behaviour. At the same time, other combinations executed well and provided new diversity with the aim of the high-yielding trait. During this research at the University of Agriculture Faisalabad, we wrote down all parameter values of moringa parents and their offspring varieties while growing in the field. We stored these data for six months as offspring varieties became the entire tree. The stored data was unstructured, which are shown in Tables I, II.

As shown in Tables I and II, we collected the field data of moringa crossing germplasm for almost three seasons. Agrarian researchers also collected these data, performed statistical analysis on these 12 parameters, and suggested the best-developed varieties.

This means these 12 parameters are essential to analyze the quality of offspring varieties.

TABLE I. RAW COLLECTED DATA OF MORINGA OLEIFERA FOR ONE MONTH

Cross Combinations	Plan Height (CM)	No. of Branches	Leaf Length (CM)	No. of Compound leaf in per leaf	Compound Leaf Length (CM)
MW X MB	38.6	5	13.22	6	5.1
MW X PKM-1	37.9	3	12.97	4	4.9
PKM-1 X MW	37.2	4	13.5	4	4.6
PKM-1 X MB	38.1	3	11.82	4	4.3
Parents					
Moringa Black	37.1	3	12.29	4	4.4
Moringa White	37.6	5	13.19	5	3.9
Moringa PKM-1	35.8	3	12.1	4	3.3

TABLE II. RAW COLLECTED DATA OF MORINGA OLEIFERA FOR ONE MONTH

Cross Combination	Leaflet Length (CM)	Total No of Leaflets per leaf (CM)	CHL A	CHL B	Phenolic	Potassium (K)	Protein
MW X MB	1.6	111	1.52	0.52	0.25	5	0.06
MW X PKM-1	1.5	112.2	1.47	0.55	0.23	4	0.05
PKM-1 X MW	1.9	105	1.45	0.49	0.22	4	0.05
PKM-1 X MB	1.93	103	1.36	0.58	0.23	3	0.05
Parents							
Moringa Black	1.6	107	1.44	0.43	0.23	3	0.06
Moringa White	1.9	110	1.39	0.35	0.22	5	0.05
Moringa PKM-1	2.4	109	1.44	0.40	0.24	4	0.05

We collected these data in the amorphous form; we performed pre-processing to make it a normalized and sense-able format so that machine learning can further process it. Now first, we look at our methodology framework.

B. METHODOLOGY FRAMEWORK & FEATURE EXTRACTION

The proposed methodology of our work is shown in Fig. 2. In the first step, we collected the moringa oleifera data from the University of Agriculture Faisalabad, which was unstructured, as described and shown above. After collecting raw data, the next step is pre-processing to make it sensible. In this study, we selected only two moringa varieties: Moringa white and black and separated the complete data of parents and offspring varieties from raw collected data. We performed different pre-processing techniques [20] to deal with the missing values in our dataset and remove noisy data.

C. FEATURE EXTRACTION

After pre-processing, the next step is feature identification; in this process, we identified distinguishing characteristics and parameters, which helps decide whether offspring variety is good. According to the research conducted at the University of Agriculture Faisalabad, we also used 12 different features/parameters of parent plants and newly developed varieties for experimental purposes. These 12 parameters are also verified by domain expert Dr Shehzad Maqsood Ahmed Basra, The chairman Department of Agronomy, University of Agriculture Faisalabad. The detailed description of these features/parameters are described below:

- Plant height: Describes the height of parents' moringa plant and newly developed varieties; these measurements are in cm (centimetre). This means what the heights of parents are and what is the height of newly developed varieties.

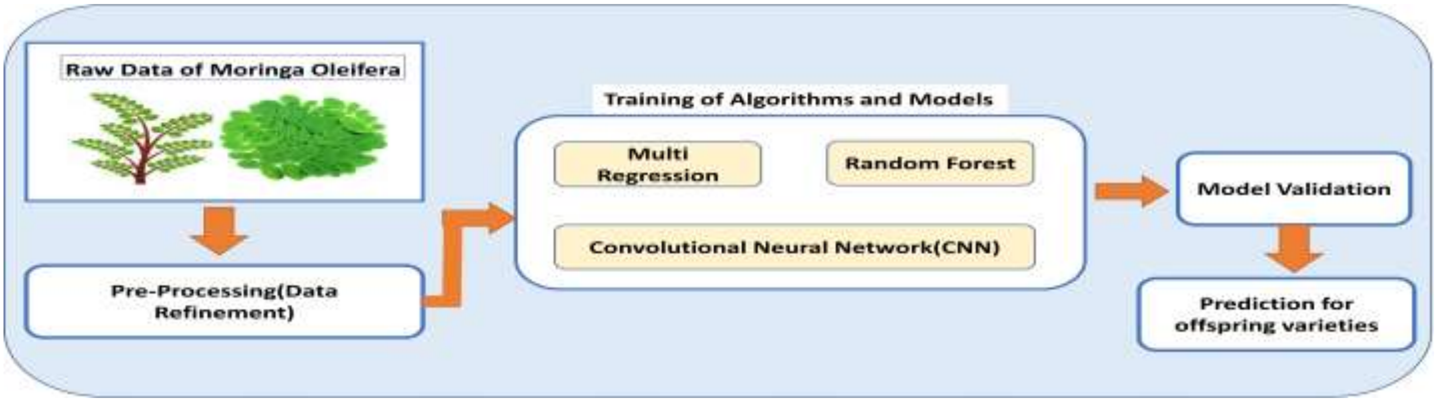

FIGURE 2: Proposed methodology framework.

- Number of Branches: Describes how many branches are of parents and how many branches are in newly developed varieties.
- Leaf Length: Describes the length of leaf in parents and also in new varieties; this feature is counted on average; we randomly pick three leaves and calculate the average leaf length.

- Compound Leaf Length: For the leaf length, measure from the tip of the entire leaf down to the base of the lowest leaflets where they meet the leaf stem if the leaf is complex (many leaflets are attached to a main leaf stem/petiole).

- Leaflet Length: The leaflet lamina of fully expanded leaves (measured from leaflet base to leaflet tip, omitting the petiole and any expansions of the venation) of compound leaves is measured in mm.

- Number of Leaflets in Per Leaf: A leaflet (occasionally called foliole) in botany is a leaf-like part of a compound leaf.
- CHL-B: Describes the amount of chlorophyll-B in parents and offspring varieties.
- Phenolic: Describes the amount of phenolic in parents and in new in developed varieties.
- Phenolic: Describes the amount of phenolic in parents and offspring varieties.
- Potassium: Describes the amount of potassium in parents and offspring varieties.

- Protein: Describes the amount of protein in parents and offspring varieties.

- CHL-A: Describes the amount of chlorophyll-A in parents and new varieties.

For calculating the values of phenolic, chlorophyll-A, chlorophyll-B, protein, and potassium, we used state of the art biochemical analysis techniques, which are described below:

- Determination of total phenolic in leaves in the unit of (µg/g of fresh wt.) [21].
- Determination of chlorophyll contents in leaf (mg/g of fresh wt.) [22].
- Determination of Protein [23].
- Determination of Potassium [24].

This is how we collected the data, performed pre-processing, and identified & specified features. The sample of our dataset is shown in Table III.

TABLE III. SAMPLE DATASET AFTER PRE-PROCESSING

Input Height (CM)				Output Height (CM)	
MB	39	MW	38.3	MW X MB	39
MB	39.6	MW	40	MW X MB	44.3
MB	42.2	MW	49.1	MW X MB	51

As shown in Table III, a sample of pre-processed dataset described the height parameter of moringa oleifera. In this table, on the left side, we have the data of parent's varieties that reported the height of parent's varieties in different time intervals of six months of growing. On the right side, we have the height of the offspring variety. After all these processes, which include collecting the data, pre-processing the data, and identification & specification of features, the next step is the training of our model. For the training and validation, we used different machine learning and deep learning algorithms like; multi regression, random forest, and convolutional neural network as a deep learning model. The base method for logistic regression is the logistic function known as the sigmoid function. As shown in Fig. 3, it is an S-shaped curve developed by statisticians who take any real-valued number and map it into a value between 0 and 1 but never precisely at those limits.

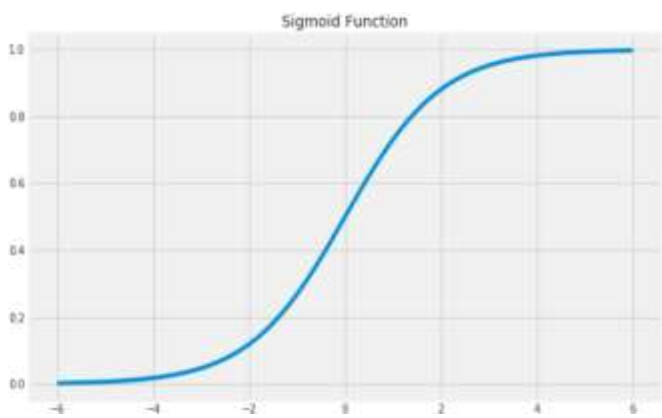

FIGURE 3: Sigmoid Function

The logistic regression equation has a very similar representation to linear regression. The difference is that the modelled output value is binary, as shown in equation 1.

$$y = \frac{e^{\beta_0 + \beta_{1x1}}}{1 + \beta_0 + \beta_{1x1}} \tag{1}$$

In equation 1, the β_0 is the intercept term, β1 is the coefficient for x1, and y is the predicted output with an absolute value between 0 and 1.

Random forest is based on bagging and feature randomness when building each tree to create an uncorrelated forest of trees whose prediction by committee is more accurate than any individual tree. While on CNN, we used sequential architecture, relu activation function, adam as an optimizer, and mean squared error as a loss function.

D. EXPERIMENT AND ANALYSIS

This study aims to leverage machine learning and deep learning technologies to save time and resources by predicting novel varieties before appropriate interbreeding and planting. To make predictions about new cultivars before they could be planted, we used various machine learning and deep learning algorithms. We used multi regression, random forest, and CNN in python. We fed the data of five months to the model for the training and used the one-month data to test and validate our model. As we know, a regression algorithm works when there are one independent variable and one dependent variable. At the same time, multiple linear regression works when there is more than one independent variable and one dependent variable, as shown in Fig. 4.

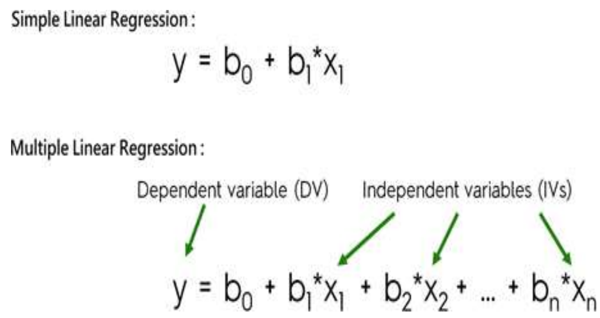

FIGURE 4: Basic Regression Algorithm

Hence, we cannot predict 12 parameters as output simultaneously. We trained a separate model for every 12 parameters. Our proposed model structure for training and prediction for each parameter is shown in Fig. 5.

As shown in Fig. 5, we trained separate models on every 12 parameters of moringa oleifera by giving the parent's plant parameter and newly developed varieties. After the training, we just input the parameter of the parent's plants. Our models predict the value of each parameter in newly developed varieties, as shown in Fig. 5. We repeated the process on all 12 parameters of the moringa plant; we took a single individual parameter, trained the separate model for this parameter and then validated these models.

IV. RESULTS AND DISCUSSION

To develop a new variety of moringa oleifera, the agrarian performs many experiments like interbreeding of Moringa oleifera germplasm with exotic germplasm. This research work employed machine learning and deep learning techniques to anticipate newly produced varieties based on the parent plant parameters. Our strategy eliminates wasting time and other resources in developing a new moringa variety. We used Multi Regression and Random Forest as machine learning algorithms and CNN as a deep learning model for experimental purposes. We trained our machine learning and deep learning models using three seasonal data of Moringa (moringa black and Moringa white), as mentioned in the above section for our experiment and analysis. The experimental results using multi regression and random forest in terms of accuracy for each parameter on test data are shown in tables IV and V, respectively.

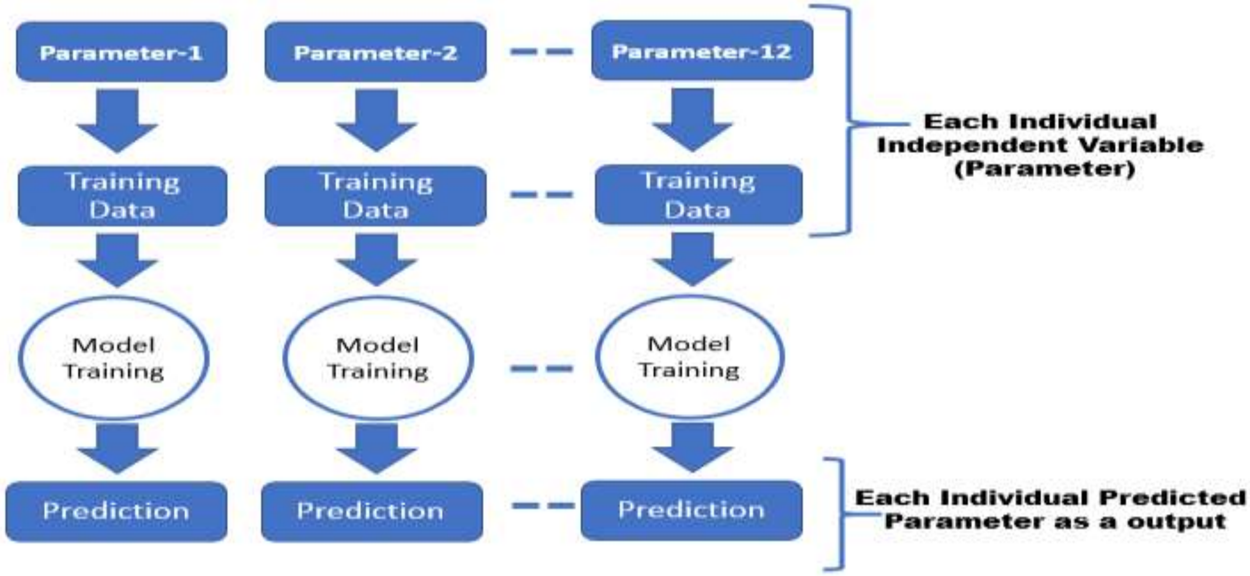

FIGURE 5: Training and prediction for each moringa parameters

TABLE IV. STATISTICS IN TERMS OF ACCURACY USING MULTI REGRESSION

Moringa oleifera Parameters	Accuracy
Total Number of Leaflets in per leaf	91.67%
Protein	75.00%
Phenolic	72.21%
Compound Leaf per leaf	66.67%
Number of Branches	70.00%
Leaflet Length	83.33%
Leaf Length	90.00%
K(potassium)	75.00%
Height of the plant	90.00%
compound-leaf-length	93.05%
chlorophyll-B	75.00%
chlorophyll-A	75.00%

TABLE V. STATISTICS IN TERMS OF ACCURACY USING RANDOM FOREST

Moringa oleifera Parameters	Accuracy
Total Number of Leaflets in per leaf	99.96%
Protein	99.09%
Phenolic	93.23%
Compound Leaf per leaf	66.67%
Number of Branches	70.00%
Leaflet Length	99.09%
Leaf Length	89.09%
K(potassium)	83.33%
Height of the plant	98.65%
compound-leaf-length	79.85%
chlorophyll-B	93.23%
chlorophyll-A	99.09%

In the same contrast of random forest and multi regression algorithms, we trained and evaluated our convolutional neural network model in terms of actual and predicted values. The experimental results of the deep neural network for each parameter are shown in Fig. 6 and Fig. 7. As shown in Fig. 6 and 7, the x-axis represents the number of observations, and the y-axis represents each observation's output or targeted value. These two graphs reported the actual and predicted value for each observation from test data. In contrast, the green line represents the actual value of each parameter in offspring variety. Similarly, the blue line represents the value from our model prediction for each parameter in offspring variety. Hence, in short, Fig. 6 and 7 compare actual value and our model predicted value.

We validated our model by feeding one-month data and evaluating our model by comparing the actual value and the predicted value for each parameter of moringa oleifera. Validation is the process of checking how accurate our model is working. As we described above, we have three seasonal data of moringa oleifera. We used data of five months for training purposes and one month for testing our model. After the training of our model, we feed one month's data to our model for testing it. As a result, our proposed approach for making predictions about new varieties before their proper plantation reported good accuracy and reliability. What were the actual values of this one-month data, and what is our model predicting? A compression is shown in Table VI.

Furthermore, the performance of the regression model is evaluated by using different computational methods such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error, as described below.

- Mean Absolute Error (MAE): The loss function for the regression models. It is the sum of absolute differences between the actual and predicted values, as shown in equation (2).

$$[\frac{1}{n}\sum_{i=1}^{n}|Actual - Predicted|] \qquad (2)$$

- MSE defines the calculating distance from the data point to the regression line and squares it. This is known as Mean Squared Error, and it is calculated as equation (3):

$$[\frac{1}{n}\sum_{i=1}^{n}|Actual - Predicted|^2] \qquad (3)$$

- Root Mean Squared Error: It is defined as the square root of the squared errors' mean. It is calculated by equation (4):

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}|Actual - Predicted|^2} \qquad (4)$$

TABLE VI. ACTUAL VS PREDICTED VALUE FOR EACH PARAMETER

Name of Parameter	Actual Value	Predicted Value
CHL A	1.339	1.442
CHL B	0.702	0.753
No. of compound leaf per leaf	3	5.0
Compound leaf length (average) cm	9.6	9.6
K (Potassium)	8	7.0
Leaf length (average) cm	20.3	28.53
Leaflet length (cm) average	3.7	3.7
no.of branches	3	5.0
Phenolic	0.269	0.292
Protien	0.086	0.084
Total no. of leaflets per leaf (average)	120	119.424
plant height (cm)	51.0	40.236

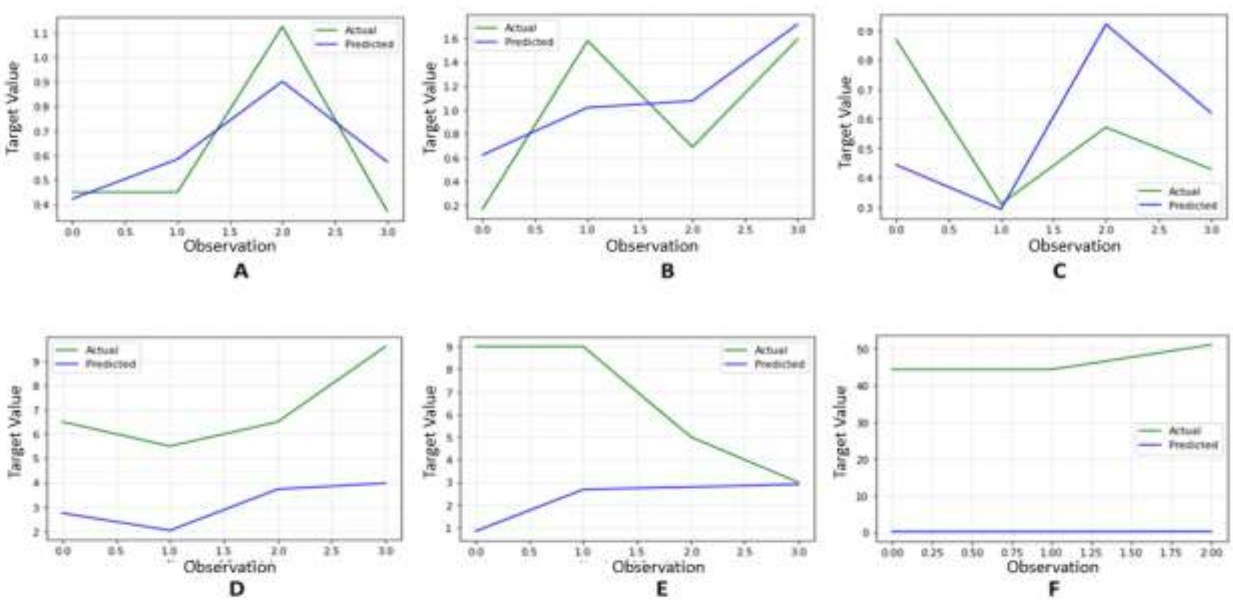

FIGURE 6: Statistics of convolutional neural network for moringa oleifera offspring variety prediction. Figures A, B, C, D, E, and F reported the actual and predicted value for the potassium chlorophyll-A, chlorophyll-B, Compound Leaf Length, Compound leaf per leaflet, and height of moringa oleifera, respectively.

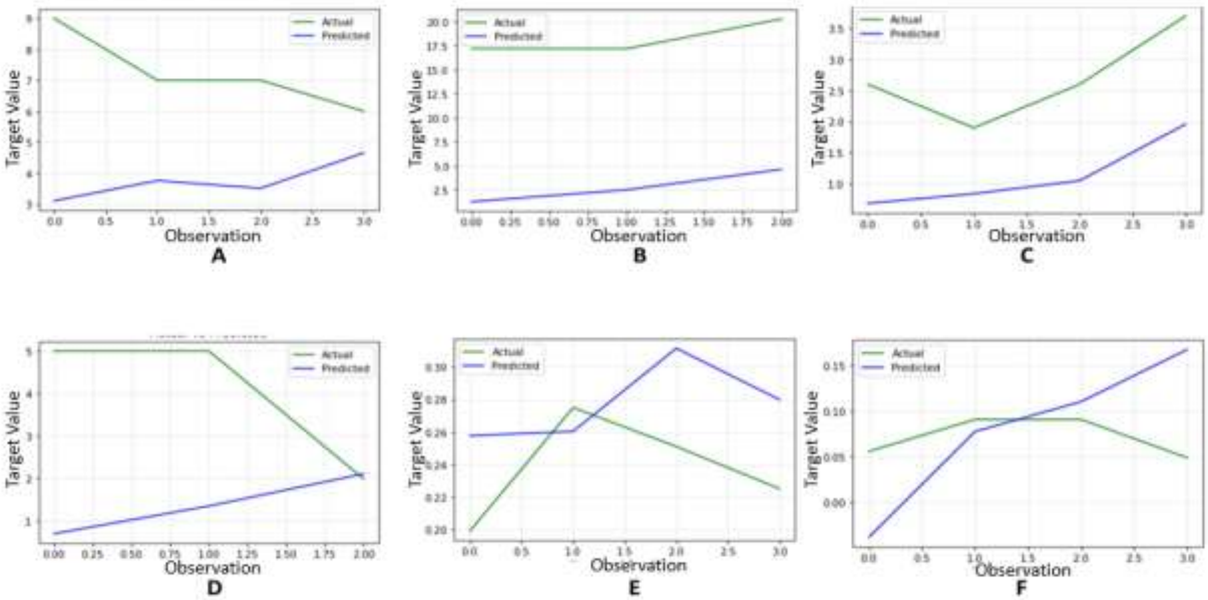

FIGURE 7. Statistics of convolutional neural network in moringa oleifera offspring variety prediction. Figures A, B, C, D, E, and F reported the actual and predicted value for leaf length, leaflet length, number of branches, phenolic, protein, and the total number of leaflets per leaf for the moringa oleifera, respectively.

V. CONCLUSION

Moringa oleifera is a tree with remarkable and significant inherent capacity as it is fast-growing and has drought tolerance. Due to the multi-benefits of Moringa, it needs time to develop new varieties. The most crucial thing is that we must wait four to six months to analyze the newly developed variety. If the newly offspring variety is not significantly better, it wastes various resources, including water, fertilizer, and various other vital and valuable resources. This research intends to employ machine learning and deep learning technologies to predict the reliability of offspring variety based on their parents and their different parameters before their proper plantation. Our proposed model takes an input of thirteen parameters of parent variety and predicts what will be the value of these thirteen parameters in offspring variety. Our proposed approach reduced the time consuming and other resources used by the agrarian to develop a new moringa variety. Our proposed model is trained using different machine learning and deep learning models. We evaluated our models in terms of accuracy and compared the predicted results with actual values. As we discussed in the above section, the prediction of our machine learning models for every parameter of moringa oleifera is almost the same as the actual value, which means our model works well. We achieved 80% and 86% of accuracy by using random forest and multi regression algorithms, respectively.

ACKNOWLEDGMENT

This research did not receive any specific grant from any funding agencies. Muhammad Ayoub: Conceptualization, Methodology, Software, Writing-Original Draft, Writing-review and Editing, Investigation, Visualization. Shabir Hussain: Writing-review and Editing, Resources, Investigation, Visualization, Software, Methodology. Muhammad Zahid: Data Curation, Writing-review and Editing, Resources. Junaid Abdul Wahid: Writing-review and Editing, Resources, Investigation.

REFERENCES

[1] George William Crosby. Soilless culture of Moringa (Moringa oleifera Lam.) for the production of fresh biomass. PhD thesis, University of Massachusetts Amherst, 2007.

[2] Lowell J Fuglie et al. The miracle tree: Moringa oleifera, natural nutrition for the tropics. 1999.

[3] Grasiele Scaramal Madrona, Geovanna Bordini Serpelloni, Ang´elica Marquetotti Salcedo Vieira, Let´ıcia Nishi, Karina Cordeiro Cardoso, and Rosangela Bergamasco. Study the effect of saline solution on the extraction of the moringa oleifera seed's active component for water treatment. *Water, Air, & Soil Pollution*, vol. 211, no. 1, pp. 409–415, 2010.

[4] S Ali and E Nasir. Flora of West Pakistan. No. 42, Dilleniaceae. University of Karachi, 1973.

[5] https://www.summerwindsnursery.com/az/inspire/blog/what-you-need-to-know-about-the-amazing-moringa-tree/

[6] P Selvakumari, V Ponnuswami, et al. Genetic diversity of Moringa (moringa oleifera lam) using ssr markers, International Journal of Tropical Agriculture, vol. 33, no. 2, pp. 943–946, 2015.

[7] Oladosu, Yusuff, Mohd Y. Rafii, Chukwu Samuel, Arolu Fatai, Usman Magaji, Isiaka Kareem, Zarifth Shafika Kamarudin, Isma'ila Muhammad, and Kazeem Kolapo. "Drought resistance in rice from conventional to molecular breeding: a review." *International journal of molecular sciences* 20, no. 14, pp. 3519, 2019.

[8] Moradi, E., Abdolshahnejad, M., Hassangavyar, M. B., Ghoohestani, G., da Silva, A. M., Khosravi, H., & Cerdà, A., "Machine learning approach to predict susceptible growth regions of Moringa peregrina (Forssk)" *Ecological Informatics*, vol. 62, pp.101267, 2021.

[9] G Sambasivam and Geoffrey Duncan Opiyo. A predictive machine learning application in agriculture: Cassava dis- ease detection and classification with imbalanced dataset using convolutional neural networks. *Egyptian Informatics Journal*, 2020.

[10] Tshabalala, T., Abdel-Rahman, E. M., Ncube, B., Ndhlala, A. R., & Mutanga, O. "Leveraging of hyperspectral remote sensing on estimating biomass yield of Moringa oleifera Lam. medicinal plant" *South African Journal of Botany*, vol. 140, pp. 37-49, 2021.

[11] Mohit Agarwal, Abhishek Singh, Siddhartha Arjaria, Amit Sinha, and Suneet Gupta. Toled: Tomato leaf disease detection using convolution neural network. *Procedia Computer Science*, vol. 167, pp.293–301, 2020.

.

[12] Kawcher Ahmed, Tasmia Rahman Shahidi, Syed Md Irfanul Alam, and Sifat Momen. Rice leaf disease detection using machine learning techniques. In 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), pages 1–5, IEEE, 2019.

[13] Kshyanaprava Panda Panigrahi, Himansu Das, Abhaya Kumar Sahoo, and Suresh Chandra Moharana. Maize leaf disease detection and classification using machine learning algorithms. *In Progress in Computing, Analytics and Networking*, pages 659–669. Springer, Singapore, 2020.

[14] Shabir Hussain, Yang Yu, Muhammad Ayoub, Akmal Khan, Rukhshanda Rehman, Junaid Abdul Wahid, and Weiyan Hou. Iot and deep learning based approach for rapid screening and face mask detection for infection spread control of covid-19. *Applied Sciences*, vol. 11, no. 8, pp. 3495, 2021.

[15] Anuja Chandgude, Nikita Harpale, Diksha Jadhav, Punam Pawar, and Suhas M Patil. A review on machine learning algorithm used for crop monitoring system in agriculture. *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 4, pp. 1470, 2018.

[16] Anna Chlingaryan, Salah Sukkarieh, and Brett Whelan. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers And Electronics in Agriculture*, vol. 151, pp. 61–69, 2018.

[17] Kavir Osorio, Andr´es Puerto, Cesar Pedraza, David Jamaica, and Leonardo Rodr´ıguez. A deep learning approach for weed detection in lettuce crops using multispectral images. *AgriEngineering*, vol. 2, no. 3, pp. 471–488, 2020.

[18] Narayanan Balakrishnan and Govindarajan Muthukumarasamy. Crop production-ensemble machine learning model for prediction. *International Journal of Computer Science and Software Engineering*, vol. 5, no. 7, pp. 148, 2016.

[19] Basavaraj S Anami, Naveen N Malvade, and Surendra Palaiah. Deep learning approach for recognition and classification of yield affecting paddy crop stresses using field images. *Artificial Intelligence in Agriculture*, vol. 4, pp.12–20, 2020.

[20] Brownlee, J. Data preparation for machine learning, 2022

[21] Chigurupati, S., Al-Murikhy, A., Almahmoud, S. A., Almoshari, Y., Ahmed, A. S., Vijayabalan, S., ... & Palanimuthu, V. R., "Molecular docking of phenolic compounds and screening of antioxidant and antidiabetic potential of Moringa oleifera ethanolic leaves extract from Qassim region, Saudi Arabia," *Saudi Journal of Biological Sciences*, vol. 29, no. 2, pp. 854-859, 2022.

[22] Sarwar, S., Akram, N. A., Saleem, M. H., Zafar, S., Alghanem, S. M., Abualreesh, M. H., ... & Ali, S. "Spatial variations in the biochemical potential of okra [Abelmoschus esculentus L.(Moench)] leaf and fruit under field conditions," *PLOS ONE*, vol. 17, no. 2, pp. e0259520, 2022.

[23] Bassogog, C. B. B., Nyobe, C. E., Ngui, S. P., Minka, S. R., & Mune, M. A. M. "Effect of heat treatment on the structure, functional properties and composition of Moringa oleifera seed proteins," *Food Chemistry*, pp. 132546, 2022.

[24] Irshad, S., Matloob, A., Iqbal, S., Ibrar, D., Hasnain, Z., Khan, S., ... & Diao, Z. H. "Foliar application of potassium and moringa leaf extract improves growth, physiology and productivity of kabuli chickpea grown under varying sowing regimes" *PLOS ONE*, vol. 17, no. 2, pp. e0263323, 2022.