

# A Cost-Effective Farmer Support System For Better Yield Prediction And Resource Management

**Priyanka Prashanth Kumar**

Department of Information Science and Engineering  
Jyothy Institute of Technology Bangalore, India

**Prajwal G A**

Department of Information Science and Engineering  
Jyothy Institute of Technology Bangalore, India

## Corresponding author.

**Correspondence: Priyanka Prashanth Kumar**

E-mail: priyankaprashanth.kumar@gmail.com

## Article info

Received 22 th February 2022 Received in revised form 31 March 2022 Accepted 3 July 2022

## Keywords

Agriculture, technology, machine learning, soil composition, climatic parameters, supervised learning methods, multi perceptron, multiclass classification algorithms.

## Abstract

Agriculture is a significant part of India's economy. Agriculture with technology can lead to groundbreaking improvements. Machine learning plays a significant role here, to help predict and formulate essential results. The main focus of this project is to predict and suggest suitable crops to sow in certain soil composition and climatic parameters based on existing historical data collected over years and later subjecting it to Supervised Learning methods and Multi - Class Classification algorithms. So, we propose to create a user-friendly smartphone application that can provide a list of top crops that can be produced in a specific type of soil, and the accuracy of our system can be confirmed using a government website called "*Soil Health Card*," which collects data on soil composition and its related data.

## I. INTRODUCTION

India is the second largest producer of wheat, rice, dry fruits, sugarcane, pulses, and a whole variety of vegetables in the world. Agriculture makes up 20.19% of the Indian GDP as of 2021. But, with limited amounts of land, water, energy and many more resources, it is highly difficult to organize and plan a successful yield for an ever growing population and with the external catastrophes, such as wars, pandemics and what not, the status of the people can stoop down to unimaginable depths.

According to the Department for Promotion of Industry and Internal Trade which has launched a StartupIndia Program [1], it has been observed that Low yield is one of the most pressing concerns confronting India's agricultural sector: India's farm production is 30-50 percent lower than that of wealthy countries.

Low agricultural production is caused by factors such as average farm size, insufficient infrastructure, a lack of utilization of farm technologies and best farming techniques, decreased soil fertility owing to excessive fertilization, and continued pesticide use. Because Indian farms are small, they have limited access to financial services, credit (or lenders), support skills,

educational services, and irrigation solutions. In the short term, yield has a direct impact on a farmer's cash flow and ability to adapt to market swings.

In the long-term, yield limits a farmer's ability to invest in their farm's future to increase productivity and reduce risks associated with their crops, but also to invest in their families in areas such as education, healthcare, training, and so on. Farmers are well-versed in their knowledge of agriculture and farming styles, but their knowledge may be limited in terms of only what they were taught from their respective ancestors, preventing them from realizing the full potential of a successful crop yield.

In times like this, technology can be used as a tool in order to bring balance. Machine Learning and Deep Learning methods can be used to build Intelligent Information Systems (IIS). According to a recently published research study [2] that examines the technology characteristics roadmap for a mission-critical intelligent systems world, 80 percent of technological ideas desire intelligent systems that can be considered successful within the next five years.

The proposed solution mentioned throughout the paper inculcates many valuable data resources. The Ministry of Agriculture and Farmers Welfare has provided valuable information about composition of the soil present all over India through their website Soil Health Card [3].

With the help of the internet, there is access to the real time global climatic conditions. Engineers can provide a tool for the public to efficiently organize and plan their resources in order to obtain the maximum yield using an abundance of such data.

In Agriculture, the use of Neural Networks will aid in the prediction of multiple alternative outcomes for the same set of data. The idea is to facilitate the farmers with the right options, which will help them plan better for a good crop yield by effective utilization of resources. The proposed strategy is to develop a mobile application that can detect a user's location using GPS or manually enter it. The data sources and our dataset include soil and weather data specific to that location. To generate a list of crops that are best suited to grow in a given location, a Machine Learning / Deep Learning model will be built using the best suitable classifiers, such as multi-class classification neural networks. The model will be used to provide the best and next best results in a smartphone application.

## II. LITERATURE SURVEY

In this section, the work done in previous published papers to predict crop yield and select the best crop is reviewed.

In order to extract and synthesize the techniques and features that have been employed in agricultural yield prediction research, a comparative Systematic Literature Review (SLR) was carried out in this paper [4]. The analysis shows that the most frequently used features in these models are temperature, rainfall, and soil type, and the most frequently used technique is

artificial neural networks. The prominent deep learning algorithms used are CNN, LSTM, and DNN. The popular models are Random Forest, Neural Networks, Linear Regression, and Gradient Boosting Tree.

The yield of almost all crops grown in India is predicted by the model created in [5]. The user can anticipate the crop output in the preferred year by using straightforward characteristics like State, District, Season, and Area. In order to estimate the yield, advanced regression techniques including Kernel Ridge, Lasso, and ENet algorithms are utilized, along with a concept known as "Stacking Regression" to improve the algorithms and produce a more accurate prediction.

Linear regression, fuzzy logic, and anfis are all used in this comparative study. The input parameters for the APSim, which was utilized to gather the data, were biomass, esw, radiation, rain, and wheat yield. The RMSE values were used to compare the models' accuracy, and the anfis model outperformed the others [6]. However, as stated in [7], anfis is only appropriate for models with a limited number of input parameters. In this study, artificial neural networks, K-Nearest Neighbors, random forests, support vector machines, and decision trees are compared. The paper first analyses these models before presenting its own crop selection algorithm for deciding the order of crops to be planted during a season.

This study [8] suggests a creative method of agricultural yield prediction and suggests the ideal environmental variables to maximize crop yield. This forecast will aid farmers in determining the ideal temperature and moisture level for the best crop yield. To predict the crop yield per acre, the method used Multivariate Polynomial Regression, Support Vector Machine Regression, and Random Forest Models. The United States Department of Agriculture's yield and weather data are used in the suggested method. Humidity, yield, temperature, and rainfall are some of the different characteristics in the dataset. At the conclusion, different metrics are calculated to compare the models, including RMSE, MAE, median absolute error, and R-square values.

This research [9] introduces a key framework for the application of machine learning systems in farming. In order to discover which machine learning algorithm is the most accurate at predicting the best harvest for a given field, several methods were compared. The best harvest is the one that outperformed previous years in terms of yield per unit area. The study focuses on six important Bangladeshi crops: potato, wheat, jute, aman rice, aus rice, and boro rice. K-Nearest Neighbor Regression and Multiple Linear Regression (MLR) were employed as algorithms (KNNR). Multiple Linear Regression (MLR), which was used in the analysis, produced the most precise results, and it was added to an Android app.

The suggested system [10] helps farmers choose the best crops based on the sowing season and region. Farmers will gain from it as well because it would boost their net profit. The system develops a model or approach that can suggest a list of crops that is particularly valuable for farmers in their decision-making by taking into account different datasets with respect to five parameters of horticulture data, such as rainfall, temperature, slope, humidity, and soil moisture. This research proposes a useful crop prediction algorithm based on

historical data. A pattern matching method is utilized to retrieve the crop according to region and season after analyzing datasets with Xarray functions.

The majority of the data used in precision agriculture is unstructured because it can be gathered from a variety of sources. Sensors can gather information about farming, including temperature, soil moisture, and nutrient content. While rainfall, heat radiation, and air pressure can all be recorded by weather stations. A statistical yearly report can be used to gather historical data from government colleges, agricultural groups, and websites, such as crop yields from prior years. [11,12,13].

The Modified Recursive Feature Elimination (MRFE), a feature selection technique used by the system [14], facilitates the identification of pertinent parameters from the obtained data set for crop prediction. The MRFE methodology chooses and ranks prominent features using a ranking method. The analysis' findings indicate that bagging is the most precise way for predicting suitable crops, while the MRFE technique is superior for feature selection. The soil properties were taken from Sankarankovil Taluk in Tenkasi, India, while the environmental dataset was obtained from the website of the Tamilnadu Agricultural University. 1000 examples with 16 attributes, 12 of which are soil characteristics, are included in the data set. The other four features are environmental characteristics. The research that was done in this paper used various existing wrapper feature selection techniques such as Boruta, SFFS, RFE, and the proposed MRFE technique.

This article's [15] author uses several machine learning techniques to study crop yield. Combinations are what they are mostly interested in. The testing outcomes demonstrated that, among all the approaches used to implement the model, the XGBoost algorithm is the most demeaning and incomparable. Different kinds of machine learning models can be applied. It is possible to add other crop varieties and geographic areas.

### III. PROPOSED METHODOLOGY

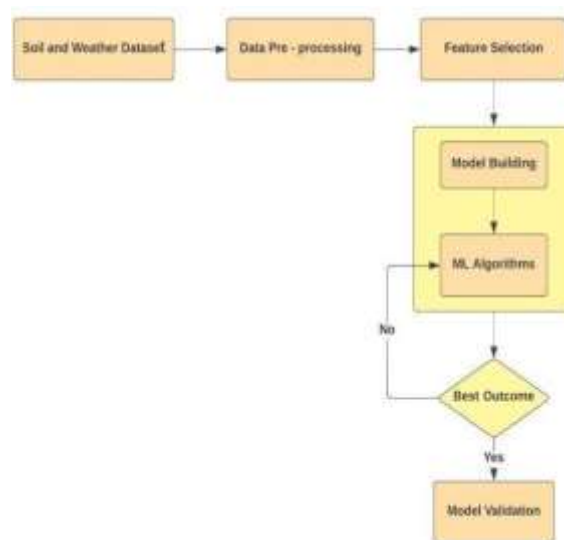


Fig 1. Model Workflow

The proposed system's details are presented in a series of steps (Fig 1). Some of these are:

- Raw dataset collection
- Date Pre – Processing
- Model Building

**A. Datasets**

To begin addressing the previously discussed problem, all of the data that represents and determines the successful growth of all of the various crops is necessary. Crop growth is influenced by a variety of factors. The proposed study will consider two major factors: soil composition and climate data.

To take note of soil composition, it is necessary to understand that soil includes parameters such as many minerals, pH levels, moisture levels, and others. Sodium, Potassium, Magnesium, and Nitrogen are a few examples. There are also rainfall levels, humidity, topography, and other climatic factors to consider.

With the above-mentioned elements present in our dataset predicting over 22 different varieties of crops, our system would be able to analyze each crop across an array of attributes and determine what levels of each element result in a successful yield for that particular crop. Our dataset makes it simpler to determine which crop is best suited for a given set of values for all of the distinct attributes. Fig 2. shows the dataset we have used that contains over 2000 rows worth of data collected and preprocessed later.

N	P	K	temperature	humidity	pH	rainfall	label
90	42	43	26.87974371	82.00274423	6.50298529	202.935536	rice
85	58	41	21.77040169	80.31964608	7.03809636	226.605537	rice
60	55	44	23.00445915	82.3207629	7.84020714	263.964248	rice
74	35	40	26.49109635	80.15816264	5.98040091	242.864034	rice
78	42	42	26.13017482	81.60487287	7.62847289	262.717341	rice
69	37	42	23.05804872	83.37011772	7.0734535	251.055	rice
09	55	38	22.70883798	82.63941394	5.70080568	271.32486	rice
94	53	40	20.27774362	82.89408619	5.71862718	241.974195	rice
89	54	38	24.51588066	83.5352163	6.68534642	230.446736	rice

**Fig 2. Structured Dataset used**

This set of data may contain various properties required to determine the crops, but the only issue is that the data may be outdated. To have updated values for these characteristics in our dataset, access to sensors and data from across the country would be required. This is a very time-consuming and tedious procedure.

However, there is a government entity; Soil Health Card [3], that provides such soil composition values observed nationwide on their platform. This data can also be accessed in a variety of languages for the convenience of farmers and others. Because the data is updated frequently, our system would be able to formulate predictions almost in real time as it would have access to the most recent data of this level of accuracy. Fig 3. represents the Soil Health Card outlook.

Fig 3. Soil health card

**B. Dataset Pre - Processing**

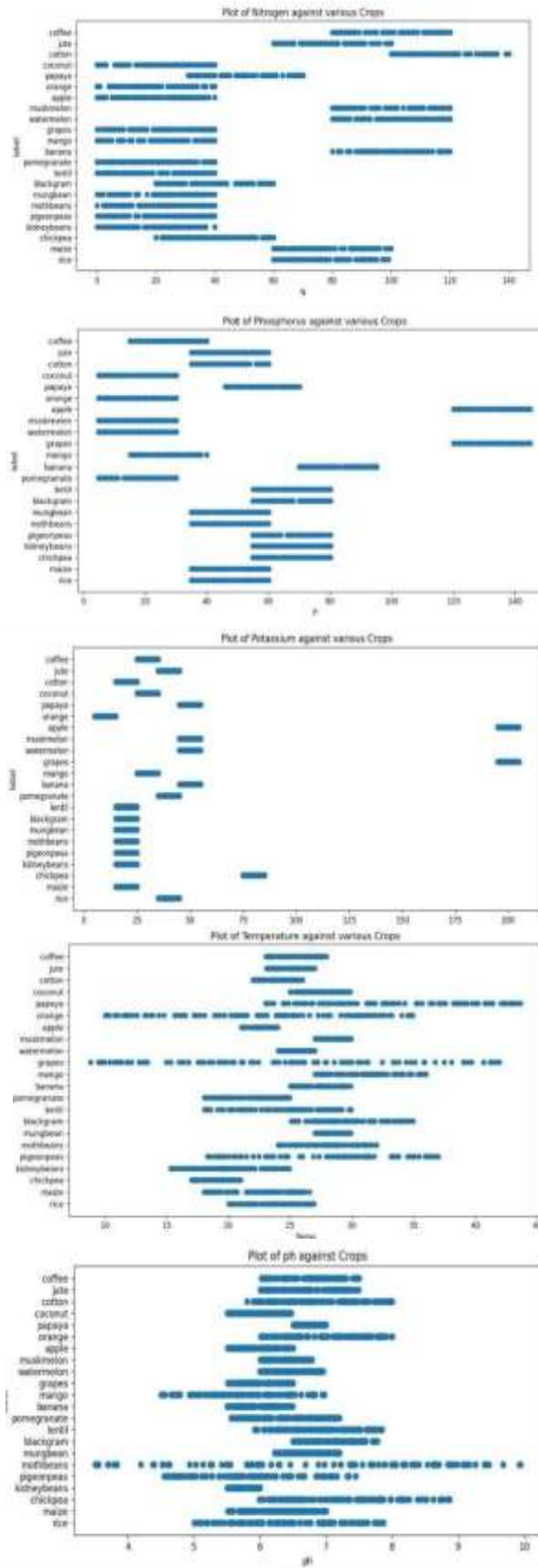
As previously stated, crop yield can be influenced by a variety of factors such as climate, soil, and so on. The data obtained contained a large number of outliers, null values, and discontinuous values. An outlier is a data point that stands out from the rest. Therefore, structuring the data is essential to build our machine learning model. Before data can be used, it must be preprocessed.

The concept of data preprocessing is the transformation of raw data into a clean data set. Data preprocessing is a vital part in Machine Learning because the quality of data and the useful information that can be extracted from it directly affects our model's ability to learn; thus, it is critical that we preprocess our data before feeding it into the model. As a result, our model's accuracy and efficiency improve.

A variety of techniques and methods are applied to the dataset during data preprocessing. For instance, Exploratory Data Analysis(EDA), Data cleaning, Data Integration, Data Transformation etcetera.

**Exploratory Data Analysis.** Exploratory Data Analysis is the critical process of performing preliminary investigations on data in order to discover patterns, spot anomalies, test hypotheses, and check assumptions using summary statistics and graphical representations /

visualizations before subjecting it to Data cleaning methods. Fig 4. represents the various EDA plots of our dataset.



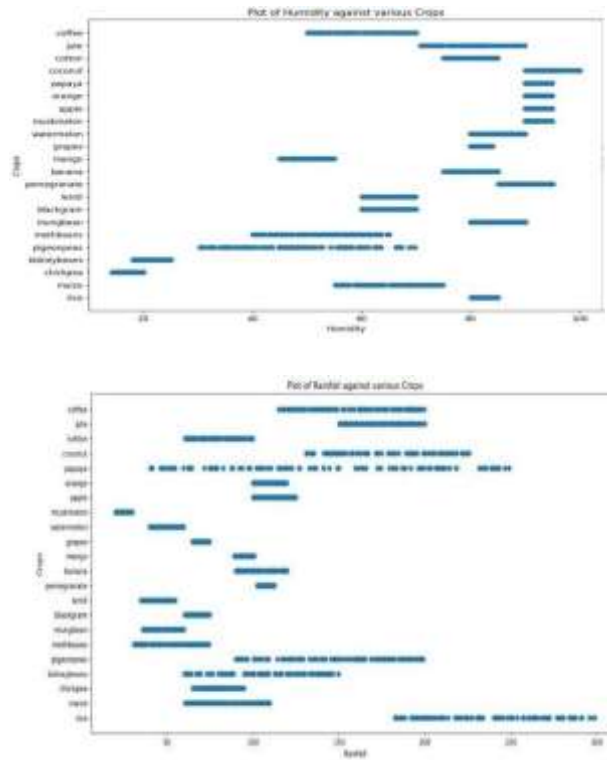


Fig 4. EDA plots of crops against (VS):

Consider one example from the Temperature plot (Plot 4); the temperature span for grapes to grow can be observed to fall in the range of 0°C to 45°C as per the plot, which covers a wide range of temperatures from cold to hot. This means that grapes are an annual plant, which can be grown all year.

**Data Cleaning.** Data cleansing is the process of detecting and correcting corrupt or inaccurate records from a dataset. It involves identifying noisy, incorrect, outliers in the data and then replacing, modifying, or deleting the dirty or coarse data. Data cleansing is also important because it improves data quality and, as a result, overall productivity. When you clean your data, all outdated or incorrect information is removed, leaving you with only the most suitable data. Fig 5 represents the Descriptive analysis and Outlier determination of our dataset. Dataframe.describe() is used to compute the statistical values for any dataset as seen in Fig 5.

	N	P	K	temperature	humidity	ph	rainfall	label
count	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000
mean	50.551918	53.362727	40.940091	25.616244	71.481779	6.494480	103.463025	10.500000
std	36.917334	32.985983	30.647931	5.063749	22.263452	0.777838	54.938389	6.345731
min	0.000000	5.000000	5.000000	8.825675	14.258040	3.504752	20.215267	0.000000
25%	21.000000	28.000000	20.000000	22.768375	60.265953	5.979893	64.551686	5.000000
50%	37.000000	51.000000	32.000000	25.598893	81.473146	6.425045	94.067624	10.500000
75%	84.250000	68.000000	40.000000	28.561654	89.948771	6.923643	124.267508	16.000000
max	140.000000	145.000000	205.000000	41.675493	99.981876	9.935091	280.562117	21.000000

Fig 5: Descriptive Analysis



After cleansing the dataset, it is transformed and integrated carefully as per our requirements.

- Plot 1: Nitrogen [mg/kg]
- Plot 2: Phosphorus [ppm]
- Plot 3: Potassium [ppm]
- Plot 4: Temperature [C]
- Plot 5: pH
- Plot 6: Humidity [g.kg-1]
- Plot 7: Rainfall [mm]

### C. Model Building

Machine learning models are developed by providing them with functionality that combines many required activities for model development and deployment. Understanding the relationships between machine learning algorithms, machine learning models, and training data is critical. A machine learning model is the result of using training data to train a machine learning algorithm. The accuracy of the model produced by a wide range of machine learning algorithms determines which model is finally used.

Initially, our model was subjected to the KNN, Logistic Regression, XGBoost classifiers in order to predict one specific crop for a given soil composition and weather conditions.

**K-Nearest Neighbors.** A supervised learning algorithm is the K-nearest neighbors (KNN) algorithm. Its applications include classification and regression problems. It classifies new data points based on feature similarity. The KNN method computes the distance (Euclidean, Manhattan, or Hamming distance) between the test and query data points and the training set data points. After calculating the distance, this method finds the k-nearest neighbors and assigns the new data point to one of the outcome classes using a voting method among neighbors. The model performance has been analyzed with a different number of neighbors, best performance is achieved with 4 neighbors. Euclidean distance is used to measure the distance in our system and *the* formula to measure it is as follows.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

**Logistic Regression.** A supervised learning algorithm is the logistic regression method. For binary classification problems, the logistic regression method models the probabilities. It is a more advanced version of the linear regression method. The outcome is based on the use of one or more numerical and categorical predictors. Logistic regression uses the logistic function (Sigmoid function). Logistic regression uses odds instead of proportions. The odds are simply

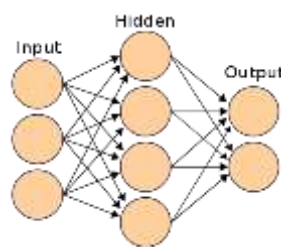
the proportion of the two possible outcomes [16].

**XGBoost.** The gradient boosting framework is used in XGBoost, a decision-tree-based ensemble Machine Learning algorithm. Decision tree-based algorithms are now the best for small-to-medium structured/tabular data. It's a well- balanced mix of software and hardware optimization strategies that produce better outcomes with fewer computer resources and in less time.

These models could easily predict a single crop, but the goal is to predict the top crops for a given set of inputs. As a result, multi-class classification / multi perceptron learning algorithms like ANN come into consideration.

*Why is ANN best suited for our proposed model?*

Neural networks adapt to changing inputs so that the network generates the best possible result without requiring the output criteria to be redesigned. Fig 6. represents the basic structure of ANN.



**Fig 6: Structure Of ANN**

*Use of One-hot Encoding for multiclass classification using ANN. (Fig 7)*

While using the previously mentioned supervised learning algorithms, Label Encoding was sufficient to achieve the highest level of accuracy. For ANN, however, we used a technique known as One hot encoding. For categorical variables with no ordinal relationship, the integer encoding is insufficient. Allowing the model to assume a natural ordering of categories and using this encoding may result in poor performance or unexpected results. In this case, a one-hot encoding can be used to encode the integer representation. The integer encoded variable is removed at this point, and a new binary variable is created for each distinct integer value. It essentially converts categorical label encoded data into binary values, thereby distributing the weightage among the entities equally.

	apple	banana	blueberry	cherry	coconut	orange	peach	pear	plum	strawberry	...	mango	nectarine	pineapple
2155	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2176	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2157	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2158	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2159	0	0	0	0	0	1	0	0	0	0	0	0	0	0

**Fig 7: One Hot Encoded data**

The most critical step after developing the model is to evaluate it for prediction. The final step of the process of model performance evaluation is based on its accuracy. The frequency of true negatives, false negatives, true positives, and false positives is represented by the Confusion Matrix. Accuracy, Recall, Precision, F1-Score, G - Mean, Cohen Kappa are some of the important measures to consider while evaluating a model. These are the following formulas to compute the aforementioned parameters.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$F1 - Score = 2 * \frac{Recall * Precision}{Recall + Precision}$$

$$GMean = \sqrt{(Recall * Precision)}$$

The actual labels of test datasets and projected values are used to construct methods which can also be imported from the scikit learn library's metrics module.

#### IV. RESULTS AND DISCUSSION

##### A. Performance Analysis

To predict one specific crop for a certain soil composition was successful using KNN, Logistic Regression and XGBoost classifiers. The dataset was initially label encoded and results were predicted. Table 1. represents the performance analysis of the various classifiers used. All the statistical parameters are computed as follows.

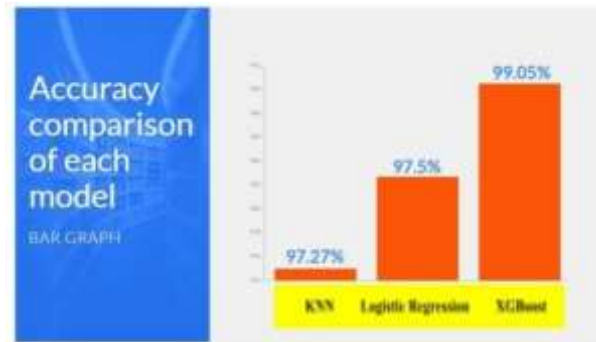
**Table 1. Statistical Performance Analysis and Comparison of classifiers**

<i>ML Classifier</i>	K- Nearest Neighbors	Logistic Regression	XGBoost
<i>Classification Report</i>			
Accuracy	0.9977	0.9975	0.9983
Precision	0.9750	0.97270	0.9820
Recall	0.9750	0.97272	0.98209
F1 - Score	0.9750	0.97270	0.98204
G - Mean	0.9750	0.97270	0.98204
Cohen - Kappa	0.97376	0.96422	0.98121

**Table 2. Comparison of Training Score and Model Accuracy**

Models	Training Score	Model Accuracy
K- Nearest Neighbors	99.77	97.5
Logistic Regression	99.75	97.27
XGBoost	99.83	99.05

As shown in the above Tables [1 and 2], the XGBoost algorithm was found to be more accurate with about 99.05% accuracy than the other 2 classifiers, but only by a small margin. Fig 8. represents the comparison of accuracies of the classifiers



**Fig 8: Accuracy Comparison**

**B. Future Methodologies**

Feature selection techniques are critical for achieving optimal accuracy because they improve the machine learning process and increase the predictive power of machine learning algorithms by selecting the most important variables and eliminating redundant and irrelevant features. We shall be looking into various feature selection techniques such as the Filter, Wrapper and Embedded methods.

We plan to implement many more machine learning algorithms to determine the accuracy and formulate a comparative study for the same with respect to feature selection techniques. For instance, training the machine learning model with algorithms such as Support Vector Machine, Random Forest, Naive Bayes etc.

To predict multiple crops for the same set of data, we will either use a traditional ANN or build our own custom neural network. This method's main goal is to obtain not only the best but also the next best crops to be sown. Farmers will benefit from this because they will have a wider range of crop varieties to grow without wasting extra farmland, pesticides and many other resources resulting in increased profits and additional benefits.

We plan to incorporate this model into our smartphone application in the future, as well as offer the app in multiple languages, which will be beneficial to farmers who are fluent in their native tongue.

## V. CONCLUSION

We discussed the importance of crop yield prediction in this paper, as well as how our model could be used to predict multiple crops for the same set of data. As a result, farmers will be able to better manage their resources, resulting in increased profits. We intend to develop and integrate our proposed application into a Smartphone application so that it can be accessed and used at any time. Also, the importance and significance of each step in our model's workflow is discussed. KNN, Logistic Regression and XGBoost algorithms were the supervised machine learning algorithms we used to predict one crop. The XGB classifier produced a higher accuracy of 99.05 percent. We also realized how critical it was to use multi-perceptron methods like ANN in our research. Finally, we can conclude that multi-class classification is critical for predicting multiple outcomes, which is why our system is still in development and will be released soon as a mobile application.

## REFERENCES

1. Department for Promotion of Industry and Internal Trade, Startup India, Govt of India. <https://www.startupindia.gov.in/content/sih/en/agriculture-challenges-det ail.html>. (accessed Jan 20, 2022)
2. Characteristics of an Intelligent Systems Future, Forbes (2021). <https://www.windriver.com/intelligent-systems>. (accessed Jan 20, 2022)
3. Department of Agriculture and Farmers Welfare, Govt of India. <https://www.soilhealth.dac.gov.in/>. (accessed Jan 20, 2022)
4. Thomas van Klompenburg, Ayalew Kassahun, Cagatay Catal. (2020). Crop yield prediction using machine learning: A systematic literature review. In: Computers and Electronics in Agriculture 177 (2020) 105709.
5. Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. Jabber (2020). Crop Yield Prediction based on Indian Agriculture using Machine Learning. In: 2020 International Conference for Emerging Technology (INCET) Belgaum, India. IEEE..
6. Aditya Shastry, H. A. Sanjay, Madhura G. Hegde (2015). A parameter based ANFIS model for crop yield prediction. In: 2015 IEEE International Advance Computing Conference (IACC), Bangalore.
7. Rakesh Kumar, M.P. Singh, Prabhat Kumar, J.P. Singh (2015). Crop selection method to maximize crop yield rate using machine learning technique. In: 2015 International Conference on Smart Technologies and Management for Computing, Communication,

Controls, Energy and Materials (ICSTM), Chennai.

8. Ayush Shah, Akash Dubey, Vishesh Hemnani, Divye Gala, D. R. Kalbande (2018). Smart Farming System: Crop Yield Prediction Using Regression Techniques. In: Proceedings of International Conference on Wireless Communication, Lecture Notes on Data Engineering and Communications Technologies 19.
9. Talha Siddique, Dipro Barua, Zannatul Ferdous, Amitabha Chakrabarty (2017). Automated Farming Prediction. In: Intelligent Systems Conference 2017, London, UK.
10. Anjana, Aishwarya Kedlaya K, Aysha Sana, B Apoorva Bhat, Sharath Kumar, Nagaraj Bhat (2021). An efficient algorithm for predicting crop using historical data and pattern matching technique. In: Global Transitions Proceedings 2 (2021).
11. Y. Jeevan Nagendra Kumar, V. Spandana, V.S. Vaishnavi, K. Neha,
12. V.G.R.R. Devi (2020). Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector. In: 5th International Conference on Communication and Electronics Systems (ICCES), 2020.
13. Ephrem Yetbarek, Richa Ojha (2020). Spatio-temporal variability of soil moisture in a cropped agricultural plot within the Ganga Basin, India. *Agricultural Water Management* 234(140):106108.
14. N. Usha Rani, G Gowthami (2020). Smart Crop Suggester. *Advances in Computational and Bio- engineering, Learning and Analytics in Intelligent Systems*.
15. Mariammal G, Suruliandi A, Raja SP, Poongothai E (2021). Prediction of Land Suitability for Crop Cultivation Based on Soil and Environmental Characteristics Using Modified Recursive Feature Elimination Technique With Various Classifiers. In: *IEEE Transactions On Computational Social Systems*.
16. Farnaz Babaie Sarijaloo, Michele Porta, Bijan Taslimi, Panos M. Pardalos (2021). Yield performance estimation of corn hybrids using machine learning algorithms, *Artificial Intelligence in Agriculture*. <https://doi.org/10.1016/j.aiia.2021.05.001>.
17. Logistic regression, Chapter 14.  
[https://nhorton.people.amherst.edu/ips9/IPS\\_09\\_Ch14.pdf](https://nhorton.people.amherst.edu/ips9/IPS_09_Ch14.pdf). (accessed Jan 20, 2022)
18. Dr. S. Jafar Ali Ibrahim, Y. Prathyusha, G. Lakshmi, M. Rajini, U. Naveen Sai, "Detection and Classification of Fruit Disease using Image Processing", *Science, Technology and Development*, ISSN: 0950-0707, Volume XI Issue II FEBRUARY 2022: 24-31. DOI:21.18001.STD.2022.V11I02.22.352703, 4-feb2022.pdf (journalstd.com)
19. S Jafar Ali Ibrahim, G. Tejaswini, T. Niharika, K. Murari, M. Bhanu Prakash, "IOT Enabled Weed Controller for Typical Agriculture", *The International Journal of Analytical and Experimental Modal Analytical and Experimental Modal Analysis*, ISSN NO:0886-9367, Volume XIV, Issue III, March/2022: 1-8