Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 12, 2025

OPTIMIZING ADABOOST FOR CROP YIELD FORECASTING: A PCA-BASED DIMENSIONALITY REDUCTION APPROACH

Jawad Hussain Shah[1], Shahnila Bibi[2], Malak Roman[*3], Awrang Zaib[4], Jamshad Ahmad Mir[5] Muhammad Azhar Uddin[6]

[1,2,5,6] BS-Computer Science, Department of Computer Science, University of Chitral, KP-Pakistan.
[3,4] Lecturer, Department of Computer Science, University of Chitral, KP-Pakistan.

[*3]**malak_5116@uoch.edu.pk**

DOI: https://doi.org/10.5281/zenodo.17936672

Keywords
Crop Yield Forecasting, AdaBoost, PCA, Machine Learning, Data Mining, Agriculture

Article History
Received: 11 October 2025
Accepted: 21 November 2025
Published: 15 December 2025

Corresponding Author: *
Malak Roman

Abstract
Agriculture is the essential human function that makes sure human existence helps various nations' economies, confrontation troubles because of the weather conditions are increasingly unpredictable moreover the soil experiences composite procedures. Through precision farming and predictive analytics, machine learning in agriculture improves crop yields and resource efficiency. Unfailingly predicting upcoming crop yield is essential for cleanse agronomic creation and guaranteeing nutrients protection. Our study deliberates an innovative MLM (machine learning model) employing PCA (Principal Component Analysis) with AdaBoost classifier. The dataset, which focuses on characteristics like soil nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall, comes from an open-source repository (Kaggle) and includes 2200 entities of 22 different crops. The MLM is employed to examine the diverse dimensional statistics and ensure precise crop choice sanctions. The PCA is used in preprocessing for the removal of noise data and feature extraction, AdaBoost algorithm was used for classification. The results of the experiment demonstrate a high accuracy of 99.2% and consistent precision, recall, and F1-scores of 0.992. All of these results suggest that the MLM (Machine Learning Model) can distinguish between different crop varieties with a low error rate. Based on these results, the researchers can conclude that Principal Component Analysis (PCA) and the ensemble machine learning algorithm AdaBoost can improve the forecasting capabilities of DM (data mining) processes in agriculture and provide information to agronomists and former for sustainable crop production.

INTRODUCTION

Computer science has revolutionized today's world by transforming industries like healthcare, finance, education, and communication, enabling rapid advancements in AI, data analysis, and connectivity [1]. Finding patterns, trends, and helpful information in vast amounts of data is known as data mining (DM). To extract information that can be helpful in making decisions, data mining uses a variety of techniques, including databases, statistical methods, and machine learning algorithms [2]. Businesses and researchers can make data-driven decisions thanks to data mining, which is essential in several industries, including marketing, healthcare, finance, and agriculture. By extracting useful insights from massive volumes of data, data mining has improved the agricultural industry [3, 4]. Predictive analytics, weather pattern analysis, crop yield prediction, disease and pest outbreak prediction, and agriculture all heavily rely on data mining. Large volumes of satellite imagery data for agricultural purposes can be analyzed by machine learning algorithms, which can also be used to extract pertinent features, classify different types of land cover, including drought-stressed areas, and forecast demand. In order to forecast future demand for agricultural products and assist farmers and suppliers in planning production and distribution, DM also examines past sales data and market trends [5]. Climate Modeling Data Mining helps analyze the data related to climate to identify trends and forecast future climate conditions and produce a way for the farmers to adapt their practices accordingly [6]. Machine learning is a crucial technique for crop yield forecasting. Machine learning has also been applied in agriculture for the past years, which helps agronomists to decide which crops to grow at which season based on a variety of datasets. Crop yield relies on various factors like climate, weather, soil, consumption of fertilizers, and types of seeds. For crop yielding different kinds of Machine learning algorithms have been applied [7].

Various Formal structures are utilized for machine learning and signal/image processing applications, including classification and clustering, regression and function approximation, image coding, recovery and enhancement, source separation, data aggregation, feature selection, and extraction [8,9]. In recent times, machine learning techniques have been applied in diverse manners to process data from both multispectral and hyperspectral remote sensing, data gathering Application of data mining techniques in gathering and aggregating data from disparate sources such as soil surveys, field experiments and remote sensing. Data cleaning: Collected data is cleaned and pre-processed to eliminate errors, inconsistencies and missing values) Feature Extraction: Extract the relevant features or variables from raw data, such as pH value in the soil, organic matter content of the soil, nutrient levels in the soil [10]. This paper contributes to an overview of the potential of machine learning in supporting sustainable agriculture and efficient land management. Agriculture is described as the process of cultivating crops and rearing livestock for food, fiber, and other products. It hence involves farming, ranching, that is raising livestock and such other activities.

Agriculture provides food and employment for people; therefore, it is an essential part of all economic systems. It provides the fundamental necessities of life for humans and is a significant source of employment for many people in developing countries such as India and Pakistan. There are three main categories of agricultural activity: pre-harvest (those that occur before a crop or livestock product is harvested), harvesting (gathering of mature crops) and post-harvest (those that occur after a crop has been harvested). This research identifies a need for both available and affordable means of collecting data to allow the application of the model to different geographical locations. Although the model was very successful, it had some constraints including specific regional data, which identified opportunities for improvement. Improving this model's applicability to other agricultural conditions will require expansion to additional climatic and geographical data sets [11].

LITERATURE REVIEW

Veeragandham, S. & Santhi, H. (2020) [12] assessed how machine learning can be applied to the agricultural sector, based on an extensive dataset which included multiple features, such as; soil classification, disease detection, plant species management, irrigation management, crop yield forecasting, crop quality assessment, and weed detection. PCA, Gaussian Median and Gaussian filters were used for removing noise and enhancing images. Data was acquired from multiple sources, i.e., UCI, MIT, Kaggle, and USDA-ARS. Different Machine Learning (ML) methods were used for disease detection in crops. The most accurate method for the three crop types was CNN (98.62%), followed by SV (94.95%), ANN (93.70%), KNN (88.75%), and FNN (88.00%).

Storm, H. et al., (2020) [13] Provides an overview of machine learning (ML) methods from an applied economics perspective, highlighting their potential in fields like farm management, economic analysis, policy simulation, processing unstructured data, and causal inference in agricultural and applied economics. Preprocessing algorithms used in this paper are Principal Component Analysis (PCA) for dimensionality reduction. CNNs and RNNs, models were implemented with PCA for optimal feature selection.

Prity, F. S. et al., (2024) [14] employed nine machine learning algorithms to enhance agricultural productivity and crop recommendations based on historical data on climate conditions, crop yield, soil properties, and farmer preferences. In total, nine machine learning algorithms were applied: Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Bagging (BG), AdaBoost (AB), Gradient Boosting (GB), and Extra Trees (ET), on a dataset characterized by the following 5 attributes: temperature, rainfall, humidity, soil pH, and nutrient levels. The highest score, 99.31%, was recorded by RF, while AB had the lowest accuracy at 14.1%.

Veeragandham, S., & Santhi, H (2020) [12] the author used 9 machine learning algorithms Among these algorithms convolutional Neural network (CNN) achieved the highest accuracy (98.62%) for disease detections in crops like wheat and rice. The paper highlighted 5 primary attributes (soil properties, Environmental conditions, crop features, Water management, and yield and quality metrics). Preprocessing involves Noise removal and dimensionality reductions

Issac, A. et al., (2022) [15] Presented their work on Dimensionality Reduction of High-throughput Phenotyping in the Fields of Cottons, obtaining data with a stereo camera, which was loaded on an automatic ground vehicle. The principal components analysis (PCA) algorithm for preprocessors scored 98% of accuracy and the OpenCV Library with full frame for the detection of cotton blooms image in the fields of cotton.

Singh, V., Et-al. have used Elemental profiling by means of EDXRF to aid their study of Agricultural Soil and to assist in the interpretation of results from data derived from the EDXRF analysis. The study is based on Soil Data obtained through Agricultural Studies conducted at the GB Pant University of Agriculture and Technology Pantnagar. Using Principal Component Analysis (PCA) on data as pre-processing prior to running through a PCA Algorithm yields a result with an accuracy of 80%. The X-3600 EDXRF Spectrometer with Liquid Nitrogen Cooled Si (Li) detector was used in the measurements of the samples.

The authors,Cruz G. B. D. et al., acquired three distinct datasets regarding agriculture from UCI datasets for analysis purposes; Furthermore, these datasets were comprised of three separate types of crops that each had varying levels of optimization prior to utilizing either PCA-GA optimizations or machine learning algorithms such as the K-NN algorithm (using J4.8 model), MLP model, and naïve bayes model. Numerous tests were conducted to measure the effectiveness of each method, and the results showed that using K-NN as well as performing PCA-GA optimization gave nearly equal accuracy (99.98%) when compared to other models.

Luo, K. (2023) [18] the dataset used in this study was obtained from Kaggle. This dataset contains 151 samples with 54,676 genes and six diagnostic classifications related to breast cancer. PCA was used as a Preprocessing for feature extraction, and the primary machine learning algorithm employed was Random Forest. The result shows the non-effectiveness of PCA in the diagnostic classification of breast cancer. After PCA, the model showed the lowest diagnostic accuracy of 0.7826.

Shastry, K. A., & Sanjay, H. A. (2021) [19] They collected dataset from numerous sources like India Stat, India Water Portal, Karnataka Atlas of Soil. For preprocessing, the author used mean imputation to fill in the missing values and normalization to scale features between 0 and 1 Genetic Algorithm (m-GA) for feature selection and Weighted Principal Component Analysis (w-PCA) were used for the purpose of feature extraction. Multiple Machine learning models are used (Regression, Artificial Neural Networks (ANN), Adaptive Neuro Fuzzy Inference System (ANFIS), Ensemble of Trees (ET), and Support Vector Regression (SVR). The Study was performed on eight-real-world Indian crop dataset. The modified version of GA shows 28% accuracy.

Chatterjee, S. et al., (2024) [20] The dataset is obtained from a private online repository, Kaggle, and EGA is to select the required features. Support vector machine (SVM), lasso, Decision Tree (DT), Random Forest (RF), and Gradient Boosting (GB), among these ML models, the forecasting accuracy of the GB with the EGA model is 93.009%.

RESEARCH METHODOLOGY:

In this research work an intelligent integrated model was developed for crop yield forecasting in agriculture. After data preprocessing, two datasets were created: one for training 70% of the datasets had been used to train the model and the other for model testing 30% of the dataset were used to test the model [21]. Principle Component Analysis (PCA) and AdaBoost classification model were applied individually using a python environment as shown in figure 1. The test data was used to calculate the accuracy of the proposed integrated model gathered from the training of the model first. The dataset is a collection of 10 key attributes and 2200 instances.

The model was executed with the Visual Code IDE. To implement the model, we use python libraries such as pandas, scikit-learn, matplotlib and seaborn.

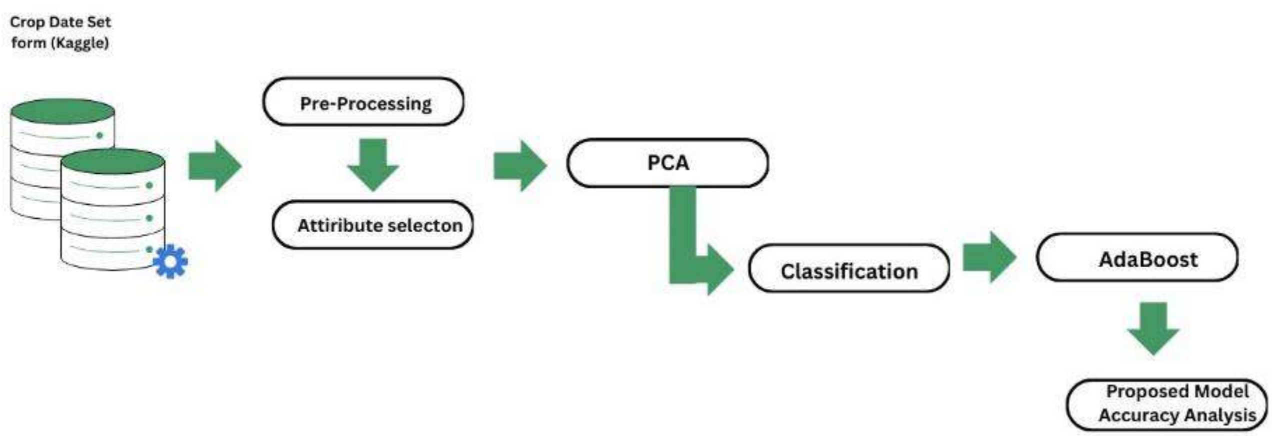

Fig. 1: Proposed Research Methodology Flow Chart

Data Set Collection

The dataset is obtained from Kaggle database that contains different crop parameters based on which users can implement predictive machine learning algorithms to recognize the most compatible crop to be grown. Rice, Maize, Jute, Cotton, Coconut, Papaya, Orange, Apple, Muskmelon, Watermelon, Grapes, Mango, Banana, Pomegranate, Lentil, Black gram, Mung bean, Moth beans, Pigeon peas, Kidney beans, Chickpea, and Coffee are among the 22 distinct crops represented in Table 1. This dataset includes meteorological characteristics such as rainfall, temperature, humidity, fertilizer, weather, and region, as well as the chemical composition of soil such as nitrogen, phosphorus, potassium, fertilizers, and soil pH. [14]. The rainfall, climate, and fertilizer data datasets that are available for India were amplified to create this dataset. [22]

Table 1: Kaggle Dataset Attributes Description.

S. No	Attribute Name	Attribute Description	Range
1	N	Total amount of Nitrogen in the soil	(0–139) kg/ha
2	P	Phosphorus amount contained in the soil	(5–145) kg/ha
3	K	The Potassium content ratio in the soil	(5–205) kg/ha
4	temperature	temperature in degrees Celsius	(10.78–43.36) K
5	Humidity	relative humidity in %	(14.69–98.80) F
6	pH	Potential of Hydrogen value of the soil	(3.55–7.45)
7	Rainfall	Rainfall in mm	(20.21–291.29) mm

8	Fertilizer	The use of fertilizers	True/False
9	Weather	It defines the weather conditions	Sunny, rainy, cloudy
10	Region	Determines the different regions	North, East, South, West

N: The total amount of nitrogen in the soil. Due to the scarcity of N from organic fertilizers, inorganic nitrogen (N) fertilizer (urea, etc.) makes up the majority of the N cycle in rice-wheat (R-W) [23].
P: The soil's total phosphorus content. After nitrogen, it is the second essential nutrient for crops. Numerous biochemical reactions and processes, including photosynthesis, respiration, energy storage, transfer, cell division, cell enlargement, and nitrogen fixation, are aided by it [24].
K: The soil's potassium content ratio. Increasing root growth, improving nutrient intake, cellulose formation, reducing crop lodging, and boosting crops' immunity against diseases are some of the crucial functions of K in crops [25].

Principal Component Analysis (PCA) Preprocessing Algorithm:

Feature selection is a significant process of data preprocessing, showing its efficiency in preparing the data-high dimensional dataset especially for several tasks involving data mining and machine learning [26]. Its basic objectives are the development of simpler and more interpretable models, enhancement in the performance of data mining, and providing spotless and well-structured data for analysis [27]. The principal component analysis is mainly aimed at reducing the dimensionality level of a dataset with several interrelated variables, retaining as much of the difference existing in the dataset. This is achieved by transformation to a new variable set-the PCs, which are uncorrelated and which are organized in such a way that the first few retain most of the variation present in all the original variables. [28]. Based on the preprocessed extracted features through Principal Component Analysis-PCA seven attributes were selected to train the classifier model.

$$PC_p = a_{1p}X_1 + a_{2p}X_2 + \ldots + a_{np}X_n \quad \text{equ. 1}$$

One technique used in statistics and biometrics is Principal Component Analysis (PCA). Equation 1 illustrates how it uses a unique transformation to transform a set of related data points into a new set of unrelated data points. By reducing the dimensions of complex data while preserving the majority of its crucial information, PCA aids in its simplification. It includes ideas from linear algebra and statistics, such as eigenvectors, covariance, and standard deviation. [29]

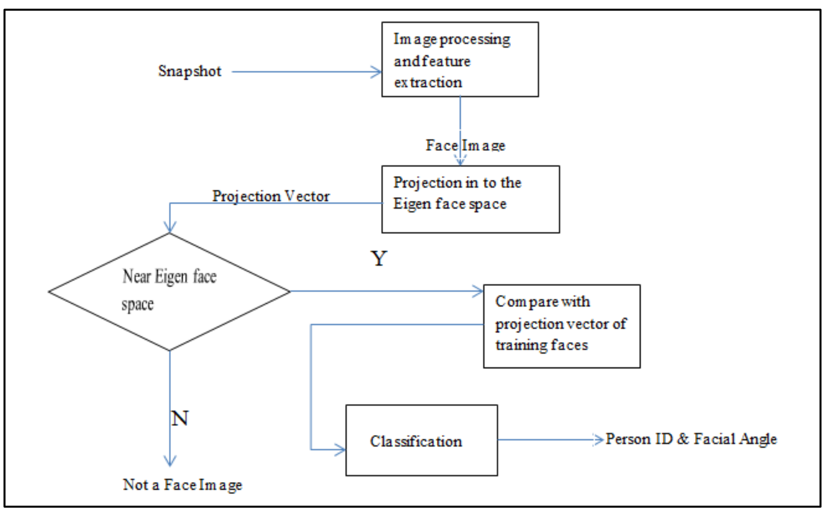

Fig. 2: Principal Component Analysis (PCA) Workflow Diagram

AdaBoost Classifier Implementation:
Adaptive Boosting, which is, in short, AdaBoost, is employed for both classification and regression of machine learning processes [30]. It requires labeled data. A variety of simple models, referred to as weak learners, are combined to form one strong and more accurate model. The contribution of the weight of those training data points increases that were relatively difficult to classify correctly in previous steps. This AdaBoost algorithm was introduced by Freund and Schapire in the year 1997 as a boosting ensemble modelling approach [31]. It is now widely used for binary classification problems. Boosting algorithms first build a model based on training data sets and then, based on that, they build a second model that fixes the issues with the first model. This process is repeated until the data set is predicted, and the mistakes are minimized.

AdaBoost Algorithm workflow
The final predicted crop yield f(x) is a weighted sum of the predictions from all the weak learners (e.g., decision trees) [32].

$$f(x) = \sum_{k=1}^{K} \alpha_k G_k(x)$$

equ. 2

Input: Initially set of n-examples are considered as the Training Set.
IPn = {(x1, y1), (x2, y2) ,... (xn, yn)}
DecisionTreeRegressor denoted as Weak Learner
Integer ITR specifying total number of iterations.
Initialize:
= 1/n, i.e. W1 = {1/n, 1/n, ... ,1/n}
The ensemble P = ∅
For itr = 1, 2, ..., ITR
Step1. Get an example Ritr from IPn using distribution Witr
Step2. Pitr is built using the training set Ritr
Step3. Calculate Eitr = and = 0.5 ln ( )
Step4. Update the weight: = normalize (. exp (- .))
Output: The ensemble P= { p1, p2, ..., pITR } and A= { α1, α2, ... , αITR }[33]

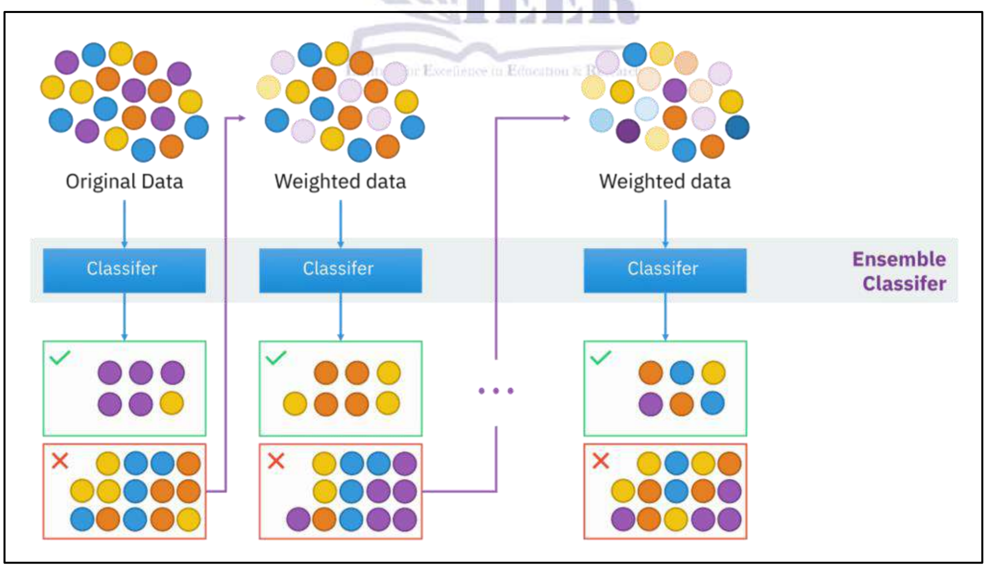

Fig. 3: Graphical Overview of Ada Boost Algorithm Workflow

RESULTS AND DISCUSSION
In portion of the study, we will elaborate on how our machine learning model performed. It will include accuracy technique along with, precision, recall, F1 score, and confusion matrix. These results will make it easier for us to evaluate the efficiency and effectiveness of this machine learning algorithm in the agricultural dataset. AdaBoost is a supervised machine learning algorithm, is used in this model

and implemented on VS Code IDE as shown in figure 4.

```
# Imports
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.ensemble import AdaBoostClassifier
from sklearn.decomposition import PCA
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns

  # AdaBoost on PCA-transformed features
  model = AdaBoostClassifier(n_estimators=50, learning_rate=1.0, random_state=42)
  model.fit(X_train, y_train)
  print("Model Trained Successfully!")
```

Fig.4: Model Implemented in VS Code IDE.

We draw our conclusions rooted on the accuracy, precision, and F1 score, which are discussed in table 2 and figure 5. These metrics and standards show us how our model performed.

Table 2: Comparative performance analysis of our proposed ML model

Model	Accuracy	Precision	Recall	F-Measure
AdaBoost with PCA	99.2%	0.992	0.992	0.992

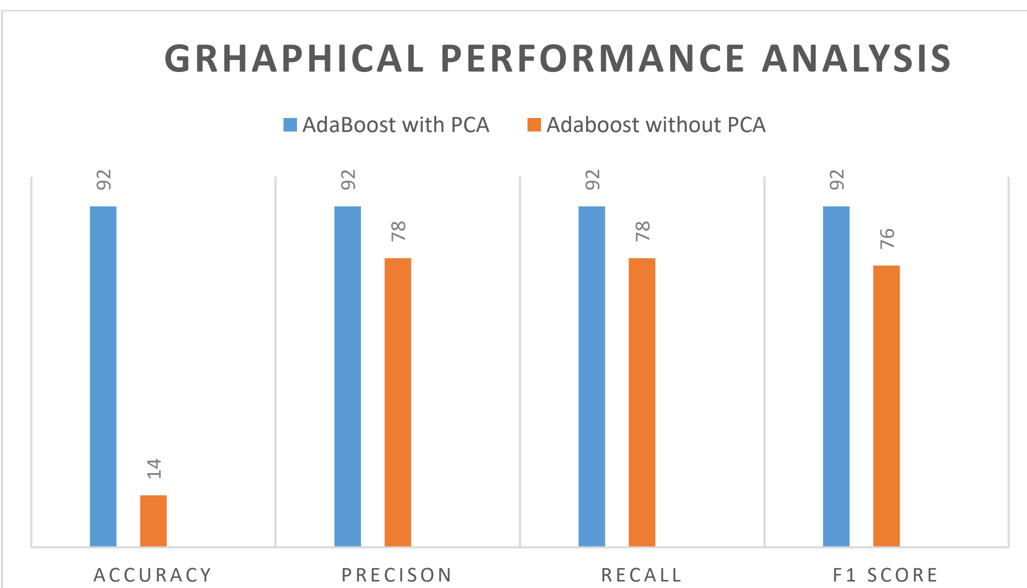

Fig.5: Graphical comparative analysis of our proposed model with and without PCA

Author	Accuracy	Precision	Recall	F-Measure
Farida Siddiqi Prity	14.1%	0.78	0.78	0.76
Our Proposed model (AdaBoost with PCA)	99.2%	0.992	0.992	o.992

Table 3: comparative performance analysis of our proposed ML model with existing studies

The comparative performance analysis shown in table 3 indicates that the proposed machine learning model, which uses AdaBoost with PCA, outperforms improvement in classification performance when AdaBoost is combined with PCA for dimensionality reduction.

The overall performance score of our model is evaluated based on the (Accuracy, Precision, Recall, and F1 Score), and the values of these attributes are calculated by the following formulas [35].

$$Accuracy = \frac{No\ of\ correct\ predictions}{Total\ no\ of\ predictions} \quad (equ. 3)$$

$$Precision = \frac{True\ Positive}{True\ positive + False\ Positive} \quad (equ. 4)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (equ. 5)$$

$$F1_{Score} = 2 \times \frac{Precision\ \times Recall}{Precision + Recall} \quad (equ. 6)$$

The results demonstrate that the AdaBoostML classifier with PCA as the preprocessing algorithm achieved remarkably high performance on the crop dataset. When investigating the detailed accuracy by class, maximum crops, such as maize, chickpea, kidney beans, mango, and apple, have been classified by following the technique of precision, recall, and F1-scores all equal to 1.0. Despite these slight errors, the weighted average scores remain nearly seamless, with an F1-score of 0.992 and an ROC area of 1.0 [36]. This specifies that the machine learning model is highly effective at differentiating between most crop classes and delivers exceptional overall classification performance, with only very small overlaps in forecast for a few crop types.

CONCLUSION

This study demonstrates how AdaBoost remarkably boosts yield prediction accuracy by successfully classifying significant crop-related features through 22 classes, as verified by a 99.2% accuracy ratio and the model by Prity, F. S., [34] across all evaluation metrics. Specifically, the proposed model achieves an accuracy of 99.2%. This demonstrates a substantial nearly impeccable firmness in the confusion matrix. The ML model's preprocessing and effective performance, with only 5 misclassifications out of 662 instances, highlights its reliability for practical agricultural applications. The conclusions emphasize the potential of AdaBoost in providing scalable solutions for trade like agriculture. Future work might focus on integrating real-time sensor data for dynamic updates, applying progressive preprocessing methods to mitigate slight misclassifications (e.g., rice and cotton), and purifying the ML model to handle exceptional situations.

REFERENCE

Rahman, M. A. U., Roman, M., Ahmad, S., Jan, M. S., & Nawab, H. U. (2021). Optimizing Collaboration: Insights into Inter-Team Coordination and Self-Management in Distributed Agile Software Development. *Webology (ISSN: 1735-188X)*, *18*(5).

Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: application of big data. *Journal of Big data*, *4*(1), 20.

Milovic, B., & Radojevic, V. (2015). Application of data mining in agriculture.

Roman, M., Nawab, H. U., Ahmad, S., & Khan, I. A. (2022). K-Nearest Neighbor and Fuzzy K-Nearest Neighbor Algorithm Performance Analysis for Heart Disease Classification. *Webology(ISSN: 1735-188X)*, *19*(1).

Roman, M., Naz, I., Luqman, M. A., Ali, J., Jan, M. S., & Nawab, H. U. (2024). Stroke Disease Prediction Using K-Nearest Neighbor and Decision Tree Algorithms with Machine Learning Pre-Processing Techniques. *Migration Letters*, *21*(S4), 2015-2027.

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*(11), 3758.

Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and electronics in agriculture*, *177*, 105709.

Do, D., & Tam, A. (2004). Formal semantic models for images and image understanding. In *Proceedings of the Fourth International Conference on Computational Semiotics for Games and New Media*.

Hussain, F., Shams, S. S., Roman, M., Anwar, M., Shah, F. A., Meraj, A. B. E., & Uddin, M. A. (2025). MACHINE LEARNING IN HEALTHCARE: PREDICTING CHRONIC KIDNEY DISEASE THROUGH FEATURE-DRIVEN HEURISTIC MODELS. *Frontier in Medical and Health Research*, *3*(7), 318-328.

Diaz-Gonzalez, F. A., Vuelvas, J., Correa, C. A., Vallejo, V. E., & Patino, D. (2022). Machine learning and remote sensing techniques applied to estimate soil indicators–review. *Ecological Indicators*, *135*, 108517.

Meshram, V., Patil, K., Meshram, V., Hanchate, D., & Ramkteke, S. D. (2021). Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences*, *1*, 100010.

Veeragandham, S., & Santhi, H. (2020). A review on the role of machine learning in agriculture. *Scalable Computing: Practice and Experience*, *21*(4), 583-589.

Storm, H., Baylis, K., & Heckelei, T. (2020). Machine learning in agricultural and applied economics. *European Review of Agricultural Economics*, *47*(3), 849-892.

Prity, F. S., Hasan, M. M., Saif, S. H., Hossain, M. M., Bhuiyan, S. H., Islam, M. A., & Lavlu, M. T. H. (2024). Enhancing agricultural productivity: a machine learning approach to crop recommendations. *Human-Centric Intelligent Systems*, *4*(4), 497-510.

Issac, A., Yadav, H., Rains, G., & Velni, J. M. (2022). Dimensionality reduction of high-throughput phenotyping data in cotton fields. *IFAC-PapersOnLine*, *55*(32), 153-158.

Singh, V., Agrawal, H. M., Joshi, G. C., Sudershan, M., & Sinha, A. K. (2011). Elemental profile of agricultural soil by the EDXRF technique and use of the Principal Component Analysis (PCA) method to interpret the complex data. *Applied Radiation and Isotopes*, *69*(7), 969-974.

Cruz, G. B. D., Gerardo, B. D., & Tanguilig III, B. T. (2014). Agricultural crops classification models based on PCA-GA implementation in data mining. *International Journal of Modeling and Optimization*, *4*(5), 375.

Luo, K. (2023, November). Application of Principal Component Analysis in the Diagnostic Classification of Breast Cancer. In *2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)* (pp. 687-692). Atlantis Press.

Shastry, K. A., & Sanjay, H. A. (2021). A modified genetic algorithm and weighted principal component analysis based feature selection and extraction strategy in agriculture. *Knowledge-Based Systems*, *232*, 107460.

Chatterjee, S., Mondal, S. K., Datta, A., & Gupta, H. K. (2024). Enhancing Feature Optimization for Crop Yield Prediction Models. *Current Agriculture Research Journal*, *12*(2).

Shah, F. A., Meraj, A. B. E., Roman, M., Anwar, M., Zaib, A., Shams, S. S., & Hussain, F. (2025). Enhancing Weather Forecasting Accuracy: A Machine Learning Approach Using Genetic Algorithm and Random Forest. *Global Research Journal of Natural Science and Technology*.

https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

Amanullah, Ullah, H., Soliman Elshikh, M., Alwahibi, M. S., Alkahtani, J., Muhammad, A., & Imran. (2020). Nitrogen contents in soil, grains, and straw of hybrid rice differ when applied with different organic nitrogen sources. *Agriculture*, *10*(9), 386.

Mwende Muindi, E. (2019). Understanding soil phosphorus. *International Journal of Plant & Soil Science*, *31*(2), 1-18.

Murphy, P. M. (2020). *Evaluating the role of scale in the sustainability of nutrient management of pasture based dairy systems* (Doctoral dissertation, University College Dublin. School of Agriculture and Food Science).

Roman, M., Ullah, A., Ullah, M. A., Hussain, F., Shams, S. S., Bint-e-Meraj, A., ... & Ali, S. (2025). Predicting Academic Success: A Machine Learning Approach Using Decision Tables and Random Forests Algorithms. *Spectrum of Engineering Sciences*, *3*(5), 205-213.

Ullah, M. A., Ullah, A., Roman, M., Anwar, M., Siddiqi, M. U. M., Jaffar, M. H., ... & Ali, J. (2025). A Comparative Study of Machine Learning Algorithms for Cardiovascular Risk Prediction: Support Vector Machine, Gradient Boosting, And Rotation Forest. *Spectrum of Engineering Sciences*, 481-491.

Mishra, S., Singh, B. R., Naqvi, A. H., & Singh, H. B. (2017). Potential of biosynthesized silver nanoparticles using Stenotrophomonas sp. BHU-S7 (MTCC 5978) for management of soil-borne and foliar phytopathogens. *Scientific reports*, *7*(1), 45154.

Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M., & Hooman, A. (2013). An overview of principal component analysis. *Journal of signal and information processing*, *4*(3), 173-175.

Taherkhani, A., Cosma, G., & McGinnity, T. M. (2020). AdaBoost-CNN: An adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning. *Neurocomputing*, *404*, 351-366.

Schapire, R. E. (2013). Explaining adaboost. In *Empirical inference: festschrift in honor of vladimir N. Vapnik* (pp. 37-52). Berlin, Heidelberg: Springer Berlin Heidelberg.

Hatwell, J., Gaber, M. M., & Atif Azad, R. M. (2020). Ada-WHIPS: explaining AdaBoost classification with applications in health sciences. *BMC Medical Informatics and Decision Making*, *20*(1), 250.

Koduri, S. B., Gunisetti, L., Ramesh, C. R., Mutyalu, K. V., & Ganesh, D. (2019, May). Prediction of crop production using adaboost regression method. In *Journal of physics: conference series* (Vol. 1228, No. 1, p. 012005). IOP Publishing.

Prity, F. S., Hasan, M. M., Saif, S. H., Hossain, M. M., Bhuiyan, S. H., Islam, M. A., & Lavlu, M. T. H. (2024). Enhancing agricultural productivity: a machine learning approach to crop recommendations. *Human-Centric Intelligent Systems*, *4*(4), 497-510.

Anwar, M., Rahman, T., & Roman, M. (2025). Voice-Activated Smart Environments: Deep Learning Approach for Pashto Speech Command Processing. *Spectrum of Engineering Sciences*, *3*(5), 541-550.

Rehman, T. U., Iqbal, J., Roman, M., Ahmad, Z., Ahmad, B., & Rafiq, M. (2025). A COMPARATIVE MACHINE LEARNING APPROACH TO DIABETES PREDICTION: INTEGRATING BEST FIRST SEARCH FEATURE SELECTION WITH SVM AND NAÏVE BAYES CLASSIFIERS. *Spectrum of Engineering Sciences*, *3*(12), 37-49.