Journal of Experimental Biology and Agricultural Sciences, April - 2023; Volume – 11(2) page 306 – 315

Journal of Experimental Biology and Agricultural Sciences

http://www.jebas.org

ISSN No. 2320 – 8694

# An Integrative Approach Towards Recommending Farming Solutions for Sustainable Agriculture

Veena Ghuriani[*#], Jyotsna Talreja Wassan[*], Pragya Deolal, Vidushi Sharma, Dimpy Dalal, Aditi Goyal

Department of Computer Science, Maitreyi College, University of Delhi, Delhi, India

Received – December 10, 2022; Revision – March 23, 2023; Accepted – April 10, 2023
Available Online – April 30, 2023

DOI: http://dx.doi.org/10.18006/2023.11(2).306.315

KEYWORDS

Sustainable agriculture

Precision Farming (PF)

Zero Budget Natural Farming (ZBNF)

Machine Learning (ML)

Classification

**ABSTRACT**

Sustainable Agriculture is rapidly emerging as an important discipline to meet societal needs for food and other resources by adopting paradigms of conserving natural resources while maximizing productivity benefits. This paper proposes an integrative methodological approach for critically analyzing Precision Farming (PF) paradigms and Zero Budget Natural Farming (ZBNF), providing sustainable farming solutions and achieving productivity and profitability. This paper analyses the productivity of crops in PF using various machine learning (ML) algorithms based on different soil and climatic factors to identify sustainable agricultural practices for maximizing crop production and generating recommendations for the farmers. When implemented on the collected dataset from various Indian states, the Random Forest (RF) model produced the best results with an AUC-ROC of 95.7%. The Juxtaposition of ZBNF and non-ZBNF is evinced. ZBNF is statistically (*$p<0.05$*) observed to be a cost-efficient and more profitable alternative. The impact of ZBNF on soil microbial diversity and micro-nutrients is also discussed.

* Corresponding author

E-mail: vghuriani@maitreyi.du.ac.in (Veena Ghuriani);
jtwassan@mitreyi.du.ac.in (Jyotsna Talreja Wassan)

Peer review under responsibility of Journal of Experimental Biology and Agricultural Sciences.

## 1 Introduction

Agriculture is a magnanimous source of livelihood for a majority of people in the world. It promotes economically viable, socially supportive, ecologically sound farming practices that help the environment replenish. The need to accommodate production without keeping sustainability at stake has paved the way to adapt and integrate new technologies into current practices. The extent and rate of change in information technology have enabled better decision-making, and shifting towards technology-driven agriculture has improved economic efficiency (National Research Council, U.S. 1997). Using sophisticated technologies such as robots, and temperature sensors, allows agriculture to be more profitable and efficient. This can be achieved by applying statistical methods and machine learning (ML) algorithms to efficiently plan experiments and interpret experimental data. Precision Farming (PF) seems to be a vision for a sustainable future in this context. It amalgamates information technology and agronomic sciences (Beluhova-Uzunova and Dunchev 2019). PF centres on data collection and analysis of measurable features to improve crop yield and quality. It reduces production costs and wastage to ensure profitability, efficiency, and sustainability (D'Antoni et al. 2012). Computer applications and technology can be used to optimize field-level management, from creating farm plans to yield maps.

The need to preserve our environment escorts us to zero-budget natural farming (ZBNF). 'Zero Budget' implies no need for credit (Bishnoi and Bhati 2017). It is thought to dramatically reduce production costs by substituting commercial fertilizers and pesticides with home-grown products such as Jeevamritha, Beejamritha, Neemastra, and others, as well as using intercropping and mulching. Currently, it is being adopted in different forms by the farmers in most of the states in India, namely, Andhra Pradesh, Himachal Pradesh, Haryana, Karnataka, Kerala, Madhya Pradesh, and Telangana (Bishnoi and Bhati 2017).

The current research proposes an integrative analysis of the profitability and usability of ZBNF coupled with the PF dealing with the soil productivity of different districts in India. A framework to generate recommendations to farmers to implement both farming methods, having known the prolificity of the soil, is proposed.

Until 1960 India was going through a massive crisis as it didn't have enough food grains to feed its growing population. But the advancement of the green revolution in 1960 changed the Indian agriculture trajectory forever. The arrival of HYV seeds, chemical fertilizers, pesticides, and tractors produced abundant food grains, making India a prominent exporter. The major catalyzing factor was chemical fertilizers. The Green Revolution may have saved the day but didn't guarantee the future because degraded land, eutrophication, expensive farm inputs, and farmers in vicious debt cycles followed. Studies found that chemical fertilizers contain heavy metals (e.g. cadmium and chromium) and high concentrations of radionuclides which lead to the accumulation of inorganic pollutants in the plants (Savcı 2012). Nitrate compounds in these fertilizers highly contaminate the surface and groundwater and cause various health issues like 'blue baby syndrome', which is fatal to infants, causes diabetes and is a precursor of carcinogens (Kostraba et al. 1992). The permissible limit of nitrate ions is 50 mg/l (Rahman et al. 2021). A study showed that from all the samples taken, nitrate concentration was between 1 and 415 mg/l, and 37% of the samples exceeded the safe limit (Jayarajan and Kuriachan 2021). They are toxic to farmers and lead to casualties. Recently, an expert committee set up by the agriculture ministry found that 66 insecticides/pesticides banned abroad are still used in India, and 27 are perilous to humans and animals (NABARD 2018). These chemical inputs are expensive and put a financial burden, especially on marginal farmers. This calls for natural alternatives. ZBNF seems perfectly tailored for such a system (Duddigan et al. 2023). It uses natural inputs and discourages deep plowing and extensive irrigation. As a result, pollution is kept to a minimum, soil fertility is restored, and the environment is preserved (FAO 2016). Four major aspects are integral to ZBNF- Bijamrita, Jeevamrutham, Mulching, and Waaphasa (Kumar et al. 2020). In 2016, the Government of Andhra Pradesh implemented ZBNF, aiming to achieve 100% chemical-free agriculture by 2024. As per the 2017-18 data, there were 17491 ZBNF farmers spread over 1000 villages across the 13 districts of Andhra Pradesh (Galab et al. 2018). The ZBNF market is further expected to grow at 20.5% in the forecast period of 2021 and 2026, to reach a value of about USD 2601 million by 2026 (Harini et al. 2021). Precision Farming (PF), on the other hand, is based on the applicability of technologies to analyze the dataset and derive "precise" results from it (Shafi et al. 2019). PF aims to instruct farmers in various perspectives, like foreseeing illness in cutting edge so that they can make moves and prevent the loss, suggesting crops reasonable for their field based on the climate and soil data, water system, and utilization of pesticides (Pierpaoli et al. 2013). Even though the adoption of PF technologies in farm management has been relatively new, the intrinsic simplicity of the crop recommendation models makes it more acquiescent. For a country like India, where agriculture is yet a prevailing occupation, accurate estimation/assessment of both the region and yield are similarly significant in guaranteeing the precise assurance of their products (Bakthavatchalam et al. 2022). The conventional cultivation techniques consequently give restricted crop yields compared to the inputs provided. Thus, to amplify the effects for a given number of inputs, various algorithms and recommender models have proven valuable in fostering a "precise" framework for smart farming (D'Antoni et al. 2012). Following PF, "recommender systems try to identify the needs and preferences of users, filter the huge collection of data accordingly and present the best-suited option before the users by using some well-defined mechanism" (Fayyaz et al. 2020). A study by Mokarrama and Arefin (2017) built a recommendation model based on factors such as physiography, crop growing period, and crop production rate. The recent advances in PF using machine learning (ML) models have allowed these models to be integrated into recommender systems with better results (Bakthavatchalam et al. 2022). This makes it possible to include ML models which can recognize favourable patterns for enhancing agricultural productivity. Thus, an integrative research framework that uses Machine Learning (ML) models for monitoring agricultural productivity and generating recommendations to farmers while adopting PF and ZBNF is proposed.

## 2 Materials and Methods

The dataset for this study was assorted from Indian government websites related to agriculture in different states of India. The dataset comprising the soil and climate attributes was accumulated for various districts in India for a comprehensive analysis. The dataset was homogenized for predicting crop productivity, depending on factors such as pH, temperature, rainfall, humidity, N-P-K and organic carbon (OC), and crop type (Table 1). The dataset consists of 9 variables and 764 observations across five years from 2015 to 2020.

The soil's acidity or alkalinity (pH) affects the amount of nutrients and chemicals soluble in soil water, thus making the nutrients available to crops. Humidity is a measure of moisture that influences stomata-related processes like evaporation and transpiration. Likewise, temperature influences most processes like transpiration and germination, directly affecting crop growth. The N-P-K levels in the soil are considered essential for optimal plant growth. Nitrogen and Phosphorus are important components of proteins and nucleic acids, and Potassium, an inorganic plant component, plays a vital role in regulating enzyme-related processes and osmosis. The N-P-K content optimizes crop growth, production and yield. The productivity matrix was derived from the crop and yield data set. Productivity as yield per hectare was classified into "High" and "Low" classes keeping the median value as a threshold. Crop types considered for the study are wheat, areca nut, bajra, banana, barley, cotton, dry chillies, garlic, ginger, jute, maize, onion, potato, rice and sugarcane.

### 2.1 Conduction of Study

An integrative analysis was conducted on the dataset repository generated by compiling information from various resources (Table 1). The workflow has been divided into two segments. The first segment analyses the productivity of crops for different districts in India. It aims to make predictions based on various factors like the soil, crop type and climate parameters influencing crop fecundity. This is done by implementing multiple ML classification algorithms to determine which is more suitable. The second segment analyses the profitability and usability of ZBNF using the statistical software STATA (Kohler and Kreuter 2005).

#### 2.1.1 Precision Farming (PF)

For PF workflow (Figure 1), firstly, the data undergoes pre-processing and is cleaned and scrutinized. Then it is processed in classification algorithms using k-folds cross-validation to predict crop productivity. Secondly, associations between these attributes using association rule mining by targeting the commonality between the parameters are explored. After that, deriving predictions and results are developed by the analysis.

Popular classification algorithms, viz Naive Bayes (NB), Support Vector Machines (SVM) and Random Forest (RF) have been compared on the compiled dataset to determine crop productivity (High or low). A comparison between Kappa statistics (McHugh 2012), AUC-ROC (Calster et al. 2008), and RMSE (Chai and Draxler 2014) values have been the selection criteria for determining the best out of these models. The RMSE measures the average error between predicted values and observations in appropriate units. A lower RMSE is preferred. The R-squared explains how much variation in the response is defined by the model. Kappa statistics measure inter-rater reliability or precision. It varies from 0 to 1, with 1 being a perfect agreement. AUC-ROC value is a performance measurement for the classifier, which determines how much the model can distinguish between classes. The higher the AUC-ROC value, the better the model is at predicting.

Table 1 Varied sources referred to attain soil and climate parameters across different Indian States

| Parameters | Source |
|---|---|
| Average pH | https://soilhealth7/gov.in/ |
| Average Temperature | https://climateknowledgeportal.worldbank.org/country/india |
| N-P-K, OC | https://soilhealth7/gov.in/ |
| Rainfall | https://mausam.imd.gov.in/imd_latest/contents/cs_anomaly_timeseries_temp_rainfall.php |
| Humidity | https://www.indiawaterportal.org/articles/district-wise-monthly-rainfall-data-2004-2010-list-india-meterological |
| Crops and Productivity | https://aps.dac.gov.in/APY/Public_Report1.aspx |

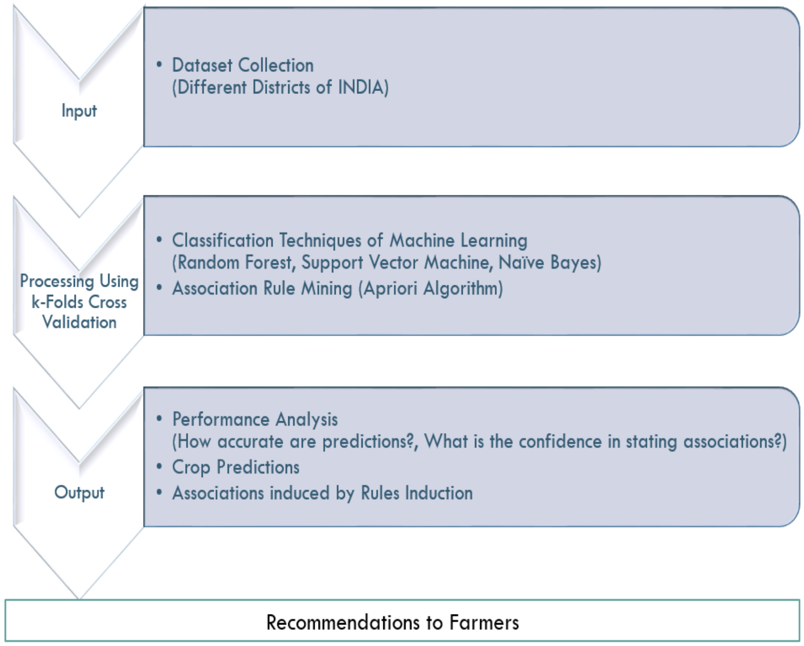


Figure 1 Workflow of Precision Farming

**2.1.2 Naïve Bayes (NB)**

NB Classifier is a simple probabilistic classifier that works based on Bayes' Theorem (Friedman et al. 1997), assuming that each input variable is independent or is unrelated to the presence of any other feature. Bayes' Theorem is used for calculating conditional probabilities, i.e., the probability of an event occurring given that another event has (assumption or assertion) occurred. According to the Theorem (Eq.1),

$$P(A|B) = P(B|A) * P(A) / P(B) \quad (1)$$

Here, B is the evidence or event, and A is the hypothesis or assumption that the predictors/features are independent. The naive Bayes Classifier calculates the posterior probability for each class (Friedman et al. 1997). It learns from training data the conditional probability of each attribute for a given class label - productivity matrix. The type with the highest posterior probability is the outcome of the prediction.

**2.1.3 Support Vector Machine (SVM)**

SVM is another ML classification algorithm used for two-group classification problems. The most straightforward formulation of SVM is the linear (Evgeniou and Pontil 2001), where the hyperplane lies in the space of the input data x (Eq.2).

$$f(x) = w\cdot x + b \quad (2)$$

In 2D, the discriminant is a line, where *w* is normal to the line known as the weight vector, and *b* is the bias. In 3D, a discriminant is a plane, and in n-dimensional space, it finds a hyperplane to classify the data points from a subset of training points, called support vectors, where n is the number of features

**2.1.4 Random Forest (RF)**

RF classifier can be described as the collection of tree-structured classifiers (Breiman 2001). It fits separate decision trees on a predefined number of bootstrapped data sets to improve the predictive accuracy and control over-fitting. RF classifier was built using the above climate and soil attributes and productivity data from 2017-19 as a class, bagging with 100 iterations. In each iteration, 10% of the data was split off as a test set.

**2.1.5 Apriori algorithm**

Association rule mining is used to identify underlying relations between different items. The Apriori algorithm is one of the approaches to finding all the association rules with the condition of minimum support and minimum confidence (Angeline 2013). The Apriori considers all the non-empty subsets of the dataset and targets the frequency of repetition and commonality of an item set. It is a bottom-up approach. The subset test starts from the bottom-most item set and is performed at each stage. The item sets with inconsistent or infrequent subsets are pruned, and the process is iterated until no more than all the successful item sets are derived. Confidence derived from the test measures how often items appear in transactions, i.e., the likelihood of a particular item appearing, provided other factors are known.

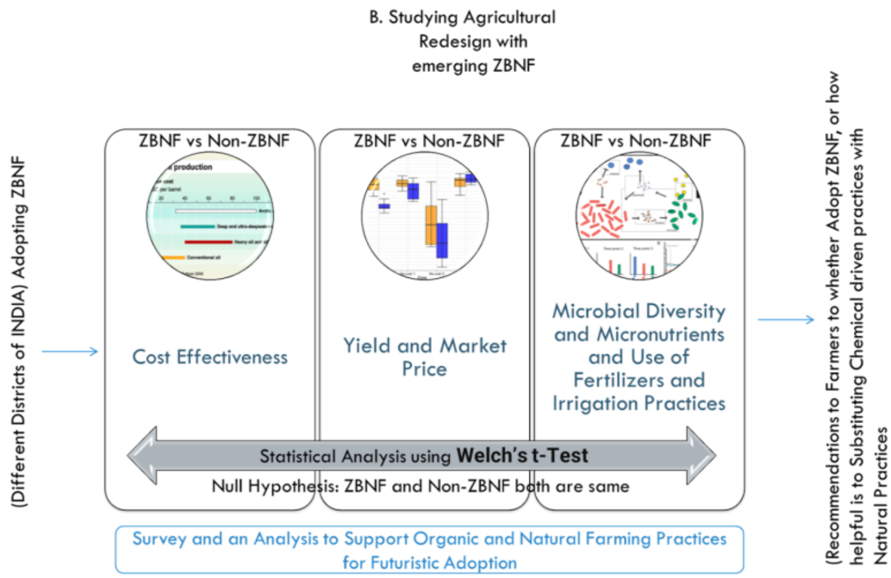


Figure 2 Analysis with ZBNF

### 2.2 ZBNF

ZBNF emerges as the beacon of hope to facilitate sustainability in an agriculture-based economy like India. The pilot experiments for ZBNF for the Indian districts have been compiled to draw statistical conclusions regarding the profitability and usability of ZBNF in real life (Figure 2).

In tandem, it is observed that medical and scientific experiments often have a small study group to prevent and minimize the extent of the inverse effects of the technology under consideration. In such situations, to record the difference in the samples before and after the experiment, different than usual statistical methods are used (Morgan 2017). In the following analysis as well since the sample size is small, Welch's t-test is used to give statistically significant results. Numerous economists have used this test in the past for similar situations. Since the sample size was small, it was suspected that the variance between the ZBNF and non-ZBNF groups was not equal; hence a Welch's t-test was conducted. Welch's t-test is the nonparametric equivalent of the conventional two-sample t-test (West 2021).

The ZBNF is explored to determine its cost-benefit, the scope of profit, and its effect on the soil's micronutrients and microbial diversity content. Welch's t-test was performed with a 95% confidence interval, under which the results are considered statistically significant if the p-value is less than 0.05. The test was conducted through the statistical software STATA.

The data for the same was collected and compiled district (Table 2) and factor-wise to interpret the results (Galab et al. 2018) efficiently. Furthermore, factors such as canals and tanks, irrigated and rainfed agricultural lands and other irrigation sources were considered to present a holistic and inclusive comparative study of the ZBNF.

The data for Cost is in rupees per acre. The net returns per acre are also measured in the same metric. The data for both Cost and net returns are taken for the states of Andhra Pradesh, Karnataka, and Maharashtra for crops, namely paddy, sugarcane, black gram, finger millet, soybean, cotton, turmeric, and chickpea. The data for micronutrients is written as mg per Kg of soil. The yield per acre is presented in Quintals.

Table 2 ZBNF adopted across various districts in Andhra Pradesh

| Mandya | Srikakulam | Visakhapatnam |
|---|---|---|
| Godavari | Guntur | Nellore |
| Prakarsa | Parbhani | Vizianagaram |
| Hingoli | Kurnool | Kadapa |
| Ananthapuramu | Chittoor | Krishna |

## 3 Results

### 3.1 PF

Machine learning algorithms RF, NB, and SVM were applied to the crop dataset and ZBNF dataset. The input parameters considered for the model were the average pH, average temperature, rainfall, humidity, N-P-K, organic carbon, crop type, and productivity matrix classified in "High" and "Low" distributions keeping the median value as the threshold. The productivity matrix was used as the label or target for the entire model.

The crop prediction accuracy of the RF model accounts for 89.14% with a 0.96 AUC-ROC value. The range values for Kappa statistics lie in the range [1,-1], with 1 presenting complete agreement and 0 meaning independence. Kappa statistics for RF is 0.78, which shows substantial agreement. The RMSE value of RF is lesser than that of NB and SVM, and hence RF model is better than other classifiers, as in Table 3.

The crop prediction model gave 96.79% accuracy for the dataset, including ZBNF. It can be inferred from Table 4 that RF was determined to be the best among the three classifiers.

Following the RF model, Association Rule Mining was applied to the dataset to put forward the predictions obtained based on support and confidence associative using the Apriori classification algorithm. The analysis gave around 96% confidence with minimum support of 0.55. After the Apriori algorithm was executed, several association rules were obtained. Out of the best rules found using the Apriori Algorithm, some of the recommendations derived are shown in Table 5.

### 3.2 ZBNF

#### 3.2.1 The cost, yield, and net returns analysis

A Welch's test is performed to determine whether there is a statistically significant difference in Cost between districts and factors, yield per acre (in quintal) for farmers that received ZBNF

Table 3 ML Classifiers for Crop Productivity

| Parameters | RF | NB | SVM |
|---|---|---|---|
| Accuracy | 89.14% | 84.32% | 84.99% |
| Kappa Statistics | 0.78 | 0.69 | 0.70 |
| RMSE | 0.28 | 0.33 | 0.39 |
| Weighted AUC-ROC | 0.96 | 0.92 | 0.85 |

Table 4 ML classifiers for ZBNF

| Parameters | RF | NB | SVM |
|---|---|---|---|
| Accuracy | 96.79% | 91.29% | 94.77% |
| Kappa Statistics | 0.83 | 0.62 | 0.71 |
| RMSE | 0.15 | 0.27 | 0.23 |
| Weighted AUC-ROC | 0.99 | 0.93 | 0.83 |

Table 5 Instances for crop recommendation using Apriori Algorithm

| pH | Temperature | Rainfall | NPK | Soil Type | Preferred Districts | Crop Recommended |
|---|---|---|---|---|---|---|
| Acidic | Hot | Low | Low to Medium | Alluvial | Panchkula, Mewat, Faridabad | Garlic |
| Acidic | Hot | Low | Low | Alluvial | Chirag, Kamrup, Nalbari, Baksa, Udalguri | Jute |
| Acidic | Hot | Low | Medium | | Bandipora, Ganderbal | Dry Chillies |
| Acidic | Hot | High | Low | Alluvial | Barnala, Pathankot, Fazilla, Haridwar | Rice |
| Acidic | Cold/Hot | Low | Very Low to Low | Alluvial | Kurukshetra, Sirsa, Jhajjar, Banka, Patna, Arwal, Tawang, Sonepur | Potato |
| Alkaline | Hot | Low | Medium | Alluvium, Colluvium | Tamenglore, Goalpara | Arecanut |
| Acidic/ Alkaline | Hot | Low | Low to Medium | Sandy Loam, Alluvial, Mountain Meadow | Muktsar, Bikaner, Jaipur, Banka, Patna, Aurangabad, Faridkot, Mansa, Jodhpur, Ajmer | Banana |
| Acidic | Hot | Low | Medium | Mountain Meadow | Kupwara, Rajauri, Samba | Garlic |
| Alkaline | Cold | Low | Low to Medium | Sandy Loam, Alluvial | Kulgam, Una, Almora, Sirsa, Sonipat | Onion |
| Acidic | Cold to slightly warm | Low | Medium to High | Alluvial | Shimla, Lohit, Aizwal, Champai, Una, West Siang, Hamirpur | Sugarcane |

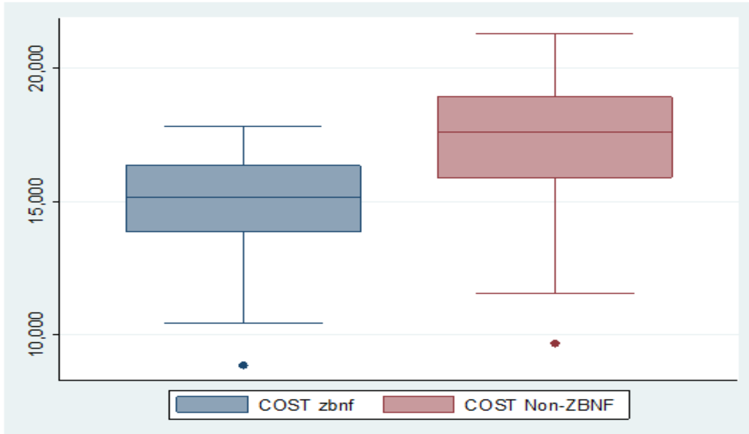


Figure 3 Boxplot of cost comparison of ZBNF and non-ZBNF

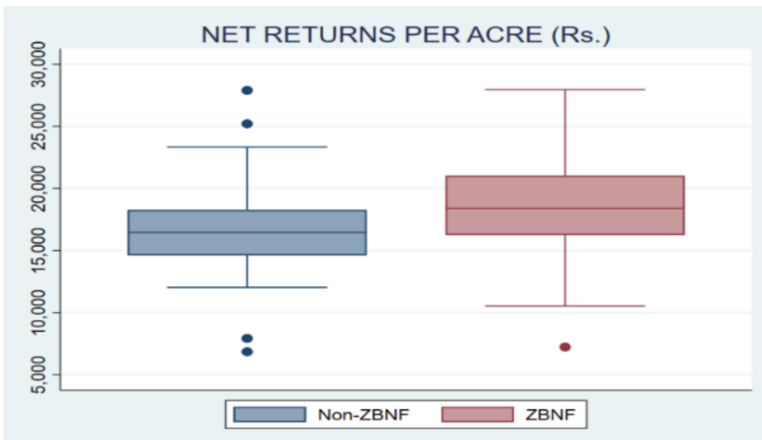


Figure 4 Boxplot of net returns per acre comparison of ZBNF and non-ZBNF

and the group that didn't. The sample size was small. Figure 3 displays two outliers, and the same reflects fitting the inefficiency in the farmer market structures, inconsistent information set, mediators, etc., which can be the possible causes for the same in the Indian context. Figure 3 showcases that ZBNF costs less than non-ZBNF and is more cost-efficient. No statistically significant difference could be recorded in the yield per acre recorded by farmers who used ZBNF for cultivation and the ones who did not, and hence is not deteriorating for them.

Additionally, Figure 4 showcases that ZBNF has slightly higher net returns per acre than the non-ZBNF alternative, which benefits the farmers. The difference in the span of boxes is evident. The ZBNF box's whiskers are longer than the non-ZBNF box.

### 3.2.2 Microbial Diversity

The natural factors also require a more extended period to show significant results. More profound research can broaden the outlook of the same. Additionally, the boxplot (Figure 5) of non-ZBNF is comparatively shorter; this implies the levels of microbial diversity are moreover uniform, whereas the ZBNF is relatively tall, which implies different levels of microbial diversity. It is shown through various studies that ZBNF enriches the soil, and hence the same is observed; more exploratory research can add more valuable insights.

### 3.2.3 Micronutrients

Similarly, a Welch's test was performed to determine whether there is a statistically significant difference in micronutrients between districts that used ZBNF and those that didn't. There is no statistically significant difference in mean values between the two groups. The mean of group ZBNF is approximately the same as that of the non-ZBNF.

Hence, usage of ZBNF does not reduce or deteriorate the micronutrient content of the soil. Without the use of supplementary nutrients in the form of fertilizers after a complete agriculture cycle, the micronutrient content of the soil is retained in the ZBNF paradigm (Figure 6). This saves the variable Cost that would have been incurred on fertilizer consumption, thus reducing the Cost of production and increasing profits.

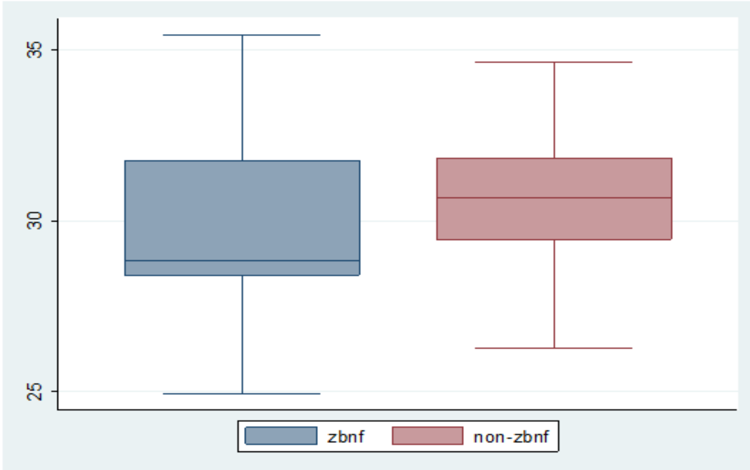


Figure 5 Boxplot of Microbial diversity comparison of ZBNF and non-ZBNF

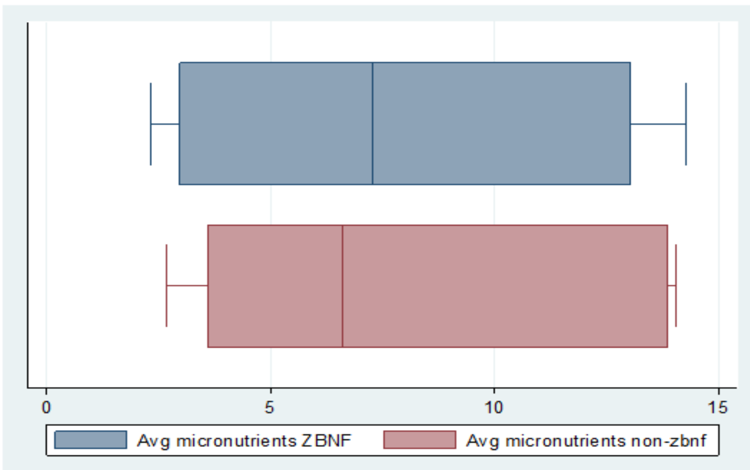


Figure 6 Boxplot of micronutrients comparison of ZBNF and non- ZBNF

## 4 Discussion

Precision farming is an evolving practice developing rapidly in the past two decades. Precision farming could be a solution to many challenges arising due to climate change. This technique could lead to an increase in agricultural output with less input. The current crop yield prediction is performed by state-of-the-art models, RF, NB, and SVM algorithms. Using ML in crop yield prediction has advantages as it provides faster and more accurate predictions. The crop yield depends on many parameters, like climatic factors, soil quality, air parameters, etc. ML-based prediction systems handle the dependency of the parameters efficiently. The Apriori algorithm produced results showing decent relations among the parameters with substantial confidence and generating recommendations to farmers. The crop recommendations are generated based on pH, rainfall, temperature, and districts. These predictions showed adequate results (> 80 % accuracy) and thus can be incorporated soon.

Welch's test provided statistically significant conclusions supporting ZBNF as an alternative to chemical-based farming. It showcased that ZBNF is more cost-efficient than non-ZBNF and gives slightly higher net returns per acre than the non-ZBNF. The research showed that ZBNF enriches the soil as the results show different levels of microbial diversity.

There was no significant statistical difference in the yield per acre recorded by farmers who opted for chemical-based farming instead of ZBNF. ZBNF does not deplete the soil of its natural content of micronutrients also. After the completion of an agricultural cycle, there is no need to rejuvenate the soil with fertilizers and supplements as the same micronutrients are retained. Hence, further decreasing the variable costs. ZBNF emerges as a more cost-efficient and profitable alternative for farmers. The advantages of using ML in integrative Agricultural Systems are (i) ML offers accurate detection of crop productivity with an accuracy of > 80%, which is better compared to manual/classical techniques, (ii) Prediction of crop productivity based on environmental factors using ML exhibit low error indices such as RMSE measuring the model accuracy for statistical analysis (iii) ZBNF is a natural alternative and enhances productivity over non-ZBNF techniques and (iv) Accurately labelled dataset was designed for implementing an ML-based agricultural system which the research community could further use.

Though ZBNF is a beneficial technique, however few of the challenges and limitations in the prediction of crop yield were identified, and these are (i) varying parameters while analyzing datasets pose a challenge to the design of the prediction model, (ii) dataset selection is critical due to the complexity; as an improper selection of data may result in underfit/overfit problem, and (iv) Accurate classification through ML seems complicated in varying geographical conditions.

Based on the results of the study, some valuable recommendations to farmers have been generated, and these are (i) Garlic crop was recommended for Hot and Acidic Climates, (ii) Sugarcane was recommended to be grown in Alluvial soils in cold and slightly warm regions such as Shimla, Lohit, Aizwal, Champai, Una, West Siang, Hamirpur, (iii) Onions were recommended in Alkaline soil Regions, and (iv) Potatoes were recommended to be grown in Acidic Soil regions

The amalgamation of two approaches, i.e., integration of the recommender model and ZBNF, can be propelled towards agriculture sustainability.

**Conclusion**

The research paper delves into the concept of an integrative model amalgamating PF and ZBNF. PF includes analysis of soil, weather, crop, and other needs to increase agricultural productivity and improve its quality. In this paper, we empirically examine the application of ML in crop productivity within Indian farming systems. Using preliminary information from varied crop growers across Indian states, classification and regression models estimated a) the differences between high and low crop production and b) the differences between ZBNF and non-ZBNF approaches. This study used three established ML supervised models - RF, NB, and SVM- and a construct of ZBNF adoption perception to analyze these agriculture tools' adoption. Holistic and intricate datasets have been generated by compiling information from various online resources available, highlighting multiple external factors influencing crop productivity in Indian states/districts and the pilot surveys. The research primarily predicts crop productivity based on various factors like the soil, crop type, and climate parameters influencing crop productivity and making beneficial recommendations for the most suitable crop.

Additionally, exploratory research has been done to ascertain the profitability and usability of ZBNF using the statistical software STATA. The study suggests that ZBNF costs less than non-ZBNF. Also, the study indicates that the micronutrient content of the soil is retained in the ZBNF paradigm even without using fertilizers. The pilot statistical experiments for ZBNF for the Indian districts have been compiled in this research to draw statistical conclusions regarding the profitability and usability of ZBNF. This article also concluded positively in favour of the integrative model with statistically significant results regarding PF and ZBNF. The article makes recommendations to farmers for an efficient and sustainable agricultural outlook.

In future work, digital platforms and chatbots can be built for farmers, and more ML algorithms can be explored in the sustainable agriculture industry.

**Acknowledgments**

The authors would like to express their gratitude towards the Centre for Research, Maitreyi College, University of Delhi, for providing the opportunity to take up this research. We highly appreciate the support provided by them throughout the research.

**Conflict of Interest**

The authors declare that the research was conducted without any commercial or financial relationships that could be constructed as a potential conflict of interest.

**Authors Contribution**

Dr. Veena Ghuriani and Dr. Jyotsna Talreja Wassan were involved in conceptualization, preparing the original draft and project administration. Pragya Deolal, Vidushi Sharma, Dimpy Dalal and Aditi Goyal worked on experimentation and writing. All authors have read and agreed to the published version of the manuscript.

**References**

Angeline D. M. D. (2013). Association Rule Generation for Student Performance Analysis using Apriori Algorithm, *The SIJ Transactions on Computer Science Engineering & its Applications*, *1*(1), 12-16.

Bakthavatchalam, K., Karthik, B., Thiruvengadam, V., Muthal, S., Jose, D., Kotecha, K., & Varadarajan, V. (2022). IoT Framework for Measurement and Precision Agriculture: Predicting the Crop Using Machine Learning Algorithms. *Technologies*, *10*, 13.

Beluhova-Uzunova, R. P., & Dunchev, D. M. (2019). Precision Farming - Concepts and perspectives. *Problems of Agricultural Economics*, *360*(3), 142-155.

Bishnoi, R., & Bhati, A. (2017). An overview: Zero budget natural farming. *Trends in Biosciences*, *10*(46), 9314-9316.

Breiman, L.(2001). Random Forests. *Machine Learning, 45*, 5–32.

Calster, B.V., Van Belle, V., Condous, G., Bourne, T., Timmerman, D., & Van Huffel, S. (2008). Multi-class AUC metrics and weighted alternatives. *In 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* , 1390-1396.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*, 1247–1250.

D'Antoni, J. M., Mishra, A. K., & Joo, H. (2012). Farmers' perception of precision technology: The case of autosteer adoption by cotton farmers. *Computers and Electronics in Agriculture*, *87*, 121-128.

Duddigan, S., Shaw, L. J., Sizmur, T., Gogu, D., et al. (2023). Natural farming improves crop yield in SE India when compared to conventional or organic systems by enhancing soil quality. *Agronomy for Sustainable Development*, *43*(2), 31.

Evgeniou, T., & Pontil, M. (2001). Support Vector Machines: Theory and Applications. *Machine Learning and Its Applications,* 249–257. DOI:10.1007/3-540-44673-7_12

FAO (2016). Zero Budget Natural Farming in India. Retrieved from Web-Link http://www.fao.org/3/a-bl990e.pdf

Fayyaz Z., Ebrahimian M., Nawara D., Ibrahim A. & Kashef R.(2020). Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities. *Applied Science*, *10*(21),7748.

Friedman, N., Geiger, D. & Goldszmidt, M. (1997). Bayesian Network Classifiers. *Machine Learning, 29*, 131–163.

Galab S., Prudhvikar Reddy, P., Sree Rama Raju, D., Ravi, C., & Rajani, A. (2018). Impact Assessment of Zero Budget Natural Farming in Andhra Pradesh – Kharif 2018-19. Retrieved from Web-Link: https://www.scribd.com/document/468784250/CESS-FINAL-KHARIF-REPORT-ZBNF-19-8-19-pdf.

Harini, N., Veni, C. P., Sailaja, A., & Lata, A. M. (2021). Zero budget natural farming (ZBNF): A critical analysis on crop wise practices, ZBNF models and cropping systems. *The Pharma Innovation Journal*, *10*(8S), 105-109.

Jayarajan, S.K.P., & Kuriachan, L. (2021). Exposure and health risk assessment of nitrate contamination in groundwater in Coimbatore and Tirupur districts in Tamil Nadu, South India. *Environmental Science and Pollution Research International, 28*, 10248 - 10261.

Kohler, U., & Kreuter, F. (2005). *Data analysis using Stata*. Stata press.

Kostraba, J. N., Gay, E. C., Rewers, M., & Hamman, R. F. (1992). Nitrate levels in community drinking waters and risk of IDDM. An ecological analysis. *Diabetes care*, *15*(11), 1505–1508.

Kumar, R., Kumar, S., Yashavanth, B.S., Meena, P.C., et al. (2020) Adoption of Natural Farming and its Effect on Crop Yield and Farmers' Livelihood in India. *ICAR-National Academy of Agricultural Research Management, Hyderabad, India.* Retrived from web link : http://www.niti.gov.in/sites/default/files/2021-03/NaturalFarmingProjectReport-ICAR-NAARM.pdf

McHugh, M. (2012). Interrater reliability: The kappa statistic. *Biochemia medica : časopis Hrvatskoga društva medicinskih biokemičara, 22*(3), 276–282.

Mokarrama, M.J., & Arefin, M.S. (2017). RSF: A recommendation system for farmers. *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC),* 843-850.

Morgan C. J. (2017). Use of proper statistical techniques for research studies with small samples. *American journal of physiology, Lung cellular and molecular physiology*, *313*(5), 873–877.

NABARD. (2018). Agriculture Credit to Farmers. Retrieved from Web-Link https://nabard.org/news-article.aspx?id= 25&cid= 552 &NID=160

National Research Council. (1997). *Precision Agriculture in the 21st Century: Geospatial and Information Technologies in Crop Management*. Washington, DC: The National Academies Press.

Pierpaoli, E., Carli, G., Pignatti, E., & Canavari, M. (2013). Drivers of precision agriculture technologies adoption: A literature review. *Procedia Technology*, *8*, 61-69.

Rahman, A., Mondal, N.C. & Tiwari, K.K.(2021). Anthropogenic nitrate in groundwater and its health risks in the view of background concentration in a semi arid area of Rajasthan, India. *Scientific Reports*, *11*(1), 9279

Savcı, S. (2012). An Agricultural Pollutant: Chemical Fertilizer, *International Journal of Environmental Science and Development*, 3(1), 77-80.

Shafi, U., Mumtaz, R., García-Nieto, J., Hassan, S. A., Zaidi, S. A. R., & Iqbal, N. (2019). Precision agriculture techniques and practices: From considerations to applications. *Sensors*, *19*(17), 3796.

West R. M. (2021). Best practice in statistics: Use the Welch *t*-test when testing the difference between two groups. *Annals of clinical biochemistry*, *58*(4), 267–269.