PLANT SCIENCE TODAY
ISSN 2348-1900 (online)
Vol 12(2): 1-8
https://doi.org/10.14719/pst.7933

HORIZON
e-Publishing Group
HePG

RESEARCH ARTICLE

Predicting small millets productivity based on machine learning models: A comprehensive study

Venkatesa Palanichamy Narasimma Bharathi[1], Kalpana Muthuswamy[1], Balakrishnan Natarajan[1*], Balamurugan Vasudevan[1], Suresh Appavu[1], Rajavel Marimuthu[2], Dhivya Rajaram[3] & Krupesh Sivakumar[4]

[1]Office of the Dean (Agriculture), Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

[2]Office of the Public Relations, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

[3]Department of Business Administration, PSGR Krishnammal College for Women, Coimbatore 641004, Tamil Nadu, India

[4]Department of Physical Sciences and Information Technology, Agricultural Engineering College and Research Institute, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

*Email: rnbkrishnan@gmail.com

OPEN ACCESS

ARTICLE HISTORY
Received: 26 February 2025
Accepted: 01 April 2025
Available online
Version 1.0 : 24 April 2025
Version 2.0 : 30 April 2025

Check for updates

Additional information
Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

CITE THIS ARTICLE
Venkatesa Palanichamy NB, Kalpana M, Balakrishnan N, Balamurugan V, Suresh A, Rajavel M, Dhivya R, Krupesh S. Predicting small millets productivity based on machine learning models: A comprehensive study. Plant Science Today. 2025; 12(2): 1-8. https://doi.org/10.14719/pst.7933

Abstract

People in rural and urban areas prefer small millet-based food products and they are more popular in markets. Before the green revolution, little millets were widely cultivated for food and fodder, but dietary preferences changed and they were no longer widely planted. Machine learning techniques have been applied in agriculture recently to analyze and predict crop yields. A major concern for farmers during the growing season is estimating their expected yield. In this study, data on area and production for small millets in Tamil Nadu are collected over 50 years. The machine learning models are used to predict productivity with the available small millets dataset. Four different machine learning models are used to estimate small millets productivity. With an accuracy of 95.46 %, a mean absolute error (MAE) of 0.0709, a root mean square error (RMSE) of 0.014 and an R-square value of 0.94, the Random Forest regressor performed better than the other models. The current research helps the farmers in mitigating potential losses, as their financial stability and productivity output are closely related. Additionally, the study provides valuable insights for better planning and implementation. Furthermore, the Random Forest regressor offers insightful information to help farmers maximize their farming techniques and make well-informed judgments.

Keywords

accuracy; machine learning model; prediction; productivity; small millets

Introduction

Millets belong to diversified family of cereal grasses and can be divided into major and small grain millets. Sorghum and pearl millet are major millets; small millets include finger millet, foxtail millet, kodo millet, proso millet, barnyard millet and little millet (1). The small millets contribute to environmental sustainability and address concerns related to health, nutrition, fodder and fibre security, which is truly remarkable.

Millets are a unique agricultural product originated in India, are becoming increasingly popular worldwide. According to projections, millets will become increasingly in demand as people emphasize healthier grains, with a predicted $12 billion in sales by 2025 (2). India is one of the world's top producers of millet, with an annual production of about 12 million tonnes (3). The Department of Agriculture and Farmers Welfare (DA&FW) is leading a Sub-Mission on Nutri-Cereals (Millets) within the National Food Security Mission (NFSM) to increase the production and productivity of millets, especially bajra.

Achieving growth in small millet production involves educating people about the nutritional benefits of small millets and raising awareness among consumers through consumption.

Furthermore, it's critical to assist farmers in implementing sustainable farming methods and to provide them with the tools and infrastructure they need to expand small-millet cultivation. This will support the general development of rural areas in addition to guaranteeing a consistent supply of millet.

Predicting crop yields is critical for farmers, food marketing business entities, lawmakers and governments concerned about food security. These important parties may devise strategies for the effective distribution of food, price stability and resource allocation by using yield prediction models. Anticipating production changes can help develop a food system that is more resilient to them. The yield of crop forecasting is challenging, however, due to several elements, such as seed types, soil composition, fertilizer types and weather. Because of this, this process requires integrating many datasets and attribute types. Small millets productivity needs to be accurately predicted for agricultural techniques to be successful and sustainable. These forecasts help farmers schedule their cultivation, allocate their resources effectively and maximize the productivity of their crops. On the other side, governments and policymakers use this data to create programs and policies that guarantee food security for their citizens. They may use effective resource allocation, distribution methods and price stabilization to prevent market swings by knowing the anticipated output of small millets.

Small millet cultivation has been an integral part of Indian agriculture for centuries, providing nutritional security and resilience to climate variability. However, in recent decades, small millet production has declined due to changing dietary preferences, urbanization and a lack of market incentives. Additionally, accurate prediction of small millets productivity remains a major challenge due to the diverse agro-climatic conditions and varying cultivation practices (4). Traditional statistical methods used for yield estimation often fail to capture complex, non-linear relationships among multiple influencing factors such as weather patterns, soil quality and farming techniques. As a result, farmers and policymakers struggle to make informed decisions regarding resource allocation, crop planning and food security measures. The need for a robust predictive framework that integrates machine learning techniques with historical data is crucial to addressing these limitations. This study aims to develop and evaluate machine learning models that can accurately predict small millets productivity, offering a data-driven approach to enhance agricultural decision-making and sustainability.

Existing research on crop yield prediction primarily focuses on staple crops such as rice, wheat and maize (5), with limited studies dedicated to small millets (4,6,7). While traditional regression-based models have been employed in past studies, they often lack the predictive accuracy required for dynamic and diverse agricultural systems. Machine learning techniques have recently gained attention for their ability to handle large datasets and complex variable interactions, but their application in small millets productivity prediction remains underexplored. Moreover, many previous studies lack a comparative analysis of multiple machine learning models to determine the most effective approach. This research bridges the gap by implementing and evaluating four machine learning models Random Forest Regressor, Decision Tree Regressor, Support Vector Regression and Linear Regression using a dataset spanning over 50 years. Assessing their predictive performance, this study contributes to the advancement of precision agriculture and provides actionable insights for farmers and policymakers. The study addresses the lack of region-specific data analysis for small millets, ensuring that the predictive models are tailored to the agro-climatic conditions of Tamil Nadu.

Review of Literature

Historically, statistical techniques like multivariate linear regression (MLR) have been used by scientists to estimate crop productivity. However, the prediction's accuracy was not as precise as anticipated (8). The efficacy of Machine Learning (ML) techniques as descriptive and predictive tools for addressing research difficulties is steadily increasing. One of the challenging problems in precision agriculture is crop production prediction, for which numerous models have been proposed and validated (4, 6). Predicting crop productivity in the early stages is a difficult task. Accurate understanding of historical crop productivity data is essential to control agricultural hazards and project future outcomes. Numerous studies used statistical models, such as regression and multivariate regression, with limited input factors to estimate crop productivity (9-11). The information below demonstrates current research on a variety of techniques and spectrums used in agricultural production prediction.

Crop production forecast techniques, such as Ordinary Least Squares (OLS) estimation, also include the use of linear regression models. In this case, the performance of the autoregressive model outperformed that of OLS, achieving a higher R^2 value (12). The study concluded that, when temperature was ignored, NDVI and precipitation had a greater impact on corn yield in Iowa. A correlation was observed between various linear and nonlinear crop yield prediction techniques (13). The comparison utilized the optimal subset of variables for each approach, determined through a comprehensive procedure and percentage split validation. The test datasets, used to measure performance, consisted of unseen samples. The study assessed multiple regression-based methods such as regression trees, stepwise regression, multiple linear regression and neural networks. The testing demonstrates that execution is considerably improved when a complete system is used for quality determination.

The controllable educational regression and classification model were developed (14). These algorithms, widely used in various industries, attempt to anticipate or detect new data input through examples. The idea of using knowledge as examples was spurred by studies utilizing supervised learning algorithms to predict future illnesses and pests in agricultural crops. Abiotic variables such as excessive heat and humidity have a detrimental effect on crop performance, increasing agricultural output's susceptibility to climate change. The date of rice production was estimated using the MLR model and results showed the highest farmer exchange rate (15). Meteorological data, sourced from the National Statistical Authority and information on the farmer exchange rate, were used to construct a regression model incorporating variables such as "Rainfall", "Average Temperature", "Average Moisture" and "Solar Radiation". Predictions were produced by examining every potential combination of factors with a low RMSE number.

The study focused on analyzing educational data to predict students' cognitive domains. Linear regression was employed along with four regularization techniques: elastic net, lasso, ridge and no regularization. Comparatively, employing cross-validation and random sampling as assessment tools, elastic net regression yielded the lowest prediction error (16). The measured Chinese medicine bone-setting modification details in a virtual environment using MLR techniques. Linear regression predictions were used to model the displacement and angle of bone manipulation information (17). This technique is also applied to optimize treatment efficiency and educational training.

The authors evaluated past sales data and forecasted future sales for a large retail establishment. Using a Deep Learning Linear Regression Algorithm, revenue data from 2011 to 2013 was used to predict 2014 statistics (18). Comparisons of estimated and actual 2014 sales data revealed an 84% accuracy rate. Descriptive analysis was used to estimate crop yields accurately by projecting future agricultural production (19). This analysis incorporated datasets on yield, rainfall and soil, applying multiple supervised algorithms to ascertain the estimated cost of different methods. The enhanced capabilities of LS-SVM over Support Vector Machines were demonstrated.

Linear regression is widely used in mathematical research techniques as it quantifies predicted effects and models them against input parameters. This method establishes linear correlations between dependent and independent variables through data analysis and modelling (20). Predicted that understanding crop yield before cultivation would help farmers make informed decisions (21). Data mining was utilized to analyze historical agricultural production data and an algorithm was implemented for forecasting.

Materials and Methods

Machine learning, as a form of artificial intelligence (AI), enables computers to learn without explicit programming. The statistical data for the study is gathered from various secondary sources like www.indiastat.com and www.milletstats.com. The dataset to be used for Small Millets (Minor Millets) includes Area (in Hectare) and Production (in Tonnes) of historical data from 1966-1967 to 2021-2022.

Machine Learning Models

In this comprehensive study, machine learning is an essential tool for decision-supporting agricultural productivity forecasting. Machine learning models can help farmers minimize agricultural losses by providing useful insights and thorough crop advise. For small millets, this study used a variety of machine learning models, such as Random Forest Regressor (RFR), Decision Tree Regressor (DTR), Support Vector Regression (SVR) and Linear Regression. These models were chosen for their ease of interpretation and robustness when applied to relatively small datasets. Using accuracy metrics, the optimal model was chosen.

Linear Regression

The Linear Regression approach was chosen because of the size of the dataset and the simple fact that the forecast was quantitative rather than categorical. With the use of linear regression, this approach investigates the distribution of a response variable Y that changes when the intervening variable X is estimated. To make a prediction, one must exclude the response variable's value based on a precise calculation of the explanatory variable. Conversely, regression is the process of applying the most appropriate straight line to ascertain the relationship between three variables, where one is the independent variable (X) and the other two are dependent variable (Y). Presented below is the regression equation.

$$Y = a + (b*X1, b*X2) + e \qquad (1)$$

where,

Y(Productivity) - Dependent variable

X1, X2 (Area, Production)- Independent variable,

a-Intercept, b-Slope, e-Residual (error)

Support Vector Regressor (SVR)

A Support Vector Regressor (SVR) is a supervised algorithm for machine learning that may be used for both classification and regression problems. Finding a function that approximates the connection between both input variables and a continuous target variable while reducing prediction error is the main goal of regression tasks. Support Vector Regressors (SVRs) work on the basic principle of finding the optimum line of fit while reducing the fitting error within a predetermined range known as the ε-tube. This idea is demonstrated in Fig. 1, where the ideal fitting line in SVR is indicated by the hyperplane with the greatest number of points.

The SVR was described by Smola and Scholkopf (2003) under the presumption that there is a set of "training data, (xj, y) with j = 1, 2, with input x = "x1, x2" and output y. For all training data, we will locate a function of f(x) with the largest deviation from the actual target using SVR. Then, when the value of is zero, perfect regression can be achieved with SVR. The SVR wants to find a function of f(x) that can approximate the output to an actual target using only minimal complexity and a tolerance error of based on the data (11). The regression function of f(x) can be communicated by the following formula:

$$f(x) = wT\varphi(x)+b \qquad (2)$$

where φ(x) signifies a point in the higher dimensional highlight space, the planning consequences of info vector x in input space which has lower aspect. By minimizing the risk function that is defined in the equation, the coefficients of w and b can be estimated:

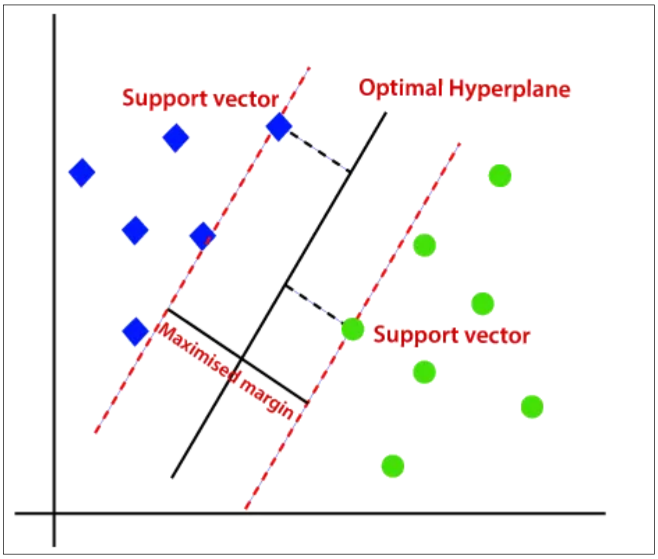

Fig. 1. Support vector regressor.

SVM classification illustrating the optimal hyperplane (black line), margin boundaries (red dashed lines) and support vectors (highlighted points). The blue and green markers represent two distinct classes. X-axis: Feature 1 (Independent Variable); Y-axis: Feature 2 (Dependent Variable).

$$min\frac{1}{2}||w||^2+C\frac{1}{\tau}\sum_{i-1}^{\tau}L_\epsilon(y_i,f(x_i)) \quad (3)$$

which fulfill,

$$y_i-w\varphi(x_i)-b\leq\varepsilon$$
$$w\varphi(x_i)-y_i+b\leq\varepsilon, i=1,2,\ldots,\tau \quad (4)$$

Where,

$$L_\varepsilon\left(y_i,f(x_i)\right)=\begin{cases}|y_i-f(x_i)|-\varepsilon|y_i-f(x_i)|\geq 0\\ 0, for\ others\end{cases} \quad (5)$$

Unlike other models that aim to minimize the difference between predicted and actual values, SVR strives to find the best line of fit within a predefined threshold value. This threshold value, also known as the distance between the boundary line and hyperplane, plays a crucial role in SVR.

Decision Tree Regressor

The decision tree regressor is an effective method that builds a model in the shape of a tree structure by analyzing the characteristics of an object. Next, this model is applied to forecast future data, producing useful and ongoing results. Results that are not restricted to a discrete, preset set of numbers or values are referred to as continuous output. Decision tree regression is a dependable approach in data analysis that may be used to comprehend intricate relationships and provide precise projections. Through the analysis of several object attributes, this approach facilitates the development of an all-encompassing model that encapsulates the complex patterns present in the data. The resulting tree structure acts as a guide, directing the process of prediction and offering significant insights into what will happen in the future.

A decision tree regressor uses a test on the characteristics to split the original dataset into smaller subsets, which then creates a tree structure. Until the subset in a node includes the same value as the target label or until splitting no longer improves predictive power, this procedure is performed recursively on each subset.

Random Forest Regressor

Random forests regressor are a powerful ensemble method that combines multiple tree predictors (100 trees), making them highly resilient to noise. To ensure that every tree in the forest has the same distribution characteristic, sets of variables and samples are chosen at random from the dataset. The random forest algorithm creates a vast number of individual trees, which enables them to vote on the most popular classes collectively, producing accurate predictions (Fig. 2). Over the past ten years, this innovative approach has shown to be useful in managing high-dimensional datasets and minimizing overfitting problems (12).

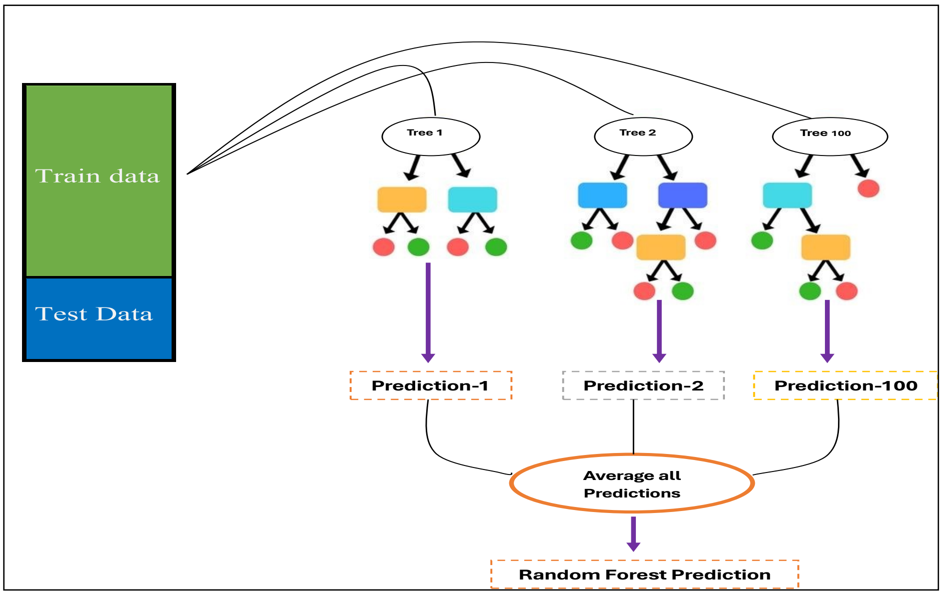

Fig. 2. Random Forest Regressor.

Working mechanism of the Random Forest model. The dataset is split into training and testing sets. Multiple decision trees (Tree 1, Tree 2, ..., Tree 100) are trained on different subsets of the training data. Each tree generates individual predictions, which are then averaged to obtain the final Random Forest prediction, improving accuracy and reducing overfitting.

Tools used

Python is the programming language of choice for this investigation, handling every phase from collecting data to assessing the model. This decision is since Python has a large library supporting machine learning and artificial intelligence, which makes it ideal for solving issues in the real world. Because Python is portable, it doesn't require any particular operating system, which increases the project's adaptability and accessibility.

Machine learning Models: Numerous studies have demonstrated the importance of machine learning as a decision-support tool for productivity forecasting. A program that uses machine learning to provide comprehensive agricultural advice and insights can help farmers lower their farming losses. The MLMs addressed in this article are the regression, Linear regression, Decision Tree regression, Random Forest regressor and Support Vector regression. These techniques were selected due to the size of the dataset and the statistical rather than category nature of the prediction.

Small millet productivity prediction using Machine learning: Using crop data on the area and production of small millet as input parameters and small millet productivity as output parameters.

Step 1: The raw data are formatted to input the machine learning models to get the desired results.

Step 2: Separate the dataset into the training and testing dataset. 80 % of the data are used as the training set and 20 % are used as the test set.

Step 3: Generate machine learning models.

Step 4: Evaluate R^2, RMSE (Root Mean Squared Error), accuracy to determine the model performance.

Step 5: The optimal model for predicting small millet productivity is one that has high accuracy and R^2 values as well as low RMSE statistics values.

Results and Discussion

This research looks at the trends in small millet production in Tamil Nadu. The largest acres and the greatest productivity of small millets occurred in Tamil Nadu in the late 1960s and early 1990s. Despite this, during the 1990s, there has been a little decline in production. It's significant to note that since 2010, the productivity of small millets has grown despite a decrease in output. This indicates that despite cultivating a smaller area, producers have been able to attain better output. However, as of 2021, small millets' productivity and area have both significantly decreased.

As Fig. 3 illustrates, small millets are still more productive now than they were in the late 1960s and early 1990s. The reasons behind the decline in small millets cultivation. It's perhaps that rapid changes in urbanization, industry and agricultural preferences shifted attention from the small millets production. Because small millets have historically generated lesser productivity than other cereal crops, farmers' decision-making processes may have been impacted by the introduction of high- productivity crop

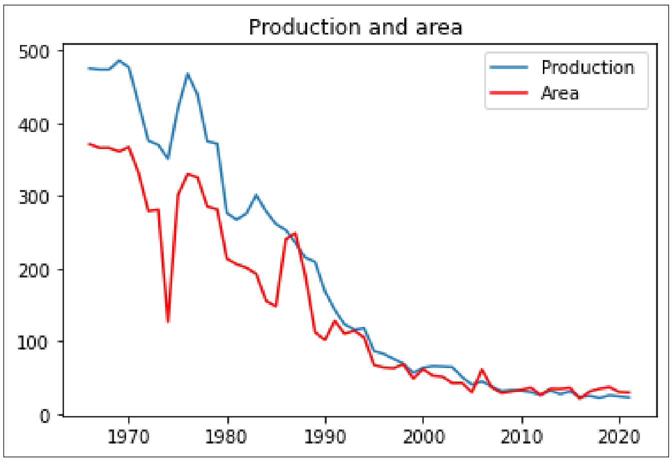

Fig. 3. Production and area for small millets.

varieties and advances in technology during the Green Revolution period.

Models Performance and their prediction

The prediction of crop productivity by machine learning is an important area of research since it helps with decision-making by offering helpful information into productivity patterns. In this study, we used four different models to analyze datasets involving small millets. Comparison of each MLMs actual and predicted crop productivity is shown in Fig. 4

A comparison of each MLMs actual and predicted crop output is displayed in Fig. 4. The Random Forest Regressor (RFR) results fit the regression line well, as seen in Fig. 4a. The results of Support Vector Regressor (SVR) predictions do not suit the regression line as well as those of LR and DTR, as seen in Fig. 4b. The findings of Linear Regression (LR) and Decision Tree Regressor (DTR) did not match the regression line as well as the RFR as seen in Fig. 4c & 4d. Random Forest Regressor (RFR) outperformed other models in my research due to its and. By tuning parameters such as the number of trees, maximum depth and minimum samples per split, RFR achieved a balance between bias and variance. Unlike Decision Trees, which overfit to training data, RFR's ensemble approach reduces variance by averaging multiple trees. It also surpasses Linear Regression and SVR by capturing complex non-linear relationships without requiring manual feature engineering. Additionally, RFR's robustness to noise and outliers contributed to its superior performance across different evaluation metrics. The findings show that predicted production has increased steadily over time. However, it is clear from comparing these predicted values to the actual values that the result is a decreasing tendency in the predicted values' alignment with the actual values. The data shown in Table 1, which shows the actual values next to their matching anticipated values from the Random Forest Regressor, further supports this fact.

Models Assessment

Examine the performance of three crop modelling approaches using the metrics R^2, RMSE and MAE. The degree to which crop data matches the productivity predicted the regression model is measured by the accuracy R^2. The difference between the expected and predicted productivity is measured using the root mean square error (RMSE), which

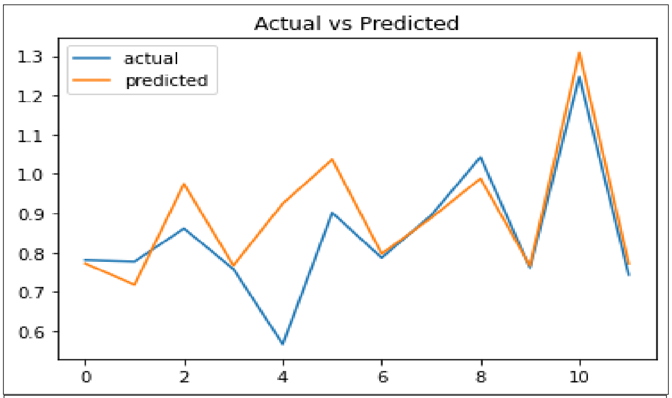

Fig. 4a. Model performance of Random Forest Regressor.

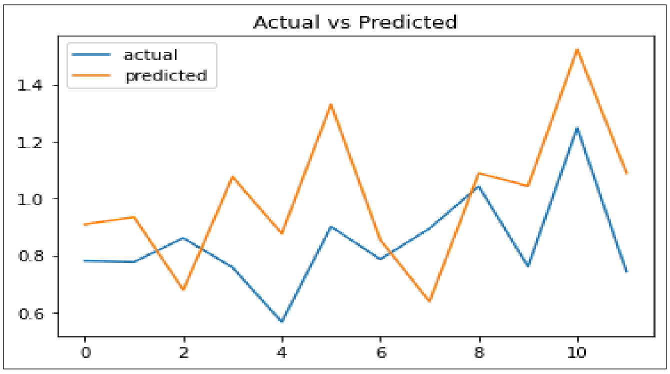

Fig. 4b. Model performance of SVR.

Table 1. Actual Vs Predicted values - Random Forest Regressor

Sl.No	Actual	Predicted
1	0.780954	0.772034
2	0.776840	0.717661
3	0.860915	0.974427
4	0.757657	0.766821
5	0.566297	0.924417
6	0.901062	1.036361
7	0.786260	0.796690
8	0.894184	0.887536
9	1.042586	0.988203
10	0.761119	0.765952
11	1.246833	1.308834
12	0.742804	0.771604

is based on Euclidean distance. The difference between the achieved productivity forecast and the predicted productivity is measured by the MAE metric. MAE averages the absolute difference throughout the full data set, represents the difference between the original value and the anticipated value. Prediction mistakes are referred to as RMSE, or residual mean square error. The relative magnitude of standard error (RMSE) and residuals are used to quantify the distance between the data points and the regression line. It demonstrates how closely the data are centered on the line of greatest fit. The degree to which the values fit one another regarding the initial values is gauged by the R-squared (Coefficient of determination). Values between 0 and 1 are used to denote the percentages. The value increases with model quality.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}| \qquad (6)$$

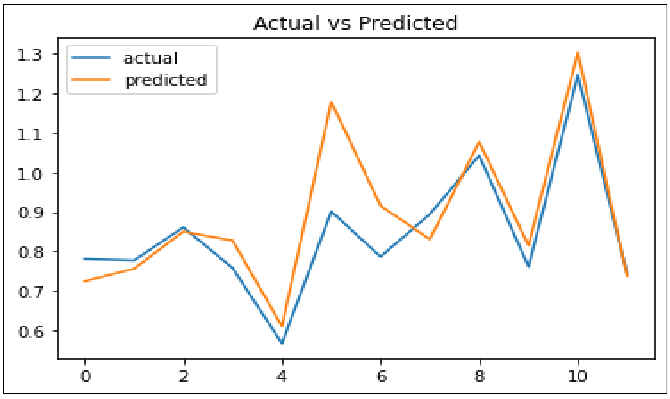

Fig. 4c. Model performance of LR.

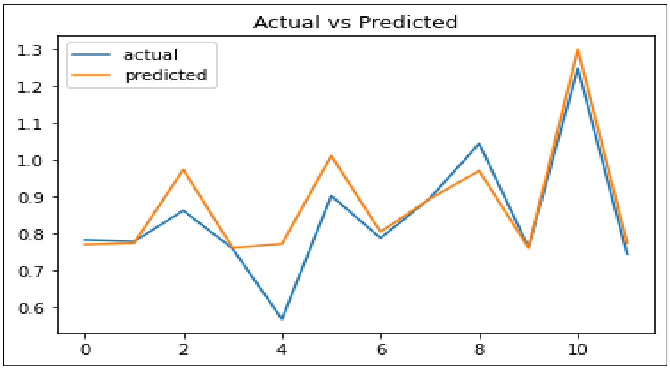

Fig. 4d. Model performance of DT.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \qquad (7)$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \overline{y})^2} \qquad (8)$$

An efficient model should have a high R^2 score and accuracy but a low RMSE value, which indicates the measure of error. Among all the models trained and tested, the Random Forest Regressor outperforms other models on the data set, including Linear Regression, Decision Tree Regressor and Support Vector Regression (SVR). The assessment findings for each crop's productivity prediction model are shown in Table 2. We found that RFR performs better with accuracy values of 95.46 % and has the greatest R^2 =0.94. As a result of this, the RF model outperforms LR, DTR and SVR in terms of crop data fit.

Additionally, we can observe that when compared to DTR and LR, RFR has the least root mean squared error. For RFR, DTR, LR and SVR, the corresponding RMSE that illustrates the difference between the anticipated and predicted productivity is 0.014, 0.062, 0.096 and 0.066. The results line up with the R^2 score and mean absolute percentage error throughout the small millets data set.

Table 2. Evaluation metrics of models trained

Models	MAE	RMSE	R-Square	Accuracy (%)
Linear Regression	0.0686 ± 0.005	0.09635 ± 0.007	0.85 ± 0.03	84.39 ± 2.5
Support Vector Regressor	0.2330 ± 0.015	0.06687 ± 0.006	0.67 ± 0.05	31.50 ± 3.0
Decision Tree Regressor	0.0513 ± 0.004	0.06269 ± 0.005	0.91 ± 0.02	91.08 ± 1.5
Random Forest Regressor	0.0709 ± 0.003	0.01424 ± 0.002	0.94 ± 0.01	95.46 ± 1.0

Conclusion

In summary, this research illustrates the shifting trends of small millet cultivation in Tamil Nadu. The agriculture sector's stakeholders and policymakers can benefit from these results. The primary outcome of the study underscores the notable advantages of applying this prediction technique in the field of agriculture to prevent agricultural losses. As everyone knows, food is essential to survival. Nevertheless, unexpected occurrences frequently provide farmers with several kinds of difficulties. To improve the productivity of crops through exact predictions, we use Machine Learning Models to determine the link between independent characteristics and their influence on small millet production. This study shows that the Random Forest Regressor performs better with accuracy values of 95.46 % and has the greatest R^2 =0.94. The accuracy of small millet production by comparing predicted outcomes with actual output. As a result, machine learning models provide farmers with long-term answers to their problems. The research's conclusions have major implications for the agricultural sector and present an acceptable path toward effective and sustainable farming methods. The adoption of machine learning-based advisory tools can provide real-time insights, helping farmers adapt to climate variability and mitigate risks. Policymakers can also integrate these predictive models into agricultural planning, enabling targeted interventions such as financial incentives, resource allocation and training programs. Extension services can use these findings to educate farmers on best practices for data-driven farming. Incorporating predictive analytics into their agricultural practices, farmers can enhance productivity, ensure sustainable land use and contribute to long-term food security. This study encompasses several areas that necessitate further investigation to enhance our understanding of agricultural productivity prediction through Machine Learning Models. In future analyses, we can employ a wider range of prediction techniques. Additionally, by acquiring diverse supplementary data, we can strive to predict small millets productivity more accurately.

Acknowledgements

The authors are thankful to the Tamil Nadu Agricultural University for supporting their research work.

Authors' contributions

VPNB contributed to the conceptualisation, framing the methodology, obtaining resources, carrying out the investigation, analysis and writing the original draft. KM helped in the statistical analysis, conducting the investigation, formulating the methodology, employing software and writing the original draft. BN has assisted with the analysis, investigation, methodological framework, software application and writing. SA, BV participated in writing, review and editing the manuscript. BV took part in framing the methodology, utilization of software, writing, review and editing. KS involved in writing, review and editing. RM and DR contributed to writing, review and editing the manuscript. All authors have read and agreed to the published version of the manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare

Ethical issues: None

References

1. Zhang L, Lei L, Yan D. Comparison of two regression models for predicting crop yield. In: IEEE International Symposium on Geoscience and Remote Sensing (IGARSS). IEEE; 2010. p. 1521-1524. https://doi.org/10.1109/IGARSS.2010.5652764
2. Sanchez AG, Solis JF, Bustamante WO. Attribute selection impact on linear and nonlinear regression models for crop yield prediction. Sci World J. 2014;2014:1-10. https://doi.org/10.1155/2014/509429
3. Corrales DC, Corrales JC, Figueroa-Casas A. Towards detecting crop diseases and pest by supervised learning. Ing Univ J. 2015;19(1):207-88. https://doi.org/10.11144/Javeriana.iyu19-1.tdcd
4. Luminto H. Weather analysis to predict rice cultivation time using multiple linear regression to escalate farmer's exchange rate. In: International Conference on Advanced Informatics, Concepts, Theory and Applications (ICAICTA). Denpasar, Indonesia; 2017. p. 16-18. https://doi.org/10.1109/ICAICTA.2017.8090974
5. Harimurti R, Yamasari Y, Asto B. Predicting student's psychomotor domain on the vocational senior high school using linear regression. In: International Conference on Information and Communications Technology (ICOIACT). IEEE; 2018. p. 448-453. https://doi.org/10.1109/ICOIACT.2018.8350768
6. Wei D, Xing M, Zhang J, Zhang C, Cao H. Applied research of multiple linear regression in the information quantification of Chinese medicine bone-setting manipulation. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2018. p. 1912-1916. https://doi.org/10.1109/BIBM.2018.8621086
7. Gopalakrishnan T, Choudhary R, Prasad S. Prediction of sales value in online shopping using linear regression. In: International Conference on Computing Communication and Automation (ICCCA). IEEE; 2018. p. 1-6. https://doi.org/10.1109/CCAA.2018.8777620
8. Vats V, Kumar A, Kumar N. Efficient crop yield prediction using machine learning algorithms. Int Res J Eng Technol. 2018;5(6):1-16.
9. Lim HI. A linear regression approach to modelling software characteristics for classifying similar software. In: IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC). IEEE; 2019. p. 942-943. https://doi.org/10.1109/COMPSAC.2019.00152
10. Sakthivel V, Kumaravel K, Suresh S, Prakash P, Suresh Kumar B. Crop yield optimization and prediction by machine learning algorithms. Int J Sci Res Eng Manag. 2022;6(6).
11. Rawat L, Karnatak AK, Bisht TS, Kukreti A. Minor millets: profile and ethnobotanical scenario. Millets and millet technology. 2021:51-80. https://doi.org/10.1007/978-981-16-0676-2_3
12. Shah P, Joshi R, Tripathy NP, Mehta N. What are the drivers of millet-based food consumption in India? A Theory of Consumption Values (TCV) perspective. Journal of International Food & Agribusiness Marketing. 2024:31:1-24. https://doi.org/10.1080/08974438.2024.2398245
13. Mehmood S, Iraqui S, Mohan M, Ahlawat YK, Sharma N, Sharma AJ, et al. Millet products in the market: Preparation and commercialization. 2024.
14. Narasimma VP, Muthusamy K, Natarajan B, Vasudevan B, Appavu S, Marimuthu R. Evaluation of machine learning models

in comparison to forecast yield for small millets. FEB-Fresenius Environmental Bulletin. 2024:1558.

15. Kuradusenge M, Hitimana E, Hanyurwimfura D, Rukundo P, Mtonga K, Mukasine A, et al. Crop yield prediction using machine learning models: case of Irish potato and maize. Agriculture. 2023:13:225. https://doi.org/10.3390/agriculture13010225

16. Venkatesa Palanichamy N, Kalpana M, Suresh A, Balamurugan V. Enhancing small millets productivity prediction through linear regression model using machine learning techniques. sensitizing the millet farming, consumption and nutritional security. 2024:62-66.

17. Palanichamy V, Muthusamy K, Natarajan B, Vasudevan B, Appavu S, Mahalingam SK. A comparison study of machine learning models to predict the bajra (pearl millet) yield prediction in Tamil Nadu. FEB-Fresenius Environmental Bulletin. 2024:1598.

18. Hu T, Zhang X, Khanal S, Wilson R, Leng G, Toman EM, et al. Climate change impacts on crop yields: A review of empirical findings, statistical crop models and machine learning methods. Environmental Modelling & Software. 2024. 19:106119. https://doi.org/10.1016/j.envsoft.2024.106119

19. Pant J, Pant RP, Singh MK, Singh DP, Pant H. Analysis of agricultural crop yield prediction using statistical techniques of machine learning. Materials Today: Proceedings. 2021;46:10922-6. https://doi.org/10.1016/j.matpr.2021.01.948

20. Ansarifar J, Wang L, Archontoulis SV. An interaction regression model for crop yield prediction. Scientific reports. 2021;11(1):17754. https://doi.org/10.1038/s41598-021-97221-7

21. Leng G, Hall JW. Predicting spatial and temporal variability in crop yields: an inter-comparison of machine learning, regression and process-based models. Environmental research letters. 2020. 20;15(4):044027. https://doi.org/10.1088/1748-9326/ab7b24