ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 17, Issue, 07, pp.33855-33861, July, 2025
DOI: https://doi.org/10.24941/ijcr.49256.07.2025

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

A COMPARATIVE STUDY OF SUPPORT VECTOR MACHINES AND OTHER MACHINE LEARNING ALGORITHMS FOR CROP YIELD PREDICTION

Pooja[1], Ravi R Saxena[1*], Sravan Kumar[1], Ritu R Saxena[2], Shilpi Verma[3] and Roopshikha Agrawal[4]

[1]Department of Agricultural Statistics and Social Science; [2]Department of Genetics and Plant Breeding; [3]Department of Foods and Nutrition, Krishi Vigyan Kendra, Neemuch; [4] Department of Botany, Govt. J Y Chhattisgarh College Raipur

ARTICLE INFO

Article History:
Received 09th April, 2025
Received in revised form
21st May, 2025
Accepted 19th June, 2025
Published online 30th July, 2025

Keywords:
SVM, LASSO, RNN, Agriculture, Crop Yield, Prediction.

***Corresponding author:**
Ravi R Saxena

ABSTRACT

The growing population has raised concerns about food security due to limited agricultural resources. Technological advancements in agriculture have improved crop management. Accurately predicting crop yield is vital for ensuring food security and informing agricultural policy decisions. With the increasing availability of large datasets and advancements in machine learning (ML) algorithms, this paper explored the application of ML algorithms for crop yield prediction. This research revealed that a Support Vector Machine (SVM) model outperforms other ML algorithms like LASSO and RNN, achieving a high prediction accuracy of 90% and the lowest RMSE value of 0.15 with an MAE of 0.107. The robust SVM model can handle complex relationships between input features and crop yield. These findings have significant implications for developing accurate crop yield prediction systems, which can inform agricultural decision-making and contribute to sustainable practices.

Citation: Pooja, Ravi R Saxena, Sravan Kumar, Ritu R Saxena, Shilpi Verma and Roopshikha Agrawal. 2025. "A Comparative Study of Support Vector Machines and Other Machine Learning Algorithms for Crop Yield Prediction". *International Journal of Current Research, 17, (07), 33855-33861.*

INTRODUCTION

Accurately predicting crop yield is essential for precision agriculture as it allows farmers to make informed decisions about planting, irrigation, fertilization, and harvest timing. Machine learning techniques are increasingly popular for crop production prediction due to the availability of massive datasets. Support Vector Machines (SVMs) have gained attention for their resilience and ability to manage complex interactions between variables (Vapnik, 1995). Various machine learning techniques, such as Neural Networks and LASSO, have been used to estimate crop yields in recent years. However, SVMs have shown better performance in several studies due to their ability to handle noisy datasets, high-dimensional data, and non-linear correlations (Kumar *et al.*, 2019; Li *et al.*, 2017; Jiang *et al.*, 2018). Despite these benefits, there is a lack of comprehensive analysis contrasting SVMs with alternative machine-learning methods for predicting agricultural yield. As demonstrated in earlier research by Mishra *et al.* (2021) and Gupta *et al.* (2022), the study compares Support Vector Machines (SVM) with other machine-learning methods for wheat crop yield prediction using an extensive dataset of crop and weather information. Through performance evaluation, the study aims to identify the strengths and weaknesses of SVMs in agricultural yield prediction and provide guidance for precision farming. The findings will benefit farmers and policymakers by aiding in the creation of more accurate crop yield forecast models.

Research Questions

- In terms of agricultural yield prediction, how well do SVMs perform compared to other machine learning algorithms?
- What are the main variables affecting SVM performance in predicting agricultural yield?
- Can SVMs be a trustworthy tool in precision agriculture for predicting crop yields?

This work thoroughly explains the effectiveness of SVMs for crop production prediction and their possible uses in precision agriculture by addressing these research objectives.

MATERIALS AND METHODS

Study Area and Data set: This study focuses on the Raipur district in the Chhattisgarh Central Plain, India, which is characterized by diverse agro-climatic conditions and is a significant wheat-producing region. Geographically, Raipur is located between 21.3691° Nlatitude and81.7787° Elongitude. The dataset includes historical weather data (rainfall, min & max temperature, relative humidity, and wind speed) and wheat yield records. The study made use of a dataset that was sourced from multiple websites between 1990 and 2023.These sources included the Directorate of Economics and Statistics (DES), ICRISAT-District level data, and the meteorological website https://power.larc.nasa.gov). From seven districts, agricultural data on the wheat crop was gathered.

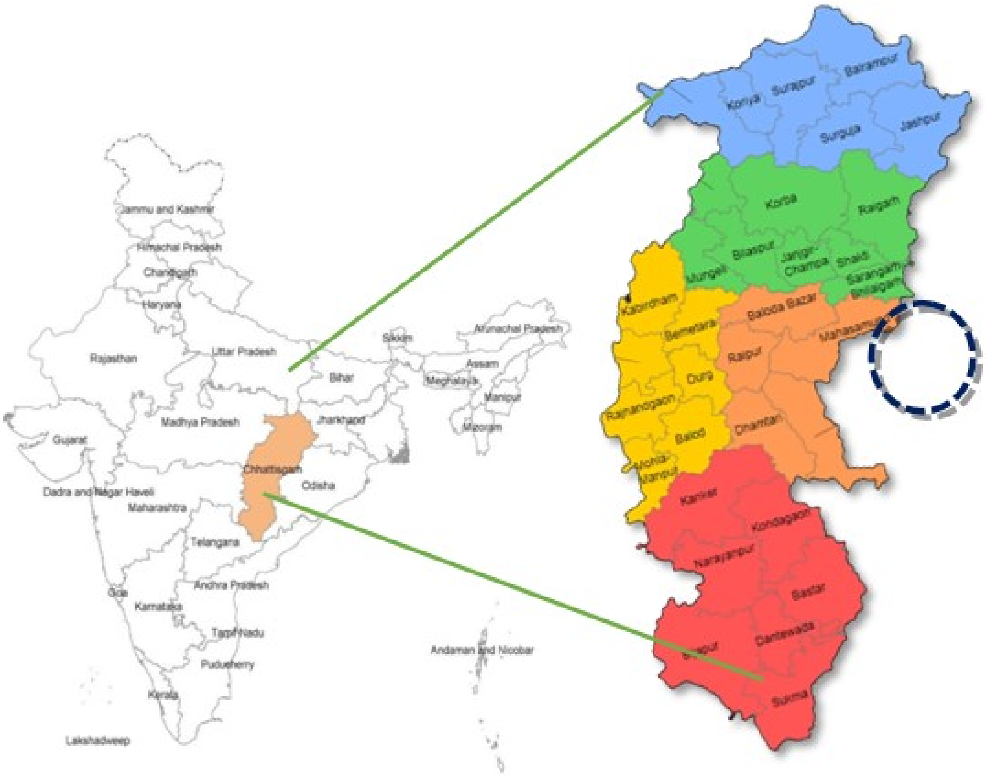

Figure 1. Spatial information about the districts located within the central plains of Chhattisgarh

Methodology Adopted: To estimate agricultural productivity, this study thoroughly compared the use of Support Vector Machines (SVMs) with various machine learning techniques, including LASSO and Recurrent Neural Networks (RNN). A collection of meteorological and agricultural parameters was used in this investigation. The dataset was carefully split into two groups: one for testing (20%) and one for training (80%). Specifically, the SVM model was rigorously trained with the RBF kernel, and its parameters were methodically improved using a grid search strategy. In a similar vein, the RNN and LASSO models underwent extensive training on the same dataset, and the appropriate hyper parameters were carefully adjusted by a thorough grid search procedure. The performance of each model was carefully assessed using fundamental metrics including the coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE). The comparison analysis results illustrated the benefits and drawbacks of using SVMs to forecast agricultural yield. In that order, Python's Numpy, Pandas, sci-kit-learn, and Matplotlib packages were used to create these models. The grid search method was applied to optimize each model's hyper parameters. This research attempts to thoroughly grasp the effectiveness of SVMs for crop yield prediction and its possible uses in precision agriculture by contrasting their performance with that of other machine learning methods..

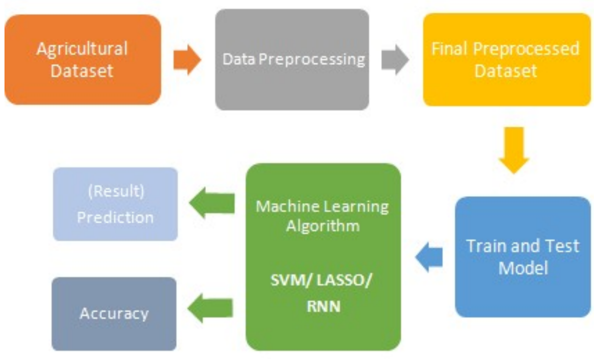

Figure 2. Data collection and processing methodology

Model

Support Vector Machine: A potent supervised machine learning method for regression and classification issues is called Support Vector Machine (SVM). It works well for high-dimensional pattern recognition, nonlinearity, and limited sample sizes. Classifying a dataset into distinct classes using the maximum marginal hyperplane (MMH) is the primary goal of support vector machines (SVM).The basic SVM model is shown below:

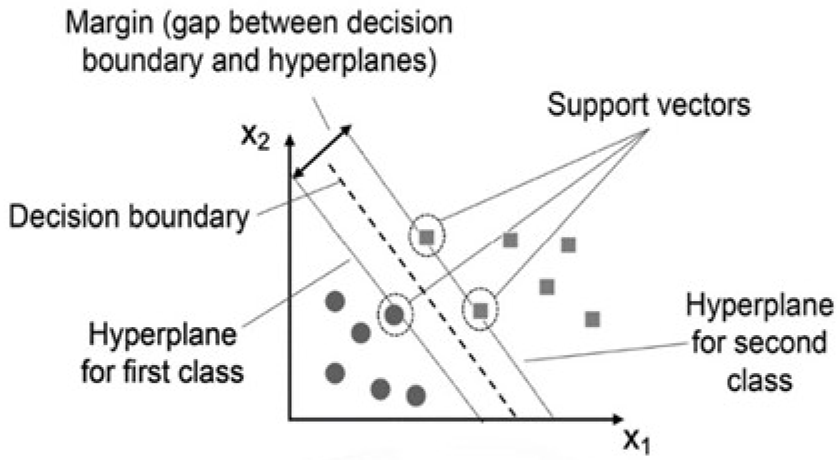

Figure 3. Basic SVM model

The foundation of SVM is determining the best hyperplane to efficiently divide data points in a classification problem or the least error in a regression task. The main goal is to maximize the margin, or the separation between the nearest data points—also referred to as support vectors—and the hyperplane.

Least Absolute Shrinkage and Selection Operator (LASSO): It is a kind of linear regression that reduces the magnitude of the coefficients and forces the model to choose the most significant features by including a penalty term in the cost function (James et al., 2013). Using L1 regularization to lessen over fitting, is helpful for high-dimensional and correlated feature datasets (Tibshirani, 1996).

Mathematical Formulation:

The LASSO algorithm can be formulated as follows:

$$\sum_{i=1}^{n} \quad (y_i - \Sigma_j x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \qquad (1)$$

Where, λ is a non-negative tuning parameter that controls the strength of the regularization. The larger the value of λ, the greater the penalty and the more coefficients are shrunk toward zero.

Recurrent Nural Network (RNN): The term "recurrent" refers to the iterative nature of recurrent neural networks (RNNs), in which the result depends on previously performed calculations that are maintained in memory. An extension of artificial neural networks (ANNs), RNNs have characteristics that retain calculations and use them as input. Data can flow in both directions within RNNs, which sets them apart from feed forward neural networks (FFNNs). However, when learning extended sequences, RNNs face disappearing and exploding gradient problems. Variants such as Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) have been created to solve these issues and provide better performance in capturing long-term dependencies (Hochreiter & Schmidhuber, 1997). LSTM has shown to be a successful remedy among them.

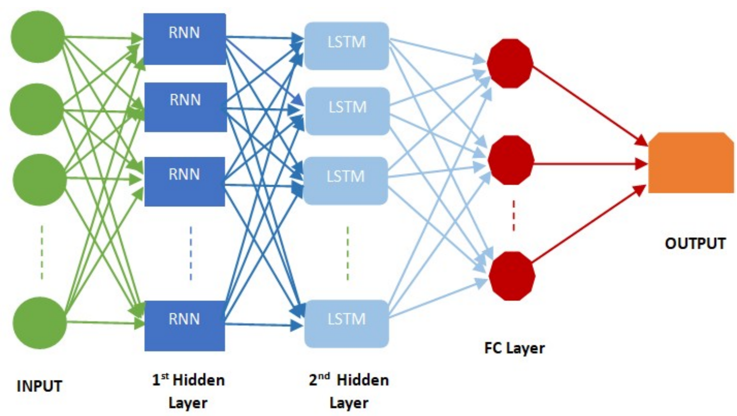

Figure 4. Hybrid RNN-LSTM base model architecture

The hybrid neural network design (Figure 4) combines RNN and LSTM networks to process sequential input. Where, Input Layer: Represents the sequence data fed into the network.

- 1st Hidden Layer: Processes the sequence using RNNs to capture dependencies between elements.
- 2nd Hidden Layer: Further processes the data using LSTMs to manage long-term dependencies.
- Fully Connected (FC) Layer: Combines outputs from the LSTM layer to prepare for the final prediction.
- Output Layer: Produces the final prediction or classification.

Methodology Implementation: Before entering the implementation stage, it is essential to carefully load various existing datasets, which may seem complex initially. Subsequently, the necessary libraries and packages are imported to facilitate the pre-processing of raw data, which is then categorized into test and trained data. Following this, a model is constructed, and AI algorithms are developed. This process aims to identify the best crop to grow on specific farmland. The methodology described in Figure 2 outlines the approach used in the study. Given that the data was collected from various sources, the primary objective was to organize it for input into Python for analysis. As a result, a CSV file was created with 10 columns and 33 rows. These 10 columns represent the wheat crop and various weather parameters. The rows correspond to the total years (33) chosen for the study. The CSV file contains information on District, Year, Area (1000ha), Production (1000t), Yield (t/ha), Relative Humidity(mm), Maximum Temperature(C), Minimum Temperature(C), Rainfall (in mm), and Windspeed (m/s).

Table 1 presents the sample dataset for several years, encompassing all parameters employed in this study. Mean values for all weather parameters (excluding Rainfall) were computed, and a summary of monthly rainfall during the crop period was compiled. Table 2 summarises the datasets for the wheat crop.

Table 1. Sample Dataset for several years, including all parameters

District	Year	Area	Production	Yield	RH	Tmax	Tmin	RF	WS
Raipur	1990-1991	17.0	18.9	1.11	44.94	38.23	15.05	8.9	1.77
Raipur	1991-1992	15.1	13.7	0.91	38.68	38.78	14.27	4.8	1.71
Raipur	1992-1993	16.7	16.5	0.99	35.34	35.11	14.81	7.37	1.84
Raipur	1993-1994	18.9	21.8	1.15	41.04	38.13	13.97	15.8	1.66
Raipur	1994-1995	18.4	22.4	1.22	46.49	37.41	13.91	7.3	1.71

Table 2. Dataset Description (Wheat Crop)

Parameters	**Area**	**Production**	**Yield**	**RH**	**T_{max}**	**T_{min}**	**RF**	**WS**
Count	33	33	33	33	33	33	33	33
Mean	11.99	15.63	1.41	45.69	37.76	14.39	11.28	1.65
Std	4.39	4.60	0.38	6.22	1.18	0.82	7.95	0.10
Min	3.56	6.78	0.91	34.13	34.47	12.13	1.06	1.40
25%	9.76	13.08	1.112	41.92	37.29	13.89	6.23	1.58
50%	12.67	16.50	1.28	45.05	37.97	14.35	8.90	1.66
75%	15.1	18.90	1.72	47.93	38.71	14.77	14.39	1.72
max	18.90	23.66	2.24	64.46	39.62	15.97	36.76	1.84

Model accuracy evaluation: The coefficient of determination (R^2) was utilized in this study to assess how well the explanatory elements in the equation fit the crop production. The accuracy of crop production estimation was evaluated using the root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).
The formulas are:

$$\boldsymbol{R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N} y_i^2}} \tag{2}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{4}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{(|y_i-\hat{y}_i|)}{y_i} \times 100\% \tag{5}$$

Where N stands for the number of data, y_i for the statistical value and $\hat{y}_i$ for the predicted model value. The larger R^2 and the lower the RMSE, MAE, and MAPE are, the better the model's performance is.

EXPERIMENTS AND RESULTS: Several experiments were carried out in Python using the Anaconda platform to evaluate the efficacy of the SVM model. The results obtained from the SVM model were subsequently compared with those derived from LASSO and RNN with LSTM. Table 3 outlines the results from all the models are discussed in the following sections.

Performance Metrics Overview: Other machine learning techniques, like LASSO and Recurrent Neural Networks (RNNs), were compared to see how well Support Vector Machines (SVM) predicted crop yields. Root Mean Squared Error (RMSE), R^2 score and Mean Absolute Error (MAE) were among the important metrics used in the evaluation (Table 3).

Table 3. Performance of the model with all parameters set

MODELS	R^2 SCORE	RMSE	MAE
SVM	0.9703	0.053	0.037
LASSO	0.9109	0.105	0.088
RNN with LSTM	0.8657	0.128	0.109

The results reveal that SVM achieved the lowest RMSE of 0.053, outperforming the other algorithms. Following SVM, LASSO obtained an RMSE of 0.088, signifying a notable improvement over RNN with LSTM, which recorded an RMSE of 0.109. The score for SVM was 0.97, demonstrating a strong correlation between predicted and actual crop yields, comparable to LASSO's and RNN with LSTM's scores of 0.91 and 0.86.

Support Vector Machine (SVM) Performance: The SVM model has displayed exceptional predictive accuracy, achieving a high R^2 score of 97%. This shows that the model can effectively explain around 97% of the variation in crop yields, demonstrating its strong explanatory capability. Furthermore, the low mean absolute error (MAE) of 0.037 for SVM suggests a minimal average deviation of predicted values from actual yields, highlighting the model's precision in forecasting crop productivity. These compelling findings emphasize the appropriateness of using SVM for crop yield prediction, particularly in regions like the Chhattisgarh Central Plain, where intricate weather patterns significantly impact crop yields. Moreover, SVM has exhibited consistent performance across cross-validation folds, demonstrating its ability to mitigate over fitting and ensure reliable performance across diverse data splits. Conversely, the Long Short-Term Memory (LSTM) in the Recurrent Neural Network (RNN) tended to over fit, with comparatively higher error values, highlighting potential challenges in its application in this context.

Comparative Analysis with Other Algorithms: The dataset has been split into 98% training and 2% testing sets, using a random state 42 for all three models. After examining the outcomes for every model separately, the following findings were noted: Compared to other models, the Support Vector Machine Algorithm has the greatest R^2 score, almost 97% when all parameters are included (figure 5). As a result, it has superior accuracy. All parameters have an RMSE of 5.3% and MAE of 3.7%. Figure 6 shows that the R^2 score for the Lasso Regression technique is 91.1%. The MAE is 8.8% while the RMSE is 10.5%and R^2 is 86.6%, the RMSE is 12.8%, and the MAE is 10.9% in the case of the Recurrent Neural Network with Long Short-Term Memory, or RNN with LSTM algorithms (figure 7).

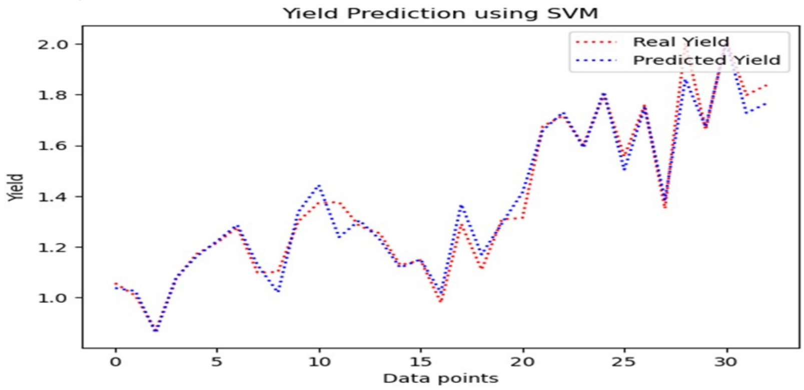

Figure 5. Model Performance of Support Vector Machine

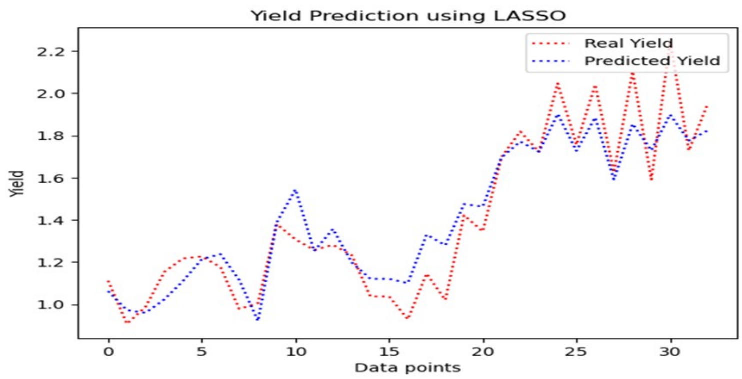

Figure 6. Model Performance LASSO

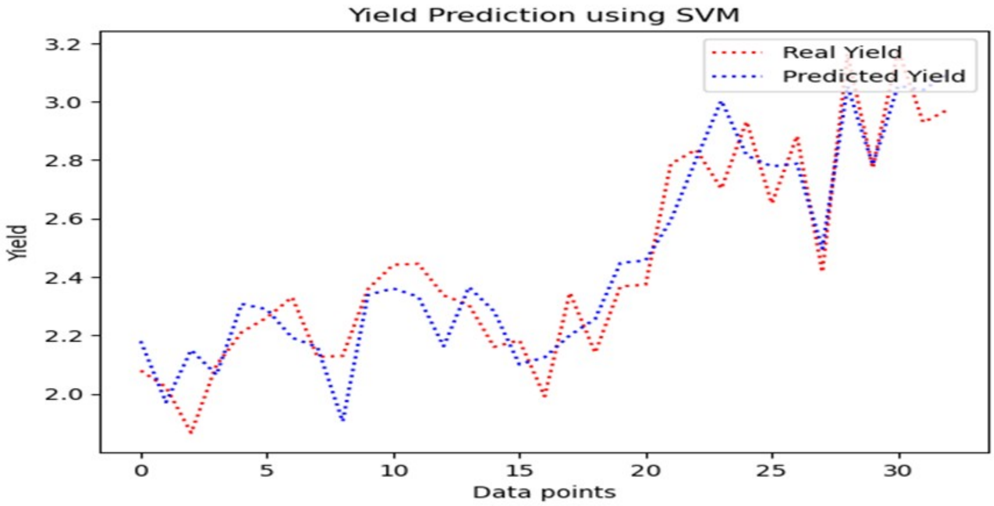

Figure 7. Model Performance RNN with LSTM

The line graphs (Fig. 5, 6, 7) indicate that the model effectively captures relationships but struggles to predict yield in some instances accurately. The SVM model (Fig. 5) performs well overall, but there is potential for improvement through exploring additional tuning techniques or incorporating more diverse features.

DISCUSSION

To compare Support Vector Machines (SVM) performance with that of other well-known machine learning algorithms, including LASSO Regression and Recurrent Neural Networks with Long Short-Term Memory (RNN with LSTM), the study carefully evaluated SVM's predictive power for agricultural yields. SVM performed better than the other models in all evaluation measures, as evidenced by its remarkable R^2 score of 97%, which indicated unmatched prediction accuracy, according to the study. Furthermore, the exceptionally low values of the MAE (3.7%) and RMSE (5.3%) demonstrated the remarkable resilience of SVM in reducing prediction errors, especially while managing intricate, non-linear interactions in agricultural data. In stark contrast, LASSO Regression achieved a respectable R^2 score of 91%, but its linear nature severely limited its ability to model the intricate, multi-dimensional factors influencing crop yields, resulting in significantly higher error rates compared to SVM. Similarly, the RNN with the LSTM model, while effective for time-series analysis, exhibited the weakest performance, attaining an R^2 score of 86.6% with the highest RMSE and MAE values. This suggests that RNN with LSTM may not be the best choice for this type of predictive task, especially with datasets lacking significant temporal structures. The findings overwhelmingly emphasize the significance of selecting machine learning models based on the dataset's nature. SVM demonstrated unparalleled efficacy in predicting crop yields owing to its unparalleled capability to capture complex interactions between environmental, soil, and management factors. On the contrary, simpler models like Lasso Regression are more suitable for datasets with linear relationships, whereas models such as RNN with LSTM are better suited for time-series data that exhibit strong temporal patterns.

CONCLUSION

Based on the study's findings, it is crucial to consider SVMS as the primary option for predicting crop yield in agricultural applications. Furthermore, it is essential to conduct additional research to investigate the use of ensemble methods and the integration of SVMs with other machine learning algorithms to further improve the accuracy of crop yield prediction.

As a result, this study highlights the superiority of SVMs in predicting crop yield, thereby establishing a solid foundation for their use in precision agriculture. To further improve the predictive power of machine learning models in agriculture, future research efforts should surely include hybrid modelling approaches, hyper parameter tuning, and the incorporation of new temporal data.

REFERENCES

Agrawal R, Jain RC and Mehta SC. 2001. Yield forecast based on weather variables and agricultural input on agro-climatic zone basis. *The Indian Journal of Agricultural Sciences.*71(7): 487-90.

Agrawal, R and Mehta, S. 2007. Weather-based forecasting of crop yields, Pests and Diseases -IASRI Models. *Journal of the Indian Society of Agricultural Statistics.* 61: 255-263.

Amaratunga V, Wickramasinghe L, Perera A, Jayasinghe J, Rathnayake U. 2020. Artificial Neural Network to estimate the Paddy Yield Prediction using climatic data. *Mathematical Problems in Engineering* 1: 8627824.

Cristianini N, Ricci E. 2008. Support Vector Machines. Encyclopedia of Algorithms. *Springer*: 928-932.

Dahikar, S S and Rode S V. 2014. Agricultural crop yield prediction using artificial neural network approach. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering.*2: 683–686.

Deo R C, Kisi O and Singh V P. 2017. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. *Atmospheric Research.* 184: 149-175.

Jha G K and Sinha K. 2014. Time-delay neural networks for time series prediction: An application to the monthly wholesale price of oilseeds in India. *Neural Computing and Applications.* 24: 563-571.

Kumar, S., Attri S D and Singh, K K. 2019. Comparison of Lasso and stepwise regression technique for wheat yield prediction. *J. Agrometeorol.* 21: 188–192.

Lu C J, Lee T S and Chiu C C. 2009. Financial time series forecasting using independent component analysis and support vector machine. *Decision Support Systems.* 47(2): 115-125.

Nishant PS, Venkat PS, Avinash BL, Jabber B. 2020. Crop Yield Prediction based on Indian Agriculture using Machine Learning. In: Proceedings of the 2020 *International Conference for Emerging Technology* (INCET), Belgaum, India: pp 5–8.

Paudel D, Boogaard H, de Wit A, Janssen S, Osinga S, Pylianidis C & Athanasiadis I N. 2021. Machine learning for large-scale crop yield forecasting. *Agricultural Systems*187: 103016.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research.*12: 2825–2830.

Mishra S, Mishra D and Santra G. 2016. Applications of machine learning techniques in agricultural crop production: a review paper. *Indian Journal of Science and Technology.* 9(38): 1-14.

Sunaina N. 2020. Analysis of Crop Prediction by using Machine Learning Algorithms. *International Research Journal of Engineering and Technology* (IRJET) 7(03).

****