Crop Yield Prediction Past and Future: HarverstInsight

Shafaq, Khan*
University of Windsor, ON

Virag Savaliya
University of Windsor, ON

Dhruvi Kharadi
University of Windsor, ON

Madhavan Vaidya
University of Windsor, ON

Meghna Singh
University of Windsor, ON

1. ABSTRACT

Agriculture stands as a linchpin for global economies and sustenance, heavily contingent on weather patterns, particularly rainfall, impacting roughly 80% of agricultural lands. Accurately forecasting crop yields, notably for staple grains like wheat, becomes paramount for effective planning and decision-making. While conventional methodologies exist, Gradient Boosting Regressor algorithms offer promising avenues for achieving heightened precision. Nonetheless, challenges persist due to the intricate dynamics of agricultural systems and the inherent uncertainties in weather patterns. Consequently, further research and development are imperative to enhance prediction accuracy. This entails leveraging innovative techniques and amalgamating diverse datasets to create robust forecasting models. The significance of accurate crop yield forecasting cannot be overstated; it underpins food security efforts, guides agricultural interventions, and fosters sustainable practices. By embracing these advancements, stakeholders can make more informed decisions, mitigate risks, and cultivate a resilient and productive agricultural sector capable of meeting the demands of a growing global population.

2. INTRODUCTION

Agriculture is not only a vital sector of the global economy but also an essential component of food security and sustainable development [1]. It relies heavily on weather conditions, with approximately 80% of agricultural lands worldwide being rainfed. Favourable weather patterns are crucial for optimal productivity, but the sector faces challenges due to fluctuations in rainfall and other meteorological variables [2].

The accurate forecasting of crop yields is paramount for agricultural planning and decision-making [3]. This involves predicting how much food crops, such as wheat, will produce under various conditions. Wheat is a major food crop, contributing nearly 40% of the global food supply [4]. Therefore, developing reliable predictive models for estimating wheat and other crop yields is critical for ensuring food security and guiding agricultural interventions [5].

Traditionally, methods like field surveys, crop models, and statistical analyses have been used for yield prediction [5]. Among these approaches, Decision Tree algorithm have shown promise, often achieving high levels of accuracy. For example, Decision Tree Algorithm have been reported to have an RMSE value of approximately 9.6 , indicating a close match between predicted and actual yields [6]. However, despite advancements in predictive analytics, accurately forecasting crop yields remains a challenge due to the complexity of agricultural systems and the uncertainties inherent in weather patterns [7].

In conclusion, accurate crop yield forecasting is essential for ensuring food security and sustainable agriculture. While traditional methods have provided valuable insights, further research and development are necessary to enhance prediction accuracy and address the challenges posed by environmental uncertainties. By leveraging innovative approaches and integrating diverse datasets, agricultural stakeholders can make more informed decisions and mitigate risks, ultimately contributing to a more resilient and productive agricultural sector.

2.1 SCOPE OF STUDY

For our study, we are harnessing an extensive Indian dataset encompassing a wide array of agricultural parameters, including crop types, crop years, seasons, areas under cultivation, production figures, annual rainfall patterns, fertilizer and pesticide usage, and crop yields. This dataset comprises a robust 19,691 rows of data, offering a comprehensive perspective on various facets of agricultural activity.

Our decision to utilize an Indian dataset stems from India's status as one of the primary contributors to global agriculture. With its vast and diverse agro-climatic zones and extensive farming practices, India presents a rich tapestry of agricultural data ripe for analysis. This dataset encapsulates agriculture-related information from all Indian states, providing a holistic view of the nation's agricultural landscape.

Importantly, the dataset spans an extensive timeframe, covering data from 22 years, from 1997 to 2020. This longitudinal coverage enables us to analyse trends and patterns in agricultural activity over time, facilitating a deeper understanding of the dynamics at play within India's agricultural sector.

By delving into this dataset, we aim to unravel the intricate dynamics of crop yield variations across different regions and climatic conditions within India. The insights gleaned from this analysis will not only enhance our understanding of agricultural processes but also facilitate the development of robust predictive models for crop yield forecasting, thereby contributing to improved agricultural planning and decision-making.

Table 1.1: Feature Description

Crop	The name of crop cultivated
Crop Year	The year in which the crop was grown
Season	The specific cropping season (e.g. Kharif, Rabi)
State	The Indian state where the crop was cultivated
Area	The total land area (in hectares)
Production	The quantity of crop production (in metric tons)
Annual Rainfall	The annual rainfall received in the crop-growing region (in mm)
Fertilizer	The total amount of fertilizer used (in Kgs)
Pesticide	The total amount of pesticide used (in Kgs)
Yield	The calculated crop yield (production per unit area)

2.2 CROPPING SEASON IN INDIA

India's agricultural landscape is characterized by distinct cropping seasons that dictate the timing and types of crops grown across the country. The two primary cropping seasons are known as Kharif and Rabi.

The Kharif season, also known as the monsoon season, typically extends from June to October. During this period, crops like rice, maize, millets, and cotton are sown. The onset of the monsoon marks the beginning of Kharif cultivation, with crops relying on rainfall for irrigation.

Following the Kharif season, the Rabi season commences around October and lasts until March. Rabi crops, which include wheat, barley, oats, and mustard, are sown during this season. Unlike Kharif crops, Rabi crops are cultivated during the winter months and rely on irrigation from canals, wells, and other sources.

Apart from Kharif and Rabi seasons, there is also a third season known as the Zaid season. Zaid crops are grown between March and June and include vegetables like cucumbers, watermelons, and muskmelons. Zaid cultivation typically takes place during the summer months when temperatures are higher

The timing of cropping seasons in India is influenced by various factors, including monsoon patterns, temperature variations, and soil moisture levels. Understanding these seasonal dynamics is essential for farmers to plan their cropping patterns effectively and maximize agricultural productivity. Additionally, policymakers and agricultural researchers utilize knowledge of cropping seasons to design and implement agricultural policies and interventions tailored to specific seasons and regions[7].

2.3. CHALLENGES

A few of the challenges we faced during this project are:

1. Selecting a suitable dataset and pre-processing makes the research more desirable to get the best results.
2. Fluctuating error rate due to continuously changing environment in the districts.

2.4. MODELS AND METHODOLOGY

In our crop yield prediction research work, we employed three distinct machine learning models: Support Vector Regression (SVR), Gradient Boosting, and Decision Tree. Each model offers unique advantages and is applied to tackle specific aspects of the prediction task.

2.4.1. Support Vector Regression (SVR)

Support Vector Regression is a powerful regression technique that utilizes the principles of Support Vector Machines (SVM) for regression analysis. SVR aims to find the hyperplane that best fits the data points while maximizing the margin, thus minimizing the error between predicted and actual values. It is particularly effective in handling high-dimensional data and is known for its robustness against overfitting.

2.4.2. Gradient Boosting

Gradient Boosting is an ensemble learning technique that combines multiple weak learners, typically decision trees, to create a strong predictive model. It sequentially trains new models to correct the errors made by the previous ones, ultimately leading to improved prediction accuracy. Gradient Boosting is highly effective in capturing complex relationships in the data and is resilient to overfitting.

2.4.3. Decision Tree

Decision Tree is a simple, yet powerful algorithm used for both classification and regression tasks. It partitions the feature space into a hierarchical structure of decision nodes, where each node represents a decision based on a feature's value. Decision Trees are intuitive to interpret and can handle both numerical and categorical data. However, they are prone to overfitting, especially with complex datasets.

These three models were selected based on their suitability for the task at hand and their complementary strengths. SVR provides robust regression capabilities, while Gradient Boosting and Decision Tree offer powerful ensemble learning techniques. By employing a combination of these models, we aimed to leverage their respective advantages and improve the overall accuracy and reliability of our crop yield predictions.

2.5. Methodology

Data Preparation: We load the dataset containing information about crop yields and relevant features.

Splitting Data: The dataset is split into training and testing sets to evaluate the model's performance.

Preprocessing: Numerical features are standardized using StandardScaler, and categorical features are one-hot encoded.

Model Training: SVR model is trained using the preprocessed training data.

Model Evaluation: The trained model is evaluated using Root Mean Squared Error (RMSE) and R-squared (R2) score on both training and testing sets.

Prediction: Next-year yield prediction is performed using the trained model.

Performance Metrics: Mean Absolute Percentage Error (MAPE) is calculated to assess the prediction accuracy.

Visualization: Model evaluation metrics are visualized using bar plots.

3. MOTIVATION

Our project aims to revolutionize crop yield prediction by integrating advanced machine learning techniques with agricultural domain expertise. The inspiration behind these endeavours stems from a deep-seated recognition of the pivotal role agriculture plays in ensuring global food security and supporting the livelihoods of farmers worldwide.

The motivation behind embarking on this project to enhance crop yield prediction using advanced machine learning techniques is deeply rooted in the urgent need to address pressing challenges facing the agricultural sector.

Fig 3.1 illustrates the dependence of crop production on September rainfall in Gujarat, India, highlights the critical role of weather patterns in agricultural productivity [8]. This dependency underscores the vulnerability of farmers to environmental fluctuations and the importance of accurate crop yield predictions in mitigating risks and optimizing resource allocation.

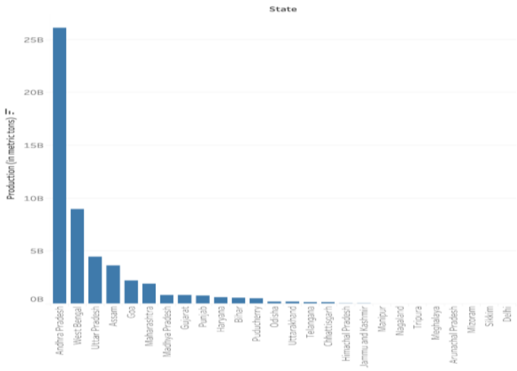

Fig. 3.1. Statewise Total Production [8].

Additionally, the concerning trend of declining GDP contribution from the agricultural sector in India underscores the economic significance of improving crop yield prediction [9]. As approximately 80 percent of farmers hail from rural areas, any downturn in crop production directly impacts their livelihoods and the broader agricultural industry. By enhancing crop yield prediction accuracy, we aim to bolster the economic resilience of farmers and contribute to sustainable agricultural development.

Moreover, the adverse effects of environmental changes on crop production underscore the urgency of proactive measures to mitigate risks and optimize agricultural practices [10]. Anticipating crop yield outcomes before harvest empowers policymakers and farmers to implement timely interventions and marketing strategies, thereby enhancing agricultural resilience and profitability. Our project endeavours to equip cultivators with actionable insights to make informed decisions and optimize crop yields, thereby fostering a more sustainable and resilient agricultural ecosystem.

In summary, our project is motivated by the imperative to address the challenges posed by environmental uncertainties, economic fluctuations, and the inherent complexities of agricultural systems. By leveraging advanced machine learning techniques, we seek to empower farmers, policymakers, and stakeholders with the tools and knowledge needed to navigate these challenges effectively and ensure a more sustainable and prosperous future for agriculture.

4. LITERATURE REVIEW

The yield of crops is influenced by a variety of factors, including weather conditions, soil quality, type of crop, fertilizer usage, and seed variety [11]. To obtain reasonable results in crop yield estimation, various models for crop simulation and yield estimation have been employed. Researchers have increasingly applied Deep Learning methods for crop yield estimations based on these factors.

In the article "Using machine learning for crop yield prediction in the past or the future" by Morales and Villalobos [12], it is stated that the use of Machine Learning (ML) in agronomy has seen exponential growth since the beginning of the century, encompassing data-driven predictions of crop yields from farm-level information on soil, climate, and management. However, the impact of data partitioning schemes on the actual performance of the models, particularly when they are constructed for yield forecast, remains less understood. In their study, the researchers examined the influence of the choice of predictive algorithm, data volume, and data partitioning strategies on predictive performance using a synthetic dataset from biophysical crop models. They simulated wheat and sunflower data using OilcropSun and Ceres-Wheat from DSSAT for the period of 2000 - 2020 across five areas in Spain.

The significance of their work lies in the algorithms they employed for predicting crop yields and the effects of data partitioning on predictive systems. They experimented with various machine learning models and different data partitioning schemes to assess the significance of data partitioning methods. The algorithms used by the researchers include Regularized linear model, Random Forest, and Artificial Neural Network (ANN) with 50% dropout. The data was partitioned in ascending order, meaning older data was used for training the models and newer data for testing. Their findings indicate that Random Forest, using the decision tree algorithm, outperformed all other algorithms with a Root Mean Squared Error (RMSE) of 35-38%, which is considered high. However, ANN performed the worst among all with an RMSE of 37-141%. Another conclusion is that ordered data partitioning is not recommended for yield prediction due to extrapolation challenges. The researchers emphasized the advantages of simpler models such as ANN and Random Forest over more complex models. They recommended using crop models for initial feasibility assessment before deploying actual data collection and complex models and advocated for considering cost-effectiveness when choosing methods for crop yield prediction.

Machine learning offers greater flexibility and provides faster predictions compared to simulation crop modelling [12]. Various machine and deep learning algorithms, such as Random Forest (RF), Support Vector Machines (SVM), and Neural Networks (NN), have been utilized for predicting crop yields under different environmental conditions. SVMs can provide robust predictions when datasets have multiple attributes, although their functionality may be limited by hardware constraints. In contrast, Random Forest algorithms have demonstrated high classification accuracy and superior predictions compared to multiple linear regression models. According to [13], simple linear and Generalized Linear Models (GLM) are constrained by restrictive assumptions with the normal distribution trained on the value of predictors with a constant variance. Typically, decision tree ensemble learning methods, such as the Random Forest Regressor and Gradient Boosting Regressor, exhibit higher accuracy. In China, ref. found that the highest estimation accuracy was achieved with the Random Forest Regression. In India, the decision tree yielded the highest accuracy (96%) for crop yield prediction [14]. The effectiveness of machine learning algorithms in prediction is evaluated based on their ability to reduce bias, variance, or both, thereby informing the recommendation of a particular algorithm. For example, if the goal is prediction with low error, then weighted ensemble models are chosen; if the objective is to detect the correct forecast direction, then a stacked LASSO regression is selected [12].

5. PROPOSED MODEL

The proposed model explores the application of machine learning techniques for predicting crop yield based on a dataset containing various features potentially influencing yield. This study employs three widely used regression models for prediction tasks: Support Vector Regression (SVR), Gradient Boosting Regressor, and Decision Tree Regressor.

5.1. Support Vector Regression (SVR)

SVR is a kernel-based method that aims to find a hyperplane in a high-dimensional feature space that maximizes the margin between data points and the hyperplane while minimizing the number of points violating the margin [15]. This approach enables SVR to handle non-linear relationships between features and the target variable (yield). The SVR model construction depends on a subset of training data, like the support vector classification model, where only training points beyond a certain boundary are considered [16]. The SVR is formulated as the minimization of a specific function, as depicted in Fig. 4, which illustrates different parameters used in the equations.

$$\frac{1}{2}||w||^2 + c\sum_{i=1}^{N} \xi_i + \xi_I^*$$

With constrains, when $y = wx + b$,

$$\begin{aligned} y_i - wx - b &\le \varepsilon + \xi_i \\ wx_i + b - y_i &\le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\ge 0 \end{aligned}$$

where ξ_i and ξ_i^* are the slack variables and $-\varepsilon$ and $+\varepsilon$ are the distance of hyper plane and the boundary line.

Fig 5.1. Support Vector Regressor Minimization [20]

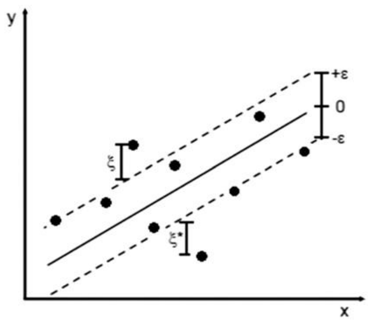

Fig 5.2. Parameter Used in SVR Minimization equation [20]

Applying Support Vector Regressor, the method included standard scaling and one hot encoding method. The diagram including that is given below:

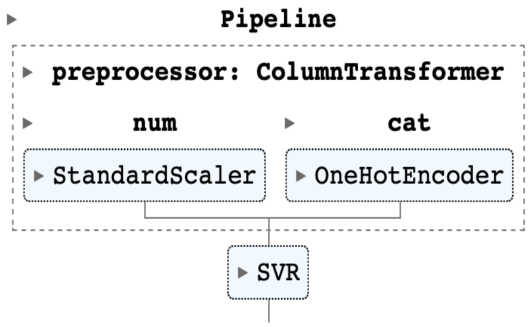

Fig 5.3. Pipeline for SVR

5.2. Gradient Boosting Regressor

This ensemble method builds a model by sequentially fitting weak learners (e.g., decision trees) in an iterative fashion to improve accuracy [17]. Each subsequent learner focuses on correcting the predictions from the previous models by minimizing residual errors. Gradient boosting is known for its flexibility and ability to handle complex relationships within the data.

1. Initialize model with a constant value:

$$F_0(x) = \underset{\gamma}{argmin} \sum_{i=1}^{n} L(y_i, \gamma)$$

2. for $m = 1$ *to* M:

2-1. Compute residuals $r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$ $for\ i = 1,\ldots, n$

2-2. Train regression tree with features x against r and create terminal node reasions R_{jm} for $j = 1,\ldots, J_m$

2-3. Compute $\gamma_{jm} = \underset{\gamma}{argmin} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$ $for\ j = 1,\ldots, J_m$

2-4. Update the model:

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jm} 1(x \in R_{jm})$$

Fig 5.4. Algorithm for Gradient Boosting Algorithm [18]

1. Feature Engineering: Incorporating temporal, soil, and agronomic features to capture long-term trends, soil fertility, and management practices' impact on yield.
2. Model Architecture: Utilizing ensemble techniques, hyperparameter tuning, and regularization to improve predictive performance and generalization.
3. Evaluation Metrics: Employing domain-specific metrics and economic indicators to accurately assess model efficacy.
4. Integration of External Data: Leveraging satellite imagery, remote sensing, and GIS data for finer resolution and spatial insights.
5. Experimental Setup: Conducting comprehensive experiments on diverse datasets with rigorous preprocessing and cross-validation.

5.3. Decision Tree Regressor

Decision trees are interpretable models that partition the data into subsets based on feature values, ultimately predicting the target variable based on the terminal leaf node reached [19]. They offer a clear understanding of feature importance and the decision-making process within the model.

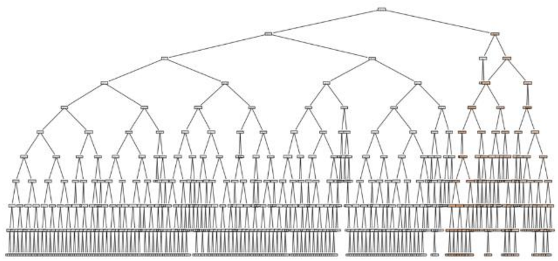

Fig 5.5. Decision Tree Regressor with Depth 10

1. **Feature Engineering**: Incorporating temporal, soil, and environmental features to capture dynamic factors influencing crop yield.
2. **Model Complexity Control**: Limiting tree depth and applying pruning techniques to improve generalization and interpretability.
3. **Visualization**: Visualizing decision tree structure to provide insights into the prediction process.
4. **Evaluation Metrics**: Assessing model performance using standard regression metrics like MSE and MAE.
5. **Experimental Setup**: Validating the model on diverse datasets with rigorous preprocessing and train-test split.

6. RESULTS

This study compared several Machine Learning algorithms to predict agricultural productivity in next seasons. Three main machine learning algorithms that outperformed others were used to increase accuracy and produce better prediction results. The Gradient Boost algorithm's assessment showed that, with a tolerance of 10, its average accuracy was 94.06%.

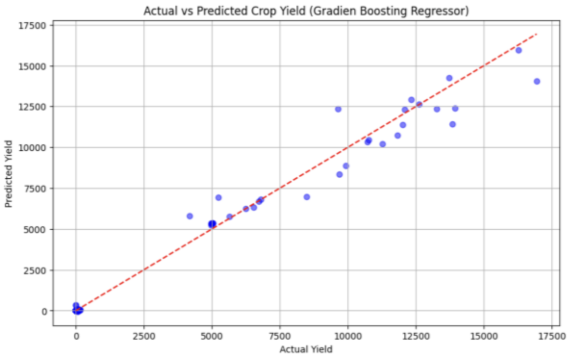

Fig 6.1. Actual vs Predicted Crop Yield using Gradient Boosting Regressor

The Gradient Boosting Regressor is used to compare the Actual and Predicted Crop Yield in Figure 6.1. The ideal value for yield prediction is shown by the red dashed line. According to our analysis, the Gradient Boosting Algorithm performs better than the others, with a mean absolute error of 9.35 and a mean squared error of 10284.27.

A depth of 10 was used for the Decision Tree Algorithm, which converted categorical input into numerical representation using one-hot encoding. With 24 random states, the dataset was divided between 80% training and 20% testing sets.

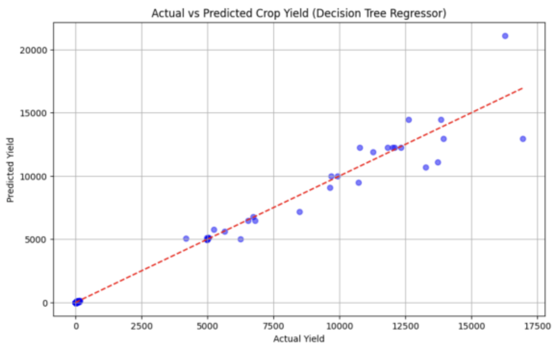

Fig 6.2. Actual vs Predicted Crop Yield using Decision Tree Regressor

The decision tree regression is used to compare the actual and predicted crop yields in Figure 6.2. The decision tree model with a max_depth of 10 and 42 random states produced the best results, with a Mean Squared Error of 16876.61 and a Mean Absolute Error of 8.41.

However, the model did not produce significant results when compared to the Support Vector Regressor (SVR) findings. Test RMSEs were 897.22 and 76.87, respectively, which were much higher than RMSEs for Decision Trees and Gradient Boosting Algorithms.

7. LIMITATIONS

7.1. MODELS

In our exploration of crop yield prediction, we employed decision tree, Support Vector Regressor (SVR), and Gradient Boosting algorithms. While decision trees yielded the best results, followed by Gradient Boosting, challenges arose primarily with Gradient Boosting. This algorithm exhibited prolonged training times, particularly with complex models, risking overfitting. Furthermore, Gradient Boosting is often regarded as a black-box model, diminishing interpretability. Despite its potential, the model's efficacy did not meet expectations, with a mean absolute error rate of 9.35. The algorithm's intricate nature necessitated higher computational resources, surpassing the available GPU processing capacity. Decision tree models, on the other hand, suffered from oversimplification, inadequately capturing complex interactions and susceptibility to overfitting. GPU limitations constrained the depth of decision trees, hindering predictive accuracy. Moreover, decision tree models proved sensitive to slight code variations, leading to disparate predictions.

SVR presented its own set of challenges, particularly in noisy data environments where performance degraded. Additionally, when the number of features surpassed the training data samples, SVR exhibited suboptimal performance. The algorithm's sensitivity to hyperparameters and limited scalability further compounded its limitations.

7.2. DATASET

Our research utilized an Indian dataset spanning from 1997 to 2019 for crop yield predictions. The unavailability of post-2019 data restricted our analysis, hindering comprehensive insights. Incorporating diverse crops and varied weather conditions would have enriched predictive capabilities. Attributes such as humidity, longitude, latitude, and months could have enhanced predictions, given India's diverse climatic conditions. Seasonal variations play a pivotal role in crop growth, making attributes like months crucial for accurate predictions.

7.3. HARDWARE

Critical hardware constraints, including CPU processing power and GPU availability, posed significant obstacles during our research. As the dataset size expanded, increased memory and processing power became imperative, exacerbating resource limitations. Although some algorithm implementations supported parallel processing, hardware lacking multiple cores or GPUs limited scalability and computational efficiency, impeding training and prediction tasks.

8. CONCLUSION

Our study on crop yield prediction using SVR, Gradient Boosting Regressor, and Decision Tree

Regressor highlighted the potential for accurate predictions in agricultural settings. The Gradient Boosting Regressor exhibited the lowest Mean Squared Error (MSE) and Mean Absolute Error (MAE), showcasing its effectiveness. The Decision Tree Regressor, while also performing well, offers interpretability advantages. Surprisingly, the SVR model achieved the lowest test MSE and MAE, demonstrating its accuracy despite computational complexity. These findings provide valuable insights for agricultural planning, enabling informed decisions to optimize crop production and yield. Future research could focus on integrating additional features, developing hybrid models, and improving model interpretability to further refine crop yield prediction models.

9. FUTURE WORK

Our research suggests several avenues for future work to enhance crop yield prediction models. Firstly, exploring additional features such as soil properties, satellite imagery, and weather data could improve prediction accuracy. Secondly, advanced machine learning techniques like deep learning models and ensemble methods could be investigated for their potential to further enhance predictions. Additionally, focusing on model explainability, uncertainty quantification, integration with decision support systems, and field validation could lead to more robust and reliable crop yield prediction models. By pursuing these directions, researchers can refine models to empower farmers with data-driven decisions, ultimately improving agricultural productivity and sustainability.

REFERENCES

[1] FAO, 2022. https://www.fao.org/home/en/

[2] Huiren T., Pengxin W., Kevin T., Dong H., Jingqi Z., Shuyu Z., Hongmei L., 2021. A deep learning framework under attention mechanism for wheat yield estimation using remotely sensed indices in the Guanzhong Plain, China. Int. J. Appl. Earth Obs.102,102375. https://doi.org/10.1016/j.jag.2021.102375.\

[3] Kuradusenge, M., Hitimana, E., anyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., Uwitonze, C., Ngabonziza, J., & Uwamahoro, A. (2023). Crop yield prediction using machine learning models: case of Irish potato and maize. Agriculture, 13(1),225. https://doi.org/10.3390/agriculture13010225

[4] Kuradusenge, M., Hitimana, E., anyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., Uwitonze, C., Ngabonziza, J., & Uwamahoro, A. (2023). Crop yield prediction using machine learning models: case of Irish potato and maize. Agriculture, 13(1),225. https://doi.org/10.3390/agriculture13010225

[5] Jiang, H., Hu, H., Zhong, R., Xu, J., Xu, J., Huang, J., Wang, S., Ying, Y., Lin, T., 2020. A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level. Glob. Chang. Biol. 26 (3), 1754–1766. https://doi.org/10.1111/gcb.14885

[6] Molden, D. Vithanage, M. de Fraiture, C. Faures, J.M. Gordon, L. Molle, F. Peden. Water Availability and Its Use in Agriculture. Treatise Water Sci. 2011, 4, 707–732.

[7] S Iniyan, V Akhil Varma, Ch Teja Naidu, Crop yield prediction using machine learning techniques, AES vol 175, doi: Nov 2022, accessed: Mar 2024, https://www.sciencedirect.com/science/article/pii/S0965997822002277?via%3Dihub

[8] V. Shah, P. Shah, Groundnut Crop Yield Prediction Using Machine Learning Techniques, IJSRCSEIT, vol 3, issue 5, doi: Jul 2018, accessed: Mar 2024, https://www.researchgate.net/publication/326112319_Groundnut_Crop_Yield_Prediction_Using_Machine_Learning_Techniques

[9] M. Kalimuthu, P. Vaishnavi, and M. Kishore, "Crop Prediction using Machine Learning," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 926-932, doi: 10.1109/ICSSIT48917.2020.9214190.

[10] R. Banu, S. Bakiyalakshmi, S.Umamaheswari, Y Rajkumar, T. S. Karthick, "Prediction of crop yielding using machine learning algorithm", 2024, AIP Conf. Proc. 2935, doi.org/10.1063/5.0198954

[11] A. Morales and F. Villalobos, "Using machine learning for crop yield prediction in the past or the future," Frontiers in Plant Science, vol. 14, 2023, doi: 10.3389/fpls.2023.1128388.

[12] A. Morales and F. Villalobos, "Using machine learning for crop yield prediction in the past or the future," Frontiers in Plant Science, vol. 14, 2023, doi: 10.3389/fpls.2023.1128388.

[13] B. M. Bolker, M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, J.-S. S. White, "Generalized linear mixed models: A practical guide for ecology and evolution," Trends in Ecology & Evolution, vol. 24, pp. 127–135, 2009.

[14] H. Pham, S. Olafsson, "Bagged ensembles with tunable parameters," Computational Intelligence, vol. 35, pp. 184–203, 2019.

[15] Cortes, C., Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

[16] Friedman, J. H. (2001). Greedy function approximation (adaptive splines) with the CART

algorithm. Journal of statistical software, 33(1), 1-23.

[17] Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and regression trees. Wadsworth international group.

[18] Curtin, R. R., Cline, J. R., Slagle, N. P., March, W. B., Ram, P., Mehta, N. A., et al. (2013). MLPACK: A scalable C ++ machine learning library. Journal of Machine Learning Research, 14(Mar), 801–805

[19] Basak, D., Pal, S., & Patranabis, D. C. (2007). Support vector regression. Neural Information Processing- Letters and Reviews, 11(10), 203–224

[20] Khosla, E., Ramesh, D., & Priya, R. (2019). Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression. Environment, Development and Sustainability, 22(6), 5687–5708. https://doi.org/10.1007/s10668-019-00445-x

