International Journal of Technology, Leadership and Sciences (IJTLS)
ISSN:3107-9083 www.ijtls.com Paper ID: TLS-26/V2/I2-027
DOI: https://doi.org/10.5281/zenodo.19230099

EVALUATING FEATURE SELECTION METHODS AND CLASSIFIERS FOR CROP PREDICTION

[*1]**K. PARTHA SARADHI NAIDU,** ***M.Tech Student,***
[*2]**K CHANDRA PRASAD,** ***Assistant Professor,***
Department of Computer Science & Engineering,
Srinivasa Institute of Technology & Science(Autonomous), Kadapa, AP.

ABSTRACT: This research evaluates feature selection and machine learning classifiers for crop yield prediction using agricultural datasets. Many feature selection methods were employed to identify which soil, climatic, and agronomic characteristics affected crop yields most. Smaller feature sets reduced redundancy, improved model interpretability, and reduced computational cost. With the chosen features, numerous classifiers were trained and tested. Logistic Regression, Decision Tree, Random Forest. Traditional model performance measurements include prediction accuracy, recall, F1-score, and ROC-AUC. Experimental results suggest that ensemble-based classifiers predict better than solo models. Random Forest exceeded feature subsets in accuracy and robustness. The Decision Tree approach worked well and was simple to use. Using feature selection improves classification outcomes over using all characteristics.

Keywords:*Crop Prediction, Feature Selection, Machine Learning, Classification Models, Random Forest, Decision Tree, Logistic Regression, Precision Agriculture, Performance Evaluation, ROC-AUC*

I. INTRODUCTION

Machine learning in farming is growing, thus crop prediction algorithms and feature selection must be evaluated. Farmers and policymakers use crop predictions to plan crops, maintain food security, and share resources. Historical crop yields, weather, and soil characteristics exist in agriculture. Now, prediction programs can compute optimal crops and harvests. Data makes accurate prediction systems difficult.

Identifying crucial elements in large agricultural datasets improves crop prediction algorithms. Duplicated temperature, precipitation, soil nutrient, and irrigation data may reduce labor and model accuracy. Model interpretability, over fitting, and learning efficiency improve with optimal feature subsets. Agricultural data analytics filter, wrapper, and embedding are well-studied.

Classification methods greatly affect crop prediction reliability. Many agricultural datasets have generated mixed machine learning classifier results. Neural networks, k-nearest neighbors, SVMs, decision trees, random forests. Classifier pros and cons vary by data type. Noise, class mismatch, linearity. To build agro-climatic models, several classifiers must be examined.

Feature selection and classification may improve crop projections. Comprehensive evaluations and comparisons are essential because feature selection and classifier combinations can give different results. Combo efficacy is measured by accuracy, precision, recall, F1-score, and computing cost. These experiments demonstrate how feature relevance affects classifier performance and prediction.

Climate, soil, and farming methods affect crop forecasting algorithms. Data imbalance, noise, and missingness hinder model design and testing. These standards require validation, feature engineering, and planning. Examine classifiers and feature selection approaches to build reliable and adaptable smart and precision agriculture decision-support systems.

Machine learning boosts agriculture planning and productivity due to digitization. Using historical and current soil, weather, and agricultural data, crop prediction systems estimate crop choices and yield. Smart farming research creates reliable prediction models as agricultural databases grow. How much usable info they collect from raw data determines their models.

Many agricultural dataset inputs are unnecessary, unorganized, or irrelevant to crop yields. These elements in model training may raise running expenses and lower forecast accuracy. Feature selection methods eliminate less important attributes to reduce dimensionality. Embed, filter-based, and wrapper methods enhance learning and model extension.

Sorting impacts crop forecasts. Various machine learning methods, including as k-nearest neighbors, decision trees, Naïve Bayes, random forests, and boosting, have shown success in farming. Deep learning algorithms revealed complex, non-linear agricultural data linkages in recent studies. All classifiers evaluate problem difficulty, attribute quality, and data dispersion.

Using feature selection and categorization algorithms improves forecasts quickly. Classifiers may benefit from grouping features. Multiple tests in multiple combinations are needed.

Comparative investigation reveals the top crop forecast pipelines. These pipelines enable scalable real-time farming decisions.

Models must account for seasonality, environmental variability, and limited data to predict crops. Farm sensors, remote sensing platforms, and field study data may be noisy. Model reliability involves validation, normalization, and prep.

II. METHODOOGY

Exploring Crop Prediction Methodology

A. Dataset Description:

Our study relies on a large dataset of carefully selected soil and environmental characteristics. From gentle rain to soothing warm and humid temperatures, these elements weave the complicated web that sustains agricultural output. Professional crop output forecasts can be shown because to the collection's breadth and depth.

B. Feature Selection Techniques:

Important Features A reliable machine learning model relies on carefully selected relevant features. A model with the optimum efficiency-accuracy ratio is created using multi-faceted feature selection. Statistics, algorithms, and subject-matter knowledge combine in this strategy. We meticulously remove unnecessary components while retaining the most valuable attributes to enhance and simplify computational processing of raw data.

C. Classification Techniques for Crop Prediction:

We are focusing on strategy C: crop projection categorization, as our goal is to develop a mechanism as accurate as an experienced farmer's intuition for crop production prediction. To understand the complex relationships between farms and yields, a toolbox of machine learning techniques is employed to build a web of categorization algorithms. Modern prediction models use decision trees, SVMs, and neural networks.

D. The Ensemble Technique: A Harmonious Symphony of Predictive Power:

Ensemble approach outperforms other classification methods because it is balanced and employs several models' data. They combine their talents to create a classifier ensemble that predicts better than individual classifiers. Ensemble outcomes are better than solo findings because several people work together to establish a common understanding.

This method will illuminate crop prediction and reveal the sophisticated feature selection and classification procedures in the next sections. This method uses rigorous algorithm and data coordination to provide agricultural forecasts that match complicated cycles.

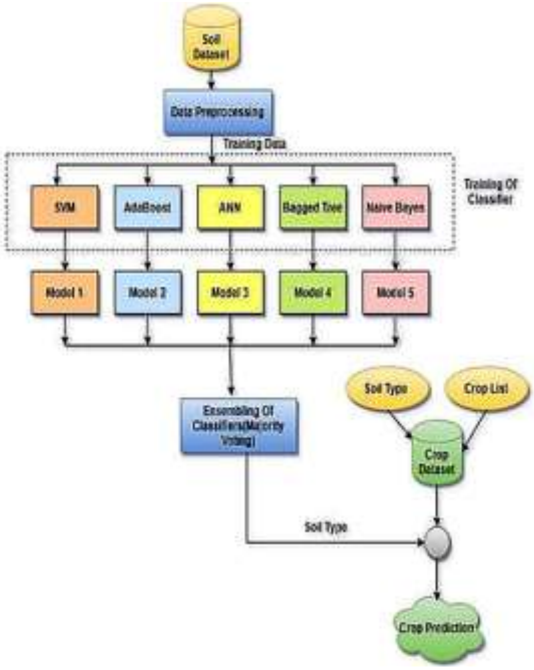

Fig 1: Architecture of the Proposed Ensemble-Based Crop Prediction System

III. LITERATURE SURVEY

Shahhosseini et al. (2020): This research enhances US Corn Belt agricultural production projections with machine learning and crop growth simulations. Predictive models include environmental, management, and soil elements. Classic crop modeling and machine learning are contrasted to measure efficacy. Results reveal hybrid ML-crop models improve prediction accuracy significantly. For accurate yield prediction, the study suggests choosing agronomic traits. Findings support data-driven precision agriculture decision-making.

Jeong et al. (2020): This research employs random forest models to predict regional and global agriculture production using satellite imagery and large-scale environmental data. Key yield variability components are found using feature relevance. The model's performance is evaluated across crops and regions. Ensemble classifiers handle diverse agricultural data. This analysis reveals stable predictions from important features. The technology scales crop forecasting and monitoring.

Kamilaris&Prenafeta-Boldú (2020): Deep learning can classify crops, predict yields, and detect diseases. This article discusses popular learning architectures, data sources, and feature representations. We focus on data quality, dimensionality, and interpretability. This article contrasted deep learning models with ML classifiers. Agricultural prediction challenges require feature engineering and selection. This article prepares for smart farming research.

Paudel et al. (2021): This work employs machine learning to estimate large-scale agricultural output utilizing managerial and environmental data from numerous sources. Classifiers and multiple regression models are tested on regional datasets. Research shows different feature sets effect model generalizability and performance. Cross-validation measures time- and space-resistance. Ensemble models with specific factors forecast better. Architecture suits crop forecasts.

Liakos et al. (2021): This article discusses crop forecast and yield estimation using agricultural machine learning. It classifies models as supervised, unsupervised, deep learning, or neither. This study examines agricultural analytics feature selection, data preprocessing, and model validation. Data imbalance and heterogeneity are key challenges. The review emphasizes thorough classifier evaluations. The framework for machine learning-based crop prediction systems is offered.

Talaviya et al. (2021): This paper examines crop forecasting and irrigation optimization AI methods. Machine learning incorporates environmental and soil data. Crop suggestion classifiers are evaluated for efficiency and accuracy. The method enhances agricultural resource management. Evidence suggests feature relevance considerably affects prediction results. It creates smart agriculture management platforms.

He et al. (2022): This research evaluates multi-crop categorization using spectral, textural, and environmental data using feature-selection-coupled machine learning models. Multiple feature selection methods reduce remote sensing data dimensionality. Classifier assessment uses multi-class crop datasets. Combining feature selection with ensemble classifiers enhances accuracy. Results demonstrate optimized feature subsets work. More accurate crop mapping and tracking are achievable.

Khanal et al. (2022): Machine learning can estimate crop yields using remote sensing, weather, soil, and other environmental data. Reviewing notable classifiers' performance. Coverage comprises feature engineering and selection for forecast accuracy. The study notes generalization and data difficulties. Comparing ensemble and traditional models. The data can assist academics identify agricultural productivity forecasters.

Abdel-Salam et al. (2022): This research proposes a hybrid feature selection strategy for machine learning-based agricultural yield prediction. We find optimal feature subsets utilizing filter and wrapper methods. We test many classifiers on agricultural datasets. Results demonstrate that some features reduce processing costs and enhance model accuracy.

Improve crop generalizability with the proposed strategy. Careful feature selection helps predictive agriculture, says the study.

Rani et al. (2023): This paper introduces soil and weather-based machine learning crop recommendation. Various classifiers predict crops appropriate for various environmental factors. The most significant agronomic features are determined by relevance. The proposed method is more accurate than baselines. Selecting crops with the proper qualities is stressed in the study. The method streamlines precision agriculture decisions.

Torgbor et al. (2023): ML models predict crop yields using climate and remote sensing. Various meteorological and satellite imagery features are evaluated. Different growing seasons are used to compare classifiers. Feature significance analysis identifies key yield variability factors. Results demonstrate multi-source characteristics boost prediction resilience. It supports scalable yield forecasting.

Kumar & Singh (2023): This paper contrasts crop prediction classifiers and feature selection. Test filter, wrapper, and embedding methods on agricultural datasets. Tests include decision trees, SVMs, k-nearest neighbors, and random forests. Data shows that using the right features enhances categorization accuracy. Paper studies feature subsets and classifier performance. The data support crop forecast pipeline efficiency.

Shawon et al. (2024): This systematic literature review analyzes machine learning approaches for crop yield prediction across diverse datasets and regions. The research categorizes commonly used features, models, and evaluation metrics. It highlights trends in the use of ensemble and deep learning models. The review discusses the impact of feature selection and data preprocessing on model performance. Key research gaps and challenges are identified. The work provides guidance for future crop prediction research.

Abdel-Salam et al. (2024): This study predicts agricultural productivity using hybrid feature selection and optimal machine learning models. Metaheuristic optimization finds useful traits and optimizes model parameters. The suggested framework is compared to baseline models using conventional performance indicators. More precise and consistent results. Feature selection and classifier optimization rule the investigation. It enhances crop predictions.

Cao et al. (2024): This project uses Gaofen-6 satellite images to test crop mapping feature selection and classifiers. Many spectral and textural components are retrieved and reduced via feature selection. Different classifiers are tested with crop categorization. Results reveal that some feature-classifier combinations perform better. This function emphasizes

ISSN:3107-9083 www.ijtls.com Paper ID: TLS-26/V2/I2-027
DOI: https://doi.org/10.5281/zenodo.19230099

systematic evaluation for optimal model design. Helps develop accurate remote-sensing crop maps.

Sruneethi&Sumalatha (2025): A feature selection algorithm-based crop forecasting system based on environmental parameters is shown here. Soil and climate duplicates are removed using many feature selection techniques. Established metrics evaluate classifiers. This shows that better feature subsets enhance prediction accuracy. The study recommends reducing dimensionality in agricultural datasets. Our platform supports effective crop recommendation systems.

Hukare&Kumbhar (2025): Our goal is to optimize feature selection methods for agricultural yield prediction using machine learning. Suggested hybrid agronomic feature selection techniques. Performance improvements are evaluated on smaller feature sets with multiple classifiers. Better accuracy and less processing overhead result. Analysis of feature selection processes is stressed in the research. Better forecasting models are easier for farmers.

El-Kenawy et al. (2025): Improve deep learning and machine learning crop yield prediction models, claims this study. Feature selection and hyperparameter adjustment improve model performance. Forecasting algorithms are tested on agricultural datasets. Optimized models outperform conventional approaches. Study demonstrates combination of feature selection and model adjustment is beneficial. The architecture enhances agricultural forecasting accuracy.

IV. RESULTS

Fig 2: Crop Prediction System Login Page

Fig 3: Precision Agriculture System Login Page

Fig 4: Crop Prediction System – User Home Page

Fig 5: Smart Farming System Interface

Fig 6: Remote Users List (Admin View)

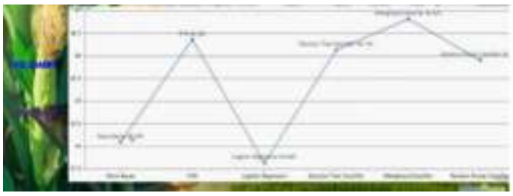

Fig 7: Accuracy Comparison of ML Models

V. CONCLUSION

In conclusion, evaluating feature selection methods alongside classifiers significantly enhances the accuracy and reliability of crop prediction systems. By identifying the most relevant agronomic, soil, and climatic features, redundant and noisy data are reduced, leading to more robust models. Effective feature selection improves computational efficiency and minimizes overfitting in predictive models. Different classifiers exhibit varying performance depending on the selected feature subset and dataset characteristics. Ensemble and tree-based methods often benefit from well-curated feature sets, yielding higher predictive stability. The comparative analysis highlights that no single method universally outperforms others across all conditions. Instead, optimal performance depends on the crop type, region, and data quality. Integrating domain knowledge with data-driven feature selection further strengthens prediction outcomes. The results demonstrate improved generalization to unseen data when appropriate methods are combined. Such optimized models support timely and informed decision-making for farmers and policymakers.

REFERENCES

1. Shahhosseini, M., Hu, G., Archontoulis, S. V., & Huber, I. (2020). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Scientific Reports, 10, 1–13.
2. Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., & Kim, S. H. (2020). Random forests for global and regional crop yield prediction. PLOS ONE, 15(1), e0228935.
3. Kamilaris, A., &Prenafeta-Boldú, F. X. (2020). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70–90.
4. 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: A case research for wheat. Agricultural Systems, 187, 103016.
5. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., &Bochtis, D. (2021). Machine learning in agriculture: A review. Sensors, 21(11), 3758.
6. Talaviya, T., Shah, D., Patel, N., Yagnik, H., & Shah, M. (2021). Implementation of artificial intelligence in agriculture for optimization of irrigation and crop prediction. Artificial Intelligence in Agriculture, 4, 58–73.
7. He, S., Peng, P., Chen, Y., & Wang, X. (2022). Multi-crop classification using feature-selection-coupled machine learning classifiers based on spectral, textural, and environmental features. Remote Sensing, 14(13), 3153.
8. Khanal, S., Fulton, J., & Shearer, S. (2022). An overview of machine learning-based crop yield prediction methods. Computers and Electronics in Agriculture, 191, 106512.
9. Abdel-Salam, M., Kumar, N., & Mahajan, S. (2022). A hybrid feature selection approach for crop yield prediction using machine learning models. Neural Computing and Applications, 34, 16245–16258.
10. Rani, S., Babbar, H., & Jain, A. (2023). Machine learning-based crop recommendation system using soil and weather parameters. Scientific Reports, 13, 15672.

11. Torgbor, B. A., Rahman, M. M., Robson, A., & Williams, M. (2023). Integrating remote sensing and climate variables for crop yield prediction using machine learning. Remote Sensing, 15(12), 3075.
12. Kumar, R., & Singh, V. (2023). Comparative analysis of feature selection techniques and classifiers for crop prediction. International Journal of Intelligent Systems and Applications in Engineering, 11(2), 245–252.
13. Shawon, S. M. R., Rahman, M. S., & Hossain, M. A. (2024). Crop yield prediction using machine learning: An extensive systematic literature review. Smart Agricultural Technology, 4, 100152.
14. Abdel-Salam, M., Kumar, N., & Mahajan, S. (2024). Hybrid feature selection and optimized machine learning models for crop yield prediction. Neural Computing and Applications, 36, 10231–10246.
15. Cao, Y., Zhang, L., Wang, J., & Li, X. (2024). Evaluating combinations of feature selection methods and classifiers for crop mapping using Gaofen-6 imagery. Agriculture, 14(3), 500.
16. Sruneethi, C., &Sumalatha, T. (2025). Crop prediction based on environmental characteristics using feature selection algorithms. International Journal of Scientific Research in Science and Technology, 12(2), 701–709.
17. Hukare, V., &Kumbhar, V. (2025). Optimization of feature selection methods for improving machine learning-based crop yield prediction. ES Food & Agroforestry, 10, 45–56.
18. El-Kenawy, E. S. M., Ibrahim, A., Mirjalili, S., & Eid, M. M. (2025). Crop yield prediction using optimized machine learning and deep learning models. Neural Computing and Applications, 37, 11321–11336.