

CROP YIELD PREDICTION USING VISION TRANSFORMER WITH BAT OPTIMIZATION ALGORITHM

J. Jency Rajakumari

Professor, Dept. of CSE,

Jayaraj Annapackiam CSI College of Engineering,
 Nazareth, India

jjensy9513034@gmail.com

N. Esakkiammal

PG Scholar, Dept. of CSE,

Jayaraj Annapackiam CSI College of Engineering,
 Nazareth, India

esekkiyammal123@gmail.com

Abstract - Crop prediction is an important task in smart agriculture. This paper proposes a crop prediction system based on data preprocessing, visualization, normalization, and a Vision Transformer model optimized using Bat Optimization. The system analyzes agricultural datasets and predicts suitable crops with high accuracy. This diagram shows the complete workflow for a crop prediction system. First, the input dataset is collected, which typically includes soil nutrients (N, P, K), temperature, humidity, rainfall, and pH values. In the preprocessing stage, Exploratory Data Analysis (EDA) is performed to understand patterns, handle missing values, and check non-numeric (object) data. Next, data visualization helps identify relationships between features and crop yield. The data is then normalized to scale all features uniformly, improving model performance. After that, the dataset is split into training and testing sets to evaluate accuracy fairly. A Vision Transformer model with Bat Optimization is applied for intelligent feature learning and parameter tuning. During model selection, the best-performing configuration is chosen. Finally, the trained model generates predictions, recommending the most suitable crop for the given environmental conditions, helping farmers improve productivity and decision-making.

Keywords - Crop Prediction, Vision Transformer, Bat Optimization, Agriculture AI

I. INTRODUCTION

Agriculture plays a vital role in food security and the economy. Selecting the right crop based on soil and environmental conditions improves productivity. Machine learning and deep learning techniques are increasingly used for intelligent crop recommendation systems. Crop prediction is the process of estimating which crops will grow best in a particular region and season based on environmental and agricultural conditions. It plays a vital role in modern farming by helping farmers make informed decisions about what to plant, when to plant, and how to manage their fields effectively. Factors such as soil quality, rainfall patterns, temperature, humidity, sunlight availability, and water resources strongly influence crop growth. By understanding these factors, farmers can select crops that are most suitable for their land, reducing the risk of crop failure and improving overall productivity.

Accurate crop prediction supports food security by ensuring stable agricultural production and minimizing losses caused by unfavorable weather or poor soil conditions. It also helps optimize the use of fertilizers, irrigation, and other resources, which lowers costs and prevents environmental damage. In regions where

agriculture is the primary source of income, reliable predictions can improve farmers' livelihoods and economic stability.

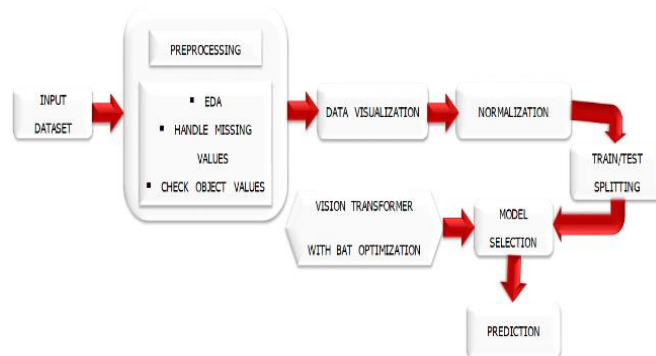


Fig 1 Proposed Diagram

II. DATASET DESCRIPTION

The input dataset for crop prediction consists of several important environmental and soil parameters that directly affect plant growth, development, and final yield. Among these, soil nutrients such as nitrogen (N), phosphorus (P), and potassium (K) are considered primary macronutrients. Nitrogen plays a crucial role in leaf development and chlorophyll formation, which is essential for photosynthesis. Phosphorus supports root growth, flowering, and energy transfer within the plant, while potassium improves overall plant health, disease resistance, and water regulation. An imbalance or deficiency in any of these nutrients can significantly reduce crop productivity.

Another key factor included in the dataset is soil pH value, which measures the acidity or alkalinity of the soil. Soil pH affects nutrient availability and microbial activity. Most crops grow best in slightly acidic to neutral soils because extreme pH levels can prevent plants.

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

Fig 2. Data Description

III. DATA PREPROCESSING

Preprocessing is a crucial step that prepares the raw dataset for further analysis and reliable prediction. It begins with Exploratory Data Analysis (EDA), which involves carefully examining the data to understand its structure, distribution, patterns, and possible anomalies. Through this process, one can identify trends, detect unusual values, and gain insights into how different variables behave and relate to each other. This understanding helps ensure that the dataset is suitable for accurate analysis.

Another important task in preprocessing is handling missing values. Real-world agricultural data often contains incomplete records due to sensor errors, manual entry issues, or unavailable measurements. Missing information can distort results if left untreated, so it must be properly addressed by either filling in reasonable estimates or removing incomplete entries when necessary. This step helps maintain consistency and prevents biased outcomes.

IV. DATA VISUALIZATION

Visualization plays an important role in understanding the characteristics of agricultural data. Techniques such as histograms show how individual variables are distributed, helping identify whether values are concentrated, spread out, or skewed. Scatter plots illustrate the relationship between two variables, making it easier to observe patterns, trends, or possible correlations between environmental factors. Heatmaps provide a color-coded view of relationships among multiple variables, allowing quick identification of strong or weak associations. Together, these visual tools make complex data easier to interpret, reveal hidden patterns, and support better insights about how different conditions influence crop growth and productivity.

Fig 3. Correlation Heatmap

V. NORMALIZATION

Normalization is an important data preparation step that adjusts numerical features to a common scale so that all variables contribute fairly during analysis. In agricultural datasets, different parameters such as rainfall, temperature, soil nutrients, and pH values may have very different ranges. For example, rainfall values may be in hundreds, while pH values typically range between 0 and 14. Without scaling, features with larger values can dominate the learning process and reduce the influence of smaller-scale variables.

By transforming all features into a similar range, normalization ensures balanced representation of each parameter. This helps the system interpret patterns more accurately and reduces bias toward any single feature. It also stabilizes calculations and prevents numerical issues that can arise from extremely large or small values.

VI. TRAIN/TEST SPLITTING

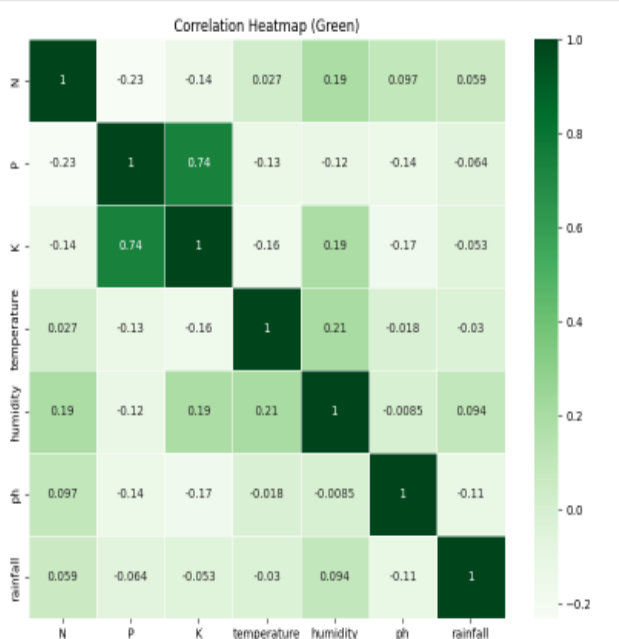
The dataset is divided into training and testing subsets to assess how well the prediction system performs on unseen data. The training subset is used to learn patterns, relationships, and underlying trends from the input variables such as soil properties and environmental conditions. During this phase, the system adjusts itself based on the examples provided, gradually improving its ability to make accurate predictions.

The testing subset, on the other hand, is kept separate and is not used during the learning process. It serves as an independent benchmark to evaluate how effectively the system can generalize to new situations. By comparing predicted outcomes with actual values in the testing data, one can measure accuracy, reliability, and potential errors.

VII. VISION TRANSFORMER WITH BAT OPTIMIZATION

Vision Transformer (ViT) is a powerful approach that captures complex and long-range relationships among input features by using a self-attention mechanism. Instead of focusing only on local patterns, it analyzes the entire input simultaneously, allowing it to understand how different environmental and soil factors interact with each other. This global perspective is especially useful in crop prediction, where variables such as nutrients, temperature, humidity, and rainfall collectively influence plant growth. By modeling these interdependencies, the system can extract meaningful patterns and produce more informed predictions.

Bat Optimization is used to refine the model's settings by searching for the most suitable parameter combinations. Inspired by the echolocation behavior of bats, it balances exploration and exploitation to find optimal solutions efficiently. Proper tuning of parameters improves stability, reduces prediction errors, and enhances overall accuracy.



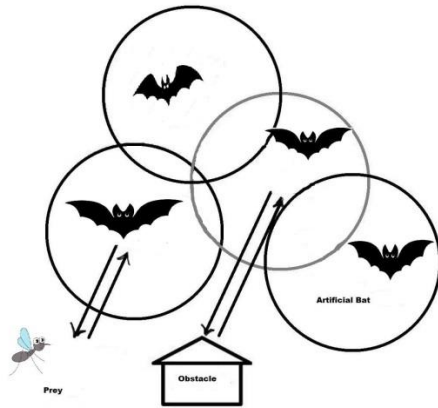


Fig 4. Bat Optimization Algorithm

VIII. RESULTS AND DISCUSSION

The proposed model demonstrates high accuracy in crop prediction by effectively analyzing complex relationships among soil characteristics and environmental conditions. By considering multiple factors simultaneously, it can identify subtle patterns that influence crop suitability, leading to more precise recommendations. This improved accuracy reduces the chances of selecting unsuitable crops, which in turn minimizes the risk of poor yield or crop failure.

Compared to traditional approaches, the model is better at handling large and diverse datasets, including variations in climate, soil composition, and seasonal changes. It adapts well to different agricultural scenarios, making it useful across various regions and farming conditions. Higher prediction reliability also helps farmers plan resource usage more efficiently, such as water, fertilizers, and labor.

Predicted Crop: RICE

Fig5. Prediction

Existing Work

This paper addresses the critical challenge of optimizing crop selection in agriculture to enhance food production sustainably. The problem is framed as a multi-class classification task where the goal is to recommend the most suitable crop based on a set of environmental and soil features. While traditional methods rely on time-consuming and labor-intensive expert knowledge, this work proposes a data-driven approach using machine learning. The novelty of our investigation lies in the comprehensive comparative analysis of seven machine learning algorithms and the development of a highly accurate neural network model. We utilize a publicly available dataset from Kaggle, which has been preprocessed to ensure data quality. We provide a detailed account of our feature engineering and hyperparameter tuning processes. Our proposed neural network model, with a specific architecture of 30–20–10 neurons, achieves a validation accuracy of 97.73%. This work also discusses the challenges of deploying such models,

including real-world data variability and the need for model interpretability. We demonstrate that our approach, particularly the neural network model, provides a robust, scalable, and adaptable solution for crop recommendation, outperforming other models (in holistic view) like Random Forest which achieved a slightly higher accuracy of 99.5% on this specific dataset but with less generalization potential. The findings of this study can empower farmers to make informed decisions, ultimately leading to improved crop yields, enhanced soil fertility, and greater profitability.

IX. CONCLUSION

This paper presents an AI-based crop prediction system that integrates a Vision Transformer (ViT) with the Bat Algorithm for optimization, aiming to support intelligent decision-making in smart agriculture. The Vision Transformer effectively captures global relationships within agricultural data through its self-attention mechanism, enabling accurate extraction of complex patterns related to soil nutrients, environmental conditions, and crop characteristics. To further enhance model performance, the Bat Algorithm is employed to optimize key hyperparameters, improving convergence speed and prediction accuracy. The proposed approach is trained and evaluated on a crop dataset containing multiple classes, demonstrating strong generalization ability across different agricultural scenarios. Experimental results show that the hybrid model outperforms traditional machine learning techniques in terms of accuracy, reliability, and robustness. This system can assist farmers and agricultural planners in selecting suitable crops, reducing risk, improving productivity, and promoting sustainable farming practices in modern precision agriculture.

REFERENCES

- [1] Shingade, Sachin Dattatraya, and Rohini Prashant Mudhalwadkar. "Analysis of crop prediction models using data analytics and ML techniques: a review." *Multimedia Tools and Applications* 83.13 (2024): 37813-37838.
- [2] Wang, Yuhan, et al. "Progress in research on deep learning-based crop yield prediction." *Agronomy* 14.10 (2024): 2264.
- [3] Mahmud, Tanjim, et al. "An approach for crop prediction in agriculture: Integrating genetic algorithms and machine learning." *IEEE Access* 12 (2024): 173583-173598.
- [4] Goel, Mukulit, and Mrinal Pandey. "Crop yield prediction using AI: A review." *2024 2nd international conference on disruptive technologies (ICDT)*. IEEE, 2024.
- [5] Durairasu, E., et al. "Precision farming for crop prediction." *Advanced Technologies for Smart Agriculture*. River Publishers, 2024. 195-217.
- [6] Saha, Sarmistha, et al. "Precision agriculture for improving crop yield predictions: a literature review." *Frontiers in Agronomy* 7 (2025): 1566201.

- [7] Shawon, Sarowar Morshed, et al. "Crop yield prediction using machine learning: An extensive and systematic literature review." *Smart Agricultural Technology* 10 (2025): 100718.
- [8] Sharma, Seema, et al. "Enhancing crop yield prediction through machine learning regression analysis." *International Journal of Sustainable Agricultural Management and Informatics* 11.1 (2025): 29-47.
- [9] Messina, Carlos, et al. "Toward a general framework for AI-enabled prediction in crop improvement." *Theoretical and Applied Genetics* 138.7 (2025): 151.
- [10] Liu, Ziao, et al. "In-season crop yield prediction: State of the art and future research direction." *International Journal of Applied Earth Observation and Geoinformation* 146 (2026): 105129.