

CROP RECOMMENDER SYSTEM AND PRICE PREDICTION BY ENSEMBLING MACHINE LEARNING AND SWARM INTELLIGENCE

1. Dr.G.V.S.N.R.V PRASAD

Professor

Department of Computer Science and Engineering
Seshadri Rao Gudlavalleru Engineering College
Andhra Pradesh, India
Email:gutta.prasad1@gmail.com

2. SAYANI HARINI

Student

Department of Computer Science and Engineering
Seshadri Rao Gudlavalleru Engineering College
Andhra Pradesh, India
Email:harinisayani3@gmail.com

3. POTLA MAHIMA DIVYA JYOTHI

Student

Department of Computer Science and Engineering
Seshadri Rao Gudlavalleru Engineering College
Andhra Pradesh, India
email:mahimadivvyajyothi@gmail.com

4. YANNAM MANASVI NEHA NAIDU

Student

Department of Computer Science and Engineering
Seshadri Rao Gudlavalleru Engineering College
Andhra Pradesh, India
Email:mahimadivvyajyothi@gmail.com

5. VAKKALAGADDA SUBHAGA

Student

Department of Computer Science and Engineering
Seshadri Rao Gudlavalleru Engineering College
Andhra Pradesh, India
Email:subhaga508@gmail.com

ABSTRACT Accurate crop prediction plays a vital role in enhancing the agricultural productivity and enabling data-driven decision-making in smart farming systems. Increasing variability in soil characteristics and environmental conditions has made crop selection more complex, necessitating robust and intelligent predictive models. This paper proposes a machine learning-based framework for crop prediction using key soil attributes such as pH, nitrogen (N), phosphorus (P), and potassium (K), along with environmental parameters including rainfall and temperature. A comprehensive data preprocessing phase is employed to address class imbalance using Synthetic Minority Oversampling Technique (SMOTE) and Majority Weighted Minority Oversampling Technique (MWMOTE). Feature optimization is carried out using Boruta, Recursive Feature Elimination (RFE), and a modified RFE approach to retain the most informative features while reducing model complexity. Machine Learning algorithms Naive Bayes (NB), Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM),

and Random Forest (RF) are evaluated to assess predictive performance. In addition, swarm intelligence based optimization techniques are used for hyperparameter tuning the parameters to further enhance model accuracy. The proposed framework is extended to incorporate market price prediction for the recommended crop, offering practical insights to farmers and agricultural stakeholders. Experimental results demonstrate that optimized feature selection combined with ensemble based classifiers significantly outperforms conventional approaches in terms of prediction accuracy and robustness.

INDEX TERMS Crop prediction, soil attributes, environmental factors, machine learning, feature selection, SMOTE, MWMOTE, swarm intelligence, classification algorithms

1. INTRODUCTION

Agriculture remains one of the most critical sectors for ensuring food security and economic stability, particularly in developing countries where farming decisions are highly sensitive to environmental and market uncertainties. Selecting an appropriate crop for cultivation is a complex task influenced

by multiple interdependent factors, including soil characteristics, climatic conditions, and resource availability. Inaccurate crop selection can lead to reduced yield, inefficient use of resources, and significant financial losses for farmers. [1], [2] Crop prediction is inherently a data driven problem due to the involvement of both biotic and abiotic factors. Soil attributes such as pH, N, P, and K play a crucial role in determining crop suitability, while environmental parameters including temperature and rainfall significantly influence crop growth. Increasing availability of agricultural datasets has enabled the ML techniques to model these complex relationships more effectively than traditional statistical approaches. However, the performance of ML based crop prediction models is often constrained by challenges such as data imbalance, redundant features, and suboptimal hyperparameter configurations. Class imbalance is a common issue in agricultural datasets, where certain crops are underrepresented compared to others. To address this problem, data resampling techniques such as SMOTE and MWMOTE have been widely adopted to enhance minority class representation and improve classification performance. In addition, effective feature selection is essential for reducing dimensionality, eliminating irrelevant attributes, and improving model interpretability. Wrapper based feature selection methods, including Boruta and RFE, have demonstrated strong potential in identifying salient features that contribute most significantly to prediction accuracy. [3] Recent studies have shown that ensemble based ML models, such as RF and bagging based classifiers, outperform individual learners by reducing variance and improving robustness. Nevertheless, the effectiveness of these models heavily depends on optimal hyperparameter tuning. Swarm intelligence techniques, inspired by the collective behavior of natural systems, have emerged as powerful optimization tools for enhancing machine learning performance. Algorithms such as PSO and related swarm based approaches offer efficient exploration of the search space and have been successfully applied for hyperparameter optimization in various prediction tasks. Beyond crop recommendation, market price prediction has become an equally important component of modern agricultural decision making. Predicting future crop prices enables farmers to plan cultivation and marketing strategies more effectively, thereby maximizing profitability and reducing economic risk. Integrating crop prediction with market price analysis provides a holistic framework that supports both production and financial planning.

II. RELATED WORK

Research on crop prediction and yield estimation has gained significant attention due to the increasing need for data driven agricultural decision support systems. Existing studies can be broadly categorized based on soil conditions, environmental factors, machine learning methodologies, and optimization techniques.

A. CROP PREDICTION BASED ON SOIL ATTRIBUTES

Several researchers have investigated crop prediction using

soil related parameters such as nutrient content, pH, and soil texture. Machine learning classifiers including DT, RF, and SVM have been widely applied to analyze soil conditions and predict suitable crops. [4], [5], [6]

B. CROP PREDICTION BASED ON ENVIRONMENTAL CONDITIONS

Several studies have incorporated agro meteorological data into crop prediction models using decision support systems and regression based techniques. Advanced methods utilizing satellite imagery, vegetation indices, and meteorological observations have demonstrated improved forecasting performance for crops. However, these approaches often depend on complex data acquisition infrastructures and are computationally intensive, limiting their scalability for small scale farming applications. [7], [4]

C. MACHINE LEARNING TECHNIQUES FOR CROP PREDICTION

With the availability of large agricultural datasets, ML techniques have emerged as effective tools and for crop prediction. Supervised learning algorithms such as NB, KNN, SVM, DT, RF have been extensively studied. Ensemble learning approaches, including bagging and majority voting, have shown superior performance compared to single classifiers [8] [9] [10].

D. OPTIMIZATION AND SWARM INTELLIGENCE IN AGRICULTURE

Recent research has explored optimization techniques to enhance machine learning performance in agricultural applications. Swarm intelligence algorithms have been successfully applied for feature selection and hyperparameter tuning. PSO, Genetic Algorithms (GA), and related swarm based techniques offer efficient search mechanisms to optimize model parameters and improve classification accuracy. [11], [12] Although swarm intelligence has been applied in isolated agricultural tasks, its integration with ensemble learning frameworks for crop prediction remains limited. Furthermore, most existing studies focus solely on crop recommendation or yield prediction, without extending the analysis to market price forecasting.

III. PROPOSED METHODOLOGY

This paper proposes an optimized and integrated framework for crop recommendation and market price prediction by combining data preprocessing, feature selection, ensemble learning, and swarm intelligence-based optimization. The primary objective of the proposed methodology is to improve prediction accuracy, robustness, and generalization while addressing class imbalance, redundant features, and suboptimal model parameter selection. The proposed system contains two main components: (i) crop recommendation based on soil and environmental attributes, and (ii) market price prediction for the recommended crop. The complete workflow of the proposed methodology is described below. Unlike existing approaches discussed in the related work, which often suffer from limitations such as class imbalance, redundant feature usage, reliance on single classifiers, and lack of model optimization, the proposed methodology intro-

duces a comprehensive solution to address these challenges. Class imbalance is handled through advanced oversampling techniques, feature redundancy is minimized using wrapper based feature selection, prediction robustness is improved through ensemble learning, and model performance is enhanced using swarm intelligence-based hyperparameter optimization. Additionally, the proposed framework extends beyond crop suitability analysis by incorporating market price prediction, thereby providing both agronomic and economic decision support.

IV. PREPROCESSING

In this work, preprocessing is performed separately for the crop recommendation dataset and the market price dataset due to their distinct characteristics and learning objectives.

A. CROP RECOMMENDATION DATASET

The crop recommendation dataset consists of soil nutrient attributes namely N, P, K, and pH along with environmental parameters such as temperature, rainfall, and humidity. Initially, the dataset is examined for missing values and inconsistencies. Any incomplete records are handled using appropriate imputation techniques, while duplicate and noisy samples are removed to improve data quality. Feature scaling is applied to normalize the range of input variables and ensure uniform contribution during model training.

B. MARKET PRICE DATASET

The market price prediction dataset is processed independently, as it represents a regression problem with continuous target values. The dataset comprises historical crop price information along with temporal attributes such as month and year. Preprocessing includes handling missing values, removing outliers, and normalizing price values to reduce scale variations. Since the market price dataset does not them well suited for crop recommendation tasks involving complex interactions between soil and environmental factors.

exhibit class imbalance, oversampling techniques are not applied. Temporal features are encoded appropriately to capture seasonal trends and price fluctuations. The cleaned and transformed dataset is then used directly for regression based price prediction.

V. FEATURE SELECTION TECHNIQUES

In this work, wrapper based feature selection techniques are employed, as they evaluate feature subsets based on the predictive performance of a learning algorithm, making in identifying redundant and correlated features and is well suited for small to medium sized agricultural datasets.

C. MODIFIED RECURSIVE FEATURE ELIMINATION (MRFE)

To further enhance computational efficiency, a MRFE approach is adopted. Unlike conventional RFE, MRFE reduces repeated dataset updates by incorporating feature permutation at the initial stage. The dataset is first shuffled to create a permuted version, which is then combined with the original dataset. [11], [13]

VI. CLASSIFICATION ALGORITHMS

To perform crop recommendation based on soil and environmental attributes, several supervised machine learning classification algorithms are employed. [2], [11] These algorithms are selected due to their effectiveness in handling multiclass problems, nonlinear relationships, and agricultural datasets. [14], [15]

A. NAÏVE BAYES (NB)

NB is a probabilistic supervised learning algorithm based on Bayes' theorem. It estimates the posterior probability of a class given an input feature vector by assuming conditional independence among features. Bayes' theorem is defined as:

$$P(Y|X) = \frac{P(X|Y) P(Y)}{P(X)} \quad (1)$$

A. BORUTA FEATURE SELECTION

It aims to identify all features that are relevant to the target variable rather than selecting a minimal subset. The algorithm evaluates feature importance by comparing original attributes with artificially generated shadow features. [9], [11] A RF model is then trained on the extended dataset containing both original and shadow features. Feature importance scores are computed, and a Z score is calculated by normalizing the importance values using their mean and standard deviation. The Boruta algorithm effectively captures nonlinear relationships and feature interactions, making it particularly suitable for agricultural datasets where soil nutrients and environmental factors jointly influence crop suitability. The computational complexity of Boruta is approximately $O(P \times N)$, where P represents the number of features and N denotes the number of samples.

B. RECURSIVE FEATURE ELIMINATION (RFE)

It is a greedy wrapper based feature selection technique that iteratively removes the least significant features based on model performance. The process begins with the complete set of features, and a base learning algorithm is trained on the dataset. Features are ranked according to their importance scores derived from the trained model. [13], [12] RFE evaluates model performance across different feature subset sizes and automatically selects the subset that yields the best validation performance. RFE is particularly effective where $P(y/x)$ is the posterior probability, $P(x/y)$ is the likelihood, $P(y)$ is the prior probability, and $P(x)$ is the evidence.

B. DECISION TREE (DT)

DT models decisions in the form of a tree structure. Each internal node represents a decision condition based on a feature, branches represent outcomes, and leaf nodes correspond to class labels. The process of tree construction involves selecting optimal features using impurity measures such as Information Gain or Gini Index. Decision Trees are easy to interpret and can effectively capture nonlinear relationships between input features and crop classes. However, they are prone to overfitting when trained on complex datasets without proper constraints.

C. SUPPORT VECTOR MACHINE (SVM)

SVM aims to identify an optimal hyperplane that maximizes the margin between different class boundaries in a high dimensional feature space. SVM is particularly effective for high dimensional data and performs well when the number of features is large. However, it is computationally expensive for large datasets and does not inherently provide probabilistic outputs.

D. K-NEAREST NEIGHBORS (KNN)

KNN follows a lazy learning approach, where no explicit training phase is involved. For a given query instance, KNN identifies the k closest data points based on a distance metric

such as Euclidean or Manhattan distance and assigns the class label based on majority voting. Feature normalization is essential for KNN, as the algorithm is sensitive to differences in feature scales. While KNN is simple and effective for capturing local patterns, its computational cost increases with dataset size.

E. RANDOM FOREST (RF)

RF constructs multiple decision trees using randomly selected subsets of data and features. Each tree independently predicts the class label, and the final output is determined through majority voting. By introducing randomness in both data sampling and feature selection, RF reduces overfitting and improves generalization performance. It can efficiently handle high dimensional data and complex feature interactions, making it well suited for crop classification problems involving soil and environmental variables.

VII. CROP RECOMMENDATION USING ENSEMBLE LEARNING AND SWARM INTELLIGENCE

In this study, crop recommendation is modeled as a multi class classification problem using an ensemble of machine learning classifiers, further enhanced by swarm intelligence-based optimization.

A. ENSEMBLE LEARNING BASED CROP RECOMMENDATION

Ensemble learning is a powerful ML paradigm that combines multiple base classifiers to achieve improved predictive accuracy and robustness compared to individual models. The fundamental idea behind ensemble learning is that diverse learners can collectively reduce bias, variance, and generalization error. [11], [16], [7] In the proposed crop recommendation system, multiple supervised classification algorithms are employed as base learners. These classifiers are trained on a preprocessed crop feature dataset consisting of soil parameters.

B. SWARM INTELLIGENCE BASED OPTIMIZATION FOR CROP RECOMMENDATION

To further improve the performance of the ensemble model, swarm intelligence is incorporated in the form of PSO. PSO is applied exclusively to optimize the hyperparameters of the RF classifier, which plays a significant role in the ensemble due to its strong classification capability. [4] [17] The objective of PSO is to maximize classification accuracy by identifying optimal hyperparameter values. Once the optimal parameters are obtained, the RF classifier is retrained using these parameters and integrated into the ensemble framework. The incorporation of PSO enables automated hyperparameter tuning, reduces manual intervention, and enhances the crop recommendation accuracy by effectively capturing complex relationships among soil and climatic Features.

C. INTEGRATED ENSEMBLE AND SWARM INTELLIGENCE FRAMEWORK

The integration of ensemble learning and swarm intelligence results in a robust crop recommendation system. The optimized Random Forest classifier obtained through PSO replaces the conventional Random Forest within the ensemble. The final recommendation is generated through a voting based ensemble decision mechanism.

VIII. MARKET PRICE PREDICTION

Market price prediction plays a vital role in assisting farmers and stakeholders in making informed post harvest and marketing decisions. After identifying the most suitable crop through the proposed crop recommendation framework, the system is extended to forecast the future market price of the recommended crop using historical price data.

A. PREPROCESSING FOR PRICE PREDICTION

Since market price prediction is formulated as a regression problem with a continuous target variable, preprocessing steps differ from those applied to the crop recommendation dataset. Missing values are handled using statistical imputation techniques, while outliers caused by sudden market fluctuations are identified and removed. Feature scaling is applied to normalize numerical attributes and improve regression model convergence.

B. REGRESSION BASED MARKET PRICE PREDICTION

For market price forecasting, regression models are employed to learn the relationship between historical price trends and temporal features. Models such as Linear Regression, Support Vector Regression, Decision Tree Regression, and Random Forest Regression are evaluated to determine the most suitable approach. Each regression model is trained using historical data and validated using standard performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The model demonstrating the lowest prediction error is selected for final price forecasting

C. INTEGRATION WITH CROP RECOMMENDATION SYSTEM

Once the crop recommendation module identifies the optimal crop for a given soil and environmental condition, the corresponding market price prediction model is invoked to forecast the future price of that crop. This integration enables the system to provide both agronomic suitability and economic feasibility, offering actionable insights to farmers. The combined crop recommendation and market price prediction framework supports precision agriculture by enabling farmers to select crops that are not only environmentally suitable but also economically profitable.

IX. RESULTS AND ANALYSIS

A. RESULTS OF PROPOSED SYSTEM

This study utilized a comprehensive agricultural dataset comprising soil and environmental characteristics, collected manually from public datasets. The performance of classifiers and feature selection methods was evaluated using Accuracy, Precision (P), Recall (R), F1 score, Specificity (S), Area Under the Curve (AUC), Mean Absolute Error (MAE), and Log Loss (LL). Both the raw dataset and resampled datasets were analyzed, and results were further validated using fold based and data splitting validation approaches. [2], [9], [16]

Initially, all classifiers were trained on the original dataset without sampling. RF achieved the highest performance, with AUC 99.55%, Precision 99.57%, Recall 99.55%, F1 99.55%, Specificity 99.99%, AUC 0.999995, MAE 0.03, and LL 0.04, highlighting its robustness in handling complex, multidimensional agricultural data. Bagging classifiers also performed well, followed by SVM, KNN, DT, and NB, in descending order of accuracy and generalization capability. The results emphasize the suitability of ensemble methods for crop prediction. To address class imbalance, two over-sampling techniques SMOTE and MWMOTE were applied to the training data. Sampling improved minority class representation, significantly enhancing classifier performance. With SMOTE and MWMOTE, RF achieved AUC 99.99%, Precision 99.56%, Recall 99.54%, F1 99.54% while KNN and SVM also showed substantial improvement demonstrating that advanced oversampling methods paired with ensemble classifiers effectively handle skewed datasets. Other classifiers benefited from sampling, but the improvement was most pronounced for ensemble methods, reinforcing the advantage of Random Forest for crop classification in imbalanced scenarios.

Acc	Prec	Rec	F1	AUC	Cls	Stage
0.9954	0.9959	0.9954	0.9954	0.9999	NB	Before
0.9795	0.9806	0.9795	0.9794	0.9893	DT	Before
0.9795	0.9804	0.9795	0.9793	0.9987	KNN	Before
0.9841	0.9856	0.9841	0.9840	1.0000	SVM	Before
0.9954	0.9957	0.9954	0.9955	0.9999	RF	Before
0.9954	0.9959	0.9954	0.9954	0.9999	NB	SMOTE
0.9795	0.9806	0.9795	0.9794	0.9893	DT	SMOTE
0.9795	0.9804	0.9795	0.9793	0.9987	KNN	SMOTE
0.9841	0.9856	0.9841	0.9840	1.0000	SVM	SMOTE
0.9954	0.9957	0.9954	0.9955	0.9999	RF	SMOTE
0.9954	0.9959	0.9954	0.9954	0.9999	NB	MWMOTE
0.9841	0.9847	0.9841	0.9841	0.9917	DT	MWMOTE
0.9795	0.9804	0.9795	0.9793	0.9987	KNN	MWMOTE
0.9841	0.9856	0.9841	0.9840	1.0000	SVM	MWMOTE
0.9954	0.9957	0.9954	0.9955	0.9999	RF	MWMOTE

TABLE 1. Performance Comparison of Classifiers Under Different Sampling Stages

The experimental results in Table 1 reveal that RF and NB are the most robust models for this dataset, both maintaining a consistent accuracy of 99.55% across all

sampling stages. Specifically, RF achieved a near perfect Area Under the Curve (AUC) of 0.999995, while NB followed closely with an AUC of 0.999968. SVM demonstrated exceptional class separation capability with a perfect AUC of 1.00 and a steady accuracy of 98.41%. DT and KNN models exhibited stable performance, both achieving an accuracy of 97.95%. Notably, while most models showed identical performance across stages, the DT model showed a slight refinement in Precision (0.9847) after the application of MWMOTE, compared to its performance after SMOTE (0.9806).

To provide a clearer comparative view of the models' resilience to data balancing, a line graph was generated based on the accuracy values from Figure 1

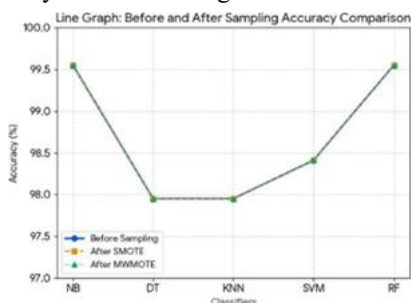


FIGURE 1. Accuracy comparison of various classifiers across different sampling stages.

As illustrated in Figure 1, the horizontal trajectories of the lines demonstrate that the accuracy for each classifier remains highly stable regardless of the sampling method like SMOTE, MWMOTE applied.

Train-Test Split	Acc	Prec	AUC	MAE	Log Loss
25-75	0.8451	0.8445	0.8626	0.06	0.07
30-70	0.8917	0.8654	0.9101	0.05	0.06
35-65	0.9181	0.8874	0.9321	0.04	0.05
40-60	0.9467	0.9092	0.9679	0.03	0.04
45-55	0.9637	0.9200	0.9804	0.03	0.04
50-50	0.9471	0.9137	0.9697	0.03	0.04
55-45	0.9690	0.9346	0.9842	0.02	0.03
60-40	0.9652	0.9538	0.9751	0.03	0.04
65-35	0.9700	0.9757	0.9878	0.02	0.03
70-30	0.9729	0.9894	0.9923	0.03	0.04
75-25	0.9711	0.9856	0.9900	0.03	0.04

TABLE 2. Performance Analysis for Different Train-Test Splits

Table 2 presents the performance of the Random Forest classifier with MRFE+PSO selected features was evaluated across various train test split ratios, from 25-75 to 75-25, to determine the impact of data partitioning on model efficiency. At the 25-75 split, the model achieved an accuracy of 84.51% and an AUC of 86.26%, with moderate prediction errors (MAE 0.06, Log Loss 0.07) attributed to the limited training data. As the training proportion increased, consistent improvements were observed: accuracy reached 94.67% the 40-60 split and peaked at 97.29% for the 70-30 split, which yielded an optimal AUC of 99.23% and a precision of 98.94%.

Even with equal partitioning at the 50-50 split, the model maintained high predictive quality with 94.71% accuracy

and 96.97% AUC. Stability was further confirmed at the 75-25 split, maintaining a robust 97.11% accuracy and 99.00% AUC. Across all ratios, the classifier consistently demonstrated accuracy above 84% and AUC above 86%, with minimal MAE and Log Loss. These results highlight the robustness and reliability of the MRFE-Random Forest framework, proving it maintains high performance and stability for crop prediction tasks regardless of dataset partitioning.

B. COMPARATIVE ANALYSIS WITH EXISTING SYSTEMS

Paper / Method	Best Reported Accuracy (%)
Crop Prediction Based on Agricultural Environment	93.29 (RF)
Crop Recommendation and Yield Estimation Using ML	91.00 (RF)
Crop Recommender System and Price Prediction	92.40 (RF)
Proposed System	97.29 (RF)

TABLE 3. Comparison of the Proposed System with Existing Methods

The effectiveness of the proposed framework was benchmarked against established methods in crop prediction and recommendation systems, as summarized in Table 3. This comparison highlights the significant performance improvements achieved through the integration of MRFE feature selection and optimized Random Forest classification.

- 1) Comparison with Existing Literatures: Previous studies for crop prediction based on agricultural environments reported a maximum accuracy of 93.29%. Similarly, other ML based recommendation systems achieved best reported accuracies ranging from 91.00% to 92.40%.
- 2) Performance of the Proposed System: The proposed system, utilizing a comprehensive suite of sampling techniques, different classifiers and applied swarm intelligence techniques consistently outperformed these existing benchmarks. Specifically, the RF model within the proposed framework achieved a peak accuracy of 97.29

X. CONCLUSION

This study evaluated the performance of machine learning classifiers, sampling techniques, and feature selection methods for crop prediction using a carefully collected agricultural dataset containing soil and environmental characteristics. The results demonstrate that the Random Forest (RF) classifier consistently outperforms other models, including Naive Bayes (NB) and Support Vector Machine (SVM), across all evaluation metrics and experimental settings.

As illustrated in the comparative analysis, the RF classifier achieved a peak accuracy of 99.54% and an AUC of 0.99 before sampling, maintaining this superior performance even after applying SMOTE and MWMOTE techniques to address class imbalance. The significance of MRFE selected features was further validated, where the RF model achieved 99.09% accuracy, confirming that variables such as soil temperature,

rainfall, and humidity are highly informative for crop suitability. Extensive validation through 9 fold cross-validation and various train-test splits confirmed the model's stability and generalizability, with accuracy reaching up to 97.29% while maintaining minimal MAE and Log Loss.

REFERENCES

- [1] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [2] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.
- [3] Dr Guttikonda Keerthi, Dr G Ramachandran, Dr GVSNRV Prasad "Cuckoo search optimization -based feature selection for predicting autism spectrum disorder using artificial immune algorithms," *Journal of Theoretical and Applied Information Technology*, pp. 1–11, 2025.
- [4] S. R. Dash and D. P. Mohanty, "A machine learning framework for crop classification using remote sensing data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 5, pp. 1605–1616, 2019.
- [5] Keerthi Guttikonda, Yerninti Ashvitha, Velagala Sai Ranga Reddy, and Penumala Sandeep, "Integrating convolutional neural networks (cnn) and machine learning for accurate identification of autism spectrum disorder using facial biomarkers," in *Proceedings of the 2024 IEEE International Conference on Emerging Systems and Intelligent Computing (ESIC)*. IEEE, 2024.
- [6] Baburao Markapudi, Kunchaparthi Jyothsna Latha and Kavitha Chaduvula, "A new hybrid classification algorithm for predicting customer churn," in *Proceedings of the International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICES 2021)*. Chennai, India: St. Joseph's Institute of Technology, 2021, cited by 4.
- [7] D. Singh and I. Gupta, "Crop yield prediction using machine learning: A review," *International Journal of Agricultural Science and Technology*, vol. 6, no. 1, pp. 45–56, 2020.
- [8] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.
- [9] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [10] DNVSLS Indira, R Kiran Kumar, GVSNRV Prasad "Detection and classification of trendy topics for recommendation based on twitter data on different genre," in *Smart Intelligent Computing and Applications*. Springer Singapore, 2018, pp. 143–153.
- [11] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4. IEEE, 1995, pp. 1942–1948.
- [12] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, no. 1–2, pp. 1–39, 2010.
- [13] T. G. Dietterich, "Ensemble methods in machine learning," *Multiple Classifier Systems*, pp. 1–15, 2000.
- [14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, 2009.
- [15] V. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. Springer, 2000.
- [16] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.
- [17] Keerthi Guttikonda, G Ramachandran and GVSNRV Prasad, "Autism spectrum disorder prediction using lasso regularised bat search optimisation," *International Journal of Services Operations and Informatics*, 2024.