A SECURED HYBRID TOOL FOR CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHM

G. Buvaanyaa[1] and Dr. M. Gobi[2]

[1]PhD., Research Scholar (FT), PG & Research Department of Computer Science, Chikkanna Government Arts College, Tiruppur, Tamil Nadu, India.
[2]PG & Research Department of Computer Science, Chikkanna Government Arts College, Tiruppur, Tamil Nadu, India.

ABSTRACT

Crop suggestions are crucial because they help farmers select the best crops for their climate and region. In the past, this procedure mostly depended on specialized knowledge, which was labor-intensive and time-consuming. Machine learning methods can be very useful in automating crop recommendations and detecting pests and diseases, allowing farmers to get the most out of their land while maintaining soil fertility and replenishing essential nutrients. The effectiveness of machine-learning algorithms for crop recommendation is investigated. To accurately predict which crops will be most suitable for a given location, the proposed method uses a variety of features, including soil nutrients and climate data. Crop suggestion could be revolutionized by this technology, which would benefit farmers of all sizes by increasing crop yields, sustainability, and overall profitability. By training and testing models with multiple configurations of machine learning algorithms, we have attained near-perfect accuracy through rigorous study of a massive historical data set. Across all models, the hybrid Algorithm achieves the best accuracy of 89.97%. This paper explores the theory, methodology, implementation, and evaluation of a hybrid RF-SVM classification model, highlighting its potential for improved accuracy and generalization.

KEYWORDS

Machine Learning, Feature Selection, Classification, Crop prediction, Hybrid Algorithm.

1. INTRODUCTION

Machine-Learning [1]-[2] is the science of enabling computer to study without explicit programming. To produce predictions or judgments, machine-learning algorithms [3] are skilled on vast volumes of information. As a significant global industry, agriculture depends on farmers producing sustainable and lucrative crops. Poor crop selection can have a major impact on agricultural yield, lowering productivity and potentially costing farmers money. Farmers may find it difficult for the selected crops to flourish and reach their maximum yield potential if they neglect important aspects like soil conditions, market demand, and climate appropriateness. Inadequate climate adaptation can cause inappropriate crops to grow poorly, become more susceptible to pests and illnesses, and produce less overall. The financial burden on farmers may be exacerbated if commodities that don't meet market demand have trouble finding customers or obtaining favourable pricing. To overcome these obstacles and make wise choices that will optimize crop yield and guarantee agricultural sustainability over the long run, farmers can use machine learning-based crop recommendation systems.

Bibhu Dash et al: ICFCSAI-2025
pp. 87-95, 2026. IJCI – 2026 DOI:10.5121/ijci.2026.150109

Various data, such as weather and soil data, can be analyzed by crop recommendation systems. Machine-learning model can be trained using this data to forecast which crops are mainly expected to succeed in a certain area. Farmers can also learn the best strategies for cultivating particular crops from crop recommendation systems. Crop recommendation systems that identify the best crops to cultivate can assist boost crop yields while using fewer resources. Crop prediction systems can also increase agriculture ability to adapt to environment change [4]. Additionally, machine-learning can help with a number of other agricultural problems [5], such as forecasting crop yield, detecting pests and illnesses, maximizing crop yield, increasing water efficiency, lowering the usage of fertilizers and pesticides, managing soil, etc.

The world's population depends heavily on crops for both food and fibre. The World Resource Institute is working to find a sustainable way to feed 10 billion people by the year 2050. Consequently, it is crucial to increase the output of high-quality crops. Crop yields and profitability can be greatly impacted by the crops that are planted. Given the location, it is difficult to forecast which crop would thrive due to climate change and other environmental conditions [6]. In this paper, we recommend crops to farmers using machine learning. Collect and pre-process the dataset first. Next, we use features like temperature, humidity, rainfall, soil type, and soil pH value to train and test models. In order to see if the model works superior with a combination of attribute and to apply it as a new attribute in the same data-set, we also tried feature engineering approaches [7]. Figure 1. shows a construction for crop forecasting in agricultural yields. This framework includes all agricultural yield information. The data-set is pre-processed using data-cleaning techniques. The dataset can be classified using machine learning techniques once extraneous characteristics have been removed using a feature selection strategy. This aids in identifying the kind of dataset that works with machine learning algorithms [8]. For a given field, this aids in determining the kind and timing of the crop forecast.

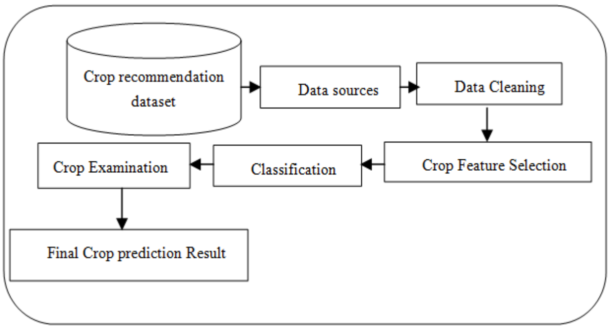

Figure 1. Diagram of the crop forecasting data-mining process

In this work, we recommend crops to farmers using machine learning. Gather and pre-process the dataset first. Next, we use features like temperature, humidity, rainfall, soil nutrients, and soil pH value to train and test models. In order to see if the model works better with a combination of features and to apply it as a new feature in the same dataset, we also tried feature engineering approaches. We therefore comprehensively emphasize the general issues in agriculture in the context of machine learning. Finally, we offer some intriguing concepts for the readers to explore.

2. LITERATURE REVIEW

G. Buvaanyaa and M. Gobi [9] present a crop recommendation system using hybrid classification algorithm. Comparing this hybrid classification algorithm with other classification algorithms such as Naive Bayes, the Random Forest algorithms reveal that it has high performance metrics.

According to selvi et al. [10], multispectral drone photos were taken during the growing season. Machine learning methods, such as SVR (support vector regression) and RFR (random forest regression), were used to aggregate cultivar information from a number of low indices. An original FS method called MRFE (modified recursive function elimination) was presented by Mariammal et al. [11]. Salient features are chosen and ranked using a classification algorithm in the suggested MRFE scheme. According to the survey results, the MRFE approaches choose the mainly accurate trait, but the bagging strategy correctly forecasts a suitable crop. Shafiulla Sheriff et al. have demonstrated a critical component for the continued existence and growth of the Indian financial system [12]. India was a significant manufacturer of a wide range of farming goods. Soil is an essential component of crop cultivation.

An analytical study was provided by Shahhosseini et al. [13] to show how crop combination model and machine-learning might develop corn output projections in the US Corn-Belt. They explained that in order to enhance performance forecasts, their recommended machine learning model need more hydrological input. Based on genotype and ecological data (soil & weather), Khaki & Wang [14] produced a deep neural network (NN)-based method to forecast maize hybrid yield, yields control, and yields variation. Through an RMSE of 12% of the mean performance and 5% of the standard variation for the test data-set utilize predict meteorological information, their representation was shown to be highly accurate.

Research on the forecast of potato tuber production was also carry out by Abbas et al. [15]. Based on relative soil and crop data gathered by a neighborhood survey, they assessed the yield of potato tubers (Solanum Tuberosum) by four machine-learning algorithms: SVR(support vector regression), (KNN)k-nearest neighbor, elastic network, and linear regression. A center for the study of agricultural yield in African nations was developed by Kaneko et al. [16]. Using district-level maize projections from satellite image data in six African nation-Ethiopia, Kenya, Malawi, Nigeria, Tanzania, and Zambia — they are developing a deep learning architecture. The most significant evaluation for producers and consumers is cropping productivity, where the consequences of climate change are most noticeable.

3. PROPOSED METHODOLOGY

Every experiment utilizing the suggested methods on a sizable benchmark dataset was victorious. The study method takes into description an efficient categorization process that is developed in this part and makes use of the data pre-processing, various machine learning algorithms, and classification.

3.1. Data Collection and Pre-Processing

The method for gathering and evaluating data from a lot of sources, like the Kaggle website, is data collecting. To acquire a system approximate dataset. The follow qualities should be there present in this dataset. Crop forecasts will take into account the following factors: Temperature, humidity, rainfall, crop data, Soil pH and NPK are the first seven factors. Data pre-processing includes data-cleaning [17]-[18]. It is necessary to pre-process the data prior to training the model. Reading the gathered data is the primary step in data pre-processing, which is followed by data cleaning. During data cleaning, several unnecessary features were eliminated from the data sets; these attributes are not engaged into description for making crop predictions. Consequently, hybrid feature selection should eliminate unnecessary attributes and datasets with missing values in order to increase accuracy. It is necessary to either eliminate these missing numbers or substitute undesired NaN (Not a Number) values [19].

3.2. Hybrid Feature Selection (HFS) Algorithm

The Recursive Feature Elimination (RFE) method, Forward Feature Selection (FFS), Support Vector Coefficient (SVC) served as the foundation for the Hybrid Feature Selection (HFS) strategy. The suggested HFS recursively eliminate the character at each stage and reclassifies the enduring features by regressing Support Vector Coefficient (SVC) on the equilibrium attribute. Only weak attributes will be eliminated using HFS. An average trait can also be useful when paired with other attributes. Therefore, the straightforward elimination of weak or redundant features can also be supported by the categorization presentation. In direct to ascertain the effects of the potentially weak feature; ERFS first assesses the categorization representation in conditions of the value of the corresponding weight vector.

Consider a data set Dt = {Dt_n = | n = 1,2,3,….,p} shown in equation (1) adapted to p quantity of data R^{nf}; $\in D_{nn}$ is the numeral of attribute $feat_i$. To define Dt^n_q as the q^{th} attribute feat, consequence of the data set D_n, each data Dt_n of the set Dt able to be symbolize by a vector.

$$D_n = \{d_n^1, d_n^2, \dots, d_n^{nf}\} \quad (1)$$

Suppose a feat set represent all features in equation (2),

$$\text{Feat} = \{\text{feat}_1, \text{feat}_2, \dots, \text{feat}_{nf}\} \quad (2)$$

Algorithm 1's feature selection procedure eliminates the weakest properties after the necessary number of features has been reached. By eliminating all potential dependencies and colinearity from the model, an ERFS seeks to recursively recall a limited number of characteristics every loop. According to the importance dimensions of coefficients or attributes of the model, the features are classified in order of importance [20]. Figure 2. below shows the selected features in the data set and Figure 3. showing the correlation matrix of the selected features. The below equation is used to calculate m_c, or the totality numeral of hyper planes, in the case where the data set contain further than two classes.

$$m_c = \frac{c\,(c-1)}{2} \quad (3)$$

Equation (3) refers the decision function.

Algorithm 1

1. Import necessary libraries
2. Load and preprocess the dataset
3. Split the data into training and testing sets
4. **Step 1: Initial Feature Ranking with Support Vector Coefficients (SVC)**

   Train an SVC on the training data

   model_svc = LinearSVC()
   model_svc.train(X_train, y_train)
   Get absolute values of feature coefficients
   sv_coefficients = abs(model_sv.coefficients)
   Rank features by descending coefficient magnitude
   ranked_features = sort_features_by(sv_coefficients)
5. **Step 2: Recursive Feature Elimination (RFE)**

   Initialize RFE with SVC as the base model

   rfe_selector = RFE (base_model=SVM, n_features_to_select='auto')
   Fit RFE on training data using the ranked features
   X_train_ranked = X_train[ranked_features]
   rfe_selector.fit(X_train_ranked, y_train)

Get the subset of features selected by RFE
rfe_features = rfe_selector.get_selected_features()

6. **Step 3: Forward Feature Selection (FFS) on RFE-selected features**
Initialize empty list: selected_features = []
For each feature in rfe_features:
Temporarily add feature to selected_features
Train SVC using selected_features only
Evaluate performance
If performance improves, keep the feature
Else, remove it
Final_features = selected_features

7. **Train Final Model**
Train a final SVC using only Final_features
model_final = SVC()
model_final.train(X_train[Final_features], y_train)

8. **Evaluate** Final Model
predictions = model_final.predict(X_test[Final_features])
score = evaluate(y_test, predictions)
print("Final model accuracy:", score)

9. **Print** the result.

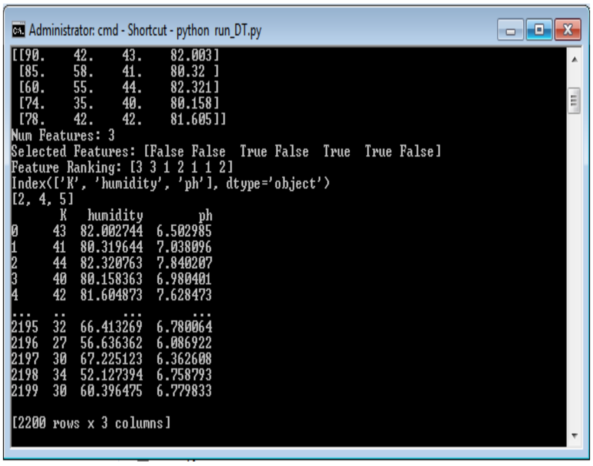

Figure 2. Selected Features

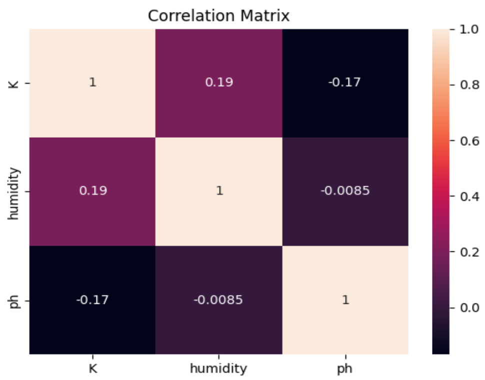

Figure 3. Selected Correlation Feature matrix

3.3. Machine Learning Algorithms Used

Despite the widespread use of numerous machine learning algorithms, we are only focusing on the following ones in this survey because that is what we employed in our research.

1. **Decision Tree:** A decision tree is a supervised learning algorithm that illustrates the link between input and output data using a structure resembling a tree. Nodes and branches make up a decision tree. Decisions are represented by nodes, and the potential results of those decisions are represented by branches.
2. **Random Forest:** An ensemble learning method made up of several decision trees is called a random forest. By training numerous decision trees on various subsets of the training data, random forests are produced. A random subset of the features is used to train each decision tree. Every decision tree in the random forest makes a prediction in order to make predictions. The mainstream vote on the forecasts from each decision tree is used to determine the final prediction. For problems involving regression and classification, random forests are frequently employed.

Machine learning algorithms like Random Forest (RF) and Support Vector Machine (SVM) have demonstrated strong predictive capabilities across various domains. Each algorithm has its own strengths and weaknesses. While RF excels at handling non-linear data and providing feature importance, SVM performs well in high-dimensional spaces with a clear margin of separation. Combining these two algorithms can lead to a robust hybrid model that leverages the advantages of both. While both Random Forest and SVM are strong classifiers, they excel under different conditions. Random Forest handles noise, missing data, and large feature spaces well, while SVM is excellent for high-dimensional, sparse datasets and decision boundaries with clear margins. Train both Random Forest and SVM separately and combine predictions via voting called Hard Voting. Hard Voting means Majority class from both models. Below algorithm 2 shows the working procedure of hybrid algorithm called random forest with support vector machine.

Algorithm 2: Hybrid RF+SVM

Input: Dataset D with features X and labels Y
Output: Predicted a labelled class

1. Data Preprocessing
    a. Handle missing values in X
    b. Split D into training set (X_train, Y_train) and test set (X_test, Y_test)
2. Feature Selection using hybrid feature selection process from algorithm 1
3. Train RF Classifier
    a. Train Random Forest classifier RF_model on (X_train, Y_train)
    b. Compute feature importance scores from RF_model
    c. Rank features by importance scores
    d. Select top K important features → X_train_selected, X_test_selected
4. Train SVM Classifier
    a. Initialize SVM with kernel
    b. Perform hyperparameter tuning (e.g., grid search on C, gamma) with cross-validation on X_train_selected
    c. Train SVM model SVM_model on X_train_selected with best hyperparameters
5. Prediction and Evaluation
    a. Predict labels Y_pred = SVM_model.predict(X_test_selected)
    b. Evaluate predictions using metrics like accuracy, precision, recall, F1-score
6. Return Y_pred and evaluation results.

4. RESULTS AND DISCUSSION

By using Python version 3.8 simulation, through experimental evaluation it is found that the recommended Hybrid Algorithm perform fine when run on a 3.21GHz x64 – based Intel I5 series pc with 8 GB memory and windows 10 OS.

The planned Hybrid method was used to compute dataset Recall, precision, F1-Score and accuracy measures. First, the dataset has been pre-processed and features have been selected using hybrid feature selection method. Then, the hybrid classification algorithm has improved for classification. By using this it has achieved a good accuracy on using hybrid classification algorithm. While comparing to other algorithms like Random Forest, Support Vector Machine, the hybrid algorithm has worked well and achieves a good accuracy level.

The performance measures have been used to validate the model. Because, have to know that our developed model is good or not. So, the performance measure has been played a role in it. The measures Precision, Recall, F1-Score and Accuracy has been calculated based on the model developed. The Results are shown in below figure and table. The figure 4. shows the confusion matrix of Hybrid Classification Algorithm. Then the table 1 shows the results for the Hybrid Algorithm.

Table 1: comparison results for the hybrid algorithm

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	86.89%	87.69%	86.89%	88.50%
SVM	87.16%	86.37%	85.58%	87.16%
Hybrid Algorithm	89.97%	88.78%	87.68%	89.03%

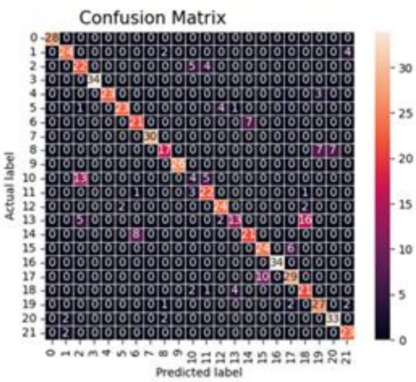

Figure 4. Confusion Matrix

5. CONCLUSION

In conclusion, this study has introduced crop recommendation models that use a diversity of sophisticated machine learning algorithms to forecast which crops will thrive. In that the hybrid Classification Algorithm has achieved 89.97%. The method is easily adaptable to new data and is scalable. The study's findings have a number of advantageous ramifications for the farming sector. First, producers can utilize the approach to make better selections about what crops to grow. Second, governments can utilize the process to create policies that help the agriculture industry. Third, companies can use the technology to develop new goods and services that benefit the agriculture sector; fourth, it will assist maintain stable prices for agricultural products. Next, it covered in detail the difficulties facing agriculture as well as some intriguing potential directions for future research. All things considered, this study has significantly advanced the subject of agriculture. The method is a useful tool for farmers, governments, and corporations since it is accurate, scalable, and simple to apply.

REFERENCES

[1] Yogesh Gandge, Sandhya, “A Study on Various Data Mining Techniques for Crop Yield Prediction”, 2017 International Conference on Electrical Electronics, Communication, Computer and Optimization Techniques (ICEEECCOT).

[2] R. Medar, V. S. Rajpurohit and S. Shweta, "Crop Yield Prediction using Machine Learning Techniques," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019, pp. 1-5.

[3] A. Nigam, S. Garg, A. Agrawal and P. Agrawal, "Crop Yield Predict ion Using Machine Learning Algorithms," 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019, pp. 125-130.

[4] G. Buvaanyaa, M. Gobi IDTC: Crop Recommendation System using Improved Decision Tree Classification Algorithm 2024 International Conference on Emerging Research in Computational Science (ICERCS), doi: 10.1109/ICERCS63125.2024.10895959.

[5] Lontsi Saadio Cedric, Wilfried Yves Hamilton Adoni, Rubby Aworka, Jérémie Thouakesseh Zoueu, “Crops yield prediction based on machine learning models: Case of West African countries”, Smart Agricultural Technology, Published by Elsevier, March 2022, https://doi.org/10.1016/j.atech.2022.100049.

[6] Dr. S. Selvi, Dr. M. Gobi, “Improving Cloud Data Security using Hyper Elliptical Curve Cryptography & Stenography”, Int. J. Scientific Research & Development, Vol. 5, Issue 4, 2017.

[7] S. Selvi, M. Gobi, “Hyper Elliptic Curve Cryptography in Multi Cloud – Security using DNA (genetic) techniques”, Int. Con. On computing Methodologies and Communication (ICCMC), July 2017.

[8] G. Buvaanyaa, M. Gobi, “Crop Recommendation System using K Nearest Neighbor Classification Algorithm”, Indian Journal of Natural Sciences, Volume 15, Issue 87, Dec 2024.

[9] G. Buvaanyaa, M. Gobi, “Crop Recommendation System Using Hybrid Classification Algorithm”, African Journal of Biological sciences”, Volume 6, Issue 14, Aug 2024, doi: 10.48047/AFJBS.6.14.2024.8913-8918.

[10] S. Selvi, M. Gobi, “Hyper Elliptic Curve based Homomorphic Encryption Scheme for Cloud Data Security”, Int. Con. On Intelligent Data Communication Technologies and Internet of Things (ICICI), 2018.

[11] G. Mariammal, A. Suruliandi, S. P. Raja, and E. Poongothai, “Prediction of land suitability for crop cultivation based on soil and environmental characteristics using modied recursive feature elimination technique with various classiers," IEEE Trans. Computat. Social Syst., vol. 8, no. 5, pp. 1132-1142, Oct. 2021.

[12] Shafiulla Shariff, R B Shwetha, O G Ramya, H Pushpa and K R Pooja, "Crop Recommendation using Machine Learning Techniques", Volume&Issue-ICEI-2022, vol. 10, no. 11.

[13] M. Shahhosseini, G. Hu, I. Huber, S. Archontoulis, Coupling machine learning and crop modeling improves crop yield prediction in the US corn belt, Sci. Rep. 11 (2021), doi: 10.1038/s41598-020-80820-1.

[14] S. Khaki, L. Wang, Crop yield prediction using deep neural networks, Front. Plant Sci. 10 (2019), doi: 10.3389/fpls.2019.00621.

[15] F. Abbas, H. Afzaal, A.A. Farooque, S. Tang, Crop yield prediction through proximal sensing and machine learning algorithms, Agronomy 10 (7) (2020), doi: 10.3390/agronomy10071046.

[16] A. Kaneko, T. Kennedy, L. Mei, C. Sintek, M. Burke, S. Ermon, D. Lobell, Deep learning for crop yield prediction in Africa, 2019.

[17] Dr. M. Gobi, R. Sridevi, “ECC Encryption and LSB Data Embedding Technique for Message Security”, Int. J. for Res. In Technological Studies, Vol.2, Issue 7, June 2015.

[18] G. Siva Brindha, Dr. M. Gobi, “CryptoData MR: Enhancing the Data Protection using Cryptographic Hash and Encryption/Decryption Through Map Reduce Programming Model”, International Conference on Innovative Computing and Communcations (ICICC) – 2023.

[19] M. Gobi, G. Buvaanyaa, “An Efficient Naïve Bayes Imputation Method for Missing Values”, IRJMETS, Vol.2, Issue 7, July-2020.

[20] R. S. Vindan, M. Gobi, “Pinning based Energy aware Computation Offloading for Mobile Cloud Computing”, First Int. Con. Computational Science and Technology (ICCST), 2022.

[21] Dr. S. Selvi, Dr. M. Gobi, “An Efficient Data Security Model using Hyper Elliptic Curve Cryptography and Stenography”, Int. J. Research and Development in Technology, Vol. 7, Issue 6, June – 2017.

[22] G. Siva Brindha, Dr. M. Gobi, “Improving the Map Reduce Performance using Symmetric Key Algorithm”, International Journal of Science and Research (IJSR), Vol. 10, Issue 4, April 2021.

Authors

G. Buvaanyaa is a Ph.D. Research Scholar in Full-Time at PG & Research Department of Computer Science, Chikkanna Government Arts College, Tiruppur, Affiliated to Bharathiar University, Coimbatore, she holds an MSC., and M.Phil., degrees, all completed at Bharathiar University. Her research focuses on Data Mining. She has published papers in International Journals and presented at conferences. She is passionate about teaching and mentoring students.

Dr. M. Gobi serves as the Associate Professor, Chikkanna Government Arts College, Tirupur. He holds MCA., M.Phil., and Ph.D., degrees, all completed at Bharathiar University. With over 29 years of teaching experience. He has published papers in International Journals and acted as reviewer for computer science Journals. He presented and attended many conferences. Acts as resource person in various institutions.