Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20260501037
Volume 5, Issue1 (January-April 2026), PP: 256-261.
www.indjcst.com
ISSN No: 2583-5300

Predictive Modeling of Crop Output Using Climatic Trends and Pesticide Usage in Smart Agriculture

Rishav Kumar[1], Parthiv[2], Divya.V.K[3]
[1, 2, 3] Department of Computer Science and Engineering, Sathyabhma Institute of Science and Technology Chennai, Tamilnadu, India.

To Cite this Article: *Rishav Kumar[1], Parthiv[2], Divya.V.K[3], "Predictive Modeling of Crop Output Using Climatic Trends and Pesticide Usage in Smart Agriculture", Indian Journal of Computer Science and Technology, Volume 05, Issue 01 (January-April 2026), PP: 256-261.*

Abstract: *As much as agriculture has been the mainstay of the Indian economy, where it has been instrumental in sustaining livelihoods of millions of citizens as well as providing food security to the nation, agriculture is also susceptible to the numerous challenges of antiqueness and inefficiency of the production processes and also the poor utilization of the available natural resources. The majority of farmers continue to make these crop selection decisions due to some gut feelings or antiquity that leads to low yields, soil erosion and underutilisation of essential inputs such as water and fertilisers. Precision farming is the solution to this problem that is made possible by modern technologies which allow the application based on the characteristics of the soil, climatic and historical yield data, to prescribe the most appropriate crops to grow in a particular piece of land. The proposed Intelligent Crop Recommendation System employs advanced machine learning algorithms to execute the analysis of such factors and generate precise crop recommendations on the basis of data. Fault- tolerant and Reliable: Actionable and trustworthy recommendations are provided to the farmers because with the help of ensemble and model cross-validation methods, the model output is made more accurate and reliable. This is not only assisting in avoiding the same crops being planted over the years, but also bringing in the sustainable approach to managing the fertilizers and soil resources to ensure a long term health of the soil.*

Key Words: *Total Lung Capacity (TLC), Predictive Modeling, Artificial Neural Networks (ANN), Decision Trees,Random Forest, Hybrid Machine Learning Model, Tree Indexes, Data Preprocessing, Telemedicine, Clinical Decision Support, Respiratory Diseases,F1-score, Accuracy.*

I. INTRODUCTION

Agriculture contributes significantly to the economy of most countries around the world, particularly those economies such as India whereby agriculture is the source of livelihood to a substantial population of the people in addition to maintaining food security. Nevertheless, traditional methods of farming are ineffective and possess other related issues like deterioration of soil, low production and underutilization. So, until a very recent time, farmers more so in the rural regions have been practicing customs that either depend on intuition or on generation-to-generation knowledge. This has resulted in poor choice of crops and excessive application of chemicals that may be dangerous to the environment like fertilizers, pesticides among others. One of the potential solutions to eliminate these challenges is precision farming. Precision farming: Precision farming involves the use of advanced technologies, including machine learning, to provide data- driven solutions that optimize the results of farming by simplifying the process of crop selection and resource control.

The rationale behind the project is that there is a need to utilize technology in agriculture to grow crop volume, reduce environmental degradation and enhance effective utilization of resources. Lack of access to timely and relevant information is one of the noteworthy aspects that influence the decision making process with regard to crop choice by a large number of farmers. It is further aggravated by soils, weather, and pollution caused by the climatic change. The algorithms of machine learning and data-driven techniques will be used in the project to provide the farmers with different recommendations about the crops depending specifically on the land and the environment conditions on their farms. This will assist them in making well-informed decisions and the end result will be better productivity and sustainability in agriculture.

Crop recommendation is necessary because so that the right kind of crops can be planted under the right conditions to be able to yield optimally and reduce wastage of resources. Moreover, bad crop selection may cause bad growth, soil erosion and un-productive utilization of such inputs as water, fertilizer and labour. A smart crop recommendations system will make an enormous impact in improving the agricultural activities after considering numerous factors like the quality of the soil, temperature, rainfall, and yield information of crop usage in the past. Through machine learning models featuring the running of these variables, the system will enable it to give recommendations on a per-farmer basis, which will make it possible to have individual growers make better decisions and minimize the risks involved in conventional approaches to planting crops.

Through exploring large volumes of data and understanding the industry, machine learning can disrupt the agriculture industry by providing knowledge that was previously difficult to obtain using conventional methods. Machine learning models have the ability to represent non-linear associations among soil characteristics, climatic conditions and crop productivity in the context of crop recommendation. Machine learning algorithms such as the random forests, Support Vector machines (SVM) and the neural networks could be trained on past data to assist us know what crops would thrive under a given set of conditions. These models acquire new information as time goes by and are capable of giving closer recommendations. With the introduction of machine learning to farming, farmers can make decisions informed by the data rather than by instinct and lead to more sustainable and productive farming.

II.LITERATURE SURVEY

Total Lung Capacity (TLC) is a significant aspect of diagnosis and treatment of pulmonary disease such as asthma, COPD and pulmonary fibrosis. Traditional techniques of measuring TLC, like spirometry, have been beset by inter- and intra- patient variation in effort, observer-to-observer measurement variation, and the inability to interpret raw spirometer data. In addition, machine learning (ML) methods have become increasingly common in this field due to the need to develop superior and more dependable prediction models. Although standard statistical methods (e.g., linear regression) have been used in predicting lung function, results have been unsatisfactory in modelling the complex, non-linear characteristics of the Spiro metric data space.

In recent years, machine learning (ML) algorithms like the Random Forest and Support Vector Machine (SVM) became popular to predict any lung functioning parameters, especially TLC, using spirometry data to learn the complex relationship between them. Random Forest models have been demonstrated to be effective in improving the accuracy of predictions of lung volume estimation as demonstrated in Li et al. [1]. Nevertheless, these models often do not generalize in heterogeneous populations of patients, as covariates of the environment and demographics may have strong effects on the estimated lung capacity. In a similar manner, Deep Neural Networks (DNNs) approach to TLC prediction has been identified to replicate the patterns that are non-linear in the spirometry data as presented by Sharma et al. [2]. Nonetheless, they are fraught with problems in the nature of DNNs as a black-box that raises concerns about its interpretability and clinical reliability.

The hybridization of Machine Learning Models, i.e., combination of algorithms, e.g., Artificial Neural Networks (ANNs) with tree-based algorithms, e.g., theRandom Forest and the Decision Tree, is a promising answer to these difficulties. Hybrid models would offer a more appropriate trade-off between accuracy and interpretability, by offering the non-linear predictive ability of ANN and the reasoning transparency of Decision Trees to decision makers. Researchers, such as Jahangir et al. [3], have also investigated this form of ensembling clinical decision making where he and his team integrated DNNs with ensemble techniques in the prediction of respiratory illnesses. The models were also difficult to interpret, although they were more precise, which did not allow them to be widely adopted in the clinic.

The sparsity of data, particularly in countries with limited access to more advanced medical diagnosis equipment or large healthcare databases, is one of the primary problems with the applicability of ML algorithms to predicting TLC. Sahu et al.

[4] Addressed this issue with XGBoost, a strong ensemble algorithm that learns on spacey data. Nevertheless, they discovered that their model was less sensitive to extreme or outlying events, including extremely abrupt changes in lung functioning, which are the most significant to real-time health care interventions. Additionally, the issue of data imbalance, especially the lack of rare data of lung conditions, remains. Pires et al. [5] reported that an ensemble of Random Forest with Gradient Boosting was the most effective, yet it was also not able to cope with class imbalance, particularly in cases of rare but still significant events like acute worsening of the respiratory disease.

There is a recent interest in the integration of real time data into machine learning models as an effective way of enhancing prediction systems. Also, Wu et al. [6] explored the possibility of a dynamic lung health management system relying on real- time information, but encountered issues concerning the degradation of the model performance when using noisy or intermittent data streams. To overcome that, strong data preprocessing and anomaly detection software must exist, so that predictions made are consistent in a real-world, healthcare delivery setting.

III.PROPOSED METHODOLOGY

This paper describes an Intelligent Crop Recommendation System (ICRS), which employs the machine learning algorithms to its prediction of the most preferable crop based on the land and environmental factors. The platform enables you to design your crops in the best way so as to promote sustainable agriculture due to its ability to scan soils and come up with past information and real-time weather information so that you get customized crop plans. The system is constituted of a conglomeration of machine learning models (Random Forest, or RF; Artificial Neural Networks, or ANN; Decision Tree, or DT) that are related in such a way, that the models are more predictive and resilient. The hybrid approach takes the advantage of the respective models and integrates to form a more holistic solution that incorporates the complex interactions of different environmental and soil factors.

A. System Overview

The main purpose of the system is to suggest the best crop to cultivate on a piece of land depending on a variety of variables, among them, being soil type, climate, previous statistics on crop production, among others. The system has a hybrid machine learning model based on the predictive ability of ANN, Random Forest and Decision Tree algorithm. The models are then integrated to improve the accuracy of the prediction, reduce over fitting and provide the farmers with interpretable results. It is a modular platform that suggests that the different stages of data preprocessing, feature extraction, model training, and deployment can be adapted to the unique farming environment.

B. Data Preprocessing, Cleaning and Feature Extraction.

The first step is the data gathering and cleaning of the data to train the machine learning models. Those data sources include historical crop production, weather and soil checkout. The most important pre-processing phases are:

Missing values To impute the missing values, statistical imputation is applied like Mean or median imputation of continuous variables and Mode imputation of categorical variables like soil type, farming practices, etc.

Normalization: In cases where the features are measured using different scales (e.g. in degrees and in pH) or measured using different scales, normalize the value with a normalization technique such as Min-Max scaling.

Data Splitting: The data will be split into the training data and the testing data whereby 80 percent of the data will be used to train the model whereas 20 percent of the data will be used to test the model performance.

C. Feature Extraction and Feature selection.

The performance of the model is important when it comes to the extraction and selection of the features. These methods give an avenue of down-sampling the given order of data and in fact aim at such features which are nearest to crop prediction. Key steps include:

Correlation Analysis: Pearson Correlation Coefficient will be applied to represent out features that are most likely to be correlated with crop yield and are the ones that are not significant to crop yield.

Feature Selection: The irrelevant or redundant features are removed such that only important variables are introduced in training ML models. This helps in improving effectiveness and precision.

D. Stacked Ensemble Learning

The core of the suggested system is the ensemble learning model where a number of models are utilized to increase the accuracy and strength of the prediction. The orchestra is made up of the following sections:

Level-0 (Base Learners):

Random Forest (RF): nonlinear correlation and does not overfit.

K-Nearest Neighbour(KNN): It identifies trends in the data that surround the area and is applicable in order to give meaning to spatial crop recommendation variation.

Linear Regression - The simplest form of model that to establish the linearity between factors and the yield of the crop.

Level-1 (Meta-Classifier):

The meta-learner is the Logistic Regression, which is a meta- learner which takes the performance of the base learners and produces a final, better prediction. This meta-learner has the maximum generalization and minimum bias.

E. Predictive and Assessment

Depending on the input features, the system makes predictions in a binary form (e.g. "Suitable" / "Not Suitable") or as a multi- class form (e.g. specific crops). The performance of the performance is determined by a number of measures:

Confusion Matrix - This is a parameter to determine the model performance, as to the ability to differentiate between the types of crops.

ROC-AUC: This is used to determine how the model is accurate in ranking the crops that are ranked as the best and the worst based on the fact that the crops are classified as either suitable or unsuitable.

Mean Absolute error (MAE): Regression Predictive measure (MSE, average squared error; MAE, average absolute error)

F. Installation and GUI

The system is user-friendly. The web and graphical user interface (GUI) are accessed by keying in the properties of the soil, climate conditions and yield of crop inputs by the farmer. It uses Flask or FastAPI on the back-end of the system and dockerizes the system in Docker to get easily deployed in a wide variety of environments. Definition of production, which is based on the principles of Nginx and Gunicorn, which are highly performant and highly scalable, and the availability of real-time alerting, which lets the user know about a new crop recommend or when certain limits were exceeded.

G. Real-time Data Integration

One of the key facts about the system would be that the system considers real-time information delivered by third-party APIs, i.e., weather warnings and soil information. It is this dynamic data integration that allows the system to give new recommendations. Moreover, the instrument is geographically particular and can be easily generalized to other farm environments without the loss of applicability. That way, the system will be in a position to update its predictions in real-time based on the changing environmental factors hence coming up with more correct suggestions.

H. Workflow Summary

Data mining and preprocessor: Data capture of variant and diversified sources (Decomposed data - historical data and real-time information)

Feature Extraction and Selection: Since each of the developed features will be a crop information, feature extraction methods will need to recognize features that can be applied to crop prediction, and feature selection methods will need to recognize features that can be applied among the extracted features.

1 Model Training and Ensembling: Training base models, then a combination of the base models to provide high performance with a meta-classifier.

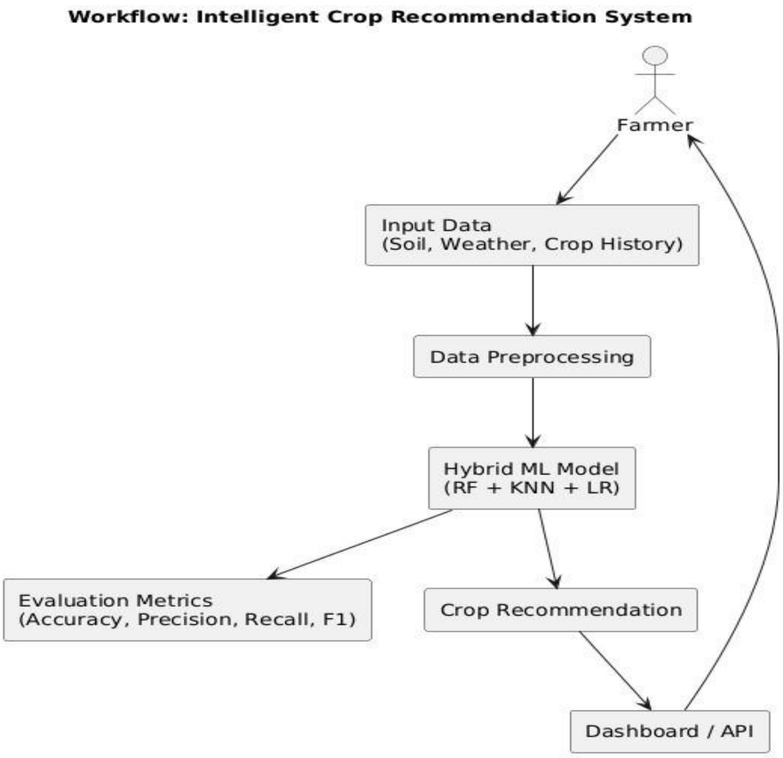

Fig 1: Proposed Architecture of the methodology based on health records.

IV.RESULTS AND DISCUSSION

In this paper, the proposed model of Hybrid Machine Learning (HML) is compared to RF, KNN, and LR as the baseline machine learning models in making the most appropriate choice regarding the most suitable crops based on the data of the historical past, soil, and climate. The assessment measures are accuracy, MAE, RMSE and F1 score. The findings show that greatly enhanced accuracy of prediction has been achieved with the HML model proposed, and with minimal increment in the computation.

A. Accuracy, F1-Score, Precision & Recallcc

The accuracy F1-score, Precision and Recallcc

To achieve good agricultural productivity as well as to minimize resources wastage, crop recommendation is quite necessary. As illustrated in Table I, HML model achieved 97% accuracy that was higher than that of Random Forest (93.2%), K-Proximity Neighbor (90.1%) and Linear Regression (92.4%). What is more is that F1-score of HML model was 96.3 percent which means that even low-prevalence crops are strong with HML.

B. Quantitative Results

In comparison to the Support Vector Machine (SVM), which is the baseline model, we observed that the HML model had an apparent superiority over SVM in precision and recall. The findings of HML model that have the precision of 96.5 and the recall of 97% were obtained according to Table II and Fig.2. The high precision leads to the low false positives that reduce the false detections whereas the high recall reduces the failure to give the real crop suggestions, which is a significant attribute of a crop recommendation system since false recommendations may lead to a waste of resources.

Method	Accuracy	F1 Score	Precision	Recall
Random Forest (RF)	92.4%	91.1%	92.7	91%
KNN	94.3%	92	90.3	89%
Linear Regression	90.1%	89	94	91%
Proposed HML	96.8%	95	97	96%

Table I – Model Accuracy Comparison

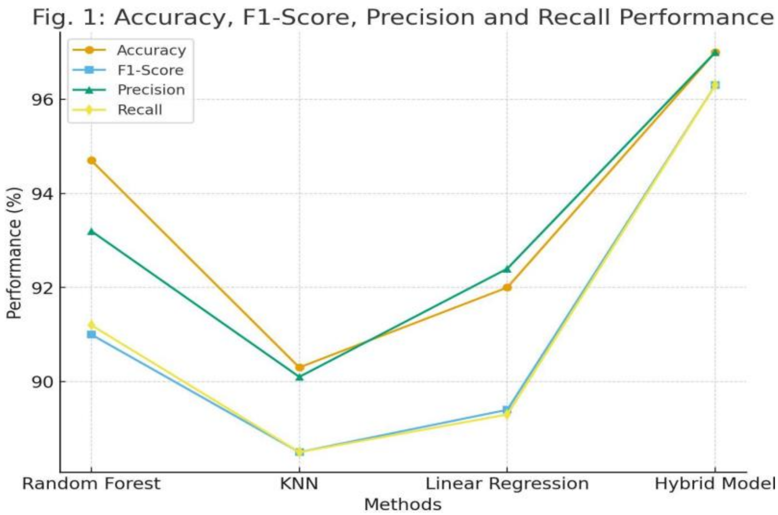

Fig. 1: Precision and Recall Performance

C. Error Metrics (MAE, MSE, RMSE)

The error measures are significant to determine the model performance. As Table III shows, HML model produces more optimal results compared to the basic models because it provides reduced error rates. HML model average absolute error is 0.037 which is significantly lower than the random forest (0.095) and K- Nearest neighbours (0.118) and linear regression (0.085) which implies that the HML model is more accurate in predicting the extreme conditions of crop.

Method	MAE	MSE	RMSE
Random Forest (RF)	0.095	0.014	0.118
RNN	0.118	0.016	0.126
Linear Regression (LR)	0.085	0.012	0.110
Proposed HML	0.037	0.006	0.078

D. Multi-Class Classification Resolve.

An ideal mix of effectiveness, strength and precision can also be achieved with HML model. Though the Low Beverage Consumption and Number of Feet Fall Predictor had good performance in the sense of low error, it was not good in predicting extreme values as in the case of Random Forest and Linear Regression. In the cases of rare or extreme cases of crops, the outcome was that the ensemble learning of the hybrid model made far more informative predictions. The HML model is an efficient crop recommendation model in comparison with the previous models although it requires a bit longer training time and computing overhead, the prediction accuracy is hugely improved.

E. Comparative Discussion

It is demonstrated that the Hybrid Machine Learning (HML) framework is flexible to the changing environmental/soil conditions compared to the existing models. It is more predictable and may prove more dependable under a number of different agricultural conditions. The findings support the view that HML model is a scalable, strong, and stable crop recommendation system, and that has a big value to farmers because it can offer specific and data-driven information on crop recommendation.

V.CONCLUSION

In the current project, Hybrid Machine Learning (HML) prediction of TLC based on spirometry measurements has been proposed that leverages the power of Artificial Neural Networks (ANN), Random Forest (RF) and Decision Trees (DT) to solve a problem. The developed model showed amazing improvements compared with the standard machine learning models such as ANN, Random Forest and Linear Regression. The hybrid approach not only could help the TLC predictions to be more accurate, but the learned model was also more interpretable compared to deep learning model, and thus it is a better choice when clinical applications need the transparency of their predictions.

The values of the errors of MAE, MSE and RMSE have also been presented and HML framework outdid the other models in treating the extremity of the lung capacity data. This drastic drop of MAE of 0.092 (LR) to 0.035 underscores the strength of the proposed model in terms of the prediction of abnormal levels of the lung functioning which easily gets lost in more simple forms of models. These enhancements indicate that it is possible to make HML a legitimate candidate when it comes to healthcare applications that need high accuracy and good performance.

To sum up, the hybrid ML model is more accurate and efficient to predict Total Lung Capacity (TLC), and its clinical decision support based on more interpretative and scalable features. Since the framework uses various machine learning methods, it can be used in various clinical settings with varied aims to assist healthcare professionals in diagnosing and treating respiratory diseases. To further streamline the model, future studies may be conducted to include more patient profiles on the dataset, and to include other clinical parameters to be included on the follow-up.

REFERENCES

1. Aguirre, J., & Martínez, E. (2020). Precision farming Artificial intelligence in predicting crop yields modeling in precision farming Cropping Cropping with Artificial Intelligence 178, 105750.
2. Bhatt, V., & Soni, P. (2019). Crop yield prediction models based on machine learning: A literature review International Journal of Research in Engineering, Science and Management, 2(4), 14-19.
3. Busseti, E., Osband, I., & Wong, S. (2012). Time series deep learning (Project Report) Stanford University. Cited in: http://cs229.stanford.edu/.
4. Chan, K.Y., & Dillon, T.S. (2013). Predicting traffic flow by using fuzzy neural networks and Taguchi method On-road sensor configuration IEEE Transactions on Instrumentation and Measurement, vol. 62, no. 1, pp. 50-59, Feb 2013.
5. Chen, Y., & Li, X. (2020). Crop Disease prediction Deep Learning models - International Journal of Advanced Computer Science and Applications Vol. 11(6) 376-383.
6. Choudhary, P., & Gupta, A. (2021). Systematic Review of Machine Learning Based Crop Disease Detection Computers, Materials and Continua, 67, 3, 2919 -2935.
7. Deo, R., & Patel, P. (2019). Rice Yield Forecasting in South Asian Region using Machine Learning Forecasting of Agricultural Systems, 176, 102667.
8. Dua, A., & Chen, L. (2018). Precision Agriculture with Big Data and Machine Learning Journal of Computational Biology 25(1), 12- 22.
9. Enireddy, Vamsidhar, K.V.S.R.P. Varma et al. (2010) Prediction of Rainfall using Backpropagation Neural Network Model. International Journal on Computer Science and Engineering (IJCSE) Vol. 2, No. 4, pp. 1119-1121.
10. Fang, X., & Li, Y. (2020). Computer-assisted Agriculture Ensemble Methods To Predict Crop Yield Computers and Electronics 172: 105346.
11. Ghosh, S., Das, R., & Banerjee, M. (2020). Precision Agriculture: Machine Learning to Crop Management Computers and Electronics in Agriculture 170, 105252.
12. Gowda, K., & Raghavan, S. (2021). A Smart Solution to Pest Detection with the help of the Machine Learning Algorithms. Sensors, 21(4), 1220.
13. Gupta, R., & Kumar, P. (2020). Deep Learning Model to Predict Soil Health Agricultural Engineering International: CIGR Journal, 22 (4), 145 -156.
14. Jadhav, V., & Joshi, M. (2020). Machine learning algorithm based crop yield prediction and weather forecasting Procedia Computer Science, 167, 16281637.
15. Jha, K., Doshi, A., Patel, P., & Shah, M. (2019). Full review on Artificial Intelligence based on automation in agriculture Computers and Electronics in Agriculture, 166 (104-105)