Kuwait Journal of Science 51 (2024) 100160

Contents lists available at ScienceDirect

Kuwait Journal of Science

journal homepage: www.sciencedirect.com/journal/kuwait-journal-of-science

Full length article

CropCast: Harvesting the future with interfused machine learning and advanced stacking ensemble for precise crop prediction

Chetan Raju [a,b,*], Ashoka D.V. [b], Ajay Prakash B.V. [c]

[a] Department of CSE, JSS Academy of Technical Education, Bengaluru, Visvesvaraya Technological University, Belagavi 590018, India
[b] Department of ISE, JSS Academy of Technical Education, Bengaluru, Visvesvaraya Technological University, Belagavi 590018, India
[c] Department of AI and ML, Raja Rajeswari College of Engineering, Bengaluru, Visvesvaraya Technological University, Belagavi 590018, India

ARTICLE INFO

Keywords:
Prediction
Ensemble
Base learners
Metamodel
Agriculture
Environment
Soil

ABSTRACT

The Agro-Ecological (AE) zone plays a vital role in determining which crops are suitable for cultivation in specific areas of land. However, changing environmental and climate conditions have made it increasingly difficult for farmers to choose the right crops, resulting in both time and financial losses. Traditional research in this field has struggled with issues like inaccurate predictions, crop variability, diversification, and a high error rate. This research introduces the Inter-fused Machine Learning with Advanced Stacking Ensemble model (IML-ASE) as a solution to improve crop prediction accuracy using AE zone data. The primary goal is to develop the IML-ASE approach, leveraging information about agricultural, environmental, and soil conditions within AE zones to help farmers make knowledgeable decisions about crop selection. The process involves collecting data from crop recommendation datasets, preprocessing it, and feeding it into the proposed model. The advanced stacking model consists of multiple layers, with the first layer using various ensemble techniques as base learners, the second layer acting as a meta-learner, and the third layer serving as the fine learner. This research addresses the challenges faced in modern agriculture by providing farmers with more accurate crop predictions based on AE zone characteristics. Prediction performance is assessed by metrics like mean absolute error, mean square error, root mean square error, accuracy (97.1%), F1-score (97.09%), precision (97.03%), recall (97.12%), and specificity (100%), allowing for comparisons across predictions. This advancement aims to empower farmers to make informed decisions and effectively adapt to the everchanging agricultural landscape.

Introduction

One of the critical occupations carried out in India is farming. It is the most varied economic sector and is vital to the country's growth (Elavarasan and Vincent, 2020). To meet the needs of 1.3 billion people, an additional 60% of the country's land is utilized for agriculture (Khaki and Wang, 2019). So, it is vital to implement current agricultural methodologies. This will direct our nation's farmers towards financial success (Khaki et al., 2020). Based on farmers' prior experience in a certain place, crop and yield predictions have previously been made. They have insufficient knowledge of the soil's nutritional contents, such as nitrogen, phosphate, and potassium (Shah Hosseini et al., 2021; Nigam et al., 2019; Palanivel and Surianarayanan, 2019). They would favour the neighbourhood's older, more established, or more fashionable crop exclusively for their property (Abbas et al., 2020). Due to the existing condition, there is a reduction in production, soil pollution (soil acidity), and damage to the top layer resulting from the lack of crop rotation and insufficient nutrient application (Gopal and Bhargavi, 2019). Also, several techniques for crop prediction do not provide sufficient accuracy and consume more time for a prediction (Patil et al., 2020).

Considering all these problems, this manuscript proposes an advanced stacking ensemble learning, in which the system is designed using machine learning (ML) classifiers interfused with ensemble learning for accurate crop prediction of various crops within low processing time (Sethy et al., 2020). The multi-disciplinary field of agrotechnology, ML, a subset of artificial intelligence, has risen along with huge data methodologies and better performance computing (Nevavuori et al., 2020). In recent years, machine learning classifiers have become increasingly important in crop prediction, a crucial component of agriculture (Bhojani and Bhatt, 2020; Khosla et al., 2020; Haghverdi et al., 2018). The agriculture industry can gain a lot from correctly applied approaches in this technology and data science age. ML, for example, is not some esoteric trick in the agricultural industry. It

* Corresponding author at: Department of CSE, JSS Academy of Technical Education, Bengaluru, Visvesvaraya Technological University, Belagavi 590018, India.

E-mail address: chetan.dhananjaya@gmail.com (C. Raju).

https://doi.org/10.1016/j.kjs.2023.11.009
Received 21 August 2023; Received in revised form 26 October 2023; Accepted 2 November 2023
Available online 1 December 2023

is a huddle of well-defined methods that gather detailed data and use classifiers to produce predicted results (Singh et al., 2018). In the field of machine learning, Ensemble methodologies, that use a set of base learners for training and integrate the predictions from all of them for final predictions, are gaining popularity. Comparing ensemble models to single models, the prediction performance is typically higher (Suresh et al., 2021). To improve model performance and boost the learning capacity of the models, ensemble learning algorithms combine several classifiers with various methods. Three common ensemble learning methodologies are boosting, bagging, and stacking. Crop prediction has not utilized the stacking-based ensemble technique, despite numerous successful implementations in computer vision and ML (Manrique-Silupu et al., 2021).

The datasets are utilized to train the model, which is then used to launch an outcome based on historical data. The dataset needs to be separated into training and test sets for prediction purposes (Brown et al., 2018). In a dataset, the training dataset is used to build a model, and the testing dataset is used to validate the model that was generated using the training dataset. Based on weather conditions and soil characteristics like soil, agricultural, climatic, and environmental conditions of the AE zones, this system would suggest the best crop for given land (Jaison, 2021; Wallach et al., 2018).

In light of the agricultural challenges faced by the traditional approaches, particularly concerning insufficient soil nutritional knowledge, crop choice, and accuracy in crop prediction, this research aims to address these issues through an advanced stacking ensemble learning approach. The proposed development involves a ML-based system that incorporates ensemble learning techniques to boost the accuracy and speed of crop prediction for various crops. This research leverages the synergy between the fields of agro-technology and ML, harnessing the potential of vast datasets and improved computational capabilities. While ML classifiers have gained prominence in crop prediction, the application of stacking-based ensemble techniques in agriculture remains largely un-explored. The primary aim is to demonstrate the success of this approach in accurately predicting crop outcomes based on historical data, weather conditions, and soil characteristics. This research aims to provide valuable insights for farmers to make knowledgeable decisions about crop selection, ultimately contributing to improved agricultural efficiency and sustainability.

Contributions

- This research work proposes the Interfused Machine Learning (IML) approach with an Advanced Stacking Ensemble (ASE) model for accurate various crop prediction, which uses the agricultural, environmental, and soil conditions of the AE zones to assist farmers in choosing the type of crop to plant.
- This model uses ML classifiers like Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), XGBoost Regression (XGBR), AdaBoost Regression (ABR), and Naive Bayes (NB) as base learners for crop prediction.
- An advanced stacking model is introduced by multiplying the weight coefficients of each classifier with the features for improving accuracy.
- This advanced ensemble stacking model acts as the meta-model (meta leaner and fine learner) which learns from the combined prediction results of base learners and provides the final prediction result.

Various research works have previously existed in the literature which are based on crop prediction using various techniques and aspects. Some of them are reviewed here. In 2023, Agriculture Production system Simulator (AP-SIM) (Ziliani et al., 2022) crop system presented a routinely utilized high-resolution CubeSat picture that was used to forecast yield. AP-SIM is used to develop a linear forecaster that links modelled yield to simulated Leaf Area Index (LAI). At the end of the growing season, a particle filter that includes Cube-Sat-LAI into APSIM is employed to ensure superior (3 m) yield forecasts. With a deep bond to separately obtained measures, yield spatial inconsistency was replicated practically well. But this method provides low spatial resolution.

A recognition technique of crops being flooded to the attached SWAT (Assessment Tackles of Soil, Water)-MOVFLOW (Flow Model of Modular Fixed Variance Groundwater) scheme (Jin et al., 2022) to precisely define and envisage crops being flooded prone areas under the AMIP6 (Attached Model Inter-comparison Project 6) climatic instances. This outcome depicted fine implementation outcome. But it is challenging to apply this crop being flooded-prone identification element to the field trial data. A smart farming system (Colombo-Mendoza et al., 2022) for crop production that is built on widely used cloud data storage and analytics services, low-cost IoT sensors, and low-cost IoT sensors is presented. The prediction of production volume from heterogeneous data sources is further supported by a novel data-mining technique that makes use of both crop production and climatic data. Using conventional machine learning methods and accessible historical data, this strategy was originally validated. This method performed effective feature extraction. But this method faces challenges in real-time implementation.

A framework for improving agricultural yield forecast accuracy that uses advanced spatial Independent Module Analysis (sI-MA) (Pham et al., 2022) and makes use of a grouping of Principal Component Analysis (PCA) and Machine Learning (ML) (i.e., PCA-ML grouping) to address the two problems. This paradigm, using Vietnam as an example, combines common Vegetation Condition Index/Thermal Condition Index (VCI/TCI) geographic variety within each of its subspaces. These two approaches flexibly process redundant input data. The PCA might not perform well with the VCI/TCI data since there might be more obvious noise present that would have an impact on the variance's maximization. A progressive ensemble regression crop yield prediction model (Iniyan and Jebakumar, 2022b) which uses precipitation, sun energy, maximum temperature, minimum temperature, and other phenotypic parameters to fore-cast crop yield. Based on phenotypic characteristics, this regression technique was used to forecast crop production for corn and soybean crops with improved prediction accuracy. Still, interpreting ensembles can be more challenging.

Problem statement

Several agricultural, soil, and environmental elements such as temperature, humidity, rainfall, moisture, and pH have an impact on agricultural production. Farmers in India continue to use the traditional methods they learned from their forefathers. However, the issue is that back then, when the climate was quite wholesome, everything went on schedule. Many things, however, have changed because of global warming and numerous other variables (Morales Alejandro and Francisco, 2023). The various existing methods have solved these issues but also have several drawbacks such as low spatial resolution, more challenges in real-time implementation, minimal accuracy rate, etc. Because of this concern, a novel technique called interfused machine learning with an advanced stacking ensemble model is introduced for accurate prediction of various crops. This framework successfully provides farmers with directions on which crops should be grown in the field.

Materials and methods

This research work utilizes IML-ASE in crop prediction. First, the environment, soil, climatic, and agricultural data from the crop recommendation dataset are given for preprocessing for cleaning and noise removal. The pre-processed data collection is split into two parts: testing and training. On the testing and training data, the proposed IML-ASE scheme is used for predicting suitable crops in the agroecological

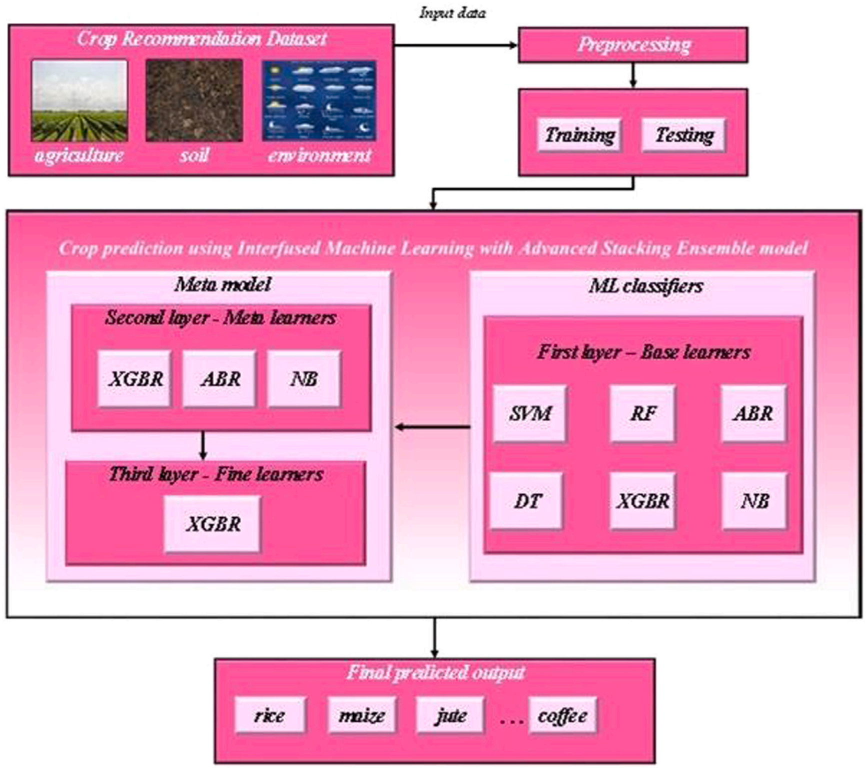

Fig. 1. The architecture of the IML-ASE model of crop prediction.

Table 1
Crop recommendation dataset.

Data	Explanation
N	Nitrogen
P	Phosphorous
K	Potassium
Temperature	in degree Celsius
Humidity	in percentage
pH	of soil
Rainfall	in millimetre

zones. In this model, advanced ensemble learning is employed, which employs six ML classifiers (SVM, DT, RF, XGBR, ABR, and NB) as base learners in its first layer. The average prediction results from the aforementioned ML classifiers are fed into the second layer, which consists of the meta-learner. The outcomes from the second layer are given to the fine learner for final prediction. These overall processes are depicted in Fig. 1. The performance of various classifiers and the proposed IML-ASE is evaluated by using several performance metrics such as Root Mean Square Error (RMSE), Mean Square Error (MSE), Root Mean Absolute Error (RMAE), accuracy, F1-score, precision, recall, and specificity.

Data acquisition

Data on rainfall, weather, soil, agriculture, and fertilizer available for India were augmented to form the crop recommendation database. The following Table 1 provides the descriptions of the dataset's various features (Ingle, 2020).

Data preprocessing

Before training the model, preprocessing of the environment, soil, and agriculture inputs from the crop recommendation dataset is required. The data sources include inappropriate, noisy, and missing data, which has a significant impact on how well machine learning models can learn. Any machine learning model's level of prediction accuracy is entirely dependent on data preprocessing. Reading the obtained dataset is the first stage of data preprocessing, which then moves on to data filtering. In data filtering, redundant characteristics and inconsistent data are removed from the dataset to improve accuracy. The dataset will be divided into a training and testing set after data filtering (Manjula and Narsimha, 2015). This pre-processed output is given to the proposed crop prediction model.

IML-ASE model for various crop prediction

A crop prediction system predicts the ideal crop which includes the information taken into account in the dataset. As in Fig. 2, the ensemble method originally began to gain information from one or more machine learning algorithms, and then compared expected and original values depending on those algorithms' forecast findings. The first layer makes use of several combiner techniques as base learners, while the second layer of training acts as meta learners, and the third layer acts as the fine learner. The meta-learner and fine learner together form the meta-model. The base learners are the machine learning frameworks. A meta-model derives information from the findings of basic learners. It compiles all of the basic learners' outcomes, completes its analysis while making errors, and gathers their experiences to make correct predictions by identifying features that were not fit for the model. With the information received from prior experience and learning skills, final predictions are made (Iniyan and Jebakumar, 2022a; Menahem et al., 2009).

Multiple base learners (i.e., the first-level predictors) are trained concurrently using the stacking technique, and the predictions generated by these base learners are subsequently employed to create new attributes for the meta-model. The meta-model undergoes refinement through advanced stacking, which comprises a meta-learner (i.e., the second-level predictor) and fine learners (i.e., the third-level predictors). Ultimately, the final prediction results of the model are determined based on the outputs of the second-level predictor.

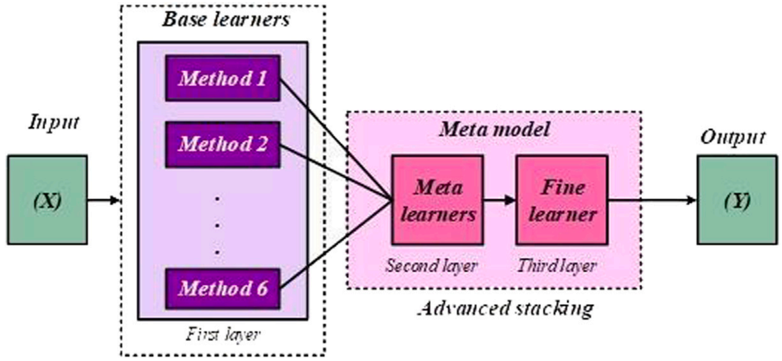

Fig. 2. Advanced stacked ensemble model.

Layer 1: Base learners

Here, the classifiers of ensemble stacking model-based crop prediction are Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), XGBoost Regression (XGBR), AdaBoost Regression (ABR), and Naive Bayes (NB).

Support Vector Machine An overlooked machine learning method called SVM is employed to address classification and prediction issues. In the case of a dataset with n attributes, SVM starts by displaying every data point in an n-dimensional area and assigning every point a position based on the rate of its attributes. From this point on, the prediction is carried out by choosing an appropriate hyperplane that, to the greatest extent possible, divides the points into two separate classes, i.e., suitable crop or not suitable crop for cultivation.

Decision Tree A decision tree is a particular kind of tree structure that looks like a flow diagram. and is frequently used in overlooked machine learning for prediction and classification tasks. Every path leading from the root node to every leaf node in a DT can be converted into a set of instructions, each of which is a rule. Every leaf node in a decision tree has a class assigned to it that is nearby if the feature meets the criteria of the outlet that leads to it. A decision tree's root node is compared to other dataset qualities and features to determine the ideal split. The outputs of one class should be spread along one side of the tree, while those of the other class should be distributed along the other, according to a perfect split.

Random Forest In terms of ensemble learning approaches, Random Forest falls under the Bagging subcategory. To obtain the most accurate forecast, the Random Forest builds and merges many decision trees. When dividing each node, the RF looks for the parameter with the highest significance before looking for the best within the collection of random attributes.

ADA Boost Regression It begins as a base learning technique with one attribute iterates several times with a similar training sample set, several poor models were developed, chooses scores based on the impact of classification, and after that balances and merges the classifiers to produce a powerful classifier which is represented in Eq. (1).

$$A(x) = \begin{cases} 1 & \text{if } \sum_{t=1}^{T} \alpha_t a_t(x) \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Where T is the total number of iterations, α_t is the base classifier produced in iteration t and its score is α_t.

XG Boost Regression It is a gradient-boosting approach that improves the efficiency and quickness of a machine-learning model that uses decision trees as its foundation. Due to the initial simplicity of the model, XGBoost will initially have lower accuracy. However, when the model iterates more, it will use the Gradient Descent technique to minimize the loss function. Up until a point when the model can no longer be optimized for loss, this process is repeated. As a result, the accuracy of the model will rise with more iterations.

Naive Bayes Widely utilized in various classification tasks is the Naive Bayes algorithm, which was derived from Bayes' principle to determine the likelihood of an event provided the presence of some other event. The Bayes' principle as shown in Eq. (2) seeks to figure out the probability that an occurrence will occur given the truth of another occurrence. The Gaussian, Bernoulli, and multinomial algorithms make up the three Naive Bayes techniques.

$$P'(b|A) = \frac{P'(A|b) \cdot P'(b)}{P'(A)} \tag{2}$$

Where, $P'(b|A)$ is the succeeding probability, $P'(A|b)$ is the chance, $P'(A)$ represents the demonstration, and $P'(b)$ denotes the preceding probability. Based on probability, naive Bayes predicts the result.

These classifiers' prediction outputs are examined and provided to the meta-model, which is then updated through advanced stacking.

Meta-model

Using the outputs of the base learners, the meta-models are developed. In this research work, the advanced stacking model acts as the meta-model. From the prediction results, XGBR, ABR, and NB are chosen as meta-learners. The main objective of the meta-learner model is to change the final output prediction to correct the inaccuracies caused by the base models. Layer one classifier predictions are output to layer two inputs. Layer 2 classifiers (meta learners) will use the predictions of layer 1 classifiers to construct their predictions. The layer 3 classifier (fine learners) will eventually produce a final prediction by combining the results from layers 2 and 3.

Layer 2: Meta learners

The second layer acts as the meta-learner. XGboost regression, Adaboost regression, and naive Bayes function as meta-learners that have learned from prior experience or base learners' findings. To increase accuracy, the advanced stacking method is introduced in this manuscript by multiplying the weights of each classifier in the meta-learners by the features. This work is completed by calculating the weight following Eq. (3).

$$W \propto \ln\left(\frac{P_{re}}{1 - P_{re}}\right) \tag{3}$$

Where W defines the classifier's weight and P_{re} signifies their level of precision. The weighted Stacking technique improves classification accuracy overall by giving different weights to various classifiers. The process has very high accuracy and excels at binary and multiple classification tasks, despite the underlying classifier's poor performance.

Layer 3: Fine learners

Making definitive predictions is the aim of this phase. The final forecast is provided by XGBoost Regression in the second layer which is informed by earlier predictions. The fine classifier analyses the circumstances under which one or more meta-classifiers make accurate

Table 2
Confusion matrix and description.

Confusion matrix	Representation	Description
True positive	t'_p	Accurately predicted as suitable crop
True negative	t'_n	Accurately predicted as not suitable crop
False positive	f'_p	Inaccurately predicted as suitable crop
False-negative	f'_n	Inaccurately predicted as not suitable crop

or inaccurate predictions during the training process. The final determination of the advanced stacking ensemble approach is based on the output of the fine classifier, which is a matrix of probabilities (one value for each class) (Dou et al., 2020; Suruliandi et al., 2021; Rao et al., 2022; Chetan et al., 2021, 2022). Finally, an evaluation of the two predictions is made.

Results and discussion

The interfused machine learning approach with an advanced stacking ensemble model of crop prediction scheme was trained and evaluated by the Python library. The crop recommendation dataset has details of soil, agricultural and environmental data. The performance of IML-ASE is compared with several existing methods such as Deep Neural Network (DNN) (Chetan et al., 2022), Boruta (Mohan and Patil, 2018) Machine Learning (ML) (Raja et al., 2022), and Ensemble Learning (EL) (Nischitha et al., 2020; Keerthana et al., 2021).

Performance metrics

The performance of prediction techniques is typically assessed using the measure of percent error, mean absolute error, mean square error, root mean error, accuracy, precision, and processing time using the following confusion matrices in Table 2.

Percent error, mean absolute error, mean square error, root mean square error, accuracy, precision, recall, and processing time are calculated using various equations.

- Percent error: The error rate in percentage is shown in Eq. (4),

$$P_e = \frac{\text{Predicted Value - Original Value}}{\text{Original Value}} \tag{4}$$

- Mean absolute error: The average absolute difference between original and forecasted values is shown in Eq. (5),

$$\text{MAE} = \frac{1}{N}\sum_{j=1}^{N}|y_j - \hat{y}_j| \tag{5}$$

- Mean square error: The average square of the absolute difference between the original and forecasted value is shown in Eq. (6),

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(y_j - \hat{y}_j)^2 \tag{6}$$

Root means square error: The square root of the mean square error is shown in Eq. (7),

$$RMSE = \sqrt{MSE} \tag{7}$$

- Accuracy: The correctly predicted value is shown in Eq. (8),

$$Acc = \frac{t'_p + t'_n}{t'_p + f'_p + t'_n + f'_n} \tag{8}$$

- Precision: The exactness of predicted values is shown in Eq. (9),

$$P_{re} = \frac{t'_p}{t'_p + f'_p} \tag{9}$$

Recall: The calculation for the recall is shown in Eq. (10).

$$R_c = \frac{t'_p}{t'_p + f'_n} \tag{10}$$

- Processing time: The time consumed for the prediction of crops is shown in Eq. (11).

$$P_t = I_c \cdot C_{pp} \cdot T_p \tag{11}$$

where, I_c represents the number of crops, C_{pp} denotes cycles per predictions and T_p processing time.

Performance evaluation

An experimental investigation of prediction using an IML-ASE methodology is presented in this section. The analysis is visualized in the form of graphs and tables. Fig. 3 depicts the confusion matrix. The confusion matrix evaluates the proposed IML-ASE performance. Rows characterize the predicted label instances and columns characterize the actual label instances. The overall accuracy of various crops is predicted as 0.971 from the confusion matrix's output. Fig. 4 illustrates the ROC curve of the introduced scheme compared with individual machine learning classifiers used in ensemble prediction. The true positive rate of the IML-ASE is higher than the several classifiers such as SVM, DT, RF, XGBR, ABR, and NB. The true positive rate which improves the prediction accuracy remains the same when the false positive rate increases.

Fig. 5(a–e) shows the performance metrics comparison of various machine learning classifiers used with advanced ensemble stacking in this manuscript. The prediction results of each classifier are calculated separately and compared with the overall prediction results. Classifiers such as XGBR, ABR, and NB perform better than SVM, DT, and RF. Therefore, SVM, DT, and RF are chosen as base learners, and the XGBR, ABR, and NB classifiers are chosen as the meta-learners in the meta-model. The weights of each classifier in the meta-learners are calculated and given to the next layer to multiply each weight with the features to improve accuracy. In the next stage, a super learner is selected from Fig. 5. XGBR performs better than ABR and NB. Therefore, XGBR is chosen as the super learner for the final prediction. The prediction results of each classifier are compared with the overall prediction results, i.e., IML-ASE prediction, and it is proven that the proposed method's prediction strategy attains a better range in multiple crop prediction with accuracy, F1-score, precision, recall, and specificity than individual classifiers.

Fig. 5(f) shows the error rate comparison of IML-ASE with ML classifiers. The average relative gap between the original and predicted measurements is represented by the MAE. It is proven that the MAE of the proposed model is 0.23%, which is lower than the individual machine learners. The mean square error represents the square of the average relative gap between the original and predicted measurements. The MSE of IML-ASE is 2.73%, showing a lower error rate for the proposed model. Then, the square root of MSE, i.e., the RMSE, is plotted, and it is proven that IML-ASE obtained a minimal error compared to individual classifiers within the range of 1.65%. Table 3 explains the performance of classifiers individually, and the base learners, meta-learners, and fine learners are selected according to the performance results.

Fig. 6 shows the evaluation of performance metrics of the IML-ASE with existing models. Compared with DNN, ML, EL, and a feature selection technique called Boruta, IML-ASE achieves a better performance in accuracy, f1-score, precision, specificity, and recall in the process of predicting various crops in the agricultural zones because in this proposed strategy machine learning classifiers are interfused

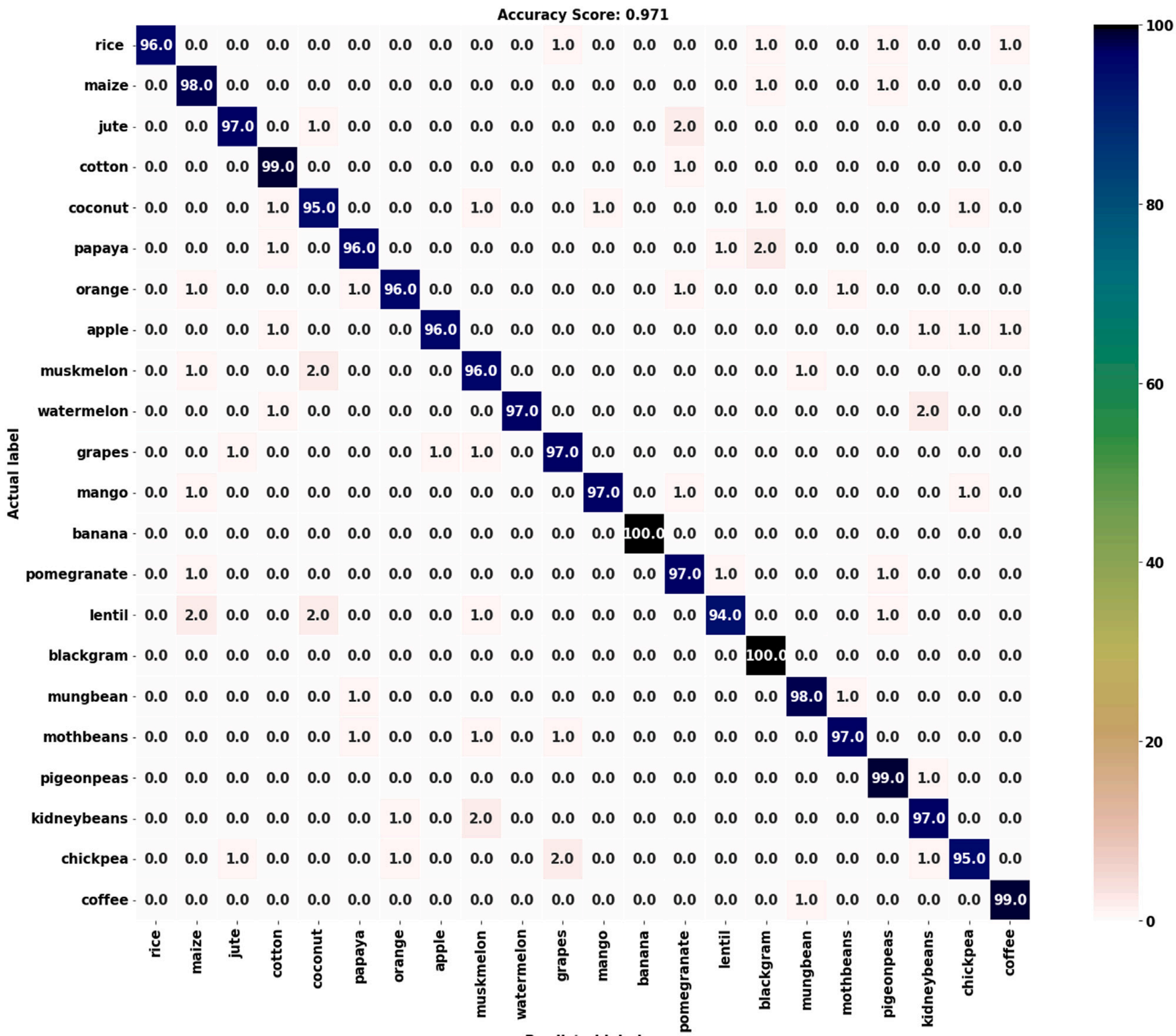

Fig. 3. Confusion matrix of the IML-ASE model.

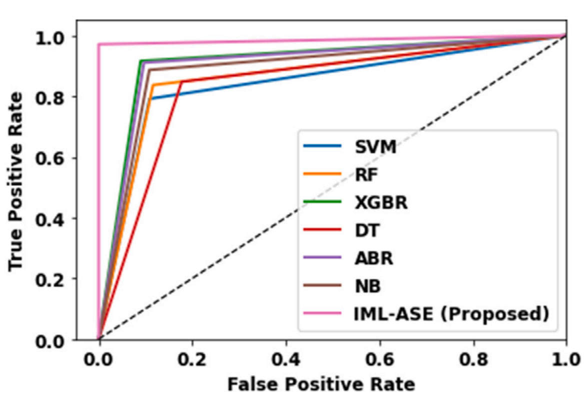

Fig. 4. ROC analysis of individual classifiers with IML-ASE model.

with ensemble advanced stacking. Table 4 shows the comparison of the proposed IML-ASE technique of crop prediction with various existing techniques.

The prediction results with and without the advanced model are depicted in Fig. 7. The advanced model uses three layers for the prediction of crops and in the second layer, the weights of each classifier are calculated. This weighted (advanced) Stacking technique increases overall classification accuracy than the normal stacking model by providing different weights to various classifiers. Therefore, the proposed advanced stacking ensemble performs better than normal ensemble stacking.

Conclusion

Thus, the prediction of various crops in agroecological zones using interfused machine learning approach with an advanced ensemble stacking model is successfully implemented using Python with an increased accuracy rate. The outcomes depict that the proposed approach outperformed several machine learning classifiers such as SVM, DT, RF, XGBR, ABR, NB, and other existing techniques such as DNN, ML, EL, and Boruta. The proposed IML-ASE performs with the minimum MAE and RMSE levels of 0.23% and 1.65% and a prediction accuracy of 97.1%. According to the agriculture, soil, and environmental factors considered, this framework successfully provides farmers with direction on which crop should be grown on the field. The overall processing

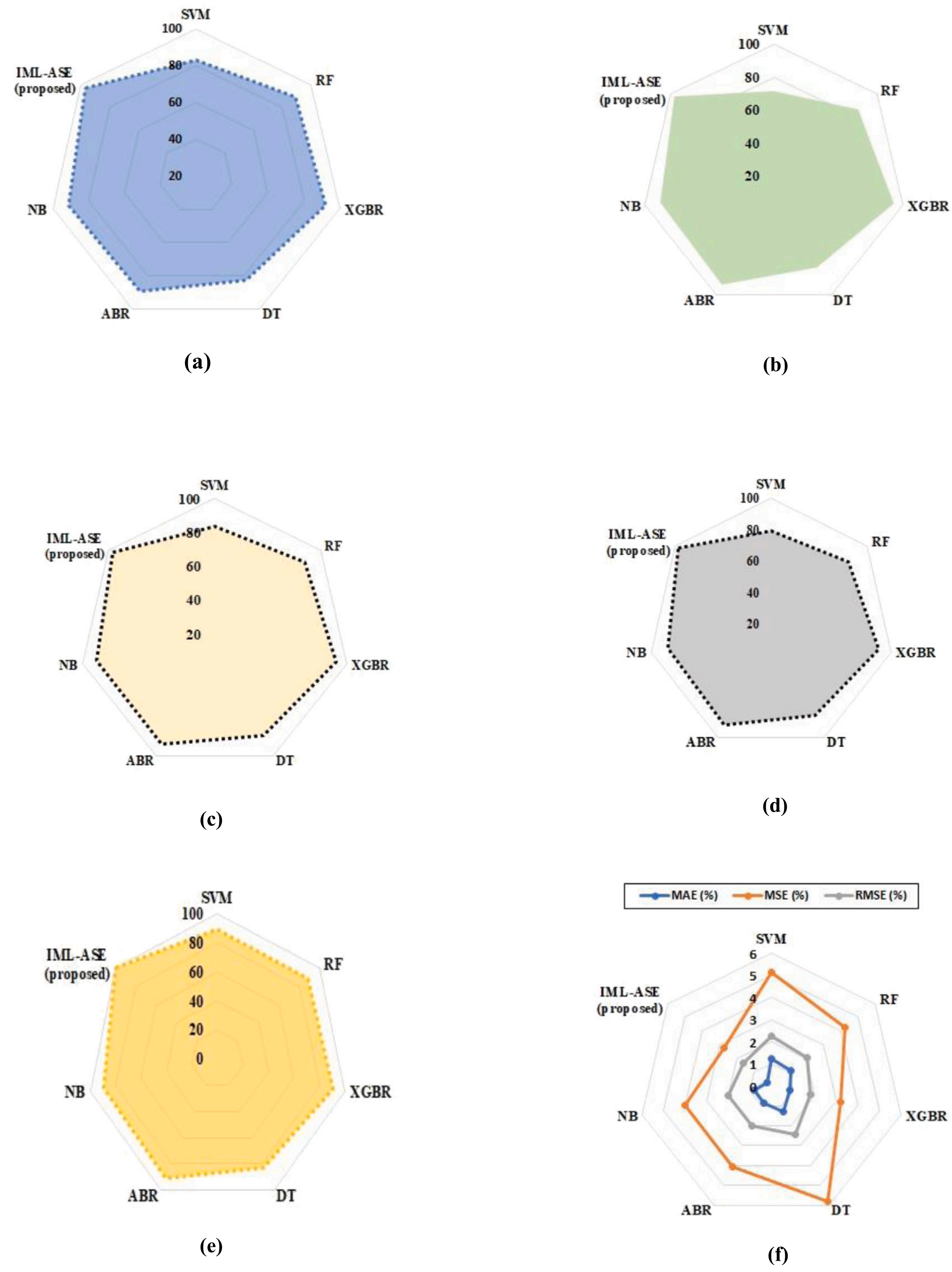

Fig. 5. (a) Accuracy (b) F1-score (c) Precision (d) Recall (e) Specificity (f) error rate comparison for various ML classifiers.

Table 3
Performance metrics comparison.

Metrics	SVM	RF	XGBR	DT	ABR	NB	IML-ASE (proposed)
Accuracy (%)	83.02	88.93	92.17	82.54	89.41	90.74	97.1
Specificity (%)	89.11	88.37	91.02	82.26	90.32	89.16	100
Precision (%)	83.78	88.15	93.34	86.54	91.74	91.28	97.03
Recall (%)	79.13	83.65	91.57	84.73	91.04	88.62	97.12
F1-score (%)	71.34	85.28	94.68	80.90	92.82	90.23	97.09
MAE (%)	1.23	1.14	0.83	1.28	0.85	0.83	0.23
MSE (%)	5.13	4.26	3.21	5.81	4.04	4.01	2.73
RMSE (%)	2.27	2.06	1.79	2.41	2.01	2.00	1.65

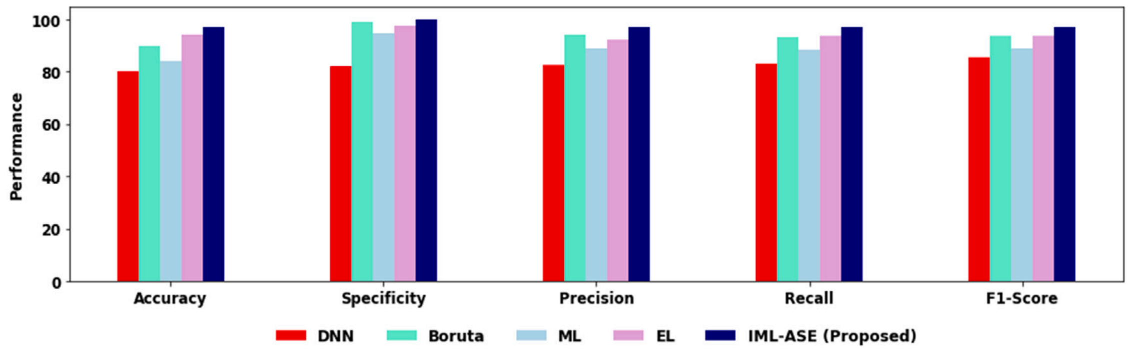

Fig. 6. Performance comparison of the IML-ASE model with existing models.

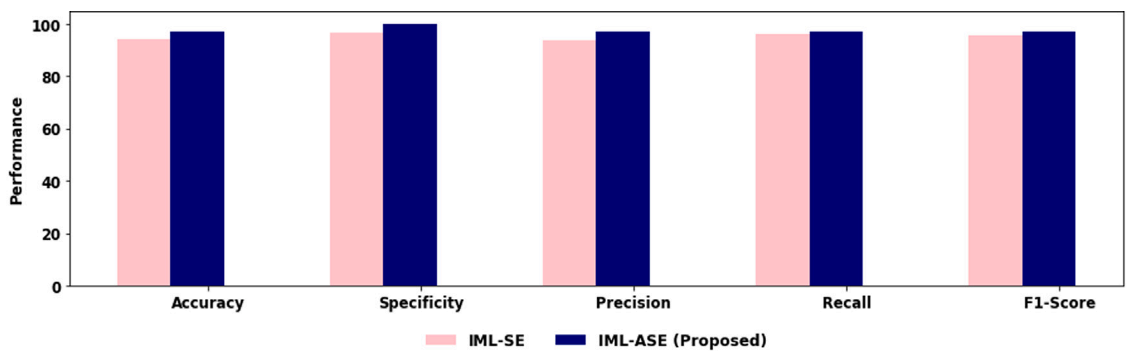

Fig. 7. Performance comparison of stacking ensemble with proposed advanced stacking.

Table 4
Performance comparison with existing techniques.

Techniques	Accuracy	Specificity	Precision	Recall	F1score
DNN	80.06	82.11	82.7	83.19	85.37
Boruta	89.7	98.83	94.14	93.24	93.68
ML	84	94.63	89.11	88.53	88.81
EL	94.43	97.68	92.37	93.64	93.62
IML-ASE (proposed)	97.1	100	97.03	97.12	97.09

time of this IML-ASE-based crop prediction is further minimized in the future by adding any optimization technique.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research work is supported by Visvesvaraya Technological University, Belagavi.

References

Abbas, F., Afzaal, H., Farooque, A.A., Tang, S., 2020. Crop yield prediction through proximal sensing and machine learning algorithms. Agronomy 10, 1–16. http://dx.doi.org/10.3390/agronomy10071046.

Bhojani, S.H., Bhatt, N., 2020. Wheat crop yield prediction using new activation functions in a neural network. Neural Comput. Appl. 32, 13941–13951. http://dx.doi.org/10.1007/s00521-020-04797-8.

Brown, J.N., Hochman, Z., Holzworth, D., Horan, H., 2018. Seasonal climate forecasts provide more definitive and accurate crop yield predictions. Agricult. Forest Meteorol. 260, 247–254. http://dx.doi.org/10.1016/j.agrformet.2018.06.001.

Chetan, R., Ashoka, D.V., Ajay Prakash, B.V., 2021. Smart agro-ecological zoning for crop suggestion and prediction using machine learning: A comprehensive review. In: Chiplunkar, N.N., Fukao, T. (Eds.), Advances in Artificial Intelligence and Data Engineering. AIDE 2019, In: Advances in Intelligent Systems and Computing, vol. 1133, Springer, Singapore, 2021, pp. 1273–1280. http://dx.doi.org/10.1007/978-981-15-3514-7_94.

Chetan, R., Ashoka, D.V., Ajay Prakash, B.V., 2022. IMLAPC: Interfused machine learning approach for prediction of crops. Revue d'Intell. Artif. 36, 169–174. http://dx.doi.org/10.18280/ria.360120.

Colombo-Mendoza, L.O., Paredes-Valverde, M.A., Salas-Zarate, MdP, Valencia-Garcia, R., 2022. Internet of things-driven data mining for smart crop production prediction in the peasant farming domain. Appl. Sci. 12, 1–19. http://dx.doi.org/10.3390/app12041940.

Dou, J., Yunus, A.P., Bui, D.T., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.W., Han, Z., Pham, B.T., 2020. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed. Japan. Landslides 17, 641–658. http://dx.doi.org/10.1007/s10346-019-01286-5.

Elavarasan, D., Vincent, P.M.D., 2020. Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. IEEE Access 8, 86886–86901. http://dx.doi.org/10.1109/ACCESS.2020.2992480.

Gopal, P.M., Bhargavi, R., 2019. A novel approach for efficient crop yield prediction. Comput. Electron. Agric. 165, 1–11. http://dx.doi.org/10.1016/j.compag.2019.104968.

Haghverdi, A., Washington-Allen, R.A., Leib, B.G., 2018. Prediction of cotton lint yield from phenology of crop indices using artificial neural networks. Comput. Electron. Agric. 152, 186–197. http://dx.doi.org/10.1016/j.compag.2018.07.021.

Ingle, Atharva, 2020. Crop recommendation dataset. https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset. (Accessed 10 June 2022).

Iniyan, S., Jebakumar, R., 2022a. Crop yield prediction on soybean crop applying multi-layer stacked ensemble learning technique. In: Manogaran, G., Shanthini, A., Vadivu, G. (Eds.), Proceedings of International Conference on Deep Learning, Computing, and Intelligence. In: Advances in Intelligent Systems and Computing, Springer, Singapore, 2022, pp. 335–348. http://dx.doi.org/10.1007/978-981-16-5652-1_29.

Iniyan, S., Jebakumar, R., 2022b. Mutual information feature selection (MIFS) based crop yield prediction on corn and soybean crops using multilayer stacked ensemble regression (MSER). Wirel. Pers. Commun. 126, 1935–1964. http://dx.doi.org/10.1007/s11277-021-08712-9.

Jaison, B., 2021. Adaptive lemuria: A progressive future crop prediction algorithm using data mining. Sustain. Comput., Inform. Syst. 31, 1–14. http://dx.doi.org/10.1016/j.suscom.2021.100577.

Jin, X., Jin, Y., Zhai, J., Fu, D., Mao, X., 2022. Identification and prediction of crop Waterlogging Risk Areas under the impact of climate change. Water 14, 1–21. http://dx.doi.org/10.3390/w14121956.

Keerthana, M., Meghana, K.J.M., Pravallika, S., Kavitha, M., 2021. An ensemble algorithm for crop yield prediction. In: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks. ICICV, pp. 963–970. http://dx.doi.org/10.1109/ICICV50876.2021.9388479.

Khaki, S., Wang, L., 2019. Crop yield prediction using deep neural networks. Front. Plant Sci. 10, 1–10. http://dx.doi.org/10.3389/fpls.2019.00621.

Khaki, S., Wang, L., Archontoulis, S.V., 2020. A CNN-rnn framework for crop yield prediction. Front. Plant Sci. 10, 1–14. http://dx.doi.org/10.3389/fpls.2019.01750.

Khosla, E., Dharavath, R., Priya, R., 2020. Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression. Environ., Dev. Sustain. 22, 5687–5708. http://dx.doi.org/10.1007/s10668-019-00445-x.

Manjula, A., Narsimha, G., 2015. XCYPF: A flexible and extensible framework for agricultural crop yield prediction. In: IEEE 9th International Conference on Intelligent Systems and Control. pp. 1–5. http://dx.doi.org/10.1109/ISCO.2015.7282311.

Manrique-Silupu, J., Campos, J.C., Paiva, E., Ipanaqué, W., 2021. Thrips incidence prediction in organic banana crop with machine learning. Heliyon 7, 1-9. http://dx.doi.org/10.1016/j.heliyon.2021.e08575.

Menahem, E., Rokach, L., Elovici, Y., 2009. Troika: An improved stacking schema for classification tasks. Inform. Sci. 179, 4097–4122. http://dx.doi.org/10.1016/j.ins.2009.08.025.

Mohan, P., Patil, K.K., 2018. Deep learning based weighted SOM to forecast weather and crop prediction for agriculture application. Int. J. Intell. Eng. Syst. 11, 167–176. http://dx.doi.org/10.22266/ijies2018.0831.17.

Morales Alejandro, J., Francisco, Villalobos, 2023. Using machine learning for crop yield prediction in the past or the future. Front. Plant Sci. 14, 1–13. http://dx.doi.org/10.3389/fpls.2023.1128388.

Nevavuori, P., Narra, N., Linna, P., Lipping, T., 2020. Crop yield prediction using multitemporal UAV data and spatio-temporal deep learning models. Remote Sens. 12, 1–18. http://dx.doi.org/10.3390/rs12234000.

Nigam, A., Garg, S., Agrawal, A., Agrawal, P., 2019. Crop yield prediction using machine learning algorithms. In: Fifth International Conference on Image Information Processing. ICIIP, pp. 125–130. http://dx.doi.org/10.1109/ICIIP47207.2019.8985951.

Nischitha, K., Dhanush Vishwakarma, M.N., Ashwini, M.M., 2020. Crop prediction using machine learning approaches. Int. J. Eng. Res. Technol. 9, 23–26. http://dx.doi.org/10.17577/IJERTV9IS080029.

Palanivel, Kodimalar, Surianarayanan, Chellammal, 2019. An approach for prediction of crop yield using machine learning and big data techniques. Int. J. Comput. Eng. Technol. 10, 110–118, http://iaeme.com/Home/issue/IJCET?Volume=10&Issue=3.

Patil, P., Panpatil, V., Kokate, S., 2020. Crop prediction system using machine learning algorithms. Int. Res. J. Eng. Technol. 7, 748–753, https://www.irjet.net/archives/V7/i2/IRJET-V7I2163.pdf.

Pham, H.T., Awange, J., Kuhn, M., Nguyen, B.V., Bui, L.K., 2022. Enhancing crop yield prediction utilizing machine learning on satellite-based vegetation health indices. Sensors 22, 1–19. http://dx.doi.org/10.3390/s22030719.

Raja, S.P., Sawicka, B., Stamenkovic, Z., Mariammal, G., 2022. Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers. IEEE Access 10, 23625–23641. http://dx.doi.org/10.1109/ACCESS.2022.3154350.

Rao, Madhuri Shripathi, Singh, Arushi, Subba Reddy, N.V., Acharya, Dinesh U., 2022. Crop prediction using machine learning. J. Phys. Conf. Ser. 2161, 1–11. http://dx.doi.org/10.1088/1742-6596/2161/1/012033.

Sethy, P.K., Barpanda, N.K., Rath, A.K., Behera, S.K., 2020. Nitrogen deficiency prediction of rice crop based on convolutional neural network. J. Ambient Intell. Humaniz. Comput. 11, 5703–5711. http://dx.doi.org/10.1007/s12652-020-01938-8.

Shah Hosseini, M., Hu, G., Huber, I., Archontoulis, S.V., 2021. Coupling machine learning and crop modeling improves crop yield prediction in the US corn belt. Sci. Rep. 11, 1–15. http://dx.doi.org/10.1038/s41598-020-80820-1.

Singh, M.C., Singh, J.P., Singh, K.G., 2018. Development of a microclimate model for prediction of temperatures inside a naturally ventilated greenhouse under cucumber crop in soilless media. Comput. Electron. Agric. 154, 227–238. http://dx.doi.org/10.1016/j.compag.2018.08.044.

Suresh, G., Kumar, A.S., Lekashri, S., Manikandan, R., 2021. Efficient crop yield recommendation system using machine learning for digital farming. Int. J. Modern Agric. 10, 906–914, https://www.modern-journals.com/index.php/ijma/article/view/688.

Suruliandi, A., Mariammal, G., Raja, S.P., 2021. Crop prediction based on soil and environmental characteristics using feature selection techniques. Math. Comput. Model. Dyn. Syst. 27, 117–140. http://dx.doi.org/10.1080/13873954.2021.1882505.

Wallach, D., Martre, P., Liu, B., et al., 2018. Multi-model ensembles improve predictions of crop environment management interactions. Global Change Biol. 24, 5072–5083. http://dx.doi.org/10.1111/gcb.14411.

Ziliani, M.G., Altaf, M.U., Aragon, B., Houborg, R., Franz, T.E., Lu, Y., Sheffield, J., Hoteit, I., McCabe, M.F., 2022. Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop model. Agricult. Forest Meteorol. 313, 1–15. http://dx.doi.org/10.1016/j.agrformet.2021.108736.