International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

An Effective Approach for Crop Recommendation with Using Features of Specific Locations and Seasons and Maximize Crop Yield Production by Using Machine Learning

Dr. Pawan*[1], Ms. Deepika Yadav[2], Mr. Ram Kumar Sharma[3], Mr. Mukesh Kumar[4], Ms. Jyoti Rani[5], Dr. Nidhi Sharma[6]

Submitted: 28/12/2023 **Revised**: 05/02/2024 **Accepted**: 11/02/2024

Abstract: Crop recommendation is a crucial task for farmers to improve their productivity and profitability. However, traditional methods of crop recommendation are often based on heuristic rules or expert knowledge, which may not be accurate or adaptive to the changing environmental and market conditions. Therefore, there is a need for a data-driven approach that can leverage the available information on soil, weather, and crop to provide optimal crop suggestions for farmers. In this paper, we propose a machine-learning algorithm that can recommend suitable crops for a given location and season. The algorithm uses a supervised learning method to train a classification model on a large dataset of historical crop data from various sources. The model takes as input the soil parameters, such as pH, nitrogen, phosphorus, potassium, etc., and the weather parameters, such as temperature, rainfall, humidity, etc., and outputs the most probable crop that can be grown in that location and season[1][8][19][20]. The results show that the algorithm can achieve high accuracy and precision in recommending crops that are suitable for the given conditions.

***Keywords:**: Crop Prediction, Machine Learning, Deep Learning, Feature Selection, Artificial Intelligence*

1. Introduction

The selection of crops in every season is a crucial task for farmers because choosing the right crop has a very significant impact on the farmer's income, food security, and environmental sustainability. However, crop recommendation is not a simple task., it involves considering multiple factors, such as soil quality, weather conditions and crop characteristics [6][]7]. Moreover, these factors are dynamic and complex and may change from year to or even from month to month. Therefore, relying on traditional methods of crop recommendation, such as trial and error, expert advice, or rule-based idea may not be sufficient or efficient.

The Machine learning helps the computer to understand Machine learning helps the computer to understand data and make final decisions and offers a new perspective on the crop recommendation problem. Machine learning can help farmers find the optimal crop for their land by analyzing large amounts of data from different sources, such as soil sensors, weather stations, satellite images, crop databases, and market trends. It also provides timely recommendations for selecting the crop and also helps to fulfil needs and requirements. Nowadays Machine learning-based techniques have been applied to various areas of agriculture, like crop yield prediction, pest detection, weed control, irrigation management, soil mapping etc[9][10]. However, crop recommendation is still a relatively new challenge for researchers. Several issues need to be addressed, such as data availability and quality, model selection and evaluation, interpretability, scalability and robustness, and ethical and social implications[21][22]. In this paper, we try to explore how machine learning can help solve the problem of crop recommendation. After reviewing the existing literature, it has been found that crop selection is a very big challenge and needs some scope for improvement. We propose a novel machine-learning algorithm that can recommend suitable crops for any location and season, based on soil and weather parameters. We evaluated the performance of our algorithm on real-world data and compared it with existing methods of crop recommendations. This research paper aims to contribute to the advancement of machine learning in the field of agriculture and to provide a useful tool for farmers to improve their productivity and profitability. We hope that this project willinspire more research and innovation in this domain, and

[1]Department of Computer Science & Engineering, Satyug Darshan Institute of Engineering and Technology, 121002,Haryana,India.
[2]Department of Computer Science & Engineering, Bharati Vidyapeeth's College of Engineering, 110063,New Delhi, India.
[3]Department of Computer Science & Engineering, Raj Kumar Goel Institute of Technology, 201003, Uttar Pradesh,India
[4]Department of Computer Science & Engineering, Satyug Darshan Institute of Engineering and Technology, 121002,Haryana,India.
[5]Department of Computer Science & Engineering, Satyug Darshan Institute of Engineering and Technology, 121002,Haryana,India.
[6]Department of Computer Science & Engineering, DPG Institute of Technology & Management, 122004, Haryana,India
** Corresponding Author Email: drpawancse@gmail.com*
Orchid Id:0009-0004-4114-0090

ultimately lead to a more sustainable and resilient agriculture system[13][14][23][25].

2. Literature Review

A related study of algorithms was given by Rashi Agarwal in his research paper. The implemented algorithm was based on decision trees, KNNs, Random Forests and neural networks. The proposed algorithm was very useful for farmers in making the right decisions about the production of crops on different environmental and geographical factors. A similar study was also given by Priyadharshini in her research article. The related work reduces crop failure, helps in choosing the right crops and increases production. Another useful related research is conducted by Mangesh Pande, in her research article. The algorithm provides a farmer-friendly and very productive forecasting system that connects all farmers via a mobile application. The location of users is identified by GPS and the main feature of the algorithm is that it gives 95%accuracy with the provided data set a very useful research was conducted by Mayank Champaneri that used the data mining technique for the prediction of crops They used a random forest classifier because it can perform classification and regression tasks[2][3][4][5]. The user-friendly website built that can be used by anyone to predict crop yield for their choice of crop by giving climate data for that area.

3. Methodologies & Experiments

Problem Definition

The selection of the best crop to grow in specific locations and seasons depends upon various parameters like quality of soil, conditions of weather, characteristics of crops, demand and profit. The solutions can be identified by using the supervised learning method, where the input is a set offeatures that describe the location and season, such as soil parameters (pH, nitrogen, phosphorus, potassium, etc.), weather parameters (temperature, rainfall, humidity, etc.) and the output is a label that indicates the optimal crop to grow in that location and season. The benefit of the problem is that it helps the farmers for maximizing their yield and income. It can also help farmers to reduce the environmental impact of agriculture, by suggesting crops that can minimize the use of water, fertilizer, pesticide, etc. It can also help farmers to adapt to the changing climate and market conditions, by suggesting crops that are resilient and profitable[6][7][17][18].

Data Collection

The study" A Machine Learning Algorithm for Crop Recommendation Using Features of a Specific Area to Maximize Crop Yield Production" will use a model to find out the best crop for given conditions.In crop recommendation using machine learning, data collection is an essential step to obtain relevant and reliable data that help to build the models that can suggest the best crop to grow in a given location and season.

The data collection process described in the study is comprehensive, well-documented, and praiseworthy. The research team has carefully selected and compiled datasets relevant to crop recommendation, ensuring a diverse and representative sample. The data sources, which include various agriculture research, farmer surveys, and environmental variables, are explained in detail, allowing readers to understand their origin and significance. Furthermore, the data preprocessing steps, such as handling missing values and data normalization, are meticulously described, ensuring transparency and replicability. The use of data from multiple sources enhances the study's robustness and increases the accuracy of the models[3][11][[4][12].

Data Pre-processing

The most critical step in any study is data pre-processing which is the process of transforming and cleaning the raw data before applying machine learning algorithms. Our study used different techniques for the data pre-processing such as removing the noise, missing values, duplicates, inconsistencies and irrelevant features from the data. It also enhances the performance and interpretability of machine learning models by scaling, normalizing, and encoding the data. Data pre-processing is an essential step for crop recommendation using machine learning, as it can help to deal with the challenges and limitations of the datasets. The data pre-processing also helps to provide the best and most accurate data to the training model and also helps to design the best training model and reduce the overfitting problems also[7][15][16][24].The distribution of some of the features is given below in Fig 1 and Fig 2.

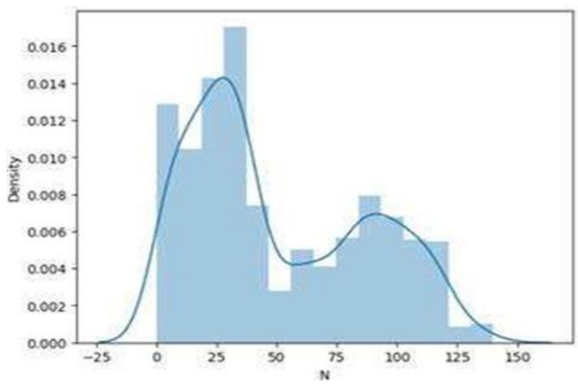

Fig 1. Temperature

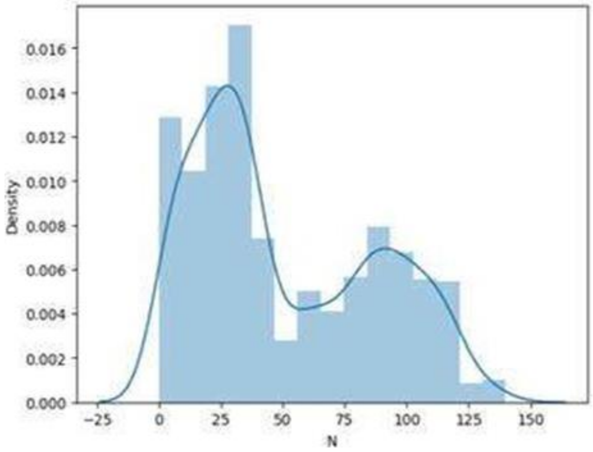

Fig2. Nitrogen

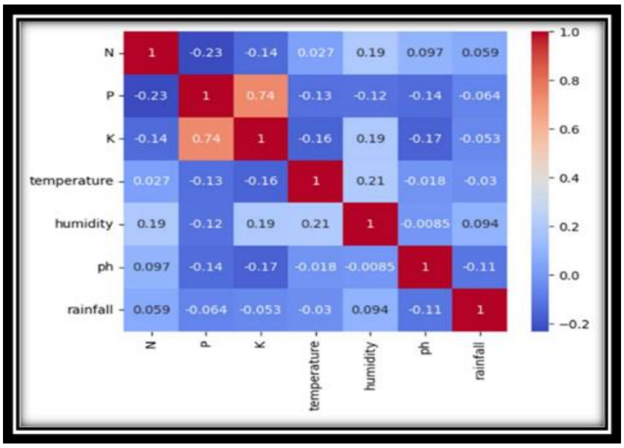

Fig 3. Co-relation between features

4. Feature Engineering

The selection of useful and important features from original data improves the performance of machine learning models. The selection of features helps to reduce the complexity and data dimensionality and eliminates the redundant, noisy, or irrelevant features that may affect the accuracy and efficiency of design models. It is the method for generating useful and meaningful features from the given data set that increases the accuracy and efficiency of designed models. Feature engineering can extract and combine information from the original data and generate new features that can capture the underlying patterns and relationships of the problem. The correlation between the features is given below in Fig 3.

5. Model Selection

The most crucial step for every machine learning study is the selection of the most reasonable model outof all. Model training or selection plays a pivotal role in achieving accurate and reliable predictions, ultimately contributing to recommending the best crop according to the situation. In our comparative analysis, we are likely exploring various existing models like decision trees, support vector machines (SVM), logistic regression, naïve –baye's classifier, KNN, random forest, K-Fold and bagging models. Model selection helps to identify which model is the most suitable for our specific dataset

and prediction task. The implemented view of some of the models is given below in Fig4 to Fig 7

Fig 4: Implemented View of Random Forest

Fig 5. Implemented view of Bagging Classifier

Fig 6. Implemented view of Decision Tree

```python
# Adaboost Classifier
from sklearn.ensemble import AdaBoostClassifier
from sklearn.datasets import make_classification
X_train, Y_train = make_classification(n_samples=1000, n_features=4,n_informative=2, n_redundant=0,random_state=0, shuffle=False)
clf = AdaBoostClassifier(n_estimators=100, random_state=0)
clf.fit(X_train, Y_train)
clf.predict([85,58,41,21.77,80.31,7.03,226.65])
```

Fig 7. Implemented view of Adaboost classifier

6. Model Training

Model training is a crucial aspect of any machine learning study, especially in the context of healthcare applications like asthma prediction. The comparative analysis of different models specifies the effectiveness of models in identifying the best crop. Model training plays a pivotal role in achieving accurate and reliable predictions, ultimately contributing to improved farming techniques. To perform a comparative analysis, multiple machine learning algorithms are typically considered. Common choices include logistic regression, decision trees, random forests, support vector machines, and more. The selection of models should depend upon various parameters like problem type, high accuracy, efficiency, and prevent overfitting. The comparative analysis aims to determine which algorithm(s) offer the best performance in terms of prediction accuracy and generalization. The comparative accuracy score of different models is given in below Fig 8 and the graphical representation is given in Fig 9.

```
Logistic Regression  with accuracy : 0.9636363636363636
Naive Bayes  with accuracy : 0.9954545454545455
Support Vector Machine  with accuracy : 0.9681818181818181
K-Nearest Neighbors  with accuracy : 0.9590909090909091
Decision Tree  with accuracy : 0.9886363636363636
Random Forest  with accuracy : 0.9931818181818182
ExtraTreeClassifier  with accuracy : 0.8863636363636364
BaggingClassifier  with accuracy : 0.990909090909091
GradientBoostingClassifier  with accuracy : 0.9818181818181818
AdaBoostClassifier  with accuracy : 0.1409090909090909
```

Fig 8. Comparative accuracy score of different models

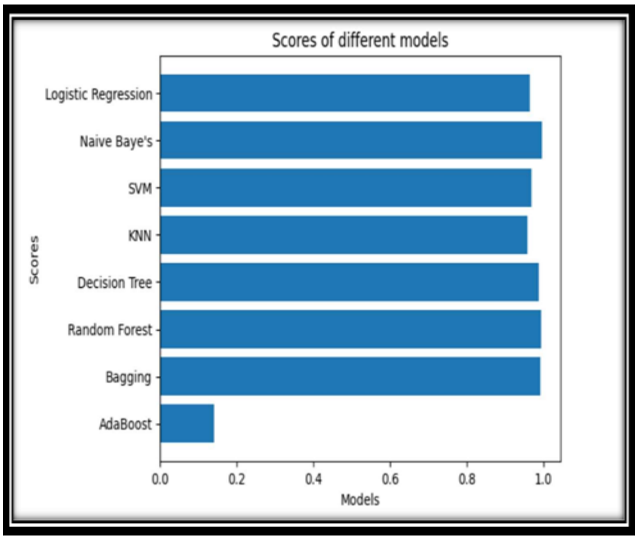

Fig 9. Graphical Representation of Comparative accuracy score of different models

7. Cross-Validation

Cross-validation is a crucial step in model selection to ensure that the model's performance is robust and not overly influenced by the specific data split and helps to identify the performance of the model on unseen data. The k-fold method ensures robustness and helps to avoid overfitting. The Cross validation scores of Logistic Regression, SVM, KNN, decision tree, Naïve Baye’s and random forest are as follows

0.92727272727274,0.9227272727272,0.9,0.9394909090909092,0.9490909090909091and0.9840909090909091respectively. The implemented representation of cross-validation of random forest is given in Fig 10.

```python
L=cross_val_score(LogisticRegression(),X_test,Y_test)
np.average(L)

0.9272727272727274
```

Fig 10. Cross Validation Score of Random Forest

The graphical representation of cross-validation scores of all various models is given in Fig11.

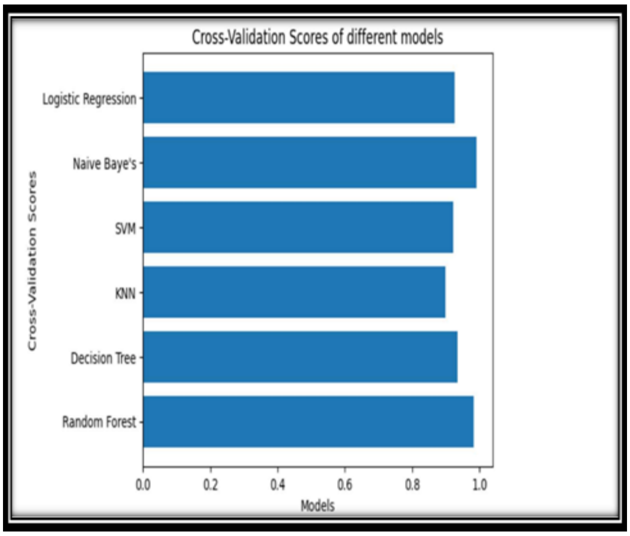

Fig 11. Graphical Representation of Cross-Validation Scores of Different Models.

Therefore, after analysing different models, it has been found that random forest by Cross-validation gives the best result, that is, 0.9840909090909091 and also overcomes the problem of overfitting and gives the best accurate result for large datasets.

8. Predictive System

We have designed predictive systems for the data of 22 crops which we included from the dataset andwe have made a dictionary to encode the 22 crops and mapped them into the prediction system. With the help of our function of random forest function, it takes the value of the parameters like nitrogen, potassium, phosphorus, rainfall etc. then we predict the best crop according to the given conditions of different parameters. Since our data is only available for 22 crops we can predict if the crop is out of them but in future we will add more crops to our dataset and at that time we hope we can predict most of the crops that are available in the market. Random Forest function for Predictive System and Predictive System information given in Fig12 and Fig 13

```
def recommendation(N,P,k,temperature,humidity,ph,rainfal):
    features = np.array([[N,P,k,temperature,humidity,ph,rainfal]])
    transformed_features = ms.fit_transform(features)
    transformed_features = sc.fit_transform(transformed_features)
    prediction = rfc.predict(transformed_features).reshape(1,-1)

    return prediction[0]
```

Fig 12. Random Forest function for Predictive System

```
N = 90
P = 42
k = 43
temperature = 20.87
humidity = 82.0
ph = 6.5
rainfall = 202.93

predict = recommendation(N,P,k,temperature,humidity,ph,rainfall)


crop_dict = {1: "Rice", 2: "Maize", 3: "Jute", 4: "Cotton", 5: "Coconut", 6: "Papaya", 7: "Orange",
             8: "Apple", 9: "Muskmelon", 10: "Watermelon", 11: "Grapes", 12: "Mango", 13: "Banana",
             14: "Pomegranate", 15: "Lentil", 16: "Blackgram", 17: "Mungbean", 18: "Mothbeans",
             19: "Pigeonpeas", 20: "Kidneybeans", 21: "Chickpea", 22: "Coffee"}

if predict[0] in crop_dict:
    crop = crop_dict[predict[0]]
    print("{} is a best crop to be cultivated ".format(crop))
else:
    print("Sorry are not able to recommend a proper crop for this environment")

Papaya is a best crop to be cultivated
```

Fig 13: Prediction System

9. Discussion

Crop recommendation is a crucial task for farmers, as it can help farmers make decisions for identifying the crops for their land and also help to increase their productivity and profitability. However, crop recommendation is not a trivial problem, as it involves many factors and constraints, such as soil properties, weather data, crop yield, market price, crop rotation, pest and disease management, etc. Therefore, traditional methods of crop recommendation, such as expert advice or trial and error, may not be sufficient or efficient to cope with the complexity and variability of the problem and machine learning plays a vital role in the predictions. We have used the data from various research, surveys and other resource to get the required outcome. We analyzed, organized, and handled the missing value to provide a well-structured view of our dataset.ML has various models which can be used for classification purposes to accurately predict the output and enhance the efficiency of the algorithm. Among the various models, it has been found that the best suitable model for accurate predictions is Random Forest with an accuracy score of 0.9840909090909091. So, crop recommendation is a very important feature of the application of ML techniques. ML can provide data-driven and intelligent solutions for crop recommendations that can improve the efficiency and sustainability of agriculture. However, there are still many issues and aspects that need to be addressed and explored in this domain. Therefore, ML-based crop recommendation is an active and promising research area that can have significant impacts on both agriculture and society.

10. Conclusion

Crop recommendation is a vital task for farmers, it can help them to choose the most suitable crops for their land, soil, and climate conditions, and thus increase their productivity and profitability. However, crop recommendation is a complex and dynamic problem, as it involves many factors and constraints, such as soil properties, weather data, crop yield, market price, crop rotation, pest and disease management, etc. Therefore, machine learning (ML) techniques can be used to provide data-driven and intelligent solutions for crop recommendation by analyzing the large amount of data available from various sources, such as soil sensors, satellite images, weather stations, crop databases, etc. In this study, we applied and compared the different models on a dataset of 2200 samples with 20 features and 22 classes of crops the dataset used 80% of data for training the model and used 20% for testing purposes. We applied standard preprocessing techniques, such as scaling, encoding, and imputation and calculated the efficiency of the models in terms of their accuracy, precision, recall, and score metrics. From the comparative result, it has been observed that the random forest algorithm was the best algorithm for crop selection among all models. Random forest achieved an accuracy of 98.18%, which was significantly higher than the other models with the highest value of accuracy, precision, efficiency, recall and high score metrics among all models. Random Forest worked on ensemble techniques in which we combined the multiple decision trees and used majority decisions to make predictions. It also handles high-dimensional and heterogeneous data, reduces overfitting and variance, and improves

generalization androbustness. Therefore, we conclude that random forest is a suitable and effective ML model for crop recommendation that can provide optimal or near-optimal solutions for farmers. Random forests can capture the complex and nonlinear relationships between the features and the classes of crops. Random forest can also deal with the uncertainty, noise, outliers, missing values, etc., in the data. Random forests can enhance the efficiency and sustainability of agriculture by helping farmers to select the best crops for their land.

References

[1] Nischitha K, Dhanush Vishwakarma, Mahendra N, Ashwini, Manjuraju M R, "Crop Prediction using Machine Learning Approaches", *International Journal of Engineering Research and Technology*, vol.9 Issue 08, August-2020 ISSN: 2278-0181.

[2] Zeel Doshi, Subhash Nadkarni, Rashi Agarwal and Neepa Shah, "AgroConsultant: Intelligent Crop Recommendation System using Machine Learning Algorithms",2018 Fourth *International Conference on Computing Communication Control and Automation.*

[3] Priyadharshini A, Swapneel Chakraborty, Aayush Kumar, Omen Rajendra Pooniwala, "Intelligent Crop Recommendation System using Machine Learning", Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021).

[4] Shilpa Mangesh Pande, Prem Kumar Ramesh, Anmol, B R Aishwarya, Karuna Rohilla and Kumar Shaurya, "Crop Recommendation using Machine Learning Approach", Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021).

[5] Mayank Champaneri, Chaitanya Chandvidkar, Darpan Chachpara and Mansing Rathod, "Crop Yield Prediction using Machine LearninKamatchi, S. Bangaru, and R. Parvathi. "Improvement of Crop Production Using Recommender System by Weather Forecasts." Procedia Computer Science 165 (2019): 724-732.

[6] Bondre, Devdatta A., and Santosh Mahagaonkar. "Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms." *International Journal of Engineering Applied Sciences and Technology* 4, no. 5 (2019): 371-376.

[7] Suresh, G., A. Senthil Kumar, S. Lekashri, and R. Manikandan. "Efficient Crop Yield Recommendation System Using Machine Learning For Digital Farming." *International Journal of Modern Agriculture* 10, no. 1 (2021): 906-914.

[8] Reddy, D. Anantha, Bhagyashri Dadore, and Aarti Watekar. "Crop recommendation system to maximize crop yield in ramtek region using machine learning." *International Journal of Scientific Research in Science and Technology* 6, no. 1 (2019): 485-489.

[9] Pudumalar, S., E. Ramanujam, R. Harine Rajashree, C. Kavya, T. Kiruthika, and J. Nisha. "Crop recommendation system for precision agriculture." In 2016 Eighth International Conference on Advanced Computing (ICoAC), pp. 32-36. IEEE, 2017.

[10] Azizkhan F Pathan, Chetana Prakash, Attention-based position-aware framework for aspect-based opinion mining using bidirectional long short-term memory, Journal of King Saud University - Computer and Information Sciences, 2021,ISSN1319-1578.

[11] Azizkhan F Pathan, Chetana Prakash, Unsupervised Aspect Extraction Algorithm for opinion mining using topic modeling, Global Transitions Proceedings, Volume 2, Issue 2, 2021, pp. 492-499, ISSN 2666-285X

[12] Dighe, Deepti, Harshada Joshi, Aishwarya Katkar, Sneha Patil, and Shrikant Kokate. "Survey of Crop Recommendation Systems." (2018).

[13] Jaiswal, Sapna, Tejaswi Kharade, Nikita Kotambe, and Shilpa Shinde. "Collaborative Recommendation System For Agriculture Sector." In ITM Web of Conferences, vol. 32. EDP Sciences, 2020.

[14] Mokhtar, U., Ali, M. A., Hassanien, A. E., and Hefny, H. (2015). "Identifying two of tomatoes leaf viruses using support vector machine," in *Information Systems Design and Intelligent Applications*, eds J. K. Mandal, S. C. Satapathy, M. K. Sanyal, P. P. Sarkar, A. Mukhopadhyay (Springer), 771–782.

[15] Raza, S.-A., Prince, G., Clarkson, J. P., Rajpoot, N. M., et al. (2015). Automatic detection of diseased tomato plants using thermal and stereo visible light images. *PLoS ONE* 10:e0123262. doi: 10.1371/journal.pone.0123262.

[16] Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117. doi: 10.1016/j.neunet.2014.09.003

[17] Tai, A. P., Martin, M. V., and Heald, C. L. (2014). Threat to future global food security from climate change and ozone air pollution. *Nat. Clim. Chang* 4, 817–821. doi: 10.1038/nclimate2317

[18] Strange, R. N., and Scott, P. R. (2005). Plant disease: a threat to global food security. *Phytopathology* 43, 83–116. doi: 10.1146/annurev.phyto.43.113004.133839

[19] Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2015). Machine learning for highthroughput stress phenotyping in plants. *Trends Plant Sci.* 21, 110–124 doi: 10.1016/j.tplants.2015.10.015

[20] Viraj Mehta, Chahat Jain, Karan Kanchan, Prof. Vinaya Sawant ―A Machine Learning Approach to Foretell the Probability of a Crop Contracting a Disease‖ , 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA).

[21] Jayraj Chopda, Sagar Nakum,Vivek Nakarani, Prof. Hiral Raveshiya ―Cotton crop disease detection using decision tree classifier‖ , IEEE paper 2015.

[22] Jitendra Kumar Jaiswal, Rita Samikannu, ―Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression‖ , IEEE paper 2017.

[23] P. P. Patel and D. B. Vaghela,"Crop Diseases and Pests Detection Using Convolutional Neural Network" , International Conference paper, 2018.

[24] J. Joo, U. Lee, S. Jeong, J. Y. Yoon, H. Jin ,S. C. Kim , " Periodontal Disease Detection Using Convolutional Neural Networks", International Conference paper, 2018

[25] B. K. Singh, R. P. Singh, T. Bisen, S. Kharayat, "Disease Manifestation Prediction from Weather Data Using Extreme Learning Machine", conference paper February 2018.