Literature Review on Chatbot used in Predicting the Cultivation of Foreign Crops in India

Dhillipkumar M

September 20, 2023

1 Introduction

Artificial intelligence (AI) has influenced how we engage in our every day activities by designing and evaluating advanced applications and devices, called intelligent agents, which can perform various functions. A chatbot is an artificial intelligence program and a Human–computer Interaction (HCI) model (Bansal and Khan, 2018). According to the dictionary, a chatbot is "A computer program designed to simulate conversation with human users, especially over the Internet" (Chatbot —Definition of chatbot in English by Lexico Dictionaries, 2019). It uses Natural Language Processing (NLP) and sentiment analysis to communicate in human language by text or oral speech with humans or other chatbots (Khanna et al., 2015). Artificial conversation entities, interactive agents, smart bots, and digital assistants are also known as chatbots.

Machine learning (ML) approaches are used in many fields, ranging from supermarkets to evaluate the behavior of customers (Ayodele, 2010) to the prediction of customers' phone use (Witten et al., 2016). Machine learning is also being used in agriculture for several years (McQueen et al., 1995). Crop yield prediction is one of the challenging problems in precision agriculture, and many models have been proposed and validated so far. This problem requires the use of several datasets since crop yield depends on many different factors such as climate, weather, soil, use of fertilizer, and seed variety (Xu et al., 2019). This indicates that crop yield prediction is not a trivial task; instead, it consists of several complicated steps. Nowadays, crop yield prediction models can estimate the actual yield reasonably, but a better performance in yield prediction is still desirable (Filippi et al., 2019a).

Machine learning, which is a branch of Artificial Intelligence (AI) focusing on learning, is a practical approach that can provide better yield prediction based on several features.

Machine learning (ML) can determine patterns and correlations and discover knowledge from datasets. The models need to be trained using datasets, where the outcomes are represented based on past experience. The predictive model is built using several features, and as such, parameters of the models are determined using historical data during the training phase. For the testing phase, part of the historical data that has not been used for training is used for the performance evaluation purpose.

An ML model can be descriptive or predictive, depending on the research problem and research questions. While descriptive models are used to gain knowledge from the collected data and explain what has happened, predictive models are used to make predictions in the future (Alpaydin, 2010). ML studies consist of different challenges when aiming to build a high-performance predictive model. It is crucial to select the right algorithms to solve the problem at hand, and in addition, the algorithms and the underlying platforms need to be capable of handling the volume of data.

To get an overview of what has been done on the application of ML in crop yield prediction, we performed a systematic literature review (SLR).

2 Related Work

Crop yield prediction is an essential task for the decision-makers at national and regional levels (e.g., the EU level) for rapid decision-making. An accurate crop yield prediction model can help farmers to decide on what to grow and when to grow. There are different approaches to crop yield prediction. This review article has investigated what has been done on the use of machine learning in crop yield prediction in the literature.

During our analysis of the retrieved publications, one of the exclusion criteria is that the publication is a survey or traditional review paper. Those excluded publications are, in fact, related work and are discussed in this section. Chlingaryan and Sukkarieh performed a review study on nitrogen status estimation using machine learning (Chlingaryan et al., 2018). The paper concludes that quick developments in sensing technologies and ML techniques will result in cost-effective solutions in the agricultural sector. Elavarasan et al. performed a survey of publications on machine learning models associated with crop yield prediction based on climatic parameters. The paper advises looking broad to find more parameters that account for crop yield (Elavarasan et al., 2018). Liakos et al. (2018) published a review paper on the application of machine learning in the agricultural sector. The analysis was performed with publications focusing on crop management, livestock management, water management, and soil management. Li, Lecourt, and Bishop performed a review study on determining the ripeness of fruits to decide the optimal harvest time and yield prediction (Li et al., 2018).

Mayuri and Priya addressed the challenges and methodologies that are encountered in the field of image processing and machine learning in the agricultural sector and especially in the detection of diseases (Mayuri and Priya, 2018). Somvanshi and Mishra presented several machine learning approaches and their application in plant biology (Somvanshi and Mishra, 2015). Gandhi and Armstrong published a review paper on the application of data mining in the agricultural sector in general, dealing with decision making. They concluded that further research needs to be done to see how the implementation of data mining into complex agricultural datasets could be realized (Gandhi and Armstrong, 2016a, Gandhi and Armstrong, 2016b). Beulah performed a survey on the various data mining techniques that are used for crop yield prediction and concluded that the crop yield prediction could be solved by employing data mining techniques (Beulah, 2019).

3 Research questions

This SLR aims to get insight into what studies have been published in the domain of ML and crop yield prediction. To get insight, studies have been analyzed from several dimensions. For this SLR study, the following four research questions(RQs) have been defined.

•RQ1- Which machine learning algorithms have been used in the literature for crop yield prediction?

•RQ2- Which features have been used in literature for crop yield prediction using machine learning?

•RQ3- Which evaluation parameters and evaluation approaches have been used in literature for crop yield prediction?

•RQ4- What are challenges in the field of crop yield prediction using machine learning?

4 Results

The table shows the publication year, title, and algorithms used in the papers.

Table 1: Publications

Source	Reference	Title	Algorithm used	Year
Scopus	Ruß et al. (2008)	Data Mining with Neural Networks for Wheat Yield Prediction	Neural networks	2008
Science Direct	Everingham et al. (2009)	Ensemble data mining approaches to forecast regional sugarcane crop production	Forward stagewise algorithm	2009
Springer Link	Ruß & Kruse (2010)	Regression Models for Spatial Data: An Example from Precision Agriculture	Clustering, random forest, support vector machine	2010
Springer Link	Baral et al. (2011)	Yield Prediction Using Artificial Neural Networks	Neural networks	2011
Springer Link	Črtomir et al. (2012)	Application of Neural Networks and Image Visualization for Early Forecast of Apple Yield	Neural networks	2012
Google Scholar	Johnson (2013)	Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods	Multiple linear regression, neural networks	2013
Google Scholar	Romero et al. (2013)	Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires	K-nearest neighbor, decision tree	2013

The below pie chart shows the distribution of types of publications. The figure shows that most of the articles we accessed are journal articles; conference papers and book chapters constitute less than 25% of the total number of papers.

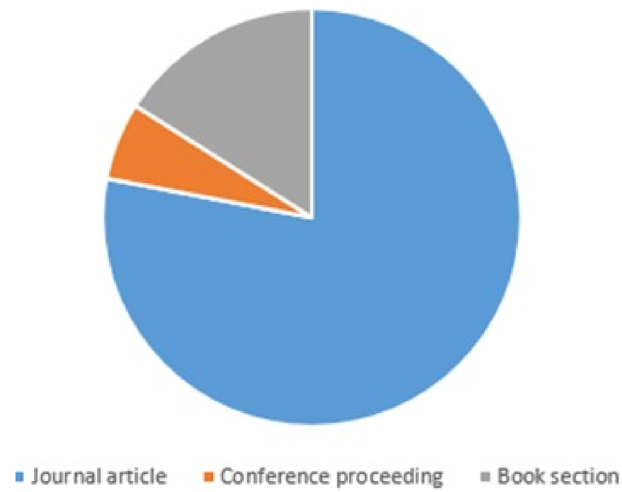

Figure 1: Distribution of the type of publications.

To address research question two (RQ2), features used in the machine learning algorithms applied in the papers were investigated and summarized.All features we were able to extract are shown in Table 2.

Table 2: Frequency of Feature Usage

Feature	**# of times used**
Temperature	24
Soil type	17
Rainfall	17
Crop information	13
Soil maps	12
Humidity	11
pH-value	11
Solar radiation	10
Precipitation	9
Images	8
Area of production	8
Fertilization	7
NDVI	6
Cation exchange capacity	6
Nitrogen	6

As shown in Table 2, the most used features are related to temperature, rainfall, and soil type. Crop yield is the dependent variable. To get a better overview of the independent variables (features), the features were grouped. The independent features can be grouped into soil and crop information, humidity, nutrients, and field management. The number of times these groups are used is presented in Table 3. As shown in this table, the feature groups that are most used are related to the soil, solar, and humidity information.

Table 3: Grouped Features

Group	**# of times used**
Soil information	54
Solar information	39
Humidity	38
Nutrients	28
Other	24
Crop information	14
Field management	12

To address the first research question (RQ1), machine learning algorithms were investigated and summarized. The algorithms used more than once are listed in Table 5. As shown in the table, Neural Networks (NN) and Linear Regression algorithms are the two algorithms used mostly. Also, Random Forest (RF) and Support Vector Machines (SVM) are widely used, according to Table 4.

Table 4: Most Used Machine Learning Algorithms

Most Used Machine Learning Algorithms	**# of times used**
Neural Networks	27
Linear Regression	14
Random Forest	12
Support Vector Machine	10
Gradient Boosting Tree	4

To address research question three (RQ3), evaluation parameters were identified. All the evaluation parameters that were used and the number of times they were used are shown in Table 5. As the table shows, Root Mean Square Error (RMSE) is the most used parameter in the studies.

Table 5: All Evaluation Parameters Used

Key	Evaluation Parameter	# of times used
RMSE	Root Mean Square Error	29
R2	R-squared	19
MAE	Mean Absolute Error	8
MSE	Mean Square Error	5
MAPE	Mean Absolute Percentage Error	3
RSAE	Reduced Simple Average Ensemble	3
LCCC	Lin's Concordance Correlation Coefficient	1
MFE	Multi-Factored Evaluation	1
SAE	Simple Average Ensemble	1
rcv	Reference Change Values	1
MCC	Matthew's Correlation Coefficient	1

To address research question four (RQ4), the publications were read to see if they stated any problems or improvements for future models. In several studies, insufficient availability of data (too few data) was mentioned as a problem. The studies stated that their systems worked for the limited data that they had at hand, and indicated data with more variety should be used for further testing. This means data with different climatic circumstances, different vegetation, and longer time-series of yield data. Another suggested improvement is that more data sources should be integrated. Finally, the publication indicated that the use of machine learning in farm management systems should be explored. If the models work as requested, software applications must be created that allow the farmer to make decisions based on the models.

RQ1-Related (algorithms) discussion: Linear Regression is the second most used algorithms, according to Table 5. Linear Regression is used as a benchmarking algorithm in most cases to check whether the proposed algorithm is better than Linear Regression or not. Therefore, although it is shown in many articles, it does not mean that it is the best performing algorithm. Table 5 should be interpreted carefully because "most used" does not mean the best-performing ones. In fact, Deep Learning (DL), which is a sub-branch of Machine Learning, has been used for the crop yield prediction problem recently and is believed to be very promising. In this study, we also identified several deep learning-based studies. There are several additional promising aspects of DL methods, such as automatic feature extraction and superior performance. We expect that more research will be conducted on the use of DL approaches in crop yield prediction in the near future due to the superior performance of DL algorithms in other problem domains.

RQ2-related (features) discussion: Groups are created for features and algorithms to visualize the main features and algorithms. Due to this decision, detailed information is lost, but clarity has been maintained. The most used features are soil type, rainfall, and temperature. Apart from those features that are used in several studies, there are also features that were used in specific studies. Those features are gamma radiation, MODIS-EVI, forecast rainfall, humidity, photoperiod, pH-value, irrigation, leaf area, NDVI, EVI, and crop information. There are also studies that use different nutrients as features, which are magnesium, potassium, sulphur, zinc, nitrogen, boron, and calcium. The most used features are not always the same kind of data. Temperature, for example, is measured as average temperature, but more features like maximum temperature and minimum temperature are also applied.

RQ3-related (evaluation parameters and approaches) discussion: There are not many evaluation parameters reported in the selected papers. Almost every study used RMSE as the measurement of the quality of the model. Other evaluation parameters are MSE, R2, and MAE. Some parameters were used in specific studies, most of these parameters look like some of the previously mentioned parameters, with a small difference. These are MAPE, LCCC, MFE, SAE, rcv, RSAE, and MCC. Most of the models had outcomes with high accuracy values for their evaluation parameters, which means that the model made correct predictions. As the evaluation approach, the 10-fold cross-validation approach was preferred by researchers.

RQ4-related (challenges) discussion: Challenges were reported based on the explicit statements in the articles. However, there might be additional challenges that were not stated in the identified papers. The challenges are mainly in the field of improvement of a working model. When more data is gathered to train and test, much more can be said about the precision of the model. Another challenge is the implementation of the models into the farm management systems. When applications are made that the farmer can use, then only can the models be useful to make decisions, also during the growing season. When specific parameters for that specific place are measured and added, predictions will have higher precision.

5 Research Gap

The cultivation of crops in India has long been a cornerstone of the country's agricultural sector, with a rich history of traditional crops such as rice, wheat, and pulses dominating the agricultural landscape. However, in recent years, there has been a growing interest and necessity to diversify the range of crops cultivated in the country. This diversification is driven by several factors, including changing consumer preferences, climate change impacts on traditional crops, and the need to enhance food security.

While this shift towards crop diversification is commendable, it is essential to note that the majority of research and attention in the Indian agricultural sector has traditionally been directed towards improving the yield and production of these conventional crops. As a result, there exists a significant research gap in the specific domain of cultivating foreign or non-native crops in India.

Several reasons underscore the importance of addressing this research gap:

Climate Adaptation: Climate change has had a substantial impact on the traditional crops cultivated in India. Rising temperatures, erratic rainfall patterns, and increased incidence of pests and diseases have created challenges for farmers. Exploring the cultivation of foreign crops that may be more resilient or adaptable to changing climate conditions can help secure food production.

Market Demand: Changes in consumer preferences and dietary habits have led to a growing demand for exotic or foreign crops, such as quinoa, avocado, and blueberries. Meeting this demand requires a better understanding of the cultivation practices, market potential, and economic viability of these crops in India.

Nutritional Diversity: Diversifying the crop portfolio can contribute to enhanced nutritional diversity in the Indian diet. Foreign crops often bring unique nutritional benefits and can address specific dietary deficiencies.

Sustainable Agriculture: Foreign crops may offer opportunities for sustainable agricultural practices, such as reduced water usage or decreased reliance on chemical inputs. Investigating their cultivation methods and sustainability implications is crucial.

International Trade: Understanding the cultivation of foreign crops is also vital for international trade and export opportunities. Meeting international standards and quality requirements is crucial for accessing global markets.

Therefore, there is a clear research gap in systematically studying the cultivation practices, challenges, opportunities, and outcomes of foreign crop cultivation in India. This includes research on crop selection, agronomic practices, pest and disease management, market analysis, and socio-economic impacts.

Addressing this research gap can not only contribute to the diversification and resilience of Indian agriculture but also align with broader sustainability and dietary goals. Future research in this area will play a pivotal role in shaping the agricultural landscape of India, enhancing food security, and meeting the changing demands of both domestic and international markets.

Researchers, policymakers, and agricultural stakeholders are encouraged to focus on this underexplored domain to unlock its potential benefits for Indian agriculture and the well-being of its populace.

6 Conclusion

Chatbots on the cultivation of foreign crops in India represents a promising avenue for the country's agricultural sector, addressing several contemporary challenges and opportunities. The existing literature has laid a foundation for understanding the importance of diversifying crop portfolios, adapting to changing climate conditions, and meeting the demands of evolving dietary preferences. However, it is evident that there exists a significant research gap concerning the systematic study of foreign crop cultivation in India.

The identified research gap underscores the need for comprehensive investigations into the cultivation practices, challenges, and potential benefits of foreign crops. Such research endeavors should encompass multiple dimensions, including agronomy, sustainability, market dynamics, and socio-economic impacts. By addressing these gaps, we can pave the way for a more resilient, diverse, and sustainable agricultural landscape in India.

Future research in this field has the potential to offer practical solutions for farmers seeking to adapt to a changing climate, enhance food security, and tap into emerging market opportunities.

In conclusion, the cultivation of foreign crops in India is not merely a matter of agricultural diversification; it is a strategic response to the evolving needs of the nation and its people. Researchers, policymakers, and agricultural stakeholders are encouraged to embark on this important journey, working collaboratively to bridge the research gap and unlock the potential benefits that foreign crop cultivation can bring to India's agricultural landscape and the well-being of its populace.

As we move forward, it is imperative that research efforts are complemented by practical initiatives and policy measures aimed at facilitating the adoption of foreign crops, promoting sustainable agricultural practices, and ensuring that the benefits of diversification reach all levels of society. By doing so, we can contribute to the resilience, sustainability, and prosperity of Indian agriculture in the face of ongoing challenges and emerging opportunities.

7 References

1. Ananthara, M. Gunasundari, T. Arunkumar, and R. Hemavathy. ”CRY—an improved crop yield prediction model using bee hive clustering approach for agricultural data sets.” 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering. IEEE, 2013.

2. Paul, Monali, Santosh K. Vishwakarma, and Ashok Verma. ”Analysis of soil behaviour and prediction of crop yield using data mining approach.” 2015 International Conference on Computational Intelligence and Communication Networks (CICN). IEEE, 2015.

3. Gandhi, Niketa, et al. ”Rice crop yield prediction in India using support vector machines.” 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE, 2016.

4. Gandhi, Niketa, and Leisa Armstrong. ”Applying data mining techniques to predict yield of rice in humid subtropical climatic zone of India.” 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2016.

5. Rahman, Mohammad Motiur, Naheena Haq, and Rashedur M. Rahman. ”Machine learning facilitated rice prediction in Bangladesh.” 2014 Annual Global Online Conference on Information and Computer Technology. IEEE, 2014.

6. Everingham, Yvette, et al. ”Accurate prediction of sugarcane yield using a random forest algorithm.” Agronomy for sustainable development 36 (2016): 1-9.

7. Khanal, Sami, et al. ”Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield.” Computers and electronics in agriculture 153 (2018): 213-225.

8. Rao, T. Venkat Narayana, and S. Manasa. ”Artificial neural networks for soil quality and crop yield prediction using machine learning.” International Journal on Future Revolution in Computer Science & Communication Engineering 5.1 (2019): 57-60.

9. Khaki, Saeed, and Lizhi Wang. ”Crop yield prediction using deep neural networks.” Frontiers in plant science 10 (2019): 621.

10. Rahnemoonfar, Maryam, and Clay Sheppard. ”Real-time yield estimation based on deep learning.” Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II. Vol. 10218. SPIE, 2017.