KJMS 2023 VOL.5 (1): 54-65

KDU Journal of Multidisciplinary Studies (KJMS)
Volume 5. Issue 1 July 2023
https://doi.org/10.4038/kjms.v5i1.62

CROP YIELD FORECASTING USING MACHINE LEARNING TECHNIQUES - A SYSTEMATIC LITERATURE REVIEW

D.M.P.W. Dissanayake[1] R.M.K.T. Rathnayake[2] and L.L. Gihan Chathuranga[3]

Department of Computing and Information Systems, Faculty of Computing,
Sabaragamuwa University, of Sri Lanka, Belihuloya[1]
Department of Physical Sciences and Technology, Faculty of Applied Sciences,
Sabaragamuwa University, of Sri Lanka, Belihuloya [2]
Department of Computer Sciences, Faculty of Sciences, University of Ruhuna Matara, Sri Lanka[3],

ABSTRACT

The utilization of machine learning has become increasingly important in the prediction of crop yields for facilitating decisions regarding crop cultivation and management during the growing season. Numerous machine learning and data mining algorithms have been developed to support research in crop yield forecasting. In this study, a systematic literature review (SLR) was conducted on research published between 2016 and 2021 to investigate the use of machine learning in crop yield forecasting. A total of 261 relevant studies were identified from five electronic databases, out of which 15 studies were selected for further analysis based on inclusion and exclusion criteria. The selected studies were thoroughly examined, and their methods and features were analyzed, to provide suggestions for future research. The results showed that evapotranspiration, temperature, precipitation, and soil type were the most commonly used features in crop yield forecasting, while RMSE, MSE, MAE, and R2 were the most commonly used evaluation parameters. The challenges include selecting appropriate input variables, handling missing data and outliers, and capturing non-linear relationships between variables. The authors discuss various techniques such as feature selection, regularization, imputation, non-linear machine learning, data preprocessing, and data augmentation to address these challenges. The Support Vector Machine, Linear Regression, Artificial Neural Network (ANN), and Long-Short Term Memory (LSTM) were identified as the most commonly used algorithms in these models.

KEYWORDS*: Crop yield forecasting, Data mining, Deep learning, Systematic review*

Corresponding Author: D.M.P.W. Dissanayake, Email: dmpwdissanayake@std.appsc.sab.ac.lk

https://orcid.org/0009-0004-6895-3980

This is an open-access article licensed under a Creative Commons Attribution 4.0 International License (CC BY) allowing distribution and reproduction in any medium crediting the original author and source.

1. INTRODUCTION

The fast growing human population in the world has increased the need for food for human survival. Around 2000 years ago most of the Earth's population started to depend on agriculture (Rutledge et al., 2011). Meeting the limited food resources available on earth is a major challenge. It is primarily located in the poorest countries of the Third World and is currently growing as its population grows. More than a billion people are suffering from food shortages. This is because population growth is higher than the development of agricultural products and agricultural technology. Therefore, the agricultural sector should be made more productive and efficient than before. People want to get more out of their fields, and therefore new techniques should be used to improve the yield. Researchers are looking for new solutions to increase crop yields.

Machine learning (ML) and Data Mining approaches are used in many fields from Developing the Profiles of Supermarket Customers (Min, H., 2006), to Predicting customer's gender and age depending on mobile phone data (Al-Zuabi et al., 2019). Machine learning has also been used in agriculture for many years (Conway, J.A. et al., 1991). Crop yield forecasting is a challenging issue in precision agriculture, and many models have already been proposed and validated. Crop yields depend on various factors such as climate, weather, soil, fertilizer application and seed variety, so it is necessary to use several data sets for this problem. This suggests that crop yield forecasting is not a trivial task. Instead, it consists of several complex steps. At present, crop yield forecasting models can reasonably estimate the true yield, but a better performance of the yield forecast is still desirable (Shahrin, F. et al., 2020).

Machine learning is a practical approach within the field of artificial intelligence (AI) that focuses on learning and has the ability to predict yields based on several features. The process involves determining patterns and correlations within datasets, training models using the data, and representing results based on past experience. During the training phase, historical data is used to build a forecast model with various features, and the parameters of the models are determined. To evaluate the model's performance, part of the historical data not used for training is reserved for testing. To understand the application of machine learning in crop yield forecasting, we conducted a systematic literature review (SLR) that identifies potential gaps in research and helps professionals and researchers who are interested in conducting new studies in this area. The SLR study provides new perspectives and insights for new researchers in the field. In this paper, we present our empirical results and responses to the research questions defined as part of this review article. The article is structured as follows: Part 2 provides background information (Related work). Part 3 explains the methodology. Part 4 presents the results of the SLR. Part 5 discusses the findings of our review, and Part 6 concludes this paper.

Related work

The development of a precise crop harvest prediction model holds significant importance for farmers in making informed decisions regarding crop selection and planting schedules. Diverse methodologies are available for crop yield forecasting. This review critically analyzes the existing literature on the application of machine learning techniques in crop yield forecasting. Although a majority of the previous research studies have not conducted a comprehensive review of the available literature, several studies have examined specific aspects of crop yield forecasting. There are a few SLR examples of reviews on this, as (summarized in Table 1)

Gandhi and Armstrong published a review paper on the application of data mining in the field of agriculture in general. They concluded that further research was needed to see how the implementation of data mining into complex agricultural datasets could be achieved (Gandhi and Armstrong, 2016). Elavarasan et al. conducted a publication survey of machine learning models related to crop yield forecasting based on climatic parameters. The survey

advises looking broadly to find other parameters that contribute to crop yield (Elavarasan et al., 2018).

Beulah conducted a survey of various data mining techniques used to predict crop yields and concluded that crop yield forecasts could be solved using data mining methods (Beulah, 2019).

Table 1: Summary of selected literature

Reference	Goal	Research questions
Van Klompenburg, T. et al., 2020	Review of Crop yield prediction using machine learning to investigate to what extent deep learning algorithms were used for crop yield prediction.	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 the challenges in the field of crop yield prediction using machine learning?
Naftali Slob. et al., 2020	Provide an overview of what has been done on the use of ML in the dairy sector.	RQ1 - What kind of problems are solved using ML and what ML tasks are these problems mapped into? RQ2 - What independent and dependent variables are used to build the ML models? RQ3 - What ML algorithms are applied for the models? RQ4 - Which evaluation parameters and which evaluation approaches are used? RQ5 - Which algorithm performs the best? RQ6 - What are the challenges reported in the identified articles?

Naftali Slob, Cagatay Catal and Ayalew Kassahun conducted a review and provided an overview of what has been done on the use of ML in the dairy sector. (Naftali Slob. et al., 2020).

Thomas van Klompenburg, Ayalew Kassahun and Cagatay Catal conducted a review of Crop yield prediction using machine learning to investigate to what extent deep learning algorithms were used for crop yield prediction. (Van Klompenburg, T. et al., 2020).

According to this SLR, the above are the important review articles presented in this section.

2. METHODOLOGY

In this research process, we followed the guidelines suggested by Kitchenham et al., (2007).

2.1. Planning the review

Initially, the research inquiries are delineated, and subsequently, relevant studies are chosen by utilizing databases. Specifically, the ACM Digital Library, IEEE Xplore, ResearchGate, SpringerLink, and ScienceDirect databases were employed for the present study. Once the pertinent research is identified, it undergoes a filtering process, and is assessed using quality standards. The pertinent data from the selected studies are then extracted and synthesized in accordance with the research questions.

2.1.1. Research questions

The objective of this systematic literature review is to obtain a comprehensive understanding of the published research on crop yield forecasting within the field of machine learning and data mining. To achieve this goal, various dimensions have been analyzed in the reviewed studies. The present study outlines five research questions (RQs) to guide the SLR analysis.

RQ1: When, where, and who have published studies?
RQ2: What are the algorithms that have been used so

far for prediction?

RQ3: What are the features/factors used in the literature to predict crop yields using machine learning?

RQ4: What are the evaluation parameters and evaluation approaches used in the literature to predict crop yields?

RQ5: What are the challenges in the field?

Table 2: Search sources

Electronic databases	ACM Digital library IEEE Xplore ResearchGate SpringerLink ScienceDirect
Searched items	Journal and conference papers
Search applied on	Full text—to avoid missing any of the papers that do not include our search keywords in titles or abstracts, but are relevant to the review object
Language	English
Publication period	From January 2016 to December 2021

Table 3: Search term of the tertiary study

Areas	Search Terms
Machine Learning	“machine learning”, “deep learning”, “data mining”
Crop Prediction	“crop yield prediction”, “crop yield forecasting”, “yield prediction”, “yield forecasting”
Review	“systematic literature review”, “systematic review”, “study”, “review”
Search string	(“machine learning” OR “deep learning” OR “data mining”) AND (“crop yield prediction” OR “crop yield forecasting” OR “yield prediction” OR “yield forecasting”) AND (“systematic literature review” OR “systematic review” OR “study” OR “review”)

2.1.2 Search strategy

After defining our research questions using the study by Kitchenham et al., (2007) as a guide to research, we began with the formulation of a formal search strategy for analyzing all existing empirical material specific to the purpose of this review.

A basic search is done by an automated search. The plan included defining the search space, including electronic databases and printed processes, as provided in Table 2. The studies were first taken from the above electronic databases and then analyzed to identify other meaningful studies through investigative search (snowballing).

The inclusion and exclusion criteria were then applied to studies obtained, involving a different number of research as described in section 2.1.4

2.1.3 Search Criteria

The search was conducted in five databases (Table 2). Search input was used to gain a broader perspective on "machine learning" and "yield forecasting" studies.

After applying the exclusion criteria and processing all the results, a more complex search sequence is built to avoid skipping relevant studies. Table 3 and Table 4 represent the search strings.

2.1.4 Inclusion and Exclusion Criteria

The following inclusion and exclusion criteria were used to determine whether a study should be included.

Table 4: Search term of the study

Areas	Search Terms
Machine Learning	“machine learning”, “deep learning”, “data mining”
Crop Prediction	“crop yield prediction” , “crop yield forecasting” , “yield prediction” , “yield forecasting”
Search string	(“machine learning” OR “deep learning” OR “data mining”) AND (“crop yield prediction” OR “crop yield forecasting” OR “yield prediction” OR “yield forecasting”)

Inclusion Criteria:

IC1: The publication is related to the agricultural sector and yield prediction combined with machine learning

IC2: It is relevant to the search terms defined in Section 2.1.3.

IC3: The study is published between January 2016

and December 2021.

Exclusion Criteria:

EC1: Publication is not written in English.
EC2: Publication that is a duplicate or already retrieved from another database.
EC3: Full text of the publication is not available.
EC4: Studies that do not meet inclusion criteria.
EC5: Publication is a review/survey paper.

2.2 Conducting the review

In this section, we obtain information from our search findings and related sources and databases.

2.2.1 Study Search and Selection

By following the search strategy (previously explained in Section 2.12), the selected electronic databases were searched and the studies were retrieved. In this original search, we retrieved 261 studies as shown in Table 5. After applying three Inclusion criteria, only 72 studies remained for further analysis, after which fourteen studies were selected for further analysis by applying all five criteria for exclusion. In Table 5, we show the number of papers obtained initially and the number of papers remaining after applying the selection criteria.

Table 5: Distribution of papers based on the databases

Database	# of initially retrieved papers	# of papers after inclusion and exclusion criteria	Percentage of Papers (%)
ACM Digital library	22	1	7
IEEE Xplore	88	4	29
ResearchGate	57	3	14
SpringerLink	54	4	29
ScienceDirect	40	3	21
Total	261	15	100

To answer the five research questions, data from the selected studies were extracted and synthesized. The selected studies that passed the inclusion and exclusion criteria are presented in Table 6. During the data synthesis, all the extracted data were consolidated and synthesized and the research questions were answered accordingly. The results are set out in Section 3.

Table 6: Selected publications

ID	Retrieved from	Reference	Title	Year
1	IEEE Xplore	#12	Rice crop yield prediction in India using support vector machines	2016
2	IEEE Xplore	#13	Rice crop yield prediction using artificial neural networks	2016
3	SpringerLink	#14	Predicting Early Crop Production by Analyzing Prior Environment Factors	2017
4	IEEE Xplore	#15	Rice yield prediction model using data mining	2017
5	IEEE Xplore	#16	Prediction of Crop Production in India Using Data Mining Techniques	2018
6	SpringerLink	#17	Yield Forecasting of Spring Maize Using Remote Sensing and Crop Modeling in Faisalabad-Punjab Pakistan	2018
7	ScienceDirect	#18	Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties	2018
8	SpringerLink	#19	Smart Farming System: Crop Yield Prediction Using Regression Techniques	2018
9	ScienceDirect	#20	Design of an integrated	2019

			climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China	
10	SpringerLink	#21	Sugarcane Yield Grade Prediction Using Random Forest with Forward Feature Selection and Hyper-parameter Tuning	2019
11	ResearchGate	#22	Prediction of Rice Yield via Stacked LSTM	2020
12	ACM Digital library	#23	Prediction of Soybean Yield using Self-normalizing Neural Networks	2020
13	ResearchGate	#24	Prediction of rice yield based on LSTM long Short and long memory network	2021
14	ScienceDirect	#25	Crop Yield Forecasting using Data Mining	2021
15	ResearchGate	#26	Review on Crop Prediction Using Deep Learning Techniques	2021

2.2.2. Data extraction and synthesis

Based on the guidelines provided by Kitchenham et al., (2007), we defined a data extraction process to identify relevant information from 15 included preliminary studies related to our research questions. Our data extraction process includes the following: First, we set up a form to report the ideas, concepts, contributions, and findings of each of the 15 studies.

The following data were extracted from each publication: (i) Research Topic; (ii) Database; (iii) Year of publication; (iv) DOI; (v) Keywords; (vi) Research Objectives Defined; (vii) Novelty of the study; (viii) Data Sample Details; (ix) Methodology; (x) Country/location of the analysis;

2.2.3. Quality Assessment Criteria and Screening Procedures

In order to assess the systematic review at hand, the following quality criteria were employed:

- Examination of the number of citations incorporated in each study sourced from existing research.
- Verification of publication venue, ensuring the inclusion of papers published in recognized conferences or journals.
- Analysis of the research inquiry that the authors have concentrated on.
- Analysis of the conclusions drawn by the researchers in response to the research questions.
- Scrutiny of adherence to standard reference formatting.
- Consideration of studies published between 2016 and 2021, prioritizing the most recent research.
- Prioritization of publications that possess well-defined discussions, results, and conclusions.

3. RESULTS

In this section, we describe the findings of our review of our research questions.

3.1. Overview of studies

Selected publications are shown in Table 6. The table shows the publication year, title, and other details of these publications.

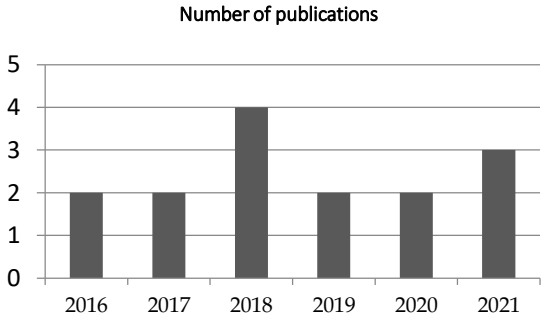

Figure 1: Distribution of the selected publications per year.

Figure 1 shows the number of publications that have

been published during the last six years from among the selected publications.

3.2. (RQ1) When, where, and who have published studies?

For this study, publications from the period 2016-2021 were extracted from various valid sources. 261 research papers were retrieved.

Out of these, 15 research papers have been selected for this systematic literature review (Table 6).

3.3. (RQ2) What are the algorithms that have been used so far for prediction?

Table 7: Most used algorithms for prediction.

Name of the algorithm	# of times used
Support Vector Machine	8
Long Short-Term Memory	3
Neural Networks	6
Linear Regression	7
Random Forest	4

To address the second research question (RQ2), machine learning algorithms were investigated and summarized. The algorithms used more than once are listed in Table 7.

3.4. (RQ3) What are the features/factors used in the literature to predict crop yields using machine learning?

Table 8: All features/factors used.

Feature/Factor	# of times used
Precipitation	10
Temperature	9
Evapotranspiration	4
Area	6
Nutrient	2
Sunshine duration	3

To address research question two (RQ3), features/factors used in the machine learning algorithms applied in the publications were investigated and summarized.

All features we were able to extract are shown in Table 8.

As shown in Table 8, The four most frequently utilized independent variables in crop yield analysis are precipitation, evapotranspiration, temperature, and area, while crop yield is considered as the dependent variable. These independent variables are part of a larger feature set that includes various other variables.

One of the feature sets is "Soil Information", which includes soil maps, soil type, pH, and product area. Another feature set is "Crop Information", which consists of variables related to crop weight, growth during the growth process, plant variety, and crop density. Additionally, there is a feature set called "Humidity", which comprises rainfall, humidity, predicted rainfall, and precipitation.

Moreover, the feature set,"Nutrients" encompasses the nutrients already present in the soil, as well as irrigation and fertilizer application. It is important to note that all of these features are used as independent variables in crop yield analysis.

3.5. (RQ4) What are the evaluation parameters and evaluation approaches used in the literature to predict crop yields?

Evaluation parameters were identified to solve the fourth research problem (RQ4). Table 9 shows all the evaluation parameters used and how they were used.

RMSE (Root mean square error), R2 (R-squared), MSE (Mean square error) and MAE (Mean absolute error) appear to be the most commonly used evaluation parameters.

Table 9: All evaluation parameters used.

Key	Evaluation parameter	# of times used
RMSE	Root mean square error	9
MSE	Mean square error	5
MAE	Mean absolute error	5
R^2	R-squared	4
WI	Willmott's Index	1
ENS	Nash-Sutcliffe efficiency coefficient	1

3.6. (RQ5) What are the challenges in the field?

To address the fifth research question (RQ5), publications were read to see if any issues or improvements were stated for future models. In several studies, data inadequacy (undersized data) has been cited as a problem. Data plays an important role in the machine learning process. One of the important issues faced here is the lack of good quality data. Dirty and useless data can make the whole process extremely tedious. A proposed improvement would be to integrate more data sources.

4. DISCUSSION

The analysis of our Selected 15 Studies underscores the significance of the authors' geographical location, as they represent various regions around the world. Despite our efforts to conduct an extensive search, it is possible that some valuable publications may have been missed, and the use of additional similar terms could have yielded new studies. Nonetheless, the search engine's ability to generate a considerable number of publications indicates a sufficiently comprehensive search.

In the context of our research question RQ1, which pertains to the temporal, geographical, and author-related aspects of the selected crop forecasting papers related to Machine Learning, we have chosen papers published between 2016 and 2021 from various reputable sources. Our systematic literature review (SLR) specifically focuses on the 15 selected papers, which we have found to be relevant and informative for our study.

In reference to RQ2, Table 7 demonstrates that a majority of experiments utilize a standard algorithm as a benchmark to evaluate the effectiveness of a proposed algorithm. Furthermore, recent studies have incorporated Deep Learning (DL) techniques, a subset of Machine Learning, for predicting crop yield. Long Short-Term Memory (LSTM) is one such DL technique that has shown promising results.

Regarding RQ3, the grouping of key features/factors can aid in visualization. The commonly employed factors for crop yield prediction include evapotranspiration, soil type, precipitation, and temperature, as shown in Table 8. Additionally, some studies have explored the use of trace elements such as magnesium, potassium, sulfur, and calcium as features. Notably, the types of features utilized in these studies differ, with temperature being measured as an average in some experiments and as maximum or minimum values in others.

In relation to RQ4, an investigation was conducted into the evaluation parameters utilized in several selected papers, as outlined in Table 9. The evaluation parameters included RMSE (Root Mean Square Error), R2 (R-squared), MAE (Mean Absolute Error), MSE (Mean Square Error), and MAE (Mean Absolute Error). The majority of the models achieved high accuracy values for their evaluation parameters, indicating that these models were capable of producing accurate predictions.

Regarding RQ5, the identified articles provided clear statements on the challenges encountered.

One of the studies focused on rice crop yield prediction in India using support vector machines (SVMs) (Gandhi et al., 2016). The study faced the challenge of selecting appropriate input variables for the SVM model. To overcome this challenge, the authors used a feature selection method to identify the most relevant input variables for the model.

Another study on rice crop yield prediction used artificial neural networks (ANNs) (Gandhi et al., 2016). The study encountered the challenge of over fitting, which occurs when the model performs well on the training data but poorly on the test data. To overcome this challenge, the authors used a regularization technique, which randomly drops out some neurons during training to prevent over fitting.

In the study on rice yield prediction using data mining (Dey et al., 2017), the authors faced the challenge of dealing with missing data. To overcome this challenge, they used the k-nearest neighbour (k-NN) imputation method to fill in the missing values.

Jambekar et al. aimed to predict crop production in India using data mining techniques. The authors

encountered the challenge of dealing with a large number of input variables, which can lead to poor model performance. To overcome this challenge, they used a wrapper feature selection method, which selects the best subset of input variables based on their performance on a validation dataset (Jambekar et al., 2018).

Kouadio et al. aimed to predict the yield of robusta coffee using soil fertility properties. The authors faced the challenge of dealing with non-linear relationships between the input variables and the target variable. To overcome this challenge, they used a non-linear machine learning technique called random forest, which can capture non-linear relationships between variables (Kouadio et al., 2018).

Shah et al. focused on crop yield prediction using regression techniques. The authors encountered the challenge of dealing with missing data and outliers in the dataset. To overcome this challenge, they used a data preprocessing method to fill in the missing values and remove the outliers from the dataset (Shah et al., 2018).

Xu et al. (2019) addressed the challenge "lack of accurate and reliable data on weather conditions, soil quality, and crop growth stages" by designing an integrated climatic assessment indicator (ICAI) that integrates various climatic factors, such as temperature, precipitation, and sunshine duration, to assess the impact of weather conditions on wheat production.

Another challenge is the complexity of crop growth processes, which require models that can capture the nonlinear relationships between input factors and crop yields. Meng et al. (2020) used stacked Long Short-Term Memory (LSTM) networks to predict rice yields. The stacked LSTM model was able to capture the nonlinear relationships between the weather factors and rice yield.

Furthermore, the performance of machine learning models depends on the quality and quantity of data available for training and testing. Mo et al. (2021) used LSTM networks to predict rice yields but faced the challenge of limited training data. To address this challenge, they applied data augmentation techniques to increase the size of the training dataset.

Kamath et al. (2021) used data mining techniques to predict crop yields, and their approach relied on the availability of historical data. They collected and processed data on weather conditions, soil quality, and crop yields to build decision trees that can be used to predict future yields.

However, it is possible that there may be additional challenges that were not mentioned in the literature.

5. CONCLUSIONS

This article presents a formal and original systematic review of the literature on crop yield forecasting using machine learning methods. The review followed established guidelines for conducting systematic literature reviews (Kitchenham et al., (2007)) to identify and categorize all relevant literature on the topic. After conducting a thorough search of a well-known electronic research database, a total of 261 original papers were identified. Through a rigorous filtration process, 15 relevant papers were selected for analysis and evaluation.

The selected papers were evaluated for their quality and analyzed based on the following research questions: (i) Who are the authors, when and where were the studies published? (ii) What are the algorithms that have been used for prediction? (iii) What are the features/factors used in the literature to predict crop yields using machine learning? (iv) What are the evaluation parameters and approaches used in the literature to predict crop yields? (v) What are the challenges in the field?

The findings of this research provide valuable insights and future directions for industry and research professionals in the field of crop yield forecasting using machine learning. The study reveals that various features are used in different research papers, depending on the scope and data availability of the selected publications. The choice of features is influenced by the availability of data and the purpose of the research. It was also observed

that models with more features do not necessarily provide better performance for yield forecasts. Several algorithms have been used in different studies, and while no definitive conclusion can be drawn as to the best model, some machine learning models have been found to perform better than others.

The results of this review indicate that neural networks and support vector machines are the most commonly applied algorithms in the existing literature. Therefore, the authors recommend that future studies explore the potential of deep learning algorithms, specifically LSTM-based models, for predicting crop yields. The study concludes by proposing further research on the development of the problem of crop yield forecasting.

In conclusion, our systematic literature review on crop forecasting papers related to Machine Learning has shed light on several key aspects, including the temporal, geographical, and author-related aspects of the selected papers. We have found that the use of Deep Learning techniques such as Long Short-Term Memory (LSTM) has shown promising results in predicting crop yield. Additionally, the commonly employed factors for crop yield prediction include evapotranspiration, soil type, precipitation, and temperature.

Evaluation parameters such as RMSE, R2, and MAE were used to assess the accuracy of the models, which produced high accuracy values. Despite the challenges encountered in crop forecasting, further data collection and exploration may lead to even better accuracy in predicting crop yield.

The article discusses various challenges and solutions encountered in crop yield prediction using machine learning techniques. Challenges include selecting appropriate input variables, dealing with missing data and outliers, and capturing non-linear relationships between variables. Solutions include using feature selection and regularization techniques, imputation methods, non-linear machine learning techniques, data preprocessing, and data augmentation. The article also emphasizes the importance of having accurate and reliable data for training and testing machine learning models, and the availability of historical data for data mining approaches. However, there may be additional challenges not covered in the literature.

The findings of this study provide valuable insights and future directions for industry and research professionals, and pave the way for further research in the field. The authors recommend the exploration of deep learning algorithms, specifically LSTM-based models, for improving crop yield forecasting.

6. REFERENCES

Ahmad, I., Saeed, U., Fahad, M., Ullah, A., Habib ur Rahman, M., Ahmad, A., & Judge, J. (2018). Yield Forecasting of Spring Maize Using Remote Sensing and Crop Modeling in Faisalabad-Punjab Pakistan. J. of the Indian Society of Remote Sensing. https://doi.org/10.1007/s12524-018-0825-8

Al-Zuabi, I. M., Jafar, A., & Aljoumaa, K. (2019). Predicting customer's gender and age depending on mobile phone data. J. of Big Data, 6(1). https://doi.org/10.1186/s40537-019-0180-9

Beulah, R., 2019. A survey on different data mining techniques for crop yield prediction. Int. J. Comput. Sci. Eng. 7 (1), pp.738–744 https://doi.org/10.26438/ijcse/v7i1.738744

Charoen-Ung, P., & Mittrapiyanuruk, P. (2018). Sugarcane Yield Grade Prediction Using Random Forest with Forward Feature Selection and Hyper-Parameter Tuning. Advances in Intelligent Systems and Computing, pp.33–42. https://doi.org/10.1007/978-3-319-93692-5_4

Conway, J. A., Brown, L. M. J., Veck, N. J., Wielogorski, A., & Borgeaud, M. (1991). A model-based system for crop classification from radar imagery. [Proceedings] IGARSS'91 Remote Sensing: Global Monitoring for Earth Management. https://doi.org/10.1109/igarss.1991.575404

Dey, U. K., Masud, A. H., & Uddin, M. N. (2017). Rice yield prediction model using data mining. 2017 Int. Conf. on Electrical, Computer and

Communication Engineering (ECCE). https://doi.org/10.1109/ecace.2017.7912925

Elavarasan, D., Vincent, D. R., Sharma, V., Zomaya, A. Y., & Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey. Computers and Electronics in Agriculture, 155, pp.257–282. https://doi.org/10.1016/j.compag.2018.10.024

Gandhi, N., & Armstrong, L. J. (2016). A review of the application of data mining techniques for decision making in agriculture. 2016 2nd Int. Conf. on Contemporary Computing and Informatics (IC3I). https://doi.org/10.1109/ic3i.2016.7917925

Gandhi, N., Armstrong, L. J., Petkar, O., & Tripathy, A. K. (2016). Rice crop yield prediction in India using support vector machines. 2016 13th Int. Joint Conf. on Computer Science and Software Engineering (JCSSE). https://doi.org/10.1109/jcsse.2016.7748856

Gandhi, N., Petkar, O., & Armstrong, L. J. (2016). Rice crop yield prediction using artificial neural networks. 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR). https://doi.org/10.1109/tiar.2016.7801222

Jambekar, S., Nema, S., & Saquib, Z. (2018). Prediction of Crop Production in India Using Data Mining Techniques. 2018 Fourth Int. Conf. on Computing Communication Control and Automation (ICCUBEA). https://doi.org/10.1109/iccubea.2018.8697446

Kamath, P., Patil, P., S, S., Sushma, & S, S. (2021). Crop Yield Forecasting using Data Mining. Global Transitions Proceedings. https://doi.org/10.1016/j.gltp.2021.08.008

Kitchenham, B., Charters, S., Budgen, D., Brereton, P., Turner, M., Linkman, S., Visaggio, G., 2007. Guidelines for performing Systematic Literature Reviews in Software Engineering. Retrieved from https://userpages.uni-koblenz.de/~laemmel/esecourse/slides/slr.pdf

Kouadio, L., Deo, R. C., Byrareddy, V., Adamowski, J. F., Mushtaq, S., & Phuong Nguyen, V. (2018). Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties. Computers and Electronics in Agriculture, 155, pp.324–338. https://doi.org/10.1016/j.compag.2018.10.014

Meng, X., Liu, M., & Wu, Q. (2020). Prediction of Rice Yield via Stacked LSTM. Int. J. of Agricultural and Environmental Information Systems, 11(1), pp.86–95. https://doi.org/10.4018/ijaeis.2020010105

Min, H. (2006). Developing the profiles of supermarket customers through data mining. The Service Industries J., 26(7), pp.747–763. https://doi.org/10.1080/02642060600898252

Mo, H., Zhang, Y., Liu, Y., & Zheng, Y. (2021). Prediction of rice yield based on LSTM long short term memory network. J. of Physics: Conference Series, 1952(4), p.042033. https://doi.org/10.1088/1742-6596/1952/4/042033

Rutledge, K., McDaniel, M., Boudreau, D., Ramroop, T., Teng, S., Sprout, E., Costa, H., Hall, H., & Hunt, J. (2011). Agriculture - National Geographic Society. In National Geographic. https://www.nationalgeographic.org/encyclopedia/agriculture

Shah, A., Dubey, A., Hemnani, V., Gala, D., & Kalbande, D. R. (2018). Smart Farming System: Crop Yield Prediction Using Regression Techniques. Proc. of Int. Conf. on Wireless Communication, 49–56. https://doi.org/10.1007/978-981-10-8339-6_6

Shahrin, F., Zahin, L., Rahman, R., Hossain, A. J., Kaf, A. H., & Abdul Malek Azad, A. K. . (2020). Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning. 2020 11th Int. Conf. on Electrical and Computer Engineering (ICECE). https://doi.org/10.1109/icece51571.2020.9393066

Shu, K. (2020). Prediction of Soybean Yield using Self-Normalizing Neural Networks. Proceedings of the 2020 5th Int. Conf. on Machine Learning

Technologies.
https://doi.org/10.1145/3409073.3409092

Slob, N., Catal, C., & Kassahun, A. (2020). Application of Machine Learning to Improve Dairy Farm Management: A Systematic Literature Review. Preventive Veterinary Medicine, p.105237. https://doi.org/10.1016/j.prevetmed.2020.105237

Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture, 177, p.105709. https://doi.org/10.1016/j.compag.2020.105709

Xu, X., Gao, P., Zhu, X., Guo, W., Ding, J., Li, C., … Wu, X. (2019). Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China. Ecological Indicators, 101, pp.943–953. https://doi.org/10.1016/j.ecolind.2019.01.059