How to Cite:
Govindasamy, S., & Jayaraj, D. (2022). A systematic study on crop yield prediction methods using deep learning. *International Journal of Health Sciences*, *6*(S2), 6809–6830. https://doi.org/10.53730/ijhs.v6nS2.6655

A systematic study on crop yield prediction methods using deep learning

S. Govindasamy
Assistant Professor / Programmer, Department of Computer Science and Information Science, Annamalai University, Annamalai Nagar, Chidambaram-608002
Corresponding author email: click2govind@gmail.com

Dr. D. Jayaraj
Assistant Professor/Programmer, Department of CSE, FEAT, Annamalai University, Annamalai Nagar, Chidambaram-608002
Email: jayarajvnr@gmail.com

***Abstract*---**DNNs (Deep Neural Networks) have estimated agricultural but lack comprehensive analysis of findings. The article gives an overview of the existing literature available in DNNs in predicting agricultural productions. This work's SLRs (Systematic Literature Reviews) were executed to assess most relevant studies. The searches resulted in 456 relevant studies based on quality assessments of which 44 primary studies were selected for this analysis. This work's examinations include data sources, key motives, targeted crops, algorithms used and features selected. Predominant usage of CNNs (Convolution Neural Networks) was found in the studies as their performances in terms of RMSEs (Root Mean Square Errors) are the best. One serious issue discovered was the absence of large training datasets which give rise to over fits of data and poor model performances. Since, researches look for gaps in studies; it is beneficial to highlight present issues and potential areas for further researches.

***Keywords*---**precision agriculture, DNNs, machine learning.

Introduction

Crop productions are influenced by a number of factors, including weather/soil conditions, crop types, application of fertilisers, and seed varieties. In [1], the researchers have constructed models for simulating the yield prediction and have resulted in satisfying outputs. In [2], Researchers have incorporated technology

Manuscript submitted: 27 Feb 2022, Manuscript revised: 18 March 2022, Accepted for publication: 09 April 2022

into their works. DNNshave estimated agricultural yields based on parameters mentioned above.

Although DNNs can improve performances, there is a paucity of literature on the limitations of applying DNNs for crop production predictions. They are dependent on the crop data, sources, and implementation frameworks.The conducted SLRs of this work were to gain overviews of the issues from literature. This study's findings can be very useful for academicians and practitioners of knowledge who would like to create new crop production prediction models. Obstacles are significant for researchers in this subject as they are aware of issues before developing their models. Creating new agricultural yield prediction models include difficult processes for practitioners and are detailed in SLRs where parameters used by models and algorithms necessitate essential ideas based on existing literature.

Hence, 456 relevant studies were gathered and examined. Ten RQs (Research Questions)which were framed for these examinations are also discussed. This study can benefit both researchers and practitioners based on solutions offered to RQs. The remaining sections of the article are structured as follows : Section 2 depicts the most prominent works in the domain of DNNs in yield prediction. The methodology of conducting this survey is elaborated in section 3 and the results were discussed in section 4. The general and research question based discussion is presented in section 5 and the arcticle is concluded with future scope in section 6.

Related Works

Very limited SLRs have been published on the use of DNNs in agricultural production predictions, though conventional crop yield predictions have been reviewed. SLRs have not highlighted the use of DNNs in agricultural yield predictions. Most of the existing methods deal with Machine Learning methods for crop yield prediction. Shallow learning of DNNs need to be separated and this work paves the way for evaluations of creations under which DNNs could be used for predicting crop productions.

In [3] ,MLTs are used for SLRs on crop yield predictions. Their study concluded that NNs (neural networks), specifically CNNs, LSTMs (Long-Short term Memories), and DNNs were commonly used to predict crop yields. The study also implied that the quantity of features varied based on studies. In some circumstances, item counts and detections were used instead of tabular data to predict yields. In [4], DNNs are widely used in image processing and a comparison of supervised methods is discussed. The study showed that yield maps obtained using GMMs (Gaussian Mixture Models) were outperformed by DNNs like U-Nets, Faster RCNNs (Recurrent CNNs) and CNNs. In[5], the authors counted fruits and estimated yields using DNNs. The study proved DNN's abilities to extract key features while providing fruit load determinations. DNNs including CNNs, deep regressions, and LSTMs were examined in the study. Using DNNs in [6] , the authors constructed self-predictable production platforms for identifying crop illnesses. The study [7]claimed that CNNs beat RCNNs and YOLO algorithms in their accurate diagnosis of crop diseases. Their CYP module's accuracieswere also higher when employing ReLU activation functions of ANNs(Artificial NNs). The applications of DNNs in dense agricultural settingsincluding recognitions and classifications, detections, counting, and yield estimates, were reviewed in [8] where DNNs outperformed most other methods according to the results of their survey.

In[9], the authors used MLTs to assess application of DNNs in predicting crop yields and measuring nitrogen status from tabular data. The study concluded that technological improvements in MLTs and specifically DNNs can provide cost-effective and comprehensive solutions for agricultural forecasting applications. In addition, hybrid systems using MLTs is predicted to play a vital role in near future. In[10], the author used DNN and found hybrid techniques and RNN-LSTMs which outperformed many other methods in predicting agricultural yields. Feedback loops and storage patterns of RNNs andLSTMswere the reasons for their great performances. They discovered that because those networks can deal with time-series data on agricultural yield, they are better at producing accurate forecasts.

Not many SLRs cater to use of DNNs in predictions of agricultural yields and objectiveanalyses are also deficient in this area. Majority of the assessments failed to specify or respond to technical topics making it imperative for practitioners to conduct additional research or provide additional support [11-14]. Furthermore, there are crops and crop-prediction technologies that have yet to be investigated. That means future study will have to be done with no prior knowledge of the features that are required or the hurdles that must be overcome [15]. A comprehensive literature review can expose this in a critical way. Hence, this study's SLRs on application of DNNs in agricultural yield predictionscan contribute in closing these gaps.

Research Methodology

The SLRs procedure is divided into three basic parts, each of which has multiple sub-processes. The methodology for this study is discussed in this section with elaborate details.A review methodology was created after the research topics were defined. Primary investigations should address the research issues, and data acquired and processed from such studies should provide answers. As a result, a search strategy for primary research with complete search phrases was devised. In addition, selection and quality assessment criterions were developed for evaluating individual studies and determining exclusion of studies from SLRs. Moreover, data extraction techniques [16-17] were designed for getting necessary information from important studies and the retrieved data was synthesized.

Research Question

SLRs are based on correct research questions which are critical factors[18]. First and foremost, this work aimed at justifying usage of DNNs to estimate agricultural yields. It was also crucial to identify crops utilised to test DNNs along with attributes that can predict crop yields. Data collections are crucial aspects of researches and hence this work investigated many implementations of DNNs that responded to the study's questions and thus frame SLRs.It's also useful to know algorithms that outperformother methods along with their ratings. Researches always face hurdles and hence the following research questions were established in the study:

RQ1 – Which motivations made the researchers to look for DNNs in Yield prediction?
RQ2 –Which crops were put to use in prediction?
RQ3 – Which data attributes were used to estimate agricultural yields while using DNNs?
RQ4 - What implementation Frameworks of DNNs were Used?
RQ5 – Which DNNs were used?
RQ6 - Which DNNs predicted crop yields better?
RQ7 – Which procedures were used in model evaluation?
RQ8 – What were the challenges faced and what were the solutions proposed?

Searching Techniques

Since, agricultural production predictions and DNNs are vast areas of studies, search techniques described above were limited based on this work's scope where specific search terms were created in obtain only potentially relevant results applicable to this study. The databases used were searched with unique search strategies and as a result, search strings were maintained, but with minor tweaks. The common search strings used were "Yield Prediction" , "DNNs" and "NNs" There were a total of 456 results.

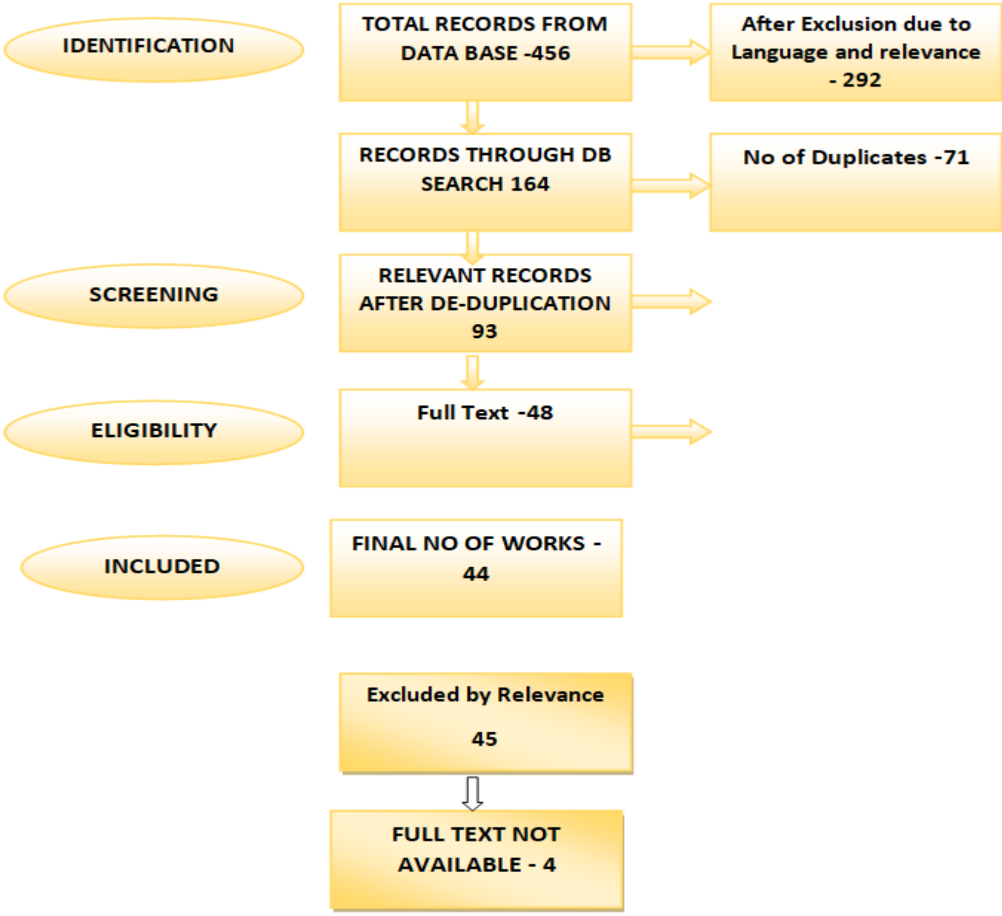


Figure 1 – Research Methodology

This study's SLRs was based on digital databases such as ACM , CAB , CABI , IEEE, IOP Science, MDPI , Science Direct Wiley , Springer Link , The selection of these databases was based on their indexed peer-reviewed publications with pre defined key words and opposed to Google Scholar which does place these constraints for both non-peer-reviewed and low-quality publications[18].

Selection criteria

Though many relevant studies could have been included or were found for this well-defined search strategy, many papers were found to be out of scope and hence were not considered for this work's SLRs. The publication's relevance to the application of DNNs in agricultural yield prediction, as well as whether it was written in English, was also exclusion criteria. As previously stated, multiple databases were used, resulting in duplicates of the papers chosen[19-20]. After eliminating duplicates, full-text editions were only selected as many publications had only abstracts. Another factor considered were publications of peer-reviewed journals. Finally, despite being referenced in the title or abstract, several articles did not incorporate any DNNs for agricultural yield prediction. As a result, these papers were not considered. After the exclusion criteria were applied, we were left with 44 papers for evaluation and study. The same is presented in figure 1.

Quality Assessment for Shortlisting

Quality assessments of this work had certain exclusionswhich resulted in published results (44) for added syntheses and meta analysis. The criterions were clarity, objectives, breadths, and contexts of the works. The solutions offered in the studies were validated for reliability. Quality assessment criterions included research process documentations, responses to established study questions, primary presentations and negative findings. These criterions were adapted from other SLRs [21] to avoid bias and enhance internal/external validities of studies.
A minimum score of 4 was set as the standard for providing high-quality data and studies scoring lesser than 4 were excluded. Figure 2 depicts the distribution of quality scores for selected studies. The mean was 7 while 6.69 was the median. Q1 reveals 25% of scores< 4.85 while 75% > 4.85 based on the box plot, In Q2, 50% <6.69 and in Q1 (mean), 50% > 6.69. In Q3, 25% >7.35 points, while 75% < 7.35. Each one of the eight criterions were assigned point values on a scale of 1 to 0, with 'Yes' equalling 1 points, 'Partially' equalling 0.5 points, and 'No' awarding 0 points. If the offered solutions were effectively presented, supported by empirical research, and cited. If the proposed answers were vague and no empirical researches were conducted for validating the outcomes, 0.5 points were deducted. No points were awarded if the proposed answerswere not properly stated or substantiated by actual evidences.

Data extraction

By using the excel sheet, required information was extracted. Citations Titles, Authors, Document/Journal Titles, Keywords, descriptions, Index Terms, Subjects, publication Years, and Publishers were all extracted and based on this database, documents were allocated unique indices. Contents matching research questions were extracted based on these indices. Multiple excel sheets were created for enhancing ease of using multiple data sets as advised in [22]. In the case of primary study raw data, cells columns holding this data were filled with '1'.

Data synthesis and reporting

This work could obtain meta-data and research questions from selected papers ad data was gathered from them. When data was extracted, it was discovered that certain information overlapped and hence they were synthesized or combined or grouped. It was possible to identify trends and variances in data using this strategy. Weather conditions, Vegetation Indices, and Satellite/Aerial data were specifically ordered by their levels of importance for 'RQ3'. Moreover, data for 'RQ8' revealed that issues highlighted had multiple solutions proposed. However, for specific challenges, a range of solutions were given. As a consequence, related jobs were linked using synonyms, and offered solutions for particular issues were put together in single cells of excel sheets for extractions.Anaconda environments with python and Jupyter Notebook frameworks were used for synthesis as in [23-24]

Results

Primary study's overall statistics were examined and this section displays the results of analyses that correlate to research objectives. It is seen that majority of the publications were listed in the year 2020 or later , according to this SLRs (Figure 2). The importance of DNNs in crop production prediction has been demonstrated by the year-by-year occurrence of primary studies during the last five years. In particular, they have increased by about 300% since 2016.

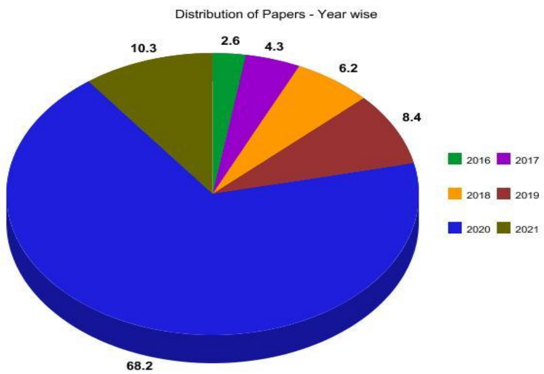


Figure 2. Distribution of published papers based on years

It is evident from Figure 4 that MDPI AG published the most primary studies (8) followed by Science Direct and Springer. Publishers with fewest publications were Korean Society of Surveying, & Taylor and Francis.

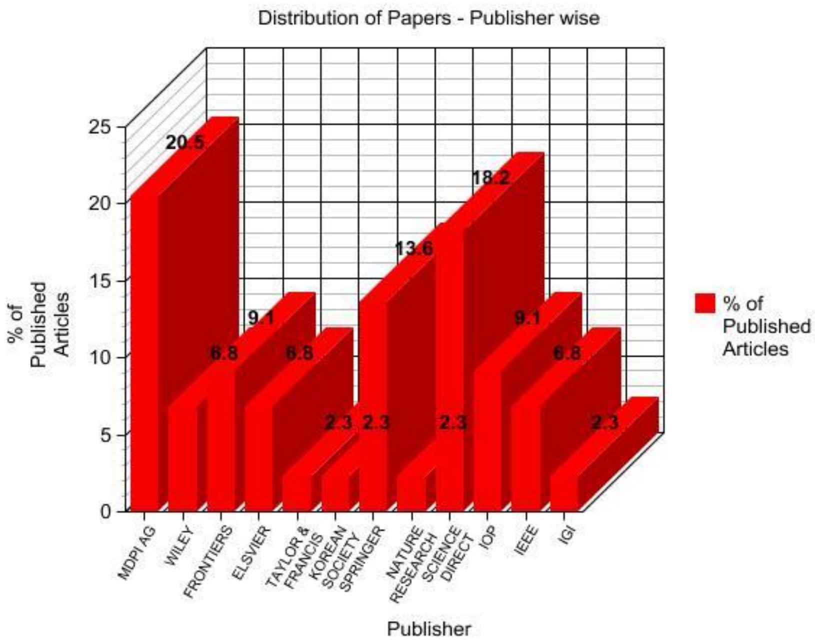


Figure 3- The distribution of papers per publisher

A number of keywords were employed in each primary study. Those keywords were isolated and entered into the data extraction form. The top occurrence of the key words in the result list are studied. The most common was “DNN” and followed by “ML” and “Yeild Prediction”. The table 1 and Figure 4 depipcts the distribution of papers based on the key words.

Keyword	Occurrence Count	Percentage of Occurrence
Deep Neural Network	21	14.8
Machine Learning	16	12.6
Crop yield prediction	13	12.8
Yield estimation	13	12.8
Precision Agriculture	11	12.1
LSTM	09	11.1
Crop Yield Estimation	16	12.6
Climate Sensing	2	1.1
SVMs	3	1.4
CNNs	7	1.9
Crop Yeild	9	11.1
Neural Networks	11	12.1
Feature selections	10	11.2

Table 1 – Distribution of Keywords

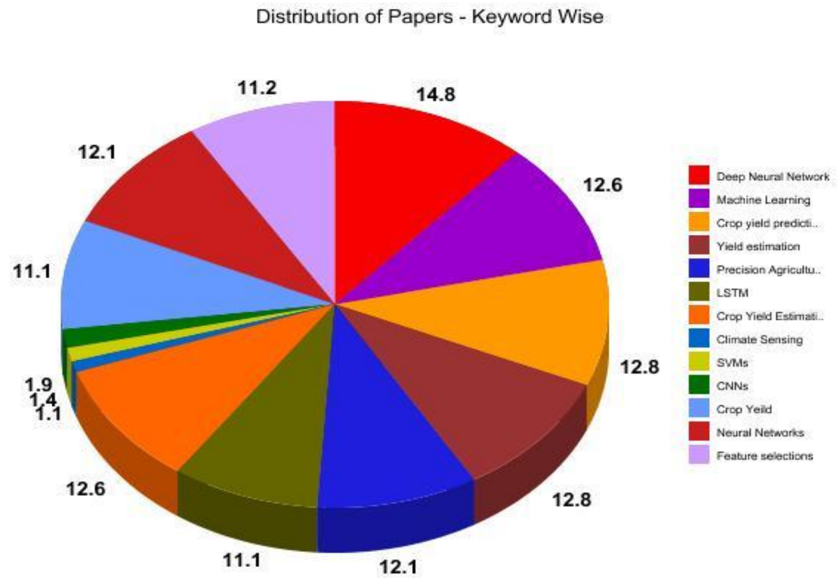


Figure 4 – Distribution based on Keyword

RQ1 - Which motivations made the researchers to look for DNNs in Yield prediction?

The first study issue concerns the primary purpose for using a DNNs technique to estimate agricultural yields. We uncovered ten important motivations, since DNNs has a number of characteristics that make it an excellent tool for predicting agricultural yields. However, depending on the objectives and limitations of each main study, the motivation for this alternative may vary. The top-5 values of important motives are shown in Table 2, demonstrating that the primary incentive for using DNNs was to 'process non-linear data' The results are depicted in figure 5.

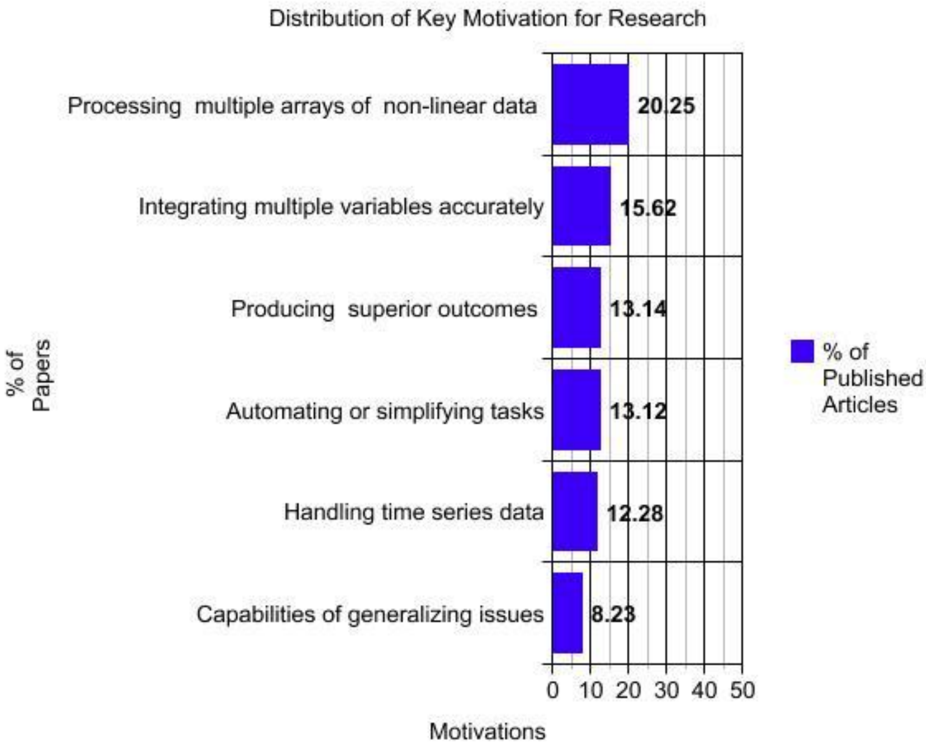


Figure 5 The distribution of papers per key motivation

Motivations	Occurrences	Occurrence percentages
Processing multiple arrays of non-linear data	21	20.25
Integrating multiple variables accurately	17	15.62
Producing superior outcomes	16	13.14
Automating or simplifying tasks	14	13.12
Handling time series data	14	12.28
Capabilities of generalizing issues	9	8.23

Table 2 –Distribution of Key Motives

When compared to the following two key reasons for adopting the DNNs approach in agricultural production prediction, processing diversified arrays in case of non-linear module that has shown significant growth after 2016 as shown in Figure-6. In fact, the trend gets tripled by the year 2022.

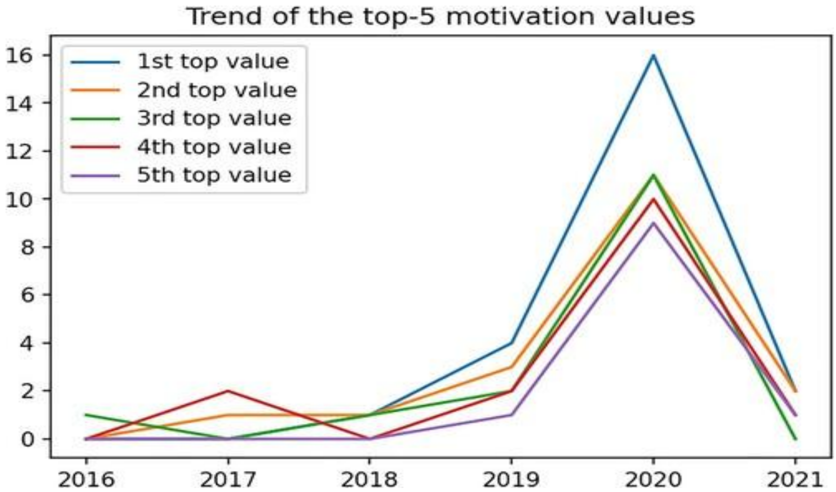


Figure 6 – Trend of top-3 values of key motivations

RQ2 - What crops were employed in yield estimations by DNNs?

During data extraction, the study discovered that DNNs has been used in a variety of crop predictions where the most common was maizeas shown in Table 3. Rice and Apple and maize are the most important crops, followed by maize, soybeans, rice, and apple orchards. DNNs retrieved 22 crops as shown in Figure 7 which indicates wide but uneven usage.

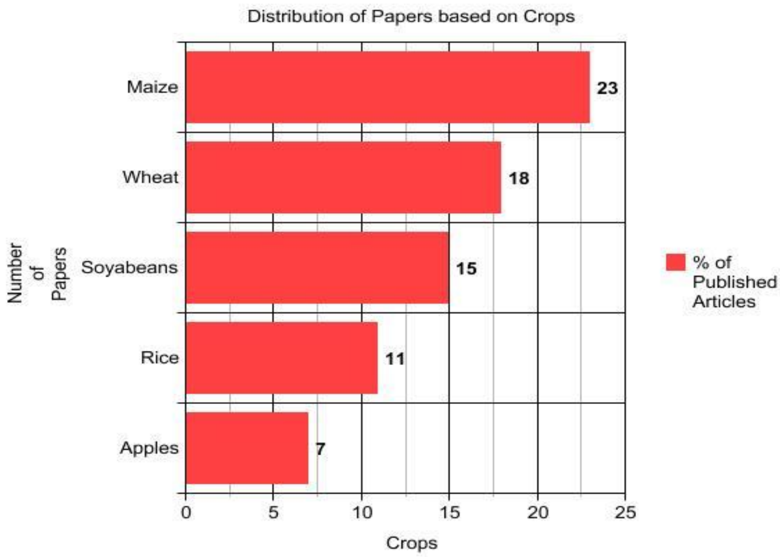


Figure 7. Distribution of papers based on crops

Crops	Occurrences	Occurrence percentages
Maize	14	23.33%
Wheat	8	13.33%
Soybean	7	11.67%
Rice	6	10.00%
Apples	5	8.33%

Table 3 – Distribution of papers based on Crops

The usage of DNNs is significantly increased and the same is clearly on the rise (Figure 8). Maize, soybean, wheat, and rice, in particular, have been studied on this topic in recent years. On this farm, maize is still the most common crop.

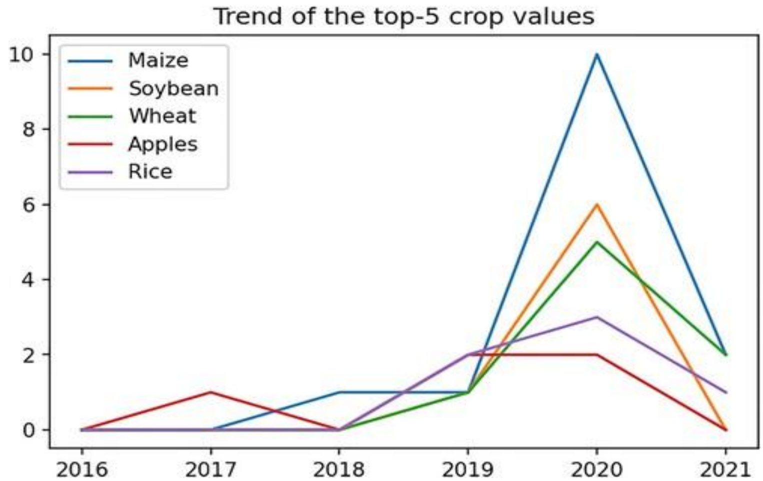


Figure 8. DNNs usage trend of top 5 crops

Throughout the original studies, the properties of the models evolved dramatically and were used in a variety of combinations As per Table 4, in more than half of the studies examined in this study are the common data source that were used in the studied literature. There were a total of 118 traits where image and precipitation data were used in common.

Features	Occurrences	Occurrence percentages
Image	29	9.58
Precipitates	27	9.02
Yeild	17	6.21
Temperature	16	5.93
Humidity	15	5.65
Rainfall	15	5.65

Table 4 – Distribution of papers based on Data groups

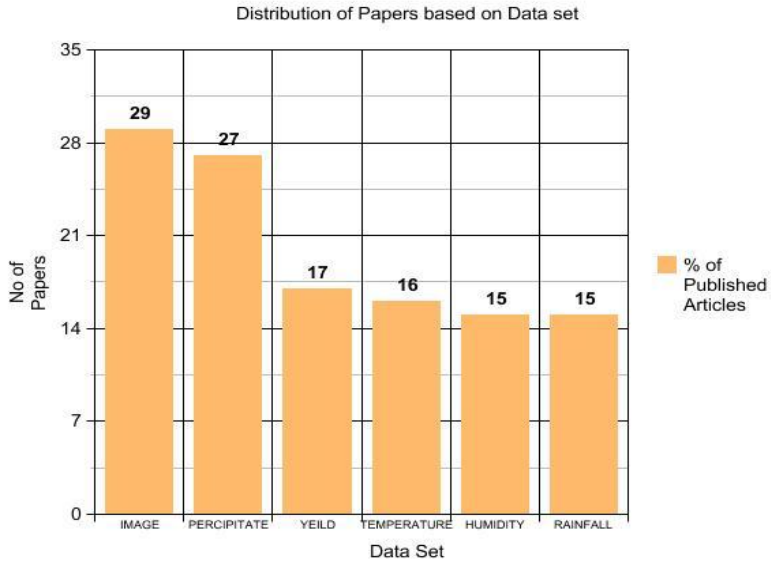


Figure 9 – Distribution of Papers based on Dataset

The data with related data features was split into 9 groups like vegetation indices, water and water conditions. Table 5lists these data features which are related to yields and water/weather conditions, the primary and common investigation parameters and depicted in Figure 9.

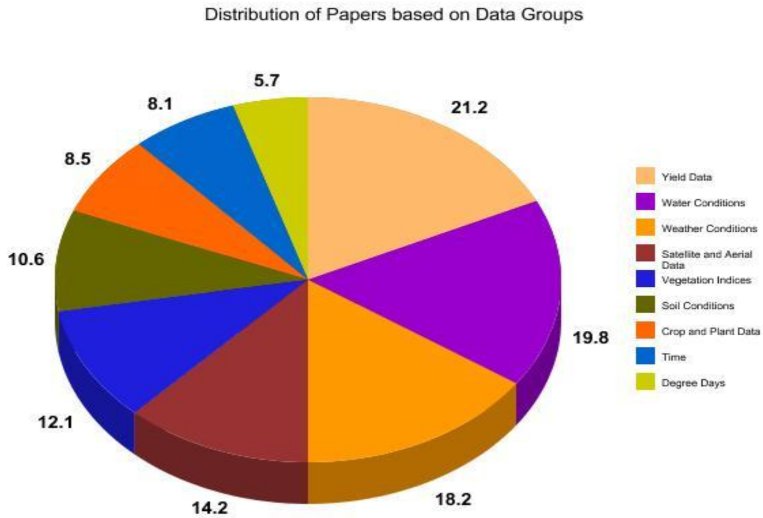


Figure 10 - Distribution of Articles based on Data Groups

RQ4 - What implementation Frameworks of DNNs were Used?

As with every technique, DNNs methods are implemented in a framework, as shown in Table 6. Major of the work did not present the framework in which they were implemented. TensorFlow and/or Keras were used by 56.86 percent of them. The figure 11 depict the distribution of frameworks used.

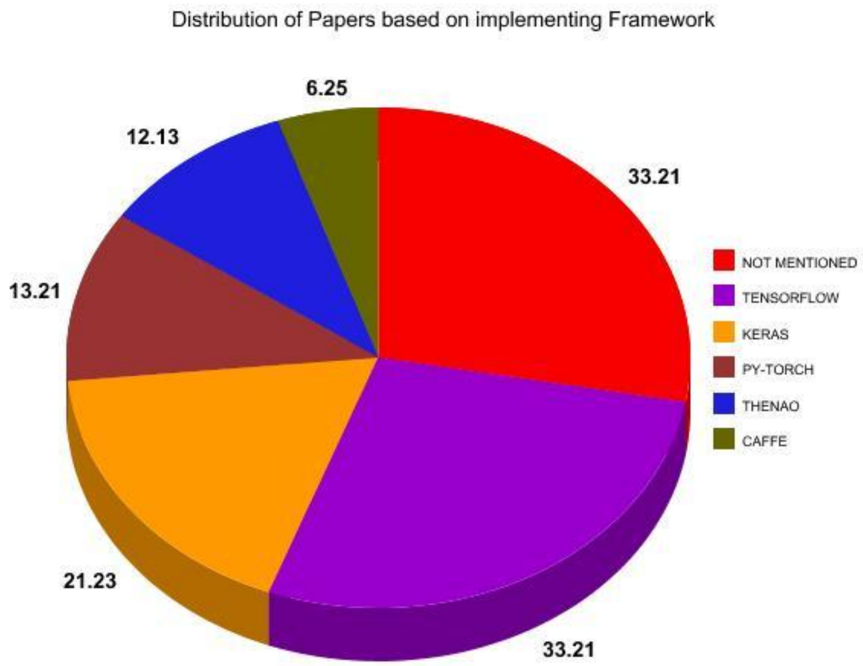


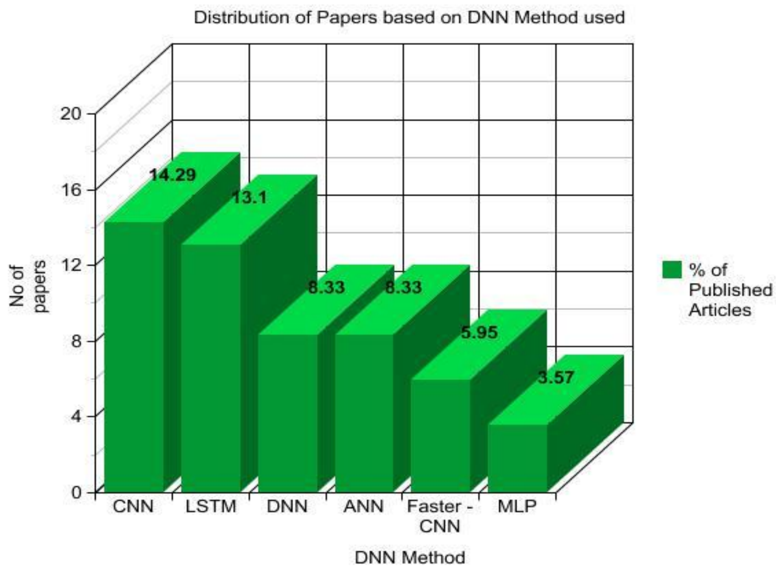
Figure 11- The distribution of papers per implementation framework

Frameworks	Occurrences	Occurrence Percentages
Not mentioned	23	33.21
TensorFlow	23	33.21
Keras	14	21.23
Py-Torch	9	13.21
Theano	8	12.13
Caffe	4	6.25

Table 6 – Distribution of research works based on Framework used

RQ5 – Which DNNs were used?

In many diverse challenges, DNNs produce highly precise outcomes. This can be accomplished by employing a DNNs method on its own, making changes, or even mixing many algorithms together, as shown in Figure 12. However, it's clear from Table 10 that DNNs methods haven't been employed together all that frequently, with CNNsas the most used methodfollowed by use of LSTM amongst 39 algorithms used.



s

Figure 12 – Distribution based on DNN used

Algorithm	Occurrences	Occurrence Percentages
CNNs	12	14.29%
LSTMs	11	13.10%
DNNs	7	8.33%
ANNs	7	8.33%
Faster R-CNNs	5	5.95%
MLPs	3	3.57%

Table 7 – Distribution based on type of Algorithm used

RQ6 - Which DNNs predicted crop yields better?

Most of the time, when choosing a DNNs to complete tasks, one or more algorithms, or even the same algorithm with tweaks, were chosen. The performance of the algorithms can then be compared for selecting the best algorithm. SLRs for agricultural yield prediction of this study showed CNNs performed best (Table 8). Image processing studies extracting features or counts to predict yields used the following methods: R-CNNs, CNNs, DNNs, and LSTMs and all were quick in executions. Bulk of these studies used CNNs or LSTMs to handle regression problemsin agricultural yield predictions. The studies proposed CNNs, DNNs, LSTMs, RNNs, or hybrid models. Figure 13 shows the algorithm based on performance.

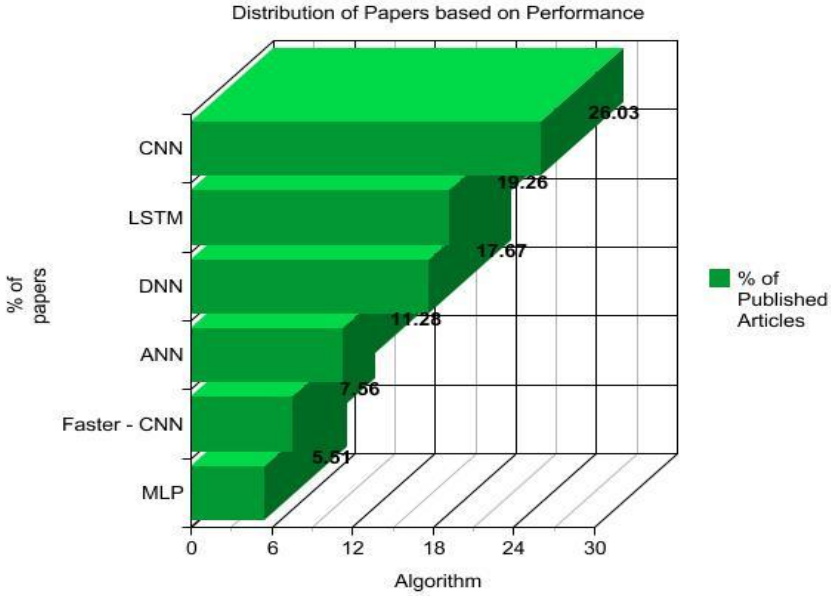


Figure 13- Distribution based on Performance

Algorithms	Occurrence	% of Occurrences
CNN	13	26.03
DNN	8	19.26
LSTM	7	17.67
Fast –CNN	5	11.28
CNN/LSTM	4	7.56
Hybrid	3	5.51

Table 8 – Distribution based on type of DNN used

It is seen that since the year 2017, the found best algorithms are being replaced by other algorithms which attains a peak in the year 2019 . In the year 2020 , the LSTM and DNN performed more significantly in prediction of crop yield. Figure 14 depict the same.

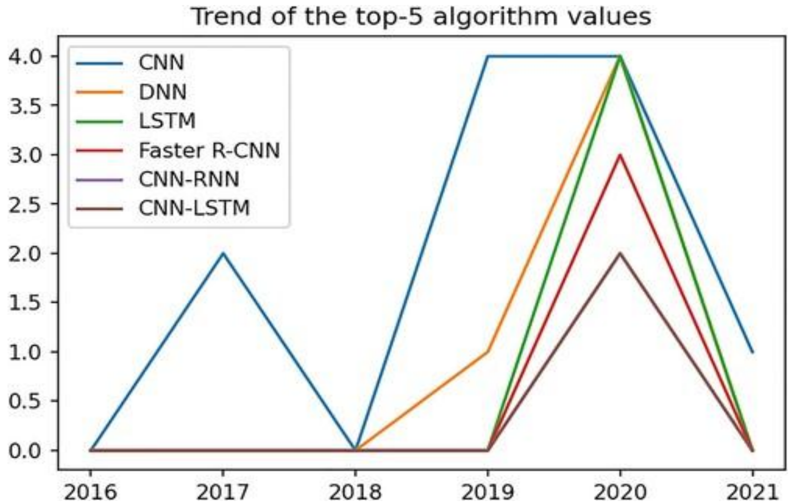


Figure 14 - Trend of Best performing algorithms.

RQ7 – Which procedures were used in model evaluation?

Different methodologies evaluated performances of DNNs in this study. Figure 15 depicts evaluation measures utilised in this work's SLRs. RMSEs were the most commonly utilised evaluation method (Table 9). In most cases, multiple evaluation approaches were utilised in primary investigations. The total number of evaluation approaches considered in these SLRs was 29.

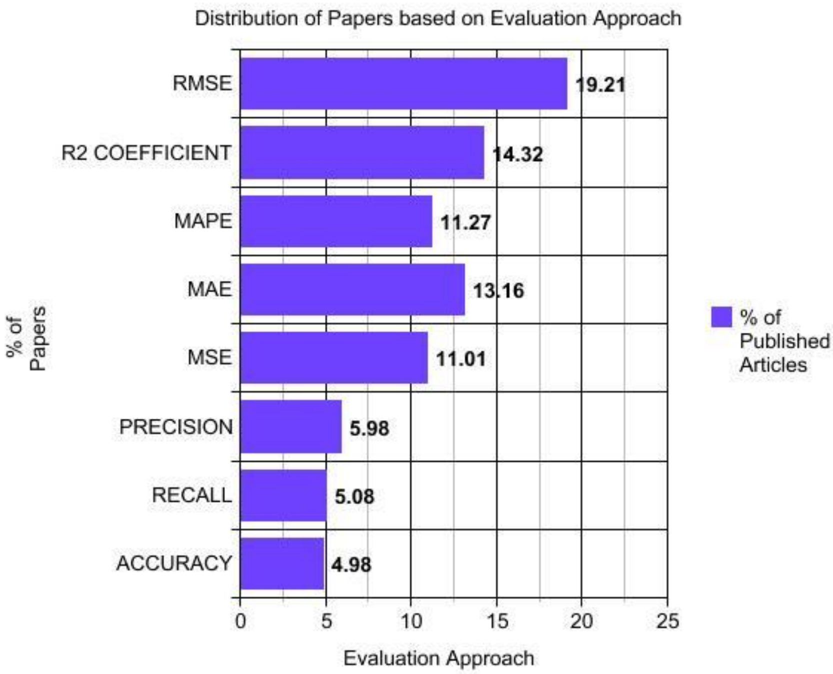


Figure 15. Distribution of works based on Evaluation Approaches

Evaluation Approaches	Occurrence	Percentage
RMSE	28	19.21
R2 Coefficient	21	14.32
MAPE	11	11.27
MAEs	15	13.16
MSEs	10	11.01
Precision	6	5.98
Recall	6	5.08
Accuracy	5	4.98

Table 9 – Distribution of Evaluation parameters

RQ8 – What were the challenges faced and what were the solutions proposed?

During the data extraction of the primary research, challenges and viable solutions were determined from the data. The most common obstacles, as well as alternative solutions, are listed in Table 10. The most difficult issue appears to be minimizing over fits and improving model performances by limiting training dataset samples. Data augmentation techniques like random crops, rotations, fancy principal analyses, modifying colour channels, adding filtering, increasing scales randomly, randomized rotations, vertical and horizontal inversions (mirroring), and colour distortions were used in primary studies to overcome this difficulty. Transfer Learning between related topics was also applied. This work's SLRs identified 158 issues and associated proposed solutions in total.

Challenges	Possible Solutions	Occurrences	Percentage
Limitations of voluminous training datasets to enhance performances while reducing over fits	random crops, rotations, fancy principal analyses, modifying colour channels, adding filtering, increasing scales randomly, randomized rotations and Transfer Learning between comparable jobs OR Digital data collections and recording	15	6.28
Capturing of Linear and non linear values of climates, and remotely sensed data withfeatures without details on non-linearity for estimating weathers, reducing dimensions, accounting for feature complications, extracting deep information	CNNs are used OR Considering ANNs especially with SSAE s (Stacked Sparse Auto-Encoders) OR MLTs OR CNN-LSTMs with 2-Dimensional CNNs (Conv2D) and LSTMs OR Using semi-parametric variants of DNNs for yield modeling	10	4.74%
Determine learning rates,	Using validation sets OR	6	2.84%

hidden unit counts, optimizers, activation functiond, and dropout settings.	evaluating models based on training processes and cost functions for reducing testing errors To increase performances, using mix of BHO s (Bayesian Hyperparameter Optimizations), early halts, and creating ensembles of models trained (i.e. bootstrap aggregations). Grid-searches with ten-fold cross-validations were used.		
For yield predictions, CNNs, RNNs and DNNs were utilized separately without fully exploiting their methodologies for combining time-series and constant data	Applying a Multilevel DNNs Network (MLDL-Nets) with various modules and CNNs for extracting the spatial features from images ,	8	3.6%
No Challenges were specified	No possible Solution specified	4	1.90%
Reducing time and efforts required to label training samples for item counts	Creating a training dataset from scratch OR Using random sub-samples, collecting subsets OR dividing images into smaller sections OR subdividing images into smaller divisions using SVM classifications at the picture level	3	1.42%
Enhancing effectiveness with which temporal patterns at diverse frequencies can be depicted.	Usage of LSTMs	3	1.42%
DNN models are less explainable due to their black box nature.	Performing feature selections on trained models using backpropagations	2	0.95%

Table 10 – Major Challenges faced

Discussion

Discussions related each research question is detailed in this section

Generic Discussion

The application of DNNs in agriculture, particularly for crop output predictions have undeniably increased four-fold since 2019. and linked with the United Nations' 2030 Agenda for Sustainable Development's prioritization of hunger relief and food security . As a result of the aforementioned prioritisations, studies published by 'MDPI AG were used as primary studies. The reason for lesser studies could be that Remote Sensing publications are open accesses. The discussions based on the research questions are as below.

RQ1-related:When it becomes clear that as more data is required, as the complexity of features grow, linear processing becomes more difficult.DNNs excel in dealing with vast amounts of complicated data in a non-linear fashion. As a result, one of the primary justifications for employing DNNs was to 'handle many array formats (refer Table 3). And similar to occurrences demonstrating the undeniable benefits of DNNs . All of these primary motives appear to be in line with future technical and labour trends that favour the use of increasingly complicated tools.

RQ2-related:It is observed that wheat and maize are largely dealt in crop yield prediction. According to the UN, sugarcane is also introduced as productive crop. It is obvious that DNN can be used to predict the yield of various other variety of crops which include grapes, fruits, etc which are lesser in production but has high nutrition benefits.

RQ3-related: The non-linear approach for agricultural yield prediction was boosted by using many features as inputs in DNNs. Images and precipitation were important elements since the former contains valuable information that Ground truths in phonological, geographical, or temporal data can be recognised by DNNs. The latter is regarded as one of the most crucial indicators of crop quality and quantity [21].As a result, groups of important data categories were formed in order to acquire a better grasp of the types of features employed in DNNs to forecast agricultural productivity This survey identified that the yield data is most significant and are used as ground input for regression task predictions where maps of crop yields were generated from satellite and aerial data.

RQ4-related:Each implementation framework has its own set of benefits, thus the choice should be made based on the study's requirements. In ierms of frameworks TensorFlows and Keras figured in most study’s implementations. Most primary studies did not explain implementation strategies as they were known frameworks.

RQ5-related: Several different DNNs methods were used in the primary investigations, probably due to the variety of features that required study and more efficient handling strategies. However, ANNs, CNNs, DNNs, RNNs, LSTMs, MLPs, RCNNs, and Faster RCNNswere used singly or in various combinations as far as geological data are concerned. Some networks were used exactly as they were intended.While others were tweaked to match the needs of other investigations. Although ANNs, DNNs, and MLPs are simplified versions of DNNs, nonetheless, they outperformed other MLTs.

RQ6-related: RQ7 anticipated that DNNs would outperform other MLTs. Furthermore, because CNNs can train features without needing to design them beforehand, and because spatial data was the most widely used feature, CNNs were the most popular DNNs technique with high performance. However, LSTMs performed better as their architecture is made up of memory cells, making it much easier to cope with such data. Algorithms such as R-CNN, performed well in primary experiments linked to fruit counting. Each algorithm takes a unique approach to a variety of problems.

RQ7-related:Each primary study used many evaluation methodologies to assess the performance of DNNs, resulting in a wide range of results. Those included inTable 12 were mostly utilised to ensure that model's outcomes were accurate where R2 (Determination Coefficients) and R (Correlation Coefficients) were used in evaluationsof yield predictionsand categorization tasks like fruit counting.

RQ8-related:The leading challenge of primary investigations are limitations of large training datasets to avoid overfits and improve model performances which addresses the issue of public data scarcity which can be overcome with data augmentations, Transfer Learning, and digital data collections and recordings. Many techniques including Transfer Learning, expanding training data, and usage of LSTMs or CNNs, have solved a wide range of problems. Despite the fact that the majority of major studies stated their issues clearly, and proposed a remedy, there were three instances where no problems were recognised and/or no potential remedies were offered. The RQ8 also opens up the avenues of introducing XAI methods as most of the DNNs have black box constraints. The future of crop yield prediction shall also lie on Explainable AIs that can over come black box properties of DNNs.

Conclusion and future work

This research work has identified a number of criteria that has a significant role in the prediction of crop yield using DNNs. Although the goals of using DNNs in the various studies may be similar, there are numerous obstacles to overcome and numerous solutions. In nonlinear modules, processing multiple array formats is insufficient to reduce overfits and improve model performance. To get around large training datasets, data augmentation and transfer learning are used. As a result, it's no longer surprising that agriculture is experiencing a boom in production prediction using DNN approaches is affected by both the type of data used and the type of crop. The majority of articles employed maize photos to employ attributes associated with water condition data, making photographs the most desired data source. TensorFlow, on the other hand, is used to implement DNN algorithms, despite the fact that the framework is rarely used. Various algorithms and DNNs methods were employed in various studies. The majority of the studies used supervised learning, according to the findings of this study. This came as a result of CNN's widespread use in agricultural yield prediction, where it outperformed other DLTs like DNNs, LSTMs, Faster R-CNNs, and hybrid models. R2s, MAPEs, MAEs, and MSEs were the most commonly used evaluation performance indicators in the selected research, followed by RMSEs.

It is clear that each case is unique when using DNNs to forecast agricultural yields. Particularly while considering problems and possible solutions for each of the case studies. Future research could focus on overcoming data constraints and providing explanations for DNN results so that they are not a "black box." For training models and improving performance, having access to data is critical. Because DNNs are 'black boxes,' it is necessary to do in-depth and to find how these models are arrived at specific discoveries. Improving the hyper-parameter tuning and labelling process is another area of future research that could have a significant impact on this issue. Processes that occur in all case studies take significant amounts of time and efforts to complete. The findings of this study will serve as a model for future research into the use of DNNs to predict agricultural productions.

The idea is to expand on the findings of these SLRs in future efforts with the goal of generating breakthrough DNNs for crop production predictions, while XAI should be considered for 'lifting' DNNs out of their black-box condition. It is also aimed to add new articles into this survey as and when they are published owing to the popularity of this research domain.

References

1. Apolo-Apolo OE, Martínez-Guanter J, Egea G, Raja P, Pérez-Ruiz M. 2020. DNNs techniques for estimation of the yield and size of citrus fruits using a UAV. European Journal of Agronomy. 115. doi:https://doi.org/10.1016/j.eja.2020.126030.
2. Apolo-Apolo OE, Pérez-Ruiz M, Martínez-Guanter J, Valente J. 2020. A cloud-based environment for generating yield estimation maps from apple orchards using UAV imagery and a DNNs technique. Frontiers in Plant Science.
3. Chlingaryan A, Sukkarieh S, Whelan B. 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and Electronics in Agriculture. 151:61–69.
4. Desa U. 2016. Transforming our world: the 2030 agenda for sustainable development. doi:https://doi.org/10.1891/9780826190123.ap02.
5. Dharani M, Thamilselvan R, Natesan P, Kalaivaani P, Santhoshkumar S. 2021. Review on crop prediction using DNNs techniques. Paper presented at the Journal of Physics: Conference Series.
6. Filippi P, Jones EJ, Wimalathunge NS, Somarathna PDSN, Pozza LE, Ugbaje SU, Jephcott TG, Paterson SE, Whelan BM, Bishop TFA. 2019. An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. Precision Agriculture. 20(5):1015–1029. doi:https://doi.org/10.1007/s11119-018-09628-4.
7. Food and Agriculture Organization of the United Nations 2019. Crops. FAO FAOSTAT
8. Fountas S, Mylonas N, Malounas I, Rodias E, Hellmann Santos C, Pekkeriet E. 2020. Agricultural robotics for field operations. Sensors. 20(9):2672. doi:https://doi.org/10.3390/s20092672.
9. Hani N, Roy P, Isler V. 2020. A comparative study of fruit detection and counting methods for yield mapping in apple orchards. Journal of Field Robotics. 37(2):263–282.

10. Jiang H, Hu H, Zhong R, Xu J, Xu J, Huang J, Wang S, Ying Y, Lin T. 2020. A DNNs to conflating heterogeneous geospatial data for corn yield estimation: a case study of the US corn belt at the county level. Global Change Biology. 26(3):1754–1766.
11. Khaki S, Wang L. 2019. Crop yield prediction using deep neural networks. Frontiers in Plant Science. 10. doi:https://doi.org/10.3389/fpls.2019.00621.
12. Khaki S, Wang L, Archontoulis SV. 2020. A CNN-RNN framework for crop yield prediction. Frontiers in Plant Science. 11. https://www.frontiersin.org/articles/10.3389/fpls.2019.01750/full.
13. Kim N, Lee YW. 2016. Machine learning approaches to corn yield estimation using satellite images and climate data: a case of Iowa State. Journal of the Korean Society of Surveying, Geodesy, Photogrammetric and Cartography. 34(4):383–390. doi:https://doi.org/10.7848/ksgpc.2016.34.4.383.
14. Kitchenham B, Charters S. 2007. Guidelines for performing systematic literature reviews in software engineering.
15. Koirala A, Walsh KB, Wang Z, McCarthy C. 2019. DNNs –method overview and review of use for fruit detection and yield estimation. Computers and Electronics in Agriculture. 162:219–234. [Crossref], [Web of Science ®], [Google Scholar]
16. Lee S, Jeong Y, Son S, Lee B. 2019. A self-predictable crop yield platform (SCYP) based on crop diseases using DNNs . Sustainability. 11(13):3637.
17. Lobell DB, Cahill KN, Field CB. 2007. Historical effects of temperature and precipitation on California crop yields. Climatic Change. 81(2):187–203. doi:https://doi.org/10.1007/s10584-006-9141-3.
18. Maimaitijiang M, Sagan V, Sidike P, Hartling S, Esposito F, Fritschi FB. 2020. Soybean yield prediction from UAV using multimodal data fusion and DNNs . Remote Sensing of Environment. 237:111599. doi:https://doi.org/10.1016/j.rse.2019.111599.
19. Oliphant PWT. 2012. Anaconda (Version 4.9.2): Anaconda Inc. https://www.anaconda.com/
20. Perez F. 2014. Jupyter Notebook (Version 6.1.5): Project Jupyter. https://jupyter.org/index.html.
21. Rossoum Gv. 1989. Python Language Reference (Version 3.7.9): Research Center Centrum Wiskunde & Informatica (CWI). https://www.python.org/. [Google Scholar]
22. Sutton RS, Barto AG. 1998. Introduction to reinforcement learning (Vol. 135). Cambridge, MA: MIT Press.
23. Van Klompenburg T, Kassahun A, Catal C. 2020. Crop yield prediction using machine learning: a systematic literature review. Computers and Electronics in Agriculture. 177:105709. doi:https://doi.org/10.1016/j.compag.2020.105709.
24. Xu X, Gao P, Zhu X, Guo W, Ding J, Li C, Zhu M, Wu X. 2019. Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu Province, China. Ecological Indicators. 101:943–953. doi:https://doi.org/10.1016/j.ecolind.2019.01.059.