Archives of Computational Methods in Engineering
https://doi.org/10.1007/s11831-021-09569-8

SURVEY ARTICLE

Emerging Trends in Machine Learning to Predict Crop Yield and Study Its Influential Factors: A Survey

Nishu Bali[1] · Anshu Singla[1]

Received: 20 August 2020 / Accepted: 9 March 2021

Abstract
Agriculture is one of the most crucial field contributing to the development of any nation. It not only affects the economy of nation but also has impact on the world food grain statistics. For agriculturist obtaining sustainable production of crop is always a challenge. Achieving optimum crop yield has always been a challenge for the farmer due to ever changing environmental conditions. The major reasons for unpredictability of crop yield are: land types, availability of resources, and changing nature of weather. Thus, the scientists all over the world are trying to discover techniques which can efficiently and accurately estimate the crop yield in much advance so that the farmers can take suitable actions to meet the future challenges. The main objectives of the study include: (a) Exploration of various machine learning techniques used in crop yield prediction; (b) Assessment of advanced techniques like deep learning in yield estimations; and (c) To explore the efficiency of hybridized models formed by the combination of more than one technique. The reviews done have shown good inclination towards hybrid models and deep learning techniques as means of crop yield prediction. The study also reviewed the works done by researchers in assessing the influence of various factors on crop yields and temperature and precipitation have been found to have maximum influence on the yields of different crops. Apart from climatic factors, agronomic practices adopted by farmers at various stages of growth of a plant also have considerable influence of the final yield of crop.

1 Introduction

Agricultural scientists all over the world are struggling with the problem of agriculture sustainability owing to the threats posed by various factors like increase in the price of food and energy, changing climates, continuous use and degradation of natural resources, alarming decrease in water availability and at the same time expected increase in population in the coming centuries. Crop yield is one amongst the important fields of agriculture that has attracted the attention of scientists owing to its strong impact on the national and international economy and to solve the problem of food scarcity. An accurate and timely forecast of crop yield can not only help the government of any country in taking various strategical decisions like planning import/export, formulating future policies like cost/selling price of crops and timely gauging the future threats but can also be a great help to a farmer whose livelihood is totally based on the expected yield of the crop [1].

✉ Anshu Singla
anshu.singla@chitkara.edu.in

[1] Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

1.1 Crop Yield Prediction: A Global Need

Agriculture is an important field that contributes to the overall development of any nation. Growing population of the world and the unexpected changes in climatic and soil conditions is forcing global researchers to uncover measures that can increase crop yield without adversely affecting our natural resources, called sustainable Agricultural practices. An accurate and timely crop yield prediction of current growing season can be an important step in this direction and can contribute in framing any Agricultural policy. Global bodies like Joint Research Centre (JRC), an initiative by European Commission, was established for Monitoring Agriculture Resources (MARS) in 1988. This body performs regular crop yield forecasting and provides monthly bulletins on expected yields to implement a Common Agriculture Policy (CAP) for the entire world. Early warnings on crop shortage or failure provides valuable information to food insecure countries and help in global food security. The system that generates this information uses real-time data like weather related observations and forecasts, data obtained from remote sensing like soil maps, crop characteristics and administrative regions in the form the inputs to the system. With these inputs, crop conditions are simulated. On regular intervals new yield statistics are added [2].

1.2 Global Demand and Supply of Crops

According to FAO (Food and Agriculture Organization), the demand and consumption of cereals has witnessed a steep growth in comparison with production in developing countries. There has been a continuous growth in the demand of rice, wheat and other coarse grains from 1964 to 2030 [3]. To meet the growing demands, there had been a considerable rise in imports of cereals in developing countries from 39 million tonnes a year (1970) to 130 million tonnes a year by 1997–1999. This rise in imports is expected to continue and may aggravate in coming years. By year 2030, these developing countries are expected to import 265 million tonnes of cereals, which amounts to almost 14 percent of their annual consumption, annually. Thus, the conditions of global market are quite volatile and are also on a falling trend in regard to real prices. These market conditions can be devastating for the progress of nations that are not thinking of taking steps to decrease their overall dependence on imports for the traditional crops. Thus it's a world challenge to change the present scenario in future and make countries more and more self-sufficient in fulfilling their food demands [4] which in turn requires a timely and an accurate estimate of yield of a crop.

1.3 Crop Yield Estimation Approaches

Crop yield estimation is essential but the involvement of complex interrelated environmental factors makes its precise measurement a very difficult and challenging task. Weather changes influence the plant growth at various stages leading to large intra-seasonal yield variations. Also the spatial variability of soil properties, farmer choices such as frequency of irrigations, pest and fertilizer application, crop rotation and land preparation practices adopted add to the complexity of accurately measuring the yield of crops. Thus, development of accurate and efficient crop yield forecast method requires correct assessment of weather and soil parameters through implementation of crop and soil management experiments. There are two diverse approaches, as shown in Fig. 1, followed to predict the preharvest yield of crops even though they are not mutually exclusive.

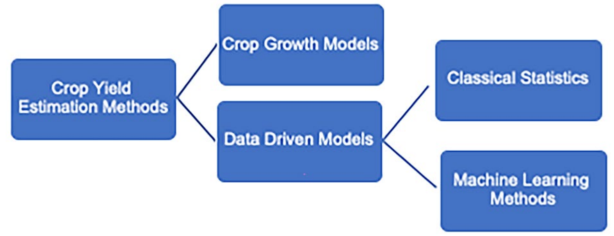

Fig. 1 Crop yield estimation approaches

Table 1 Crop specific mechanistic models [5]

References	Crop	Model
Wilkerson et al. [6]	Legumes	SOYOGRO
Jones and Kiniry [7]	Corn	CERES-Maize
Porter [8]	Wheat	AFRCWHEAT2
Jamieson et al. [9]	Wheat	Sirius

1.3.1 Crop Growth Model

Crop yield is always affected by the environmental factors. The impact of these factors varies at different stages of crop growth. Mathematical models can be used to represent these diverse interactions of plant physiological processes with environment to predict the final production or yield of a crop. Such mathematical models are called crop process models or crop growth models. These mathematical models use daily crop growth simulator to estimate biomass production potential and provide an abstract view of the implementation of dynamic behaviour of the plant's physiological phases. The actual data and various speculations about soil types, solar radiations, various management practices adopted, rainfall and temperature changes serve as an input fed through the models of seed formation and plant growth. Mostly, the mechanistic models are crop-specific [5] and some of the models for specific crops are shown in Table 1. In recent studies also, quite good results are obtained by using these semi empirical crop models [10]. Although efficient but these kind of models usually turn out to be expensive in terms of time and money, and are proven to be impractical for massive applications and agricultural planning.

1.3.2 Data Driven Model

Another approach, called empirical approach, is comparatively more practical and easy to use than crop growth model. In this approach, crop yield data for several years is considered and the set of parameters most effective or contributing to the yield variations are determined. Accepting these efficacious parameters as independent and harvest yield as dependent variables, empirical equations are formulated to compute the coefficients of these parameters. These coefficients are used to estimate the final crop yields. Every statistical model determines one set of parameters. Such techniques are relatively less expensive and easy in application and also they do not need any prior information on the various physiological processes involved in the growth of the plant or predefined structure of the model [5].

Thus, both the approaches have their own pros and cons and the need of the hour is a united framework that has the capability of modelling nonlinear relationship between soil factors, weather conditions and biomass and the yield of crop [5].

2 Methodology

This section presents the methodology adopted for an extensive literature review on the topic. A thorough and complete analysis of the domain required two steps to be taken (a) Collection of related literature, (b) Analysis of the final selected work. For accomplishing the first step, appropriate keywords were selected for searching related conference & journal papers from scientific databases and scientific indexing services like IEEE explore and Google scholar. The search keywords included words like, machine learning, crop yield estimation, neural networks, influential factors in crop yield. The collection consisted of more than 100 papers which were studied and the papers having close relevance with the domain were selected. It was found that collected papers were having related information but can be further categorized into separate sub domains relevant to the topic under study. The steps taken for planned literature study are as shown in the Fig. 2.

Thus, the filtering of papers as per the categories mentioned above led to identification of 17 papers studying the role of machine learning in quantifying the effect of various environmental factors and other agronomic practices on the final yield of various crops, 43 papers examining the role of machine learning in the field of crop yield prediction and 15 papers studying the role of deep learning in the field. Among the final selected papers, 28 papers belonged to Scopus indexed journals and rest of the papers also belonged to reputed journals. The final identified papers were thoroughly studied to find answers to following questions:

1. What type of machine learning techniques have been used in quantifying the effect of various environmental factors on final yield for various crops?
2. How machine learning techniques have contributed to the study of crop yield estimation?
3. What is the efficiency of neural networks in the field as compared to other machine learning techniques?
4. Which deep learning techniques have been explored in the field and how efficient are they in predictions?
5. What type of data sets have been used by the authors?
6. How remote sensing data has contributed to the study?

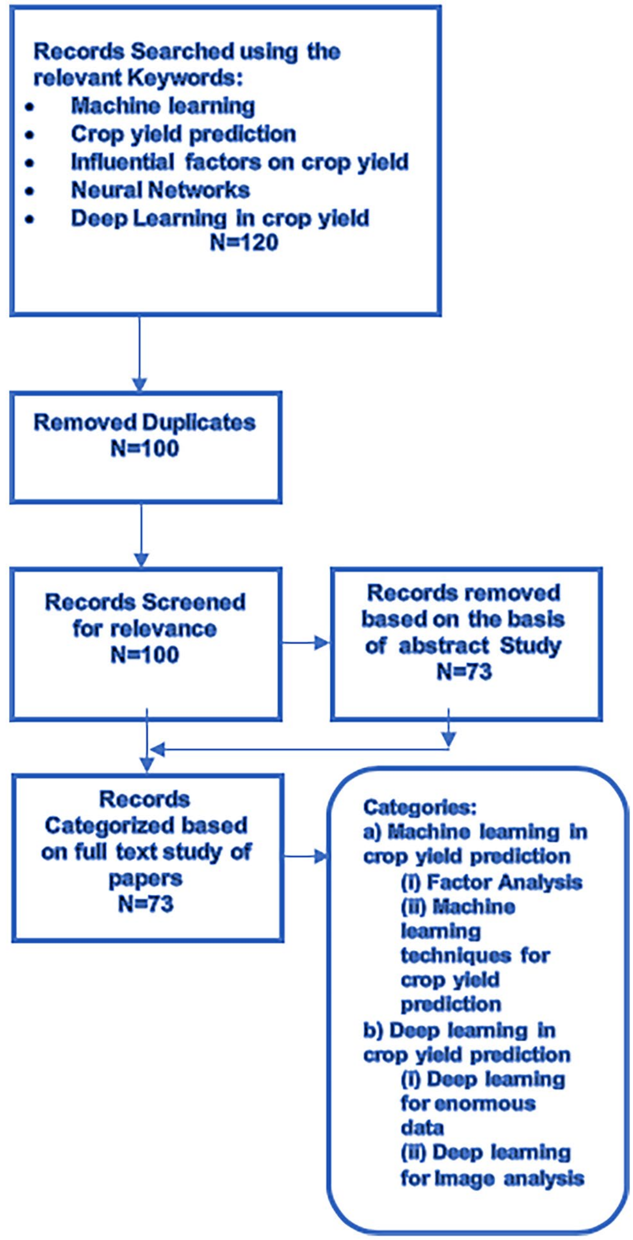

Fig. 2 Methodology for Systematic Review

The detailed study of various sections has been summarized in Tables 2 and 3 in Sect. 3 and Sect. 4 respectively and final results have been discussed in conclusion section.

3 Emergence of Machine Learning Techniques in Yield Prediction

Machine learning, an established field of computer science, has shown a promising future in different research areas. It is largely used by the data scientists and researchers who

Table 2 Existing machine learning techniques for crop yield prediction on different crop varieties

References	Anshal Savla et al. [11]	Ahamed et al. [32]	Lamba and Dhaka [33]	Zhang et al. [16]	Gonzalez-Sanchez et al. [31]	Nari et al. [37]	Bose et al. [38]	Pantazi et al. [39]	Kaul et al. [40]	Chlinga-ryan et al. [41]
Crop type										
MT					✓					
CB					✓					✓
P					✓					
W			✓				✓	✓		
C						✓			✓	✓
R		✓								
T					✓					
SB	✓			✓					✓	✓
M										
S										
A										
CP					✓					
Machine learning technique used										
SVM						✓				✓
RT	✓									
REPT	✓									✓
BG	✓									✓
BA	✓									
LR		✓								
KNN		✓		✓	✓					
ANN	✓		✓		✓				✓	
SVR					✓					
RF						✓				✓
DL						✓				
SNN							✓			
CP-ANN								✓		
XYF								✓		
SKN								✓		
MR					✓				✓	
MP					✓					
ERT						✓				

References	J.H. et al.[42]	Dai et al. [43]	Ji et al. [45]	Gandhi et al. [47]	Uno et al. [48]	Cheng et al. [50]	Ghodsi et al. [51]	Singh [52]	Shastry et al. [56]
Crop type									
MT									
CB									
P	✓								
W	✓						✓		✓
C					✓				
R			✓	✓					
T									
SB									
M	✓							✓	
S		✓							
A						✓			
CP									

Table 2 (continued)

References	J.H. et al.[42]	Dai et al. [43]	Ji et al. [45]	Gandhi et al. [47]	Uno et al. [48]	Cheng et al. [50]	Ghodsi et al. [51]	Singh [52]	Shastry et al. [56]
Machine learning technique used									
SVM									
RT									
REPT									
BG									
BA									
LR									
KNN									
ANN		✓	✓	✓	✓	✓	✓	✓	✓
SVR									
RF	✓								
DL									
SNN									
CP-ANN									
XYF									
SKN									
MR	✓	✓	✓						✓
MP									
ERT									

SVM: Support Vector Machine; *RT*: Random Tree; *REPT*: REP Tree; *BG*: Bagging; *BA*: Bayes; *LR*: Linear regression; *KNN*: K Nearest Neighbour; *ANN*: Artificial Neural Network; *ARIMA*: Auto regressive Moving Average Model; *ANFIS*: Adaptive neuro-fuzzy Inference System; *SVR*: Support Vector Regression; *RF*: Random Forest; *DL*: Deep Learning; *SNN*: Spiking Neural Network; *CP-ANN*: Counter Propagation-ANN; *XYF*: XY Fused Networks; *SKN*: Supervised Kohonen Networks; *MR*: Multiple regression; *MP*: M5-Prime Regression Tree; *ERT*: Extremely randomized Tree; *MT*: Mexicon Tomato ;*CB*: Common Bean; *P*: Potato; *W*: Wheat; *C*: Corn; *R*: Rice; *T*: Tomato; *SB*: Soya Bean; *M*: Maize; *CP*: Chick Pea; *A*: Apple; *S*: Sunflower

want to predict or find trends in the raw data. As data pertaining to agriculture is quite vast and is ever growing day by day, so machine learning techniques can be of great help in analysis of such huge data. The machine learning techniques are characterized into two broad categories: supervised and unsupervised learning techniques. In Supervised learning, machine is made to learn from the data provided and is trained to make decisions on new or unseen data. ANN, Bayesian network, decision tree, support vector machines, ID3, k-nearest neighbour, hidden Markov model are some examples of supervised learning. Unsupervised machine learning is a technique in which machine is made to infer based on the patterns identified in a dataset without prior knowledge of any referenced or labelled outcomes. Self organizing map, and partial based clustering, hierarchical clustering, k-means clustering are examples of unsupervised machine learning techniques. Recent progressions in the field of machine learning has inspired researchers all over the world to explore the potential of emerging techniques in different areas related to crops like yield prediction and quantification of factors affecting the crop yield. Following section describes the findings of prominent research works done in mentioned areas.

3.1 Effectiveness of Factors Affecting Crop Yield

Crop yield is a complex phenomenon involving contribution of various weather and soil parameters. Apart from uncontrolled factors related to climate and soil, there are many controlled factors contributing to the variations in the yield of a crop like farm practices employed by the farmers, type and quantity of fertilizers applied, frequency of irrigations applied on field etc. In this scenario, it becomes essential to quantify the contribution of various factors responsible for crop yield. Following section provides a review of works done in the said area.

Vashisht et al. piloted experiments to assess the contribution of various agriculture practices like planting time and date, crop varieties and crop irrigation schedules on variations in bread wheat yield under different climatic variations. Field experiments were conducted using six seasons for time slice (PTS; 2008–2013). Simulation studies revealed a reduction in crop yield with rise in temperatures (maximum and minimum) [11].

In another study, increase in growing season average temperature was found to have negative impact on the winter wheat yield whereas growing season precipitation (GSP)

Table 3 Deep learning techniques for crop yield prediction on different crop varieties

References	Villanueva and Salenga [82]	Oliveira et al.[12]	Wang et al. [81]	You et al. [80]	Kuwata and Shibasaki [85]	Fourie et. al. [83]	Bargoti and Underwood [6]	Mohan and Patil [8]	Jiang et. al. [9]
Crop Type									
BM	✓								
CSB				✓					
C					✓				
OR						✓			
A							✓		
W								✓	
R								✓	
M								✓	
CC									✓
Machine Learning Technique used									
CNN	✓					✓	✓		
DNN			✓	✓	✓				✓
LSTM				✓					
DCNN									
RNN		✓							
SOM								✓	

CNN: Convolutional Neural network; *DNN*: Deep Neural Network; *LSTM*: Long Short term memory; *DCNN*: Deep Convolutional neural Network; *RNN*: Recurrent Neural Network; *SOM*: Self Organizing Map; *CNN*: Convolutional Neural network; *BM*: Bitter Melon; *SB*: Soya Bean; *CSB*: County Soya Bean; *C*: Corn; *OR*: Orchard; *A*: Apple; *W*: Wheat; *R*: Rice; *M*: Maize; *CC*: County Corn

showed encouraging impact on the yield. Cobb–Douglas production function was used to quantify the effect of various factors involved in the study [14].

Linear Regression model was examined for quantifying the effect of different meteorological parameters on the rice yield in district Raipur, India. Particular stages of plant growth such as seedling, tillering, 50% flowering and maturity, were selected and effect of parameters on plant during these stages was analysed. Different correlations (positive and negative) were exhibited between various parameters and growth stages [15].

Zhang et al. explored the utility of remote sensing technique in collecting and analyzing data obtained at different stages of plant growth to find the contribution of growth stages on final yield of winter wheat crop. Hyperspectral information was gathered at three different stages (jointing, heading and grain filling) and its effect on final crop yield was analysed. The study proposed an enhanced 2D correlation spectral analysis method to identify the perceptive wavebands. The contributions of different phases of crop growth to the estimated yield was determined by the models based on coefficients of partial least square method using complete spectral information. Support Vector machine model was found to perform well with satisfactory accuracy and robustness [16].

The effect of different land use practices on the overall variations in surface temperature and its adverse effect on the yield of rice and wheat crop was studied. The study was limited to three different geoclimatic regions of Punjab. The satellite data under study was categorized into four major LULC classes: water, vegetative, built-up and plain soil. There was found to be increase in temperature in areas where the land use was transformed from agriculture, plain soil and forest to urban. Normalized difference vegetation index (NDVI) of the area was found to have positive correlation with rice and wheat yield, but significantly negative correlation with LST [17].

In another study, authors explored the impact of weather parameters and technological advancements like improved pesticide use and adoption of high yielding varieties on crop produce in various regions of Haryana. Principle component analysis (PCA) was carried out for preharvest yield estimation in Haryana on various agro-climatic zones. Four climatic zones were used for the study include different regions of Haryana, India. The estimated yield(s) as per the designed models were in good match with DOA wheat yield estimates in most of the districts [18].

Mukherjee et al. reviewed the influence of various climatic factors on yield of wheat in the states of Northwest India. Various important factors were included in study such as daily air temperature, standard precipitation and evapotranspiration index and ground water variability. The rise in count of days having temperature above 35^{o} C during maturity period led to loss in yield of wheat crop. Also there was depletion in ground water and surface water for irrigation owing to less rainfall in the wheat growing season (November–March). Thus, high temperatures along with acute scarcity of water and less irrigations led to an overall increase in yield reduction [19].

In another study, the authors reviewed the effect of date of sowing along with climate features on distinctive stages of growth of wheat crop in Uttar Pradesh, India. Effect of the two factors, date of sowing and climatic factors, was studied primarily on time of germination, plant height and number of tillers per spike. Correlation study revealed relationship that was directly proportional with temperature, rainfall and indirectly proportional with humidity for plant height and number of tillers [20].

Jiayu et al. conducted a study on effect of meteorological factors on different growing stages of rice and wheat crops. The study used counties of China from 1980 to 2012 and 2481 weather stations for the day-degree data. The AIC method was used to optimize the combination of various meteorological factors and to select the most influencing parameters. This method, outweighs the previous researches in identifying the factors with less influence on the yield and in exploring the relationship between meteorology and the yield [21].

Epule et al. compared the impact of climatic and non-climatic factors on the yield of 31 food and cash crops in Uganda. Multiple linear regression model was used for the analysis. Non climatic factors were found to have more effect on the yields of the crop as compared to climatic factors. Among climatic factors, temperature played major role in affecting the yield of crop which was followed by precipitation and CO_2 emanations due to deforestation. Among non-climatic factors, forest area dynamics (t value: -11.11; p value: 0.012 (1.20%); R: -0.5), wood fuel (t value: -9.40; p value: 0.032 (3.16%); R: 0.3) and tractors used (t value: 8.46; p value: 0.041 (4.09%); R: 0.2) showed their importance in the order given [22].

Another study explored some of the key climatic and agronomic factors affecting production of quality bread wheat seed production. It being difficult to identify all production factors at once, some important factors were initially selected for the study. The selected factors included rainfall and temperature among climatic factors while seed rate and nitrogen fertilization were the agronomic factors. The conclusions from the study stated that temperature, amount of rainfall and nitrogen fertilization of the soil were some of the most important influential factors affecting the state of the physiological processes in seeds and finally affected the overall yield and quality of seeds [23].

Authors in a study quantified the effect of different weather parameters on yield of wheat phenophase and grain yield. The variations in climate was studied from 1981 to 2014. The data collected from 10 regions of Mongolia was analysed using Agricultural Production Systems Simulator (APSIM) model. Owing to significant climate warming from year 1981 to 2014, there was a considerable reduction in spring wheat yield, with an average of 3564 kg ha^{-1}. The air and surface temperature variations were found to be the major weather parameters affecting the phenophase of spring wheat in Inner Mongolia. Between maximum and minimum average temperatures, the former had more pronounced effect than the latter. This was trailed by the relative dampness and sun powered radiation. Precipitation, wind speed and reference crop evapotranspiration were found to be least affecting among various climatic factors. Regarding spring wheat yield, temperature, solar radiation and air relative humidity were the major contributing climatological factors affecting in the eastern and western Inner Mongolia [24].

Meng et al. conducted a study for 20 districts over the time slice of 1987–2010 to explore the effect of climatic factors like precipitation and temperature on canola and spring wheat yield. The moment-based methods were used to analyse asymmetric associations between climate and crop yields. There was a rise in crop yield with the rise in growing season degree-days and pre-growing season precipitation whereas extreme temperatures during the growing season adversely affected the crop. Also the effect of variations in temperature was found to be more impactful than precipitation on yield distribution [25].

Safa et al. explored the efficiency of ANN in wheat crop yield prediction in New Zealand. Out of 140 factors, 6 were found to be showing considerable effect on the yield and were used as input to the model. The study explored indirect factors affecting the crop yield like conditions of farm land, area of wheat and irrigation frequency, machinery conditions and farm inputs in the form of N and fungicides consumptions. The final ANN model could predict wheat production with very low error margin of $\pm 9\%$ (± 0.89 t ha^{-1}) [26].

An assessment of climatic variables as factors affecting the yield of corn and soybean crop was done by Johnson, D.M. Regression analysis was used to find correlation between the factors and the final yield of crop. The study concluded that NDVI and daytime land surface temperature (LST) were positively and negatively correlated with the yield of both the crops whereas precipitation and night LST did not show prominent impact on the yield [27].

Parekh and Suryanarayana studied the effect of different combinations of weather parameters on crop yield using ANN technique in Vallabh Vidyanagar for the period 1981–1999. The model was further used to validate the dataset from 2000 to 2006. It was concluded that the combined effect of all the three parameters, is essential for accurate prediction of the crop yield [28].

Ruß et al. explored the role of IT in precision agriculture using Neural Networks in the field of wheat crop yield prediction. Among different factors, topology of network is found to be the most important factor that affects the efficiency of the network model, It was found that by increasing the amount of available data, the prediction accuracy of the model increases but size of network is not found to be having much effect on the effectiveness of the model [29].

The reviews done have shown that temperature and precipitation have been important influential factors affecting the yield of most of the crops. Agronomic practices adopted by farmers like amount of irrigation, time of sowing and pesticides used have also been found to be prominent in the studies. Also, machine learning techniques like linear regression, and neural networks have shown good potential in rating the factors by exploring different combinations of factors.

3.2 Machine Learning Techniques for Crop Yield Prediction

An accurate and timely forecasting of crop yield before harvest is extremely important. Due the involvement of diverse natured factors in the study, it has always been a challenge for researchers to predict the yield accurately. The advances in the field of machine learning has shown promising future in the field. Following section reviews the application of various machine learning techniques for crop yield prediction for different crop varieties.

Wheat crop yield estimation was done using support vector regression model. Various models were tested which included nine base learner models and two ensemble models. The results showed that out of the nine models, SVR showed the best learning efficiency and also ensemble models, in spite of considerable increase in cost, did not report much improvement in accuracy. Also, an increase in training data led to better results for all the models [30].

An extensive comparison of various machine learning approaches in crop yield prediction was done for multiple crops. The authors in this paper compared the machine learning techniques for predictive accuracies. Four accuracy rubrics [root relative square error (RRSE), correlation factor (R), normalized mean absolute error (MAE), and root mean square error (RMS)] were used to check the accuracy of the models. Results favoured M5-Prime and k-NN techniques with lowest RMSE errors (5.14 and 4.91), the smallest RRSE errors (79.46% and 79.78%), very low average MAE errors (18.12% and 19.42%), and at the same time showing highest correlation factors (0.41 and 0.42) [31].

In another study, Ahamed et al. explored data mining techniques to see the impact of various environmental (weather) factors including biotic factors and production area on the crop produce in different districts of Bangladesh. Clustering technique was applied to divide regions on the basis of attributes to be studied and suitable classification techniques (Linear Regression, KNN and neural network) were applied for predicting crop yield. Results of RMSE proved ANN as being better in predicting for some of the crops with missing values i.e. wheat, potato and aus. Linear Regression gave good results for boro and amo [32].Another study done by Lamba and Dhaka also found superiority of neural network model over other models (Statistical, Metrological, Simulation, Agronomic, Remote Satellite Sensed, Synthetic and Mathematical) [33].

Nath et al. explored the efficiency of Box Jenkin's Autoregressive Integrated Moving Average model, ARIMA (1, 1, 0), a time series modelling approach to predict wheat production for India. The forecast was done using previous yield data from 1949–1950 to 2016–2017 (68 years) and an effort was made to do the predictions for ten leading years. As model can be used only on stationary data, the data was converted to stationary data by differencing the time series. It was concluded that ARIMA (1, 1, 0) provided a reasonable predictive model and proposed a raise in production for duration of 10 years (2017–2018 to 2026–2027) [34].

Another study was done in Ukraine by Kogan et al. to explore the application of remote sensing for forecasting yield of winter wheat crop. Three approaches of forecasting, empirical regression based model, with Moderate Resolution Imaging Spectroradiometer (MODIS), empirical regression model based on weather-related parameters and CGMS were compared. The most reliable and accurate predictions for 2010 were obtained using the CGMS system while performance of all three approaches was same for 2011 (0.6 t ha^{-1} in April) [35].

In another study, the accuracy of calendar-day based approach in forecasting the phonological growth of soybean was explored. Prediction models were built using various machine learning techniques as artificial neural network (ANN), k-nearest neighbour (kNN) and regression. Results proved the approach under study to be feasible in predictions as all three methods achieved acceptable levels of accuracy for vegetative and reproductive stage using ANN, kNN and Regression models [36].

Kim and Lee studied remote sensing data for predicting corn yield in Iowa State using four machine learning techniques (SVM, RF, ERT and DL). Results proved in favour of machine learning techniques as a means for estimating the yield especially for DL which showed more stable results [37].

Bose et al. performed remote sensing spatiotemporal analysis using spiking neural network for crop yield valuation. The study made preharvest yield prediction six weeks prior to harvest with an accuracy of 95.4% and average error of prediction of 0.236 t/ha and correlation coefficient of 0.801 using a nine-feature model [38].

Pantazi et al. studied the variations in the wheat yield based on online multilayer soil data and satellite imagery crop growth characteristics. Three machine learning techniques, counter-propagation artificial neural networks (CP-ANNs), XY-fused Networks (XY-Fs) and Supervised Kohonen Networks (SKNs), were compared in performance. Results indicated that yield prediction in case of cross validation generated yield for low yield class is more than 91% which is highly significant in reference to intricate relationship between controlling aspects and the yield. For average and high-pitched yield class, accuracies obtained were 70% and 83% respectively. Among the three models, SKN, CP-ANN and XYF, SKN showed highest accuracy of 81.65% proving it to be the best model [39].

Kaul et al. compared ANN models in terms of their prediction capabilities and their performance with multiple linear regression models. Results indicated that ANN consistently produced more accurate predictions as compared to linear regression model [40].

Chlingaryan et al. reviewed various machine learning methods for predicting yields in remote sensed data. Reviewers discussed researches done in last 15 years on machine learning techniques for yield prediction. The study finds a promising future of remote sensing and machine learning techniques in providing good solutions for better crop estimation and decision making [41].

A comparative study of Random Forest models with Multiple linear regression models for yield estimation of varied crops (wheat, maize, and potato) was done by Jeong et al. Better results were shown by Random Forest models in all performance metrics. For RF models, the root mean square errors (RMSE) ranged from 6 to 14% of the mean obtained yield for all test cases whereas the values varied from 14 to 49% for MLR models. Thus, the study proved RF to be an efficient and useful technique for yield prediction not only at regional but also on global scale [42].

Apart from climate, soil is another important and contributing factor to the variation in yields of a crop. Dai et al. studied the effect of two soil properties, soil moisture and salt content of soil, on the yield of sunflower crop. Two machine learning techniques ANN and MLR were compared in the study for efficiency in analysing the effect of different input variables related to soil as soil moisture and soil salinity during different phases of crop growth. Connection weight method was adopted to measure crop sensitivity to soil moisture and salt stress at different growth stages. In this method, the connection weights of input-hidden and hidden-output were taken into consideration. Compared with MLRs, both ANN models (ANN-10 and ANN-6) showed better precision according to RMSE, RE and R2 values. Overall, the models based on neural network showed good performance in exploring diverse relationships between yield of the crop and and soil features during different stages of crop growth [43].

Usually the model proposed for crop yield predictions in an area is applicable for only that specific area but in a study, a generalized regression model was proposed for doing predictions in multiple areas. The model was first used for doing predictions in Kansas and the same model was tested in Ukraine. The results gave good accuracy of the proposed model by reporting 7% error in Kansas and 10% in Ukraine [44].

In another study, Ji et al. investigated the capability of artificial neural network on rice crop yield prediction in mountainous regions of The Fujian province of China where continuous weather aberrations such as typhoons, floods and droughts constantly impend the rice production. The objective of the study was to test the feasibility of ANN model to predict rice yield in typical climatic conditions and to compare it with multiple linear regression models. ANN model parameters such as learning rate and number of hidden layers were found to have significant effect on the accuracy of the model in rice yield predictions. Smaller data sets were found to require less hidden nodes and lower learning rates for model optimization. Comparative analysis of ANN models and MLR for accuracy of predictions favoured ANN models over MLR models [45].

Four BPN models of ANN were studied for corn yield prediction based on topographic features, vegetation indices and textural indices. Results confirmed that the use of topographic data along with vegetation and textural indices can greatly improve the prediction accuracies. Also, efficiency of ANN as a prediction model was highlighted in study [46].

Another study explored the potential of ANN in rice yield prediction in various districts of Maharashtra state in India. The study was based on different environmental predictor variables including temperature and precipitation on the yield of the Kharif season for the years 1998 to 2002. A multilayer perceptron neural network model was used on the current dataset. The results gave high accuracies of 97.5% with a sensitivity of 96.3 and specificity of 98.1. WEKA tool was used for the analysis purpose [47].

In recent times, the remote sensing has shown great potential in vegetation growth analysis and has good prospects in helping in the field of crop yield prediction. Uno et al. studied the importance of accurate crop yield maps in precision farming. The ability of remote sensing systems to acquire information for a large area in a brief span of time makes it more beneficial over harvester-mounted crop yield monitoring units for yield map creation. Both statistical and ANN models were used to develop yield estimation models. In order to reduce large amount of unnecessary information, usually generated in hyper spectral imagery, and to deal with the problem of over fitting, PCA was used. Results showed greater accuracy in prediction with ANN models than with either of three traditional empirical models [48].

Balaghi et al. studied the efficiency of least square regression models to estimate the yields of wheat crop in Morocco at province and national level. The predictions used NDVI/

AVHRR, rainfall sums and average monthly air temperatures as input parameters. The study tried to explore the future of NDVI in the field of crop yield prediction. Yields associated with provinces were assessed with errors between 80 to 762 kg ha^{-1}, depending on the province. whereas at national level, yield prediction showed 73 kg ha^{-1} error. The study recommended that proposed model can be useful for early forecast of wheat yield in Morocco [49].

A new artificial neural network approach was proposed for early yield prediction in fruit crops by means of image analysis. Two BPNN models were established for two phases of the season: opening period and the ripening period. Results were analysed in terms of various measuring factors such as correlation coefficients (R2), mean forecast error (MFE), mean absolute percentage error (MAPE), and root mean square error (RMSE). For early periods, the values came out to be 0.81, −0.05, 10.7%, 2.34 kg/tree, respectively whereas for the ripening period, these measures were 0.83, −0.03, 8.9%, 2.3 kg/tree, respectively [50].

Among various factors affecting the yields of a crop, climatic factors play an important role. Ghodsi et al. examined the effect of different climatic parameters on the yield of wheat crop in Iran. ANN models were used for analysing the data. ANNs are an encouraging substitute to econometric models. Eight important factors were considered in the study. In order to select the best ANN model, 11 varied ANN models with different number of neurons in hidden layers were tried and the optimum model was selected. ANN-MLP model based on conjugate gradient back propagation algorithm reported lowest MAPE making it the preferred or optimum model. The results reported have also supported the proposed ANN model as the suitable way for wheat yield prediction [51].

Another study was conducted to explore Multi-layered feed forward model of artificial neural network. Two learning algorithms namely gradient descent algorithm (GDA) and conjugate gradient descent algorithm (CGDA) were explored to train ANN for maize crop yield forecasting. Three layered MLFANN with two hidden layers containing (11, 16) units was found to be the best for training, validation and test sets [52].

Alvarez studied the effect of soil and climatic parameters on wheat yield in Argentina pampas to propose an accurate and efficient model for crop yield estimation. The region under study was split into 10 geographical areas. Two techniques, surface regression and ANN were studied and compared for efficiency. ANN has been found to be much better and accurate in predictions (RMSE = 0.05) as compared to surface regression method [53].

Park et al. compared three adaptive techniques, artificial neural networks (ANNs), general linear models (GLMs), and regression trees (RTs) in forecasting maize crop yield in eleven dissimilar land management tests in southern Uganda. GLM showed poorest results whereas RT showed the best results. ANN also showed promising results [54].

In another study, authors explored the potential of regression model in making the preharvest estimate of what crop yield in Ludhiana district Punjab. The efficiency of regression models was found to be highly significant at 5% significance level. The regression models developed on weather parameters justified 69% variations in crop yield [55].

A default ANN consist of an input layer, a hidden layer and one output layer. Shastry et al. compared the efficiency of default ANN with customized ANN (C-ANN) and MLR techniques for wheat crop yield prediction using some of the important influential factors for the study like amount of rainfall, crop biomass, soil evaporation, transpiration, Extractable Soil Water (ESW) and amount of fertilizer applied (NO3). Results showed significant improvement in yield prediction in C-ANN with higher R2 statistics and lower percentage errors as compared to MLR and D-ANN techniques [56].

Bhangale et al. proposed a methodology for crop yield estimations. The study was pertaining to various states of India. Each state was divided into different agriculture zones according to the amount of rainfall received and geographical location. The proposed model used SVM for analysing weather changes, K means approach to classify the soil and plants and ANN for crop yield prediction on the basis of data collected from previous techniques. The proposed agricultural DSS framework provided a means to predict the cropping information in advance from a set of inputs [57].

Bejo et al. explored the efficiency of ANN in various aspects related to crop yield estimation. These aspects included ANN in study of Environmental factors, soil and soil–plant Hydrology, Sensing Techniques, Biomass factor prediction, and controlled environment studies. The authors emphasized the need to explore the current trends in crop yield prediction for precision agriculture. ANN has been found to be showing great future in all the studied areas and can definitely prove to be an asset for agriculture researchers [58].

In this paper, Dahikar and Rode studied the effect of various predictor variables such as type of soil, PH value, presence of various nutrients in soil, depth temperature, rainfall and humidity on yields of various crops using ANN models. Study varied the model architectures by wavering the hidden layers used. The results of study showed a promising future for ANN in crop yield estimation even in rural areas [59].

Application of ANN for prediction of crop yield for rice, sugarcane and wheat crops was studied. The proposed model used weather variables as input whereas crop yield was the output produced. Different MLP algorithms have been explored. Among various algorithms and architectures explored, forecasts done using MLP architecture based on Conjugate gradient descent algorithm was found to be closest to the actual observations in almost all the cases. The study authenticates ANN as a promising technique for crop yield prediction [60].

In the area of green house operations, crop yield prediction is still a manual affair. Qaddoum, Hines and Iliescu proposed an efficient technique to estimate yield of tomato crop to help human operators in anticipating and accordingly take care of problems of both over demand and over production accurately. The influential parameters selected for the study comprised of different ecological variables inside greenhouse, such as, temperature, vapour pressure deficit (VPD), CO_2 and radiation, as well as yields from previous years. The proposed system was an intelligent system called EFuNN (evolving fuzzy neural network). Results reported gave weekly predictions with an accuracy of 90% [61].

Khoshnevisan et al. considered the impact of different energy input contributions on wheat harvest yield utilizing a consolidated model of ANN and fuzzy framework, ANFIS. Several ANFIS models were structured and trained for the study each utilizing the learning ability of ANN to formulate if-then rules of fuzzy system and to develop suitable functions designed from training pairs to generate inferences. The models were compared for accuracy with ANN for prediction. The most promising ANFIS structure found in the study reported R, RMSE and MAPE as 0.976, 0.046 and 0.4, respectively proving that ANFIS model has better prediction capability than ANN [62].

Naderloo et al. proposed adaptive neuro fuzzy system for crop yield prediction using different energy inputs. Owing to such a large number of inputs, the input vector was clustered into two groups and two networks were trained. The RMSE and R2 values were found 0.013 and 0.996 for ANFIS 1 and 0.018 and 0.992 for ANFIS 2, respectively. The values predicted by ANFIs1 and ANFIS 2 models were utilized as input for third ANFIS model. Results showed that parameters used in first ANFIS model had more profound effect on yield than other energy inputs. Also, the ANFIS 3 model using outputs of ANFIS 1 and ANFIS 2 models showed RMSE and R2 values as 0.013 and 0.996, respectively [63].

Kouchakzadeh and Ghahraman explored the effect of varying weather conditions on the wheat crop yield using ANN and ANFIS models. Various weather related parameters such as evapotranspiration, precipitation, daily temperature (max, min, and dew temperature), daily average relative humidity for twenty-two years at nine stations and net radiation, were considered as part of the study. The results of ANFIS model were found to be consistently more precise in terms of various statistical indices (R2 = 0.67, RMSE = 151.9 kg ha^{-1}, MAE = 130.7 kg ha^{-1}), when temperature (max, min, and dew temperature) data was used as an independent variable [64].

Pandey et al. compared the efficiency of ANN in predicting wheat crop yield in comparison to fuzzy time series method. Authors found fuzzy methods quite subjective as their output has an element of user's interpretation which leads to different results interpreted by different analysts. This drawback is not there in ANN which is objective in nature as in ANN the prediction is done solely by the designed network and there are multiple interpretations of the results [65].

Balakrishnan and Muthukumarasamy compared efficiency of various machine learning techniques for various crops yield predictions. The significant classification measures used in the study were Support Vector Machine (SVM) and Naive Bayes. Two proposed ensemble models, AdaSVM and AdaNaive were explored to estimate the crop production over a particular selected time slice. The results showed great perfection by reporting very few errors in prediction and also there was appreciable amount of fall in the classification error for both the proposed techniques [66].

Priya et al. studied the efficiency of random forest technique for crop yield prediction for kharif and rabi seasons for rice crop. Data analysis was done using R-Tool and the results showed a promising future for random forest technique for massive crop yield predictions [67].

In another study, authors proposed a model for crop yield prediction employing data mining with association rules as a tool for prediction in different districts of Tamil Nadu in India. The study revealed that the projected model is efficient in predicting the yield [68].

Shree studied the impact of parameters related to soil and atmosphere on variations in crop yield. This paper predicts the crop yield and also suggests the best crop that should be sown for improving the quality and profits incurred from the agricultural sector. The soil and climatic factors were taken as inputs for the study including type of soil, temperature variations, relative humidity, ground water level, spacing, depth, pH of the soil, seasonal variations, fertilizers used and months of cultivation. This prediction was aimed at helping the farmers in determining whether the specific crop is suitable for a given soil type. The Bayesian algorithm was employed for predictions for achieving high accuracy and speed [69].

Ingole et al. explored the efficiency of Fuzzy logic in crop yield prediction and in selecting suitable crop under particular conditions. Authors generated a decision support system that gets values of inputs and provides the name of the crop that can be sown for the given inputs. Two sensing mechanisms were used light sensors and temperature sensors [70].

Use of Fuzzy logic for crop yield prediction was explored in which different Fuzzy models based on different partitions of Universe of Discourse and their effect on wheat crop yield prediction was studied. This paper proposed a method for wheat crop forecasting by using actual production as the universe of discourse and intervals based partitioning. The proposed method was found to be optimal and gave high precision with insignificant mean square and average forecasting error rate. The proposed fuzzy approach has proved to be an inerrant and efficient way to estimate wheat production [71].

Efficiency of Neuro fuzzy model (ANFIS) was explored for rice crop yield prediction. Various meteorological parameters were included in study and Gamma test was done to find the factors closely related with the yield of crop. The prediction was done on a time series data of 27 years and results have shown good efficiency of ANFIS model for rice crop yield prediction [72].

Table 2 provides a summarized view of the works done by researchers on crop yield prediction in terms of different crop type and machine learning techniques used. The columns in table containing less markings shows the need of further exploration in those crop types and ML techniques. Also maximum markings are in the column of ANN showing that neural networks have been a preferred choice of many researchers for the study.

Recent advancements in the field of ANN in the form of Deep learning has opened new avenues in the field. Following section has reviewed works done in the field of deep learning for crop yield prediction.

4 Deep Learning: An Emerging Trend in the Field of Agriculture

Among various machine learning techniques, neural networks have shown good results owing to their efficiency in dealing with both linear and non linear aspects of the data. Deep learning is an extended version of neural networks. Deep Learning extends classical machine learning by contributing more “depth” into the model and transforms the data by using several functions providing data a hierarchical representation, with several levels of abstraction [73]. The ability of feature learning, extracting information from raw data, inherent in deep learning architecture makes it particularly advantageous for solving complex problems. DL models provide classification accuracy even on very large datasets which are difficult to be dealt with by other machine learning techniques. DL architecture comprises of various different components, based on the overall network architecture adopted like Unsupervised Pre-trained Networks, Convolutional Neural Networks, Recurrent Neural Networks, Recursive Neural Networks. Deep learning has achieved considerable popularity in analyzing raster based data like videos and images although it can be used on any form of data, such as audio, speech, and natural language, and more commonly to continuous or point data like weather data, soil chemistry and population data [74]. Apart from yield estimation, deep learning based approaches have also shown good results in other agriculture based fields like early plant disease detection in different crops including fruit crops like banana [75, 76]. Following section discusses the contribution of deep learning in the field of crop yield estimation.

4.1 Deep Learning in Crop Yield Prediction

Deep learning models consist of a highly complex hierarchical structure with a large learning capacity. These features make these models particularly suitable for dealing with classification and prediction challenges. Various architectures of deep learning have been used depending upon the type of data involved. Newlands et al. studied the role of deep learning in assessing the potential risk in agricultural insurance or management. In this paper, the authors emphasized the need of accurate crop yield prediction for a proper index insurance of agricultural products. In the study, the authors compared the forecasting power of deep learning by gauging its functioning with other established predictive techniques. The study revealed a good potential of deep learning in the field by showing highest predictive accuracies. Also the authors inferred that use of deep learning avoids underestimating the inefficiencies and costs of insurance coverage occurring due to the use of vague metrics of real risk exposure [77]. In another study, deep learning was used for estimation of corn yield and was compared with SVR technique. The results favored DL as an efficient and accurate technique for yield estimation [78].

4.1.1 Deep Learning Approach to Cater Enormous Data Requirements

Application of any machine learning technique requires enormous amount of data. More the amount of data, more is the accuracy in predictions. In case of deep learning, this requirement becomes one of the bottlenecks for some areas where large data acquisition is quite difficult. Yield prediction in developed countries like USA is largely facilitated by easy access to large scale survey data and common variables related to crop growth. The developing countries have to struggle to get right amount of information for accurate yield predictions. Researchers are trying to propose techniques through which this data requirement can be fulfilled.

Cunha et al. proposed a new method of yield prediction for a crop before the start of the season called pre season prediction using data from multiple sources. Normalized Difference Vegetation Index (NDVI) has commonly been used as an indicator of vegetation activity of an area and is obtained through remote sensing of the farm. Although these indices provide a good insight of the conditions but they usually come with a cost i.e. cost of data acquisition and data processing to generate required analytics. This study proposes a new ML based method using data from multiple sources to perform crop yield estimation of soybean crop before the start of the season. The system comprises of RNN, Recurrent Neural Network model trained using precipitation, temperature, soil properties and previous years observed soybean yield as data for 1500 + cities in Brazil and USA. The two main highlights of the study were that system could perform yield predictions with much lesser amount of data as compared to existing yield forecasting systems and also the predictions could be done before the beginning of the crop season. Results have shown that it is possible to obtain dependable yield forecasts with very less data requirements as the proposed neural network model has the ability to identify and exploit unnecessary information inherent in soil and weather related data. Also the results showed that the prediction accuracy also varies with the type of crop into consideration as they vary in their physiological properties [79].

You et al. proposed a new approach based on modern representation learning ideas to predict county-level soybean yield in U.S. The lack of sufficient training data was accomplished through a new dimensionality reduction technique. In this, the raw images were treated as histograms on which deep learning architectures, including CNNs and LSTMs were trained to predict the crop yield. The account for spatio-temporal dependencies between data points was compensated by incorporating a Gaussian process layer ahead of Neural Network model. Experimental results have shown that proposed model has outperformed the customary remote sensing centred techniques by 30% in terms of RMSE and USDA national estimates by 15% in terms of Absolute Percentage Error (MAPE) [80].

Wang et al. introduced the concept of transfer learning approach to predict the yields with less available data. The author emphasized the need of abundant ground truth training data for success of any deep learning model. This approach is particularly useful when deep learning needs to be applied in region with little training data. In such cases, fine tuning the pre-trained models can be of considerable help. In this study, model proposed by You et al. for predicting crop yield on remote sensed data has been fine tuned to predict the yield of soya bean crop in Brazil. Unlike the approach used by You et al. who tested and trained the model in same region, the authors tested the ability to transfer a model trained in one region to another. The study has shown a monotonic increase in the accuracy of prediction with the increase in the amount of data in most of the cases. Also the results have favoured the approach of transfer learning as an exciting new approach specifically helpful in the regions with less available data [81].

4.1.2 Convolutional Neural Network for Image Data Analysis

Advancements in the field of computers has made images a good source of input for researches in the field of agriculture. Nowadays, various remote sensing devices can capture images from areas which were considered to be unapproachable and difficult to study. The need of the hour is to explore techniques which can efficiently extract information from these images and use this information for accurate future predictions. Convolutional neural network, a class of deep learning neural network, has been a breakthrough in the area of image recognition and analysis. Following section reviews some of the works done by researchers in the said area.

Villanueva and Salenga studied the fruit bearing ability of bitter melon or bitter gourd crop using CNN method. The study was based on the scrutiny of the health of leaves of plant gathered from Ampalaya farms. The fruit bearing capability of the plant was judged on the basis of color and shape of leaves as small size, deformed shape and dark green, yellow or brown color of leaves signified bad class with no fruit bearing capability whereas normal size, green or light green color of leaves signified good fruit bearing capability of plant. Training of data was done through Keras, tensor Flow and Python worked together. The study concluded that testing of at least 293 images was required to train the CNN model to correctly predict the fruit bearing capability of the plant. On increasing the count of training images, the neurons of the model will also increase which will in turn increase the prediction capability of the model [82].

Accurate crop yield prediction in orchards is particularly useful for efficient load management. Traditionally, this is done by manually taking a measure of important features of the fruit trees (wood, buds, flowers, fruitlets, and fruit) during various stages of growth that can affect the crop yield. This is quite laborious and expensive process and may lead to inaccuracies. Fourie, Hsiao and Werner in this paper proposed an automated yield prediction system that optically estimates crop yield during various stages of growth. Deep convolutional neural networks (DCNN) was used to build object detectors that extract regions from the image that represents the fruit. InceptionV3 model, pretrained on ImageNet database, was used as an image feature extractor and was customized using own classifier for classifying parts of an image containing fruit or background features. The same framework can be applied to detect leaves, branches or other parts of orchard canopy [83].

Another study for fruit detection and counting on orchard image dataset was done by Bargoti and Underwood. Two feature learning algorithms, CNN and MLP were used along with an image segmentation approach. Metadata pertaining to the methods of capturing image data was added to the networks. Results showed improvement in the fruit image segmentation with the inclusion of metadata. The count estimate results produced using CNN and WS gave a squared correlation coefficient of $r^2 = 0.826$ [6].

Kuwata and Shibasaki explored deep learning techniques and machine learning technique SVR for estimation of

Illinois corn yield. In this study, the researchers proposed a methodology for crop yield estimation using deep learning to unearth features prominently affecting crop growth and yet to be quantified. Various environmental factors selected for the study included NDVI (Normalized Vegetation Index), APAR (Absorbed Photosynthetically Active Radiation), canopy surface temperature and water stress index. Convolutional Architecture for Fast Feature Embedding (Caffe) was used to implement the estimation model of deep learning. Results showed highest accuracy in prediction with two InnerProductLayer model. Correlation coefficient was reported as 0.810 and RMSE was 6.298. For single InnerProductLayer, the values were 0.727 and 7.427 respectively. For SVR, correlation coefficient was found to be 0.644 and RMSE was 8.204. Results indicated that two InnerproductLayer model with trained model of Caffe can estimate crop yield index more accurately as compared to SVR model which overestimates once the crop yield index goes below 0.8 [85].

Accurate weather forecasting is one of the key features in success of any yield prediction in agriculture. Mohan and Patil proposed a dimensionality decreasing strategy, Self organizing Map (SOM), along with Latent Dirichlet Allocation (LDA) for predicting appropriate season and crop for agriculture purpose. Suitable season for an appropriate crop was decided with the help of deep neural network classification system. The Results of the study claims that proposed approach when compared to the other approaches for weather and crop prediction, proved to be more effective(7–23%) according to accuracy, sensitivity, specificity, precision and recall, than the previous methods [8].

Jiang et al. proposed deep LSTM, a special form of RNN model, to predict County level Corn yields. A large data comprising of county level corn yield and hourly weather data necessitated the need for using deep learning model. This paper claims to be the first to employ LSTM for corn yield prediction. The model gave quite good predictions which shows a promising future for LSTM in the field of crop yield prediction [9].

Table 3 shows the summarized view of deep learning techniques studied on different crops emphasizing the areas yet to be delved into. The success rate of deep learning techniques in the area of crop yield estimation can be an inspiration for researchers to do more exploration in the area.

5 Conclusion

Crop yield estimation is one of the important areas of agriculture which is essential for agriculture planning involving proper crop selection, giving farmers the correct estimates of their gains for the crop they are planning to sow and also for deciding the import and export decisions of the government of any nation. Many studies have already been done on different types of crops and different techniques for the estimations. Traditional ways are quickly becoming obsolete with the intervention and success of machine learning as a tool in various practical fields.

Machine learning techniques have the ability to extract information and identify patterns from structured as well as unstructured data and that too without the intervention of any human intelligence. These properties make it well suited for studies requiring future predictions from raw data. Also its ability to cater enormous amount of varied nature data further help in areas like agriculture where climatic and soil data is involved having spatial and temporal variations.

The accuracy of a machine learning technique is measured in terms of different evaluation metrics like Root Mean Square Error (RMSE), Correlation factor (R2), Root relative Squared Error (RRSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). Figure 3 shows values of R and RMSE reported for multiple techniques by different authors in their study.

Some inferences drawn from the study:

(a) The variations of same technique in different studies is owing to the use of different crops and diverse parameters of study for the models. Apart from weather and soil parameters, agronomic practices adopted by farmers have also shown considerable effect on the final yield of crop. The evidence shows that there is no standardization of parameters and how the parameters are being tuned. The focus is required to study the optimized deduction of parameters on which the model is based.
(b) As is evident from the graph, best results are shown by Neural Network technique, ANN and ANFIS. Hybridized model based on fuzzy and ANN has shown good accuracies proving to be efficient in the area. We can explore more such hybridized techniques in future.
(c) Advanced techniques of machine learning like deep learning has shown good potential in dealing with huge amounts of data. Also their efficiency in learning through pattern recognition in the data without any outside training has made them particularly suitable for the area of crop yield predictions.

6 Future Work

As is evident, among machine learning techniques, ANN based techniques hold a very promising future in the field of crop prediction. The enormous amount of data and that too from varied sources force the need of more efficient models than ANNs. Deep learning is a very recent

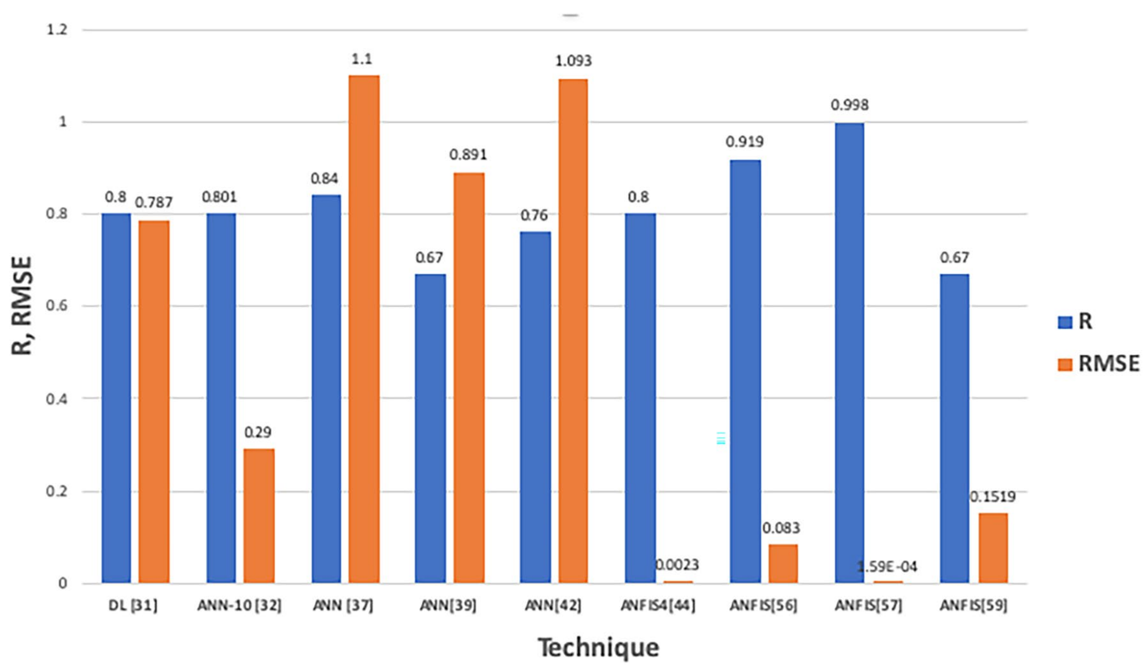

Fig. 3 Values of various evaluation metrics for various machine learning and hybrid techniques

advancement in the field of ANN. The technique has shown promising results in various research areas like speech recognition, medical diagnosis, drug design and many other important areas. Recently, its contribution in solving the food problems of the world has shown tremendous success. The advanced techniques like LSTM and RNN have shown good results paving way for more exploration for deep learning techniques in the field. Also, Fuzzy techniques have shown good results both in crop yield prediction and in parameter evaluation. Fuzzy theory is an important concept in decision making theory and science. However fuzzy logic is characterized by its membership function lying between 0 and 1 but not a non membership function. To overcome this, the concept of intuitionistic FS (IFS) was introduced for both membership and non membership functions between 0 and 1 which was extended to a more generalized form as Interval-valued intuitionistic FS (IVFS). As IFS and IVFS cannot cover indeterminate and inconsistent information, Neutrosophic sets have been a new trend to express indeterminate and inconsistent information which can be widely explored in the area of crop yield prediction as it involves factors of indeterminant nature.

Author contributions All authors contributed to the study as follows: NB: conceptualization, original draft preparation, literature search, editing the drafts. AS: supervision and guidance, suggestions for improvements, directions, critically reviewed and revised the work. All authors read and approved the final manuscript.

Funding This study was not funded by any organization, institution or research centre.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest Authors Nishu Bali, Anshu Singla have received no research grants or honorarium from any Company, or are members of any committee. The authors declare that they have no conflict of interest.

References

1. Lipper L et al (2014) Climate-smart agriculture for food security. Nat Clim Change 4(12):1068–1072. https://doi.org/10.1038/nclimate2437
2. Atzberger C (2013) Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. Remote Sens 5(2):949–981. https://doi.org/10.3390/rs5020949
3. Prospects by Major Sector (2020, April 10). http://www.fao.org/3/Y3557E/y3557e08.htm
4. Wright BD (2012) International grain reserves and other instruments to address volatility in grain markets. World Bank Res Obs 27(2):222–260. https://doi.org/10.1093/wbro/lkr016
5. Basso B, Cammarano D, Carfagna E (2013) Review of crop yield forecasting methods and early warning systems. In: The first meeting of the scientific advisory committee of the global strategy to

improve agricultural and rural statistics, pp 1–56. https://doi.org/10.1017/CBO9781107415324.004
6. Wilkerson GG, Jones JW, Boote KJ, Ingram KT, Mishoe JW (1983) Modeling soybean growth for crop management. Trans ASAE 26(1):63–73
7. Jones CA, Kiniry JR (1986) CERES-Maize: A simulation Model of Maize Growth and Development. Texas A&M Press, College station
8. Porter JR, (1993) AFRCWHEAT2: A model of the growth and development of wheat incorporating responses to water and nitrogen. European J Agronomy 2(2):69–82
9. Jamieson PD, Semenov MA, Brooking IR, Francis GS (1998) Sirius: a mechanistic model of wheat response to environmental variation. European J Agronomy 8(3–4):161–179
10. Chen Y, Donohue RJ, McVicar TR, Waldner F, Mata G, Ota N, Houshmandfar A, Mata G, Lawes RA (2020) Nationwide crop yield estimation based on photosynthesis and meteorological stress indices. Agric For Meteorol 284:107872
11. Savla A, Israni N, Dhawan P, Mandholia A, Bhadada H, Bhardwaj S (2015, March) Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture. In 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS) (pp. 1-7). IEEE
12. Oliveira I, Cunha, RL, Silva B, Netto MA (2018) A scalable machine learning system for pre-season agriculture yield forecast. arXiv preprint arXiv:1806.09244
13. Vashisht BB, Maharjan B, Jalota SK (2019) Management practice to optimize wheat yield and water use in changing climate. Arch Agron Soil Sci 65(13):1802–1819. 10.1080/03650340.2019.1578957
14. Geng X et al (2019) Climate change impacts on winter wheat yield in Northern China. Adv Meteorol. https://doi.org/10.1155/2019/2767018
15. Jain A et al (2019) Developing regression model to forecast the rice yield at Raipur condition. J Pharmacogn Phytochem 8(1):72–76
16. Zhang L et al (2010) Simulation and prediction of soybean growth and development under field conditions. Am-Euras J Agric Environ Sci 7(4):374–385
17. Majumder A et al (2020) Influence of land use/land cover changes on surface temperature and its effect on crop yield in different agro-climatic regions of Indian Punjab. Geocarto Int 35(6):663–686. https://doi.org/10.1080/10106049.2018.1520927
18. Jeev S, Verma P, Verma U (2018) Development of weather based wheat yield forecast models in Haryana. Int J Curr Microbiol App Sci 7(12):2973–2978. https://doi.org/10.20546/ijcmas.2018.712.340
19. Mukherjee A, Wang SYS, Promchote P (2019) Examination of the climate factors that reduced wheat yield in northwest India during the 2000s. Water (Switzerland) 11(2):1–13. https://doi.org/10.3390/w11020343
20. Agrawal DK, Nath S (2018) Effect of climatic factor and date of sowing on wheat Crop in Allahabad condition, Uttar Pradesh. Int J Curr Microbiol App Sci 7(09):1776–1782. https://doi.org/10.20546/ijcmas.2018.709.214
21. Jiayu Z et al (2018) The influence of meteorological factors on wheat and rice yields in China. Crop Sci 58(2):837–852. https://doi.org/10.2135/cropsci2017.01.0048
22. Epule TE et al (2018) The determinants of crop yields in Uganda: what is the role of climatic and non-climatic factors? Agric Food Secur 7(1):1–17. https://doi.org/10.1186/s40066-018-0159-3
23. Nadew BB (2018) Effects of climatic and agronomic factors on yield and quality of bread wheat (*Triticum aestivum* L.) seed: a review on selected factors. Adv Crop Sci Technol 06(02):356. https://doi.org/10.4172/2329-8863.1000356
24. Zhao J et al (2017) Assessing the combined effects of climatic factors on spring wheat phenophase and grain yield in Inner Mongolia, China. PLoS ONE 12(11):1–17. https://doi.org/10.1371/journal.pone.0185690
25. Meng T et al (2017) Analyzing temperature and precipitation influences on yield distributions of canola and spring wheat in Saskatchewan. J Appl Meteorol Climatol 56(4):897–913. https://doi.org/10.1175/JAMC-D-16-0258.1
26. Safa M, Samarasinghe S, Nejat M (2015) Prediction of wheat production using artificial neural networks and investigating indirect factors affecting it: case study in Canterbury Province, New Zealand. J Agric Sci Technol 17(4):791–803
27. Johnson DM (2014) An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sens Environ 141:116–128
28. Parekh FP, Suryanarayana TMV (2012) Impact of climatological parameters on yield of wheat using neural network fitting. Int J Mod Eng Res 2(5):3534–3537
29. Ruß G et al (2008) Data mining with neural networks for wheat yield prediction. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), 5077 LNAI, pp 47–56. https://doi.org/10.1007/978-3-540-70720-2_4.
30. Kamir E, Waldner F, Hochman Z (2020) Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. ISPRS J Photogramm Remote Sens 160:124–135
31. Gonzalez-Sanchez A, Frausto-Solis J, Ojeda-Bustamante W (2014) Predictive ability of machine learning methods for massive crop yield prediction. Span J Agric Res 12(2):313–328. https://doi.org/10.5424/sjar/2014122-4439
32. Ahamed ATMS et al (2015) Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh. In: 2015 IEEE/ACIS 16th international conference on software engineering, artificial intelligence, networking and parallel/distributed computing, SNPD 2015—proceedings. https://doi.org/10.1109/SNPD.2015.7176185
33. Lamba V, Dhaka VS (2014) Wheat yield prediction using artificial neural network and crop prediction techniques (A Survey). Int J Res Appl Sci Eng Technol 2:330–341
34. Nath B, Dhakre D, Bhattacharya D (2019) Forecasting wheat production in India: An ARIMA modelling approach. J Pharmacogn Phytochem 8(1):2158–2165
35. Kogan F et al (2013) Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models. Int J Appl Earth Obs Geoinf 23(1):192–203. https://doi.org/10.1016/j.jag.2013.01.002
36. Zhang Y et al (2018) Optimal hyperspectral characteristics determination for winter wheat yield prediction. Remote Sens 10(12):1–18. https://doi.org/10.3390/rs10122015
37. Kim N, Lee YW (2016) Machine learning approaches to corn yield estimation using satellite images and climate data: a case of Iowa State. J Korean Soc Surv Geod Photogramm Cartogr 34(4):383–390. https://doi.org/10.7848/ksgpc.2016.34.4.383
38. Bose P et al (2016) Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series. IEEE Trans Geosci Remote Sens 54(11):6563–6573. https://doi.org/10.1109/TGRS.2016.2586602
39. Pantazi XE et al (2016) Wheat yield prediction using machine learning and advanced sensing techniques. Comput Electron Agric 121:57–65. https://doi.org/10.1016/j.compag.2015.11.018
40. Kaul M, Hill RL, Walthall C (2005) Artificial neural networks for corn and soybean yield prediction. Agric Syst 85(1):1–18. https://doi.org/10.1016/j.agsy.2004.07.009

41. Chlingaryan A, Sukkarieh S, Whelan B (2018) Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. Comput Electron Agric 151(November 2017):61–69. https://doi.org/10.1016/j.compag.2018.05.012
42. Jeong JH et al (2016) Random forests for global and regional crop yield predictions. PLoS ONE 11(6):1–15. https://doi.org/10.1371/journal.pone.0156571
43. Dai X, Huo Z, Wang H (2011) Simulation for response of crop yield to soil moisture and salinity with artificial neural network. Field Crops Res 121(3):441–449. https://doi.org/10.1016/j.fcr.2011.01.016
44. Becker-Reshef I, Vermote E, Lindeman M, Justice C (2010) A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote Sens Environ 114(6):1312–1323
45. Ji B et al (2007) Artificial neural networks for rice yield prediction in mountainous regions. J Agric Sci 145(3):249–261. https://doi.org/10.1017/S0021859606006691
46. Serele CZ, Gwyn QHJ, Boisvert JB, Pattey E, McLaughlin N, Daoust G (2000) Corn yield prediction with artificial neural network trained using airborne remote sensing and topographic data. In: IGARSS 2000. IEEE 2000 international geoscience and remote sensing symposium. Taking the Pulse of the Planet: the role of remote sensing in managing the environment. Proceedings (Cat. No. 00CH37120), vol 1. IEEE, pp 384–386
47. Gandhi N, Petkar O, Armstrong LJ (2016) Rice crop yield prediction using artificial neural networks. In: Proceedings—2016 IEEE international conference on technological innovations in ICT for agriculture and rural development, TIAR 2016 (Tiar), pp 105–110. https://doi.org/10.1109/TIAR.2016.7801222
48. Uno Y et al (2005) Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. Comput Electron Agric 47(2):149–161. https://doi.org/10.1016/j.compag.2004.11.014
49. Balaghi R et al (2008) Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. Int J Appl Earth Obs Geoinf 10(4):438–452. https://doi.org/10.1016/j.jag.2006.12.001
50. Cheng H et al (2017) Early yield prediction using image analysis of apple fruit and tree canopy features with neural networks. J Imaging 3(1):6. https://doi.org/10.3390/jimaging3010006
51. Ghodsi R, Yani RM, Jalali R, Ruzbahman M (2012) Predicting wheat production in Iran using an artificial neural networks approach. Int J Acad Res Bus Soc Sci 2(2):34
52. Singh RK (2008) Artificial neural network methodology for modelling and forecasting maize crop yield. Agric Econ Res Rev 21(347-2016–16813):5–10
53. Alvarez R (2009) Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. Eur J Agron 30(2):70–77. https://doi.org/10.1016/j.eja.2008.07.005
54. Park SJ, Hwang CS, Vlek PLG (2005) Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions. Agric Syst 85(1):59–81. https://doi.org/10.1016/j.agsy.2004.06.021
55. Bal SK et al (2004) Wheat yield forecasting models for Ludhiana district of Punjab state. J Agromet 6(January):161–165
56. Shastry KA, Sanjay HA, Deshmukh A (2016) A parameter based customized artificial neural network model for crop yield prediction. J Artif Intell 9(1–3):23–32. https://doi.org/10.3923/jai.2016.23.32
57. Bhangale PP, Patil PYS, Patil PDD (2017) Improved crop yield prediction using neural network. IJARIIE 3(2):3094–3101
58. Bejo S, Mustaffha S, Wan Ismail W (2014) Application of artificial neural network in predicting crop yield: a review. J Food Sci Eng 4(1):1–9
59. Dahikar SS, Rode SV (2014) Agricultural Crop Yield Prediction Using Artificial Neural Network Approach. Int J Innov Res Electr Electron Instrum Control Eng 2(1):2321–2004
60. Laxmi RR, Kumar A (2011) Weather based forecasting model for crops yield using neural network approach. Stat Appl 9(1):55–69
61. Qaddoum K, Hines EL, Iliescu DD (2013) Yield prediction for tomato greenhouse using EFuNN. ISRN Artif Intell 2013:1–9. https://doi.org/10.1155/2013/430986
62. Khoshnevisan B et al (2014) Development of an intelligent system based on ANFIS for predicting wheat grain yield on the basis of energy inputs. Inf Process Agric 1(1):14–22. https://doi.org/10.1016/j.inpa.2014.04.001
63. Naderloo L et al (2012) Application of ANFIS to predict crop yield based on different energy inputs. Meas J Int Meas Confed 45(6):1406–1413. https://doi.org/10.1016/j.measurement.2012.03.025
64. Kouchakzadeh M, Ghahraman B (2011) 'Ar', 13, pp 627–640
65. Pandey AK, Sinha AK, Srivastava VK (2008) A comparative study of neural-network & fuzzy time series forecasting techniques-case study: wheat production forecasting. Int J Comput Sci Netw Secur 8(9):382–387
66. Balakrishnan N, Muthukumarasamy G (2016) Crop production—ensemble machine learning model for prediction. Int J Comput Sci Softw Eng 5(7):148–153
67. Priya P, Muthaiah U, Balamurugan M (2018) Predicting yield of the crop using machine learning algorithm. Int J Eng Sci Res Technol 7(1):1–7
68. Manjula E, Djodiltachoumy S (2017) A model for prediction of crop yield. Int J Comput Intell Inform 6(4):298–305
69. Preethaa KS, Nishanthini S, Santhiya D, Shree KV (2016) Crop yield prediction. Int J Eng Technol Sci III:111–116
70. Ingole K, Katole K, Shinde A, Domke M (2013) Crop prediction and detection using fuzzy logic in MATLAB. Int J Adv Eng Technol 6(5):2006
71. Garg B, Aggarwal S, Sokhal J (2018) Crop yield forecasting using fuzzy logic and regression model. Comput Electr Eng 67:383–403. https://doi.org/10.1016/j.compeleceng.2017.11.015
72. Kumar P (2011) Crop yield forecasting by adaptive neuro fuzzy inference system. Math Theory Model 1(3):1–7
73. Schmidhuber J (2015) Deep Learning in neural networks: an overview. Neural Netw 61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003
74. Kamilaris A, Prenafeta-Boldú FX (2018) Deep learning in agriculture: a survey. Comput Electron Agric 147(February):70–90. https://doi.org/10.1016/j.compag.2018.02.016
75. Amara J, Bouaziz B, Algergawy A (2017) A deep learning-based approach for banana leaf diseases classification. Datenbanksysteme für Business, Technologie und Web (BTW 2017)-Workshopband
76. Francis M, Deisy C (2020) Mathematical and visual understanding of a deep learning model towards m-agriculture for disease diagnosis. Arch Comput Methods Eng 1–17
77. Newlands N, Ghahari A, Gel YR, Lyubchich V, Mahdi T (2019) Deep learning for improved agricultural risk management. In: Proceedings of the 52nd Hawaii international conference on system sciences
78. Kuwata K, Shibasaki R (2015) Estimating crop yields with deep learning and remotely sensed data. In: 2015 IEEE international geoscience and remote sensing symposium (IGARSS). IEEE, pp 858–861
79. Cunha RLF, Silva B, Netto MAS (2018) A scalable machine learning system for pre-season agriculture yield forecast. In:

Proceedings—IEEE 14th international conference on eScience, e-Science 2018, pp 423–430. https://doi.org/10.1109/eScience.2018.00131.
80. You J et al. (2014) Deep Gaussian process for crop yield prediction based on remote sensing data, pp 4559–4565
81. Wang AX, Lobell D, Ermon S (2015) Deep transfer learning for crop yield prediction with remote sensing data
82. Villanueva MB, Salenga MLM (2018) Bitter melon crop yield prediction using Machine Learning Algorithm. Int J Adv Comput Sci Appl 9(3):1–6. https://doi.org/10.14569/IJACSA.2018.090301
83. Fourie J, Hsiao J, Werner A (2017) Crop yield estimation using deep learning. In: 7th Asian-Australasian conference on precision agriculture, pp 1–10
84. Bargoti S, Underwood JP (2017) Image segmentation for fruit detection and yield estimation in apple orchards. J Field Robot 34(6):1039–1060. https://doi.org/10.1002/rob.21699
85. Kuwata K, Shibasaki R (2016) Estimating Corn Yield in the United States With Modis Evi and Machine Learning Methods. ISPRS Ann Photogramm Remote Sens Spat Inf Sci 3(8):131–136. https://doi.org/10.5194/isprsannals-iii-8-131-2016
86. Mohan P, Patil KK (2018) Deep learning based weighted SOM to forecast weather and crop prediction for agriculture application. Int J Intell Eng Syst 11(4):167–176. https://doi.org/10.22266/ijies2018.0831.17
87. Jiang Z et al (2018) Predicting county level corn yields using deep long short term memory models. http://arxiv.org/abs/1805.12044