Importance of Deep learning models in crop yield prediction

Aravind K S

Division of Agricultural Physics

ICAR-Indian Agricultural Research Institute

New Delhi, India

Aravindwwt28@gmail.com

Ananta Vashisth

Division of Agricultural Physics

ICAR-Indian Agricultural Research Institute

New Delhi, India

ananta.iari@gmail.com

P Krishanan

Division of Agricultural Physics

ICAR-Indian Agricultural Research Institute

New Delhi, India

prameelakrishnan@yahoo.com

Monika Kundu

Division of Agricultural Physics

ICAR-Indian Agricultural Research Institute

New Delhi, India

mkkundu07@gmail.com

ABSTRACT

Accurate crop yield prediction is a challenging task and very much important in decision-making at global, regional and field levels. For the creation of crop yield prediction models, machine learning (ML) has emerged as a key technology. Advanced machine learning techniques, in particular deep learning (DL), can precisely capture the complex data necessary for crop mapping and yield estimates by taking into account the nonlinear interactions between variables. The DL algorithm has gained outstanding success in different fields of agriculture and its application in crop yield prediction. This chapter aims at shedding light on machine learning in agriculture and provide a concise summary of major Deep learning algorithms, including concepts, implementation, training processes, limitations.

**Keywords**—yield; machine learning; Deep learning;

1. **INTRODUCTION**

Agriculture is the field which plays an important role in improving our countries economy. India is an agrarian country and its economy largely based upon crop productivity. Environmental changes have a significant impact on the growth and development of agricultural crops. Numerous parameters related to crop, environmental factors, and management practices influence crop production [1]. Climate, temperature, rainfall, vegetative index, soil type, texture, and major and minor nutrients etc. are the key important factors involved in prediction of crop yield. Yield prediction in crop production is essential for attaining effective and sustainable crop production and managing crops in future. It also provides valuable contributions to policy makers, agronomists, farmers and commodity traders [2-3]. Timely and accurate crop yield predictions are vital for the farmers to make decisions regarding planting, intercultivation, irrigation scheduling, fertilizer application, harvesting, and trading. Prediction of crop yield done through various methods ie, Field surveys, statistical models, crop simulation models, remote sensing, and their combinations. In the ancient times people had the knowledge about predicting the weather conditions and based on that weather and the monsoon of that particular year they had selected the crop and they had a chance of predicting the yield before the period of harvest. Ground truthing is done at farmers field level through conducting field surveys. Crop growth models can be used to simulate crop growth, development and yield, by integrating information’s from soil, plant, environment and management interactions [4]. Remote sensing methods is another application tool which rely on satellite imagery to capture the current state of crops and then to estimate the final yield [5]. Statistical models use weather variables and the outputs of three previous methods as predictors successively incorporated to derive liner relationship between the variables and crop yield.

Recent studies explored combinations of different statistical, simulation and machine learning methods for building crop yield forecasting models. Machine learning (ML) has emerged as an important method for the creation of prediction models. Machine learning offers a viable method for enhancing yield predictions by using a data-driven or empirical modelling approach to identify useful patterns and connections from input data [7]. An approximation function that connects features or predictors to labels, such crop yield. Machine learning algorithms can use the results of other techniques as features, like statistical models do. In addition, machine learning algorithms have some distinct benefits: they can model non-linear relationships between multiple data sources [8]; their performance generally improves when more training data is available [9]; and they can become robust to noisy data by using regularization techniques that help decrease the variance and the generalization error [9-10]. The main concept of ML is to build a prediction model using historical data, evaluate it against new observations, and then use it in an efficient manner [11]. In the last few years, there has been a sharp rise in the number of agricultural yield prediction techniques using machine learning technology. Freely available ML algorithms, remote sensing methods, and improved smart agricultural modelling of genotypes, soil, weather, crop management, and other environmental variables were helped to amplify the crop yield prediction [12]. Machine learning is one of application of AI that make system able to learn automatically or developing programs that can be fed into computer systems. These programs can access data from different objects and use it to learn for themselves. There are mainly three types of machine learning techniques: supervised learning, unsupervised learning, reinforcement. Using these ML techniques, decision can be made and actions can be planned that can be applied in read world situations without any human participation.

Artificial intelligence (AI) and machine learning (ML) are very popular tools used in various applications in the recent times. Artificial intelligence (AI) is computer software that imitates human thought processes to carry out complicated tasks including analysis, reasoning, and learning. The majority of AI nowadays is carried out using machine learning, therefore both terms are often used interchangeably. However, AI actually refers to the general idea of developing human-like cognition using software and systems, whereas ML merely refers to one technique for doing so.

**What is artificial intelligence?**

Artificial intelligence (AI) is computer software that imitates cognitive functions of humans in order to carry out complicated activities that were previously only able to be performed by humans, such as language translation, data analysis, and decision-making. AI is programming for computers that has been specifically built to carry out tasks that require for human reasoning. AI-powered machines and systems can learn from their interactions to enhance their performance and efficiency, in contrast to automated ones that only obey a set of instructions and do them without modification. The word "AI" serves as a generic term for a number of related but unique subfields.

Some of the most common fields within the broader field of artificial intelligence include:

* **Machine learning (ML):** A subset of AI in which algorithms are trained on data sets to become machine learning models capable of performing specific tasks.
* **Deep learning:** A subset of ML, in which artificial neural networks (AANs) that mimic the human brain are used to perform more complex reasoning tasks without human intervention.
* **Natural Language Processing (NLP):** A subset of computer science, AI, linguistics, and ML focused on creating software capable of interpreting human communication.
* **Robotics:**A subset of AI, computer science, and electrical engineering focused on creating robots capable of learning and performing complex tasks in real world environments.

**Machine learning**

The goal of machine learning, a subfield of artificial intelligence (AI) and computer science, is to develop machine learning models that are capable of performing complex tasks like sorting images, making forecasts, or analyzing large amounts of data. They mimic human learning by using data and algorithms, gradually increasing prediction accuracy. The rapidly expanding discipline of data science includes machine learning as a key element. Algorithms are trained using statistical techniques to produce classifications or predictions and to find important insights in data mining projects.

**Machine learning, Deep learning and neural network**

Neural networks, deep learning, and machine learning are all sub field of artificial intelligence. However, deep learning is a branch of neural networks, which itself is a branch of machine learning. The method by which each algorithm learns determines how deep learning and machine learning differ. Traditional or "non-deep" machine learning is more reliant on human input. In order to comprehend the distinctions between different data inputs, human specialists choose a set of features, which typically requires more structured data to learn. While supervised learning, sometimes referred to as labelled datasets, can be implemented in "deep" machine learning to direct its algorithm, it is not a must. Deep learning can automatically identify a number of features that separate several categories of data from one another after processing unstructured information in its raw form (such as text or photos). This reduces the need for human interaction and makes it possible to handle bigger data sets. Artificial neural networks (ANNs), often known as neural networks, consist of three node layers ie, an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, is connected to others and has a weight and threshold that go along with it. Any node whose output exceeds the defined threshold value is activated and starts feeding data to the network's next layer. Deep learning simply denotes the total number of hidden layers in a neural network by the use of the word "deep". Deep learning algorithms or deep neural networks can be defined as neural networks with more than three layers, inclusive of the input and output. A basic neural network is one with three node layers present in their architecture. Deep learning and neural networks are credited with accelerated advancements in fields including speech recognition, computer vision, and natural language processing.

Machine learning algorithm is divided into three main parts, that are:

1. A Decision Process: Machine learning algorithms are typically used to create a forecast or classify information. It will generate an estimate about a pattern in the input data based upon specific input data, which may be labelled or unlabeled.
2. An Error Function: This assesses how well the model predicted. If there are known instances, an error function can compare them to determine how accurate the model is.
3. A Model Optimization Process: Weights are modified to reduce the difference between the known example and the model estimate if the model can fit the training data set precisely. Up until a specific accuracy threshold is reached, the algorithm will iteratively continue this "evaluate and optimise" process, updating the weights on its own.

**ML Methods: Machine learning models fall into three primary categories**.

**Supervised machine learning**

The term "supervised learning," which is used to denote supervised machine learning, refers to the process of learning algorithms to correctly classify data or predict outcomes using labelled datasets. The model adjusts its weights as input data is fed into it until it is perfectly fitted. This manifests as part of the cross-validation process to make sure the model does not overfit or underfit. Numerous difficulties in the real world can be resolved through supervised learning. Neural networks, naive bayes, linear regression, logistic regression, random forests (RF), and support vector machines (SVM) are a few techniques used in supervised learning.

**Unsupervised machine learning**

Unsupervised learning, commonly referred to as unsupervised machine learning. It analyses and clusters unlabeled datasets using machine learning algorithms. These algorithms identify hidden patterns or data clusters without the assistance of a human. This strategy is useful for exploratory data analysis, cross-selling strategies, consumer segmentation, and imagine and pattern recognition since it can find similarities and differences in the data. It can also be employed to perform dimensionality reduction on a model to lower its number of features. Two popular methods for dimensionality reduction are singular value decomposition (SVD) and principal component analysis (PCA). K mean clustering, neural networks and probability clustering are some of the unsupervised learning algorithms.

**Semi-supervised learning**

A satisfying agreement between supervised and unsupervised learning is provided by semi-supervised learning. It facilitates classification and feature extraction from a larger, unlabeled data set during training using a smaller, labelled data set. Lack of sufficient labelled data for a supervised learning system can be resolved via semi-supervised learning. It also helps if labelling enough data is too expensive.

The process of supervised learning use a set of training data to train models to produce the desired results. This training dataset has both the right inputs and outputs, enabling the model to learn over time. The loss function serves as a measure for the algorithm's correctness, and iterations are made until the error is sufficiently reduced.

Supervised learning can be separated into two types of problems when data mining-classification and regression:

1. Classification: In order to accurately classify test data into different categories, classification uses an algorithm. It identifies particular entities in the dataset and makes an effort to determine how those things should be defined or labelled. The following classification techniques are frequently used: decision trees, k-nearest neighbour, random forest, support vector machines (SVM), linear classifiers, and SVM.
2. Regression: To comprehend the relationship between dependent and independent variables, regression is used. Regression methods like logistic regression, polynomial regression, and linear regression are frequently used to predict outcomes, such as yield of a crop with weather variables.

**Challenges of supervised learning**

* For supervised learning models to be accurately structured, a certain level of training may be needed.
* Supervised learning model training may require quite a lot of time.
* Datasets may have greater probabilities of human error, which could cause algorithms to learn the incorrect information.
* Supervised learning cannot cluster or classify data on its own, in contrast to unsupervised learning models.

able 1. **Difference between Supervised learning and Unsupervised learning**

|  |  |  |
| --- | --- | --- |
| Parameter | Supervised learning | Unsupervised learning |
| Input Data | Uses Known and Labeled Data as input | Uses Unknown Data as input |
| Computational Complexity | Less  | More  |
| Real Time | Uses off-line analysis | Uses Real Time Analysis of Data |
| Number of Classes  | Known number | Not known |
| Accuracy of Results | Accurate and Reliable Results | Moderate Accurate and Reliable Results |
| Output data | Desired output is given | Desired output is not given |
| Model | It is not possible to learn larger and more complex models | It is possible to learn larger and more complex models |
| Training data | training data is used to infer model | training data is not used. |
| Another name | Supervised learning is also called classification | Unsupervised learning is also called clustering |
| Test of model | We can test our model | We cannot test our model |
| Example | Optical Character Recognition | Find a face in an image |

**Reinforcement machine learning**

Unlike supervised learning, which uses sample data to train the algorithm, reinforcement machine learning does not. Model learn the data by trial-and-error method. Most appropriate successful outcome generated from the learning is reinforced to develop the model.

**Need of crop yield prediction**

Prior to harvest, agricultural yield prediction is an extremely important subject mainly because the crop yield variations from year to year affect international trade, the availability of food, and the pricing on the world market. Crop yield forecasting at the within-field level has become more common recently. Weather conditions exert a significant effect on crop productivity. Farmers can be informed well in advance so that the large loss can be reduced and would be beneficial for growth in the economy if the weather-based prediction is made more accurate. The forecast will also help farmers make judgements on the selection of alternative crops or the early discarding of a crop in emergency circumstances. Furthermore, yield forecasting can help farmers to manage precisely the timing as well as cultivation of seasonal crops more effectively. For effective crop management and to achieve the desired results, it is therefore required to simulate and predict the crop yield before cultivation. Techniques based on machine learning may be effective for yield predictions by considering the nonlinear relationship between crop yield and  the dependent variables involved in it. In this chapter, we are discussing various Machine Learning Models used for crop yield prediction and their importance.

**Deep Learning**

For the prediction of crop yield, many research have utilised machine learning techniques such regression trees, random forests, multivariate regression, association rule mining, and artificial neural networks [4,13]. The result, crop yield, is taken into account by machine learning models as an implicit function of the input variables, such as meteorological factors and soil conditions, both of which may be quite complex [14]. Since deep learning can analyse enormous datasets, discover correlations between different variables, and learn nonlinear functions, it is widely applied in the agriculture sector. In an unsupervised context, these techniques can extract features for big datasets. Deep learning approaches outperform conventional machine learning methods in feature extraction [15]. Deep learning has a significant ability to extract features from the existing data because the accuracy of agricultural yield prediction depends on the variables influencing crop growth. The nonlinear layers in deep neural networks transform the untested input data into an extracted form at each layer [16]. Finding the nonlinear relationship between the input and response variables requires the use of deep neural networks with an array of hidden layers [16]. However, they are challenging to train and require modern technology and optimisation techniques [9]. Utilising the network's residual skip connections can help to mitigate the vanishing gradient problem in deeper neural networks [9,17]. Furthermore, by implementing several techniques including stochastic gradient descent (SGD), batch normalisation, and dropout, the performance of deep learning systems has been enhanced.

**Support Vector Machines**

SVMs were first used in the work of [18] on the conceptual foundations of statistical learning theory. SVM is fundamentally a binary classifier that builds a linear separation hyperplane in order to classify data instances. The "kernel trick" can be utilised to transform the original feature space into a feature space of a greater dimension, significantly enhancing the classification skills of conventional SVMs. SVMs have been applied to classification, regression, and clustering. SVMs are useful in many applications because they handle overfitting issues that arise in high-dimensional spaces according to their global optimization-based design [19,20]. Most used SVM algorithms include the support vector regression [21], least squares support vector machine [22], and successive projection algorithm-support vector machine [23].

**Decision Trees**

Classification or regression models with a tree-like topology are called decision trees (DT) [24]. With DT, the dataset is gradually divided into smaller, homogenous subsets (sub-populations), and a corresponding tree graph is also produced. Each branch of the tree structure represents the outcome of this comparison, whereas each internal node represents a different pairwise comparison on the selected feature. Following the path from root to leaf, leaf nodes represent the decision or prediction (represented as a classification rule). Three parts make up a decision tree: decision nodes, leaf nodes, and a root node. A training dataset is split up into branches by its algorithm, which then divides those branches further. This process keeps going until a leaf node is reached. It is impossible to further divide the leaf node. The attributes that are utilised for predicting the outcome are represented by the nodes in the decision tree. The classification and regression trees [25], the chi-square automatic interaction detector [26], and the iterative dichotomiser [27] are the most popular learning algorithms in this category.

**Random Forest**

Popular machine learning algorithm Random Forest is a part of the supervised learning methodology. It can be successfully used for regression as well as classification in ML. It is based on the principle of ensembling, which is the act of merging multiple classifiers to address a complex problem and enhance the model's performance. The bagging method is extended by the random forest algorithm, which uses feature randomness in addition to bagging to produce an uncorrelated forest of decision trees. Low correlation between decision trees is ensured by feature randomness, sometimes referred to as feature bagging, which generates a random subset of features. The most significant difference between decision trees and random forests is this. Random forests merely choose a portion of those feature splits, whereas decision trees take into account all possible feature splits. Greater number of trees in the random forest model leads to higher accuracy and it also prevents the problem of overfitting.

**RF features**

* Compared to the decision tree algorithm, it is more accurate.
* When a significant amount of the data is absent, accuracy can still be maintained.
* Even with the enormous dataset, it operates effectively and predicts the outcome with a high degree of accuracy.
* Without hyper-parameter adjustment, it can generate a reasonable prediction.
* It fixes the overfitting problem with decision trees.
* In every random forest tree, a subset of features is selected randomly at the node’s splitting point.

A random forest algorithm's building blocks are decision trees. The fundamental distinction between the random forest method and the decision tree algorithm is that the latter randomly selects the root nodes and groups the nodes. To produce the necessary forecast, the random forest uses the bagging approach. Bagging entails using multiple samples of data (training data) as opposed to a single sample. Predictions are made using features and observations from a training dataset. Depending on the training data that the random forest algorithm receives, the decision trees generate a variety of results. The highest ranking of these outputs will be chosen as the final output.

**Artificial Neural Networks (ANN)**

Artificial neural networks are basic neural networks that were built on the neural architecture of the human brain [16]. The neurons are organised into layers at the nodes of the neural network, which are connected to one another. The input layer, the hidden layer, and the out layer are the three layers that jointly make up the network. The input layer of neurons receives the inputs; the hidden layer, which also consists of linked neurons, carries out the function and then sends the output to the output layer [28]. Additionally, initial weights are assigned at random to initiate the process.

**Deep Neural Networks (DNN)**

A DNN is a particular sort of feed-forward neural network with numerous fully connected hidden layers. The hidden layers are typically employed with activation functions like ReLU (Rectified linear unit) and loss functions like L2 (Ridge regression), regularisation, and mean squared error [29].

**Bayesian Neural Networks (BNN)**

Probability distributions are employed as weights in BNN, a neural network with Bayesian inference. The issue of overfitting without the required validation data for evaluating the regularisation parameter can be avoided by using a Bayesian neural network [30]. Training a BNN with a large dataset can be useful for improving accuracy.

**Convolution Neural Network (CNN)**

A CNN incorporates layers including convolution layers, pooling layers, and fully connected layers in comparison to conventional neural network approaches, which aids in effectively finding notable features within the data. Feature extraction is carried out by the activation function and convolution operation which build up the convolution layer [31]. A filter and feature map are included in the convolution procedure. A filter is a set of weights that are applied to the input, and a feature map is the filter's corresponding output. Additionally, a pooling operation is employed to carry out down-sampling since it improves the efficiency of feature detection [31]. After that, a nonlinear activation function is applied to the outputs in order to introduce nonlinearity. Convolution layers are followed by fully connected (FC) layers, and by adding more FC layers, the network can learn how to map a feature to its target [32].

**2D-CNN and 3D-CNN**

A prime example of a spatial approach is a 2D-CNN. While 3D-CNN is referred to be a spatio-temporal approach [28]. The input data for a 2D-CNN are referred to as the spatial-spectral volume, and the kernel moves along the two spatial dimensions that are across width and height. In a 3D-CNN, a temporal component is included in addition to the two spatial dimensions. Three-dimensional kernels are used by a 3D-CNN to create a 3D feature map by sliding along width, height, and depth [33]. The implementation of 3D convolutional layers led to the development of the 3D-CNN technique [33].

**Faster R-CNN**

Object localization and object detection are two applications that the region-based convolutional neural network (R-CNN) is most frequently used to perform [34]. R-CNN comes in four different kinds: R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN. The R-CNNs are unique and their process is quicker due to the differences in pooling methods and region proposal methods.

**Long Short-Term Memory (LSTM)**

The LSTM is a particular type of recurrent neural network (RNN) that, when used with an appropriate gradient-based algorithm, can learn time-dependent information. An input layer, one or more LSTM layers, and the output layer are the three layers in the chain structure of the LSTM. The input gate, forget gate, and output gate are the three gates that the LSTM utilises to regulate cell state and output. These gates are more likely to function as layers in neural networks that can regulate the transmission of information [16]. Three gates are present in each LSTM layer cell: the input gate selects the information to be retained, the forget gate selects the amount of previous information to be forgotten and the amount of current input to be reserved, and the output gate selects the final output by combining the current input and previous output [35].

Table 2. **comparison of widely used prediction algorithm in crop yield prediction.**

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm | Objectives | Advantages | Limitations |
| SVM | Used for both classification and regression tasks. The main idea behind SVMs is to find a hyperplane that maximally separates the data points of one class from those of the other. | Handles nonlinear relationships. That is important in yield prediction because weather patterns and other environmental factors often have nonlinear relationships with crop yields [37]. Deals with high dimensional data sets, which are common in yield prediction [38] and water management applications [39]. | Sensitive to outliers in the data set. This means that supervision must be taken to ensure that the data used for training the model are representative of the real conditions that will be encountered during the application. |
| RF | Integrates multiple decisiontrees and outputs the most frequentor average prediction from those trees. Used for classification and regression tasks due to their high accuracy and robustness against overfitting. | Relatively easy to tune and does not require extensive data preprocessing like some other ML algorithms [40]. | Computationally expensivewhen working with largedatasets. Needs more time for training as it integrates a lot of decision trees to determine the class. |
| ANN | Creates an AI system capable of making predictions or decisionswithout human external programming or instruction. | Ability to learn from data, recognizepatterns, and generalize from examples, handle noisy data; and deals with nonlinear problems. | Complex structure. Can be easily overfit on training data if it is not carefully designed. Requires a large amount of training data to learn effectively. |
| DNN | Helps generate a flexible and effective detection approach | Used for both classification and regression; Automatically learns high level features, handles complex data with high accuracy. | Requires large amounts of data to train the model, which can be difficult to obtain in some cases. Can be computationally intensive, making it impractical for real-time applications. |

**Yield estimation using machine learning models**

SVR is commonly employed to forecast agricultural yield [41, 42]. Because nonlinear problems connected to the input space are described by being matched with linear problems of high-dimension feature space, one benefit of this approach is that mathematical analysis is relatively simpler [43]. To improve the performance of the predictive model in SVR, the radial basis function kernel is frequently set [44]. Artificial Neural Networks (ANN) is the most used algorithm that is mainly emphasized in agriculture for crop yield prediction. Deep learning, a branch of machine learning, has recently shown cutting-edge results in a variety of fields, including face recognition and image classification. These Deep Neural Networks (DNN) algorithms are built on the same principles as ANN algorithms, but instead of having only one hidden layer, they have several hidden layers, including diverse hidden layer types including convolutional layers and pooling layers. Deep neural networks are a particular class of representation learning models that can determine the fundamental structure of a set of data without the need for manually created feature input. Deep neural networks consist of multiple stacked non-linear layers, each of which transforms the initial raw input data into a higher-level representation at each stacked layer [16]. More complex characteristics features are extracted when the network gets deeper, and it will leads to improve the accuracy of the output. Deep neural networks are known to be universal approximator functions, which means that they can estimate practically any function given the proper parameters, although it may be quite challenging to identify the right parameters [9]. Deep neural networks with multiple hidden layers are more effective at demonstrating the fundamental nonlinear relationship between input and response variables than artificial neural network models, which were shallow networks with a single hidden layer [16], but they also require more sophisticated hardware and training methods. Deep learning models called CNNs, or convolutional neural networks, are particularly efficient at handling grid-like input. Images or rows of multi-column data can be included in such data. Models with numerous layers are referred to as deep learning models. Convolutional neural networks (CNN) outperform all other DL techniques in image classification and regression tasks [45, 46]. Some studies have demonstrated excellent performance of DL algorithms in pixel-based classification [47] and scene understanding [48] by using high spatial resolution satellite images. The application of DL in agriculture by using high spatial resolution imagery was reviewed by [49]. The basic CNNs architectures receive images of size Width *×* Height *×* Depth as input, then highly abstract features can be extracted through a series of operations such as convolution and pooling. The first successful architecture of CNN (named LeNet) was used for zip code digit recognition [50]. A convolutional layer of CNNs performs the convolution operation as the first among multiple transformation operations. Generally, the convolution operation can be described as calculating the sum of products between a set of input values and values of a convolutional kernel, also called a filter.To determine the best characteristics from a data collection for predicting agricultural yield, CNN's kernel values are trained in a sophisticated way.

Reference [51] described the development of ANN models as an accurate technique for corn and soybean yield prediction in Maryland nutrient management planning. The results showed that ANN yield prediction is more accurate than the MLR-based yield model. Reference [52] implemented Back-propagation Neural Network (BPNN) modelling to test the efficiency of the spectral vegetation indices: NDVI, green vegetation index (GVI), soil adjusted vegetation index (SAVI) and perpendicular vegetation index (PVI) in corn crop yield prediction. The results showed that the corn yield was best predicted using BPNN models that used the means and standard deviations of PVI grid images. Reference [53] presented a comparison of two machine learning techniques (BRT and SVM) for prediction of winter wheat yield in Henan province of China. The results of comparison, based on a cross-validation error (RMSE), showed that Boosted Regression Trees (BRT) model consistently outperforms SVM. Reference [54] aimed to produce accurate and timely predictions of grassland LAI for the meadow steppes of northern China, using different regression approaches and hybrid geostatistical methods. The results showed that the RF model provides the most accurate predictions among the regression models. RFs can provide better resistance to the over-fitting problem and to noise in the data compared with other regression methods. Reference [55] presented a comparative study of ANN, SVR, M5-Prime, kNN ML techniques and Multiple Linear Regression for crop yield prediction in ten crop datasets. Results of that study ranked the techniques from the best to the worst, according to RMSE, RRSE, R, and MAE results, in the following order: M5-Prime, kNN, SVR, ANN and MLR. In another study [56] applied four ML techniques, SVM, Random Forest (RF), Extremely Randomized Trees (ERT) and Deep Learning (DL) to estimate corn yield in Iowa State. Comparisons of the validation statistics showed that DL provided more stable results by overcoming the overfitting problem.

**Most used regression metrices**

For crop yield prediction, many machine learning methods are being applied. Although there are numerous machine learning algorithms that can be used, the type of application and the algorithm's accuracy in making predictions that determines which method is to be adopted. Different measures, such as Mean Square Error (MSE), Percent Deviation, and Root Mean Square Error (RMSE), are used to verify the predictability of classifiers.

* 1. **Root mean square error (RMSE)**

This is frequently used to quantify the discrepancy between actual observed values from the experiment that is being modelled and expected values from the model. This test allows for the determination of model performance both during the calibration and validation periods. Additionally, it helps to compare each model's performance to that of other prediction models.

$$RMSE=\sqrt{\frac{1}{N}\sum\_{i=1}^{N}(Pi-Oi)^{2}}$$

Where RMSE is absolute root mean square error, Pi is the predicted value, Oi is the observed value and N is the number of observations

* 1. **Normalized mean square error (nRMSE)**

If Pi, Oi, N and M are notated as predicted value, observed value, number of observations and mean of observed value,nRMSE can be written as the formula given below. Numbers close to 0 suggest superior model performance when the normalised mean square error is reported as a percentage. nRMSE is a percentage indicator of the differences between estimated and observed data. The prediction is considered outstanding if the nRMSE <10%, good if the nRMSE <20%, fair if the nRMSE <30%, and poor if the nRMSE> 30%. (Jamieson et al., 1991)

$$nRMSE=\frac{100}{M}\*\sqrt{\frac{1}{N}\sum\_{i=1}^{N}(Pi-Oi)^{2 }}$$

* 1. **Percent Deviation**

Percent Deviationis calculated using following formula:

% Deviation = $\frac{ Pi-Oi}{Oi} $\*100

Pi is the predicted value and

Oi is the observed value

* 1. **Mean Square Error (MSE)**

Mean squared error (MSE) or mean squared deviation (MSD) of an estimator measures the average of the squares of the errors—that is, the average squared difference between the estimated values and what is estimated. MSE is a risk function, corresponding to the expected value of the squared error loss. The fact that MSE is almost always strictly positive (and not zero) is because of randomness or because the estimator does not account for information that could produce a more accurate estimate.

$$MSE= \frac{1}{N}\sum\_{i=1}^{N}(Pi-Oi)^{2 }$$

* Where Pi is the predicted value, Oi is the observed value, N is the number of observations.
* where *Oi* is the actual value and *Pi* is the forecast value. The difference between *Pi* and *Oi* is divided by the actual value *Oi* again. The absolute value in this calculation is summed for every forecasted point in time and divided by the number of fitted points *N*. Multiplying by 100% makes it a percentage error.

**Conclusion**

Now a days an increasing number of machine learning applications in agriculture are reported in various literature. An enormous amount of data that is already available from many different sources can be analysed to uncover any hidden knowledge in the agricultural field. This is a highly explored area that is anticipated to expand in the future. The combination of computer science and agriculture improves in crop yield predictions. Building on an objective methodology is necessary for pre-harvest crop forecasts. An effective ML model will offer advantages above the conventional forecasting approach. This chapter highlights the value of using advanced ML techniques to make more accurate predictions and real-time applications in the field of agricultural forecasting. The importance of collaboration and partnerships between farmers, researchers, and industry stakeholders in ensuring the widespread adoption of ML-based solutions in the agriculture sector. As with this it was clear that agriculture was completely moving towards the field of artificial intelligence. So, the researchers can concentrate their research on verities of crops and real time dataset for better crop yield estimation.

**REFERENCES**

1. Fischer, R. A. (2015). Definitions and determination of crop yield, yield gaps, and of rates of change. *Field Crops Research*, 182, 9-18.
2. Basso, B., and Liu, L. (2019). Seasonal crop yield forecast: Methods, applications, and accuracies. *advances in agronomy*, 154, 201-255.
3. Chipanshi, A., Zhang, Y., Kouadio, L., Newlands, N., Davidson, A., Hill, H., and Reichert, G. (2015). Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape. *Agricultural and Forest Meteorology*, 206, 137-150.
4. Basso, B., Cammarano, D., and Carfagna, E. (2013, July). Review of crop yield forecasting methods and early warning systems. In Proceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics, FAO Headquarters, Rome, Italy (Vol. 41, pp. 1-56).
5. López-Lozano, R., Duveiller, G., Seguini, L., Meroni, M., García-Condado, S., Hooker, J., and Baruth, B. (2015). Towards regional grain yield forecasting with 1 km-resolution EO biophysical products: Strengths and limitations at pan-European level. *Agricultural and Forest Meteorology*, 206, 12-32.
6. Bussay, A., van der Velde, M., Fumagalli, D., and Seguini, L. (2015). Improving operational maize yield forecasting in Hungary. *Agricultural Systems*, 141, 94-106.
7. Willcock, S., Martínez-López, J., Hooftman, D. A., Bagstad, K. J., Balbi, S., Marzo, A., and Athanasiadis, I. N. (2018). Machine learning for ecosystem services. *Ecosystem services*, 33, 165-174.
8. Chlingaryan, A., Sukkarieh, S., and 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.
9. Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
10. James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.
11. Jordan, M. I., and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
12. Wolfert, S., Ge, L., Verdouw, C., and Bogaardt, M. J. (2017). Big data in smart farming–a review. *Agricultural systems*, 153, 69-80.
13. Horie, T., Yajima, M., and Nakagawa, H. (1992). Yield forecasting. *Agricultural systems*, 40(1-3), 211-236.
14. Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., and Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PloS one*, 11(6), e0156571.
15. Islam, N., Rashid, M. M., Wibowo, S., Wasimi, S., Morshed, A., Xu, C., and Moore, S. (2020, August). Machine learning based approach for Weed Detection in Chilli field using RGB images. In *The International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery* (pp. 1097-1105). Cham: Springer International Publishing.
16. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
17. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
18. Cortes, C., and Vapnik, V. (1995). Support vector machine. *Machine learning*, *20*(3), 273-297.
19. Suykens, J. A., and Vandewalle, J. (2000). Recurrent least squares support vector machines. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, *47*(7), 1109-1114.
20. Chang, C. C., and Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3), 1-27.
21. Smola, A. J. (1996). *Regression estimation with support vector learning machines* (Doctoral dissertation, Master’s thesis, Technische Universität München).
22. Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., and Vandewalle, J. (2002). Least squares support vector machines, World Scientific Publishing, Singapore.
23. Galvao, R. K. H., Araujo, M. C. U., Fragoso, W. D., Silva, E. C., Jose, G. E., Soares, S. F. C., and Paiva, H. M. (2008). A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm. *Chemometrics and intelligent laboratory systems*, 92(1), 83-91.
24. Belson, W. A. (1959). Matching and prediction on the principle of biological classification. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 8(2), 65-75.
25. Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification and regression trees. (Milton Park, Abingdon, Oxfordshire, UK.
26. Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(2), 119-127.
27. Quinlan, J. R. (1992, November). Learning with continuous classes. In *5th Australian joint conference on artificial intelligence* (Vol. 92, pp. 343-348).
28. Khaki, S., Wang, L., and Archontoulis, S. V. (2020). A cnn-rnn framework for crop yield prediction. *Frontiers in Plant Science*, *10*, 1750.
29. Islam, N., Rashid, M. M., Wibowo, S., Wasimi, S., Morshed, A., Xu, C., and Moore, S. (2020, August). Machine learning based approach for Weed Detection in Chilli field using RGB images. In *The International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery* (pp. 1097-1105). Cham: Springer International Publishing.
30. Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., and Bédard, F. (2016). Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agricultural and forest meteorology*, 218, 74-84.
31. Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9, 611-629.
32. Nevavuori, P., Narra, N., and Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. *Computers and electronics in agriculture*, 163, 104859.
33. Fernandez-Beltran, R., Baidar, T., Kang, J., and Pla, F. (2021). Rice-yield prediction with multi-temporal sentinel-2 data and 3D CNN: A case study in Nepal. *Remote Sensing*, 13(7), 1391.
34. Chen, Y., Lee, W. S., Gan, H., Peres, N., Fraisse, C., Zhang, Y., and He, Y. (2019). Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages. *Remote Sensing*, 11(13), 1584.
35. Tian, H., Wang, P., Tansey, K., Zhang, J., Zhang, S., and Li, H. (2021). An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China. *Agricultural and Forest Meteorology*, 310, 108629.
36. Tian, H., Wang, P., Tansey, K., Han, D., Zhang, J., Zhang, S., and Li, H. (2021). A deep learning framework under attention mechanism for wheat yield estimation using remotely sensed indices in the Guanzhong Plain, PR China. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102375.
37. Chen, H., Wu, W., and Liu, H. B. (2016). Assessing the relative importance of climate variables to rice yield variation using support vector machines. *Theoretical and Applied Climatology*, 126, 105-111.
38. Shi, L., Duan, Q., Ma, X., and Weng, M. (2012). The research of support vector machine in agricultural data classification. In Computer and Computing Technologies in Agriculture V: 5th IFIP TC 5/SIG 5.1 Conference, CCTA 2011, Beijing, China, October 29-31, 2011, Proceedings, Part III 5 (pp. 265-269). Springer Berlin Heidelberg
39. Deka, P. C. (2014). Support vector machine applications in the field of hydrology: a review. *Applied soft computing*, *19*, 372-386.
40. Chen, Y., Hou, J., Huang, C., Zhang, Y., and Li, X. (2021). Mapping maize area in heterogeneous agricultural landscape with multi-temporal Sentinel-1 and Sentinel-2 images based on random forest. *Remote sensing*, 13(15), 2988.
41. Gu, Y. H., Yoo, S. J., Park, C. J., Kim, Y. H., Park, S. K., Kim, J. S., and Lim, J. H. (2016). BLITE-SVR: New forecasting model for late blight on potato using support-vector regression. *Computers and electronics in agriculture*, 130, 169-176.
42. Su, Y. X., Xu, H., and Yan, L. J. (2017). Support vector machine-based open crop model (SBOCM): Case of rice production in China. *Saudi journal of biological sciences*, 24(3), 537-547.
43. Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., and Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4), 18-28.
44. Zhang, G., and Ge, H. (2013). Support vector machine with a Pearson VII function kernel for discriminating halophilic and non-halophilic proteins. *Computational biology and chemistry*, 46, 16-22.
45. Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
46. Liu, X., Li, S., Kan, M., Zhang, J., Wu, S., Liu, W., ... and Chen, X. (2015). Agenet: Deeply learned regressor and classifier for robust apparent age estimation. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 16-24).
47. Yu, S., Jia, S., and Xu, C. (2017). Convolutional neural networks for hyperspectral image classification. *Neurocomputing*, 219, 88-98.
48. Han, X., Zhong, Y., Cao, L., and Zhang, L. (2017). Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. *Remote Sensing*, 9(8), 848.
49. Kamilaris, A., and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147, 70-90.
50. LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, *2*.
51. Kaul, M., Hill, R. L., and Walthall, C. (2005). Artificial neural networks for corn and soybean yield prediction. *Agricultural Systems*, 85(1), 1-18.
52. Panda, S. S., Ames, D. P., and Panigrahi, S. (2010). Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote sensing*, 2(3), 673-696.
53. Stas, M., Van Orshoven, J., Dong, Q., Heremans, S., and Zhang, B. (2016, July). A comparison of machine learning algorithms for regional wheat yield prediction using NDVI time series of SPOT-VGT. In *2016 fifth international conference on agro-geoinformatics (agro-geoinformatics)* (pp. 1-5). IEEE.
54. Li, Z., Wang, J., Tang, H., Huang, C., Yang, F., Chen, B., ... and Ge, Y. (2016). Predicting grassland leaf area index in the meadow steppes of northern china: A comparative study of regression approaches and hybrid geostatistical methods. *Remote Sensing*, 8(8), 632.
55. Gonzalez-Sanchez, A., Frausto-Solis, J., and Ojeda-Bustamante, W. (2014). Predictive ability of machine learning methods for massive crop yield prediction. *Spanish Journal of Agricultural Research*, 12(2), 313-328.
56. Nari, K., and Yang-Won, L. (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, Photogrammetry and Cartography| Korea Science.