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

ABSTRACT

Reliable and timely crop-yield prediction and crop mapping are crucial for food security and decision making in the food industry and in agro-environmental management. Machine learning (ML) has become an important technology for the development of crop yield prediction models. Advanced machine-learning methods, particularly deep learning (DL), can accurately represent the complex features essential for crop mapping and yield predictions by accounting the nonlinear relationships between variables. The DL algorithm has attained remarkable 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. As changes in environment, it has a great impact on production and maintenance of agriculture crops. Crop yield is influenced by many crop-specific parameters, environmental conditions and management decisions [1]. In crop yield prediction environmental factors plays a major role, which includes climatic conditions, temperature, rainfall, vegetative index, soil type, texture and nutrients. Yield prediction has an important role in crop farming aimed at efficient and sustainable production. Estimating of crop yield is crucial in agriculture field because it will help in managing crops in future. Crop yield forecasts are valuable to many stakeholders in the agri-food chain, including farmers, agronomists, commodity traders and policymakers [2-3]. Accurate and timely predictions are important for farmers’ decision-making regarding planting, irrigation, fertilization, harvesting, and trading. Field surveys, crop growth models, remote sensing, statistical models and their combinations have been commonly used to predict crop yield. 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. Field surveys are done at farmers field level in order to capture the ground truth. Crop growth models can be used to simulate crop growth and development according to agronomic principles of 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 the three previous methods as predictors successively incorporated to derive linear relationships between the predictors and crop yield [6].

Recent studies have combined different methods in innovative ways to build yield forecasting models. For the development of prediction models, machine learning (ML) has become a key technology. Machine learning takes a data-driven or empirical modeling approach to learn useful patterns and relationships from input data and provides a promising avenue for improving crop yield predictions [7]. Machine learning algorithms approximate a function that relates features or predictors to labels, such as crop yield. Similar to statistical models, machine learning algorithms can utilize the outputs of other methods as features. 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]. Therefore, machine learning could combine the benefits of other methods, such as crop growth models and remote sensing, with data-driven modeling to make reliable crop yield predictions. The principal idea of ML is to learn a prediction model from past data, evaluate the model based on new observations, and ultimately deploy the model in a productive environment [11]. The number of applications of ML technology for crop yield prediction have increased rapidly in the past few years. This growth has been amplified by freely available ML algorithms, improved remote sensing techniques, and the enhanced provision of smart farming data that represents genotypes, soil, weather, crop management, and other environmental parameters that affect crop growth [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. In simplest terms, AI is computer software that mimics the ways that humans think in order to perform complex tasks, such as analyzing, reasoning, and learning. Today, most AI is performed using machine learning, so the two terms are often used synonymously, but AI actually refers to the general concept of creating human-like cognition using computer software and systems, while ML refers to only one method of doing so.

**What is artificial intelligence?**

Artificial intelligence **(**AI**)** is computer software that mimics human cognitive abilities in order to perform complex tasks that historically could only be done by humans, such as decision making, [data analysis](https://www.coursera.org/articles/what-is-data-analysis-with-examples), and language translation. AI is code on computer systems explicitly programmed to perform tasks that require human reasoning. While automated machines and systems merely follow a set of instructions and dutifully perform them without change, AI-powered ones can learn from their interactions to improve their performance and efficiency. AI is an umbrella term covering a variety of interrelated, but distinct, 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**

Machine learning is a branch of Artificial intelligence (AI) and computer science which focuses on training machine learning algorithms with data sets to produce machine learning models capable of performing complex tasks, such as sorting images, forecasting or analyzing big data. They use of data and algorithms to imitate the way that humans learn, gradually improving its prediction accuracy. Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects.

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

Machine learning, deep learning, and neural networks are all sub-fields of artificial intelligence. However, neural networks are actually a sub-field of machine learning, and deep learning is a sub-field of neural networks. The way in which deep learning and machine learning differ is in how each algorithm learns. Classical, or "non-deep", machine learning is more dependent on human intervention to learn. Human experts determine the set of features to understand the differences between data inputs, usually requiring more structured data to learn. "Deep" machine learning can use labeled datasets, also known as supervised learning, to inform its algorithm, but it doesn’t necessarily require a labeled dataset. Deep learning can ingest unstructured data in its raw form (e.g., text or images), and it can automatically determine the set of features which distinguish different categories of data from one another. This eliminates some of the human intervention required and enables the use of larger data sets. Neural networks, or artificial neural networks (ANNs), are comprised of node layers, containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network by that node. The “deep” in deep learning is just referring to the number of layers in a neural network. A neural network that consists of more than three layers—which would be inclusive of the input and the output—can be considered a deep learning algorithm or a deep neural network. A neural network that only has three layers is just a basic neural network. Deep learning and neural networks are credited with accelerating progress in areas such as computer vision, natural language processing, and speech recognition.

Learning system of a machine learning algorithm divides into three main parts.

1. A Decision Process: In general, machine learning algorithms are used to make a prediction or classification. Based on some input data, which can be labeled or unlabeled, algorithm will produce an estimate about a pattern in the data.
2. An Error Function: which evaluates the prediction of the model. If there are known examples, an error function can make a comparison to assess the accuracy of the model.
3. A Model Optimization Process: If the model can fit better to the data points in the training set, then weights are adjusted to reduce the discrepancy between the known example and the model estimate. The algorithm will repeat this “evaluate and optimize” process, updating weights autonomously until a threshold of accuracy has been met.

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

**Supervised machine learning**

Supervised learning, also known as supervised machine learning, is defined by its use of labeled datasets to train algorithms to classify data or predict outcomes accurately. As input data is fed into the model, the model adjusts its weights until it has been fitted appropriately. This occurs as part of the cross-validation process to ensure that the model avoids overfitting or underfitting. Supervised learning helps to solve a variety of real-world problems. Some methods used in supervised learning include neural networks, naïve bayes, linear regression, logistic regression, random forest (RF), and support vector machine (SVM).

**Unsupervised machine learning**

Unsupervised learning, also known as unsupervised machine learning, uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention. This method’s ability to discover similarities and differences in information make it ideal for exploratory data analysis, cross-selling strategies, customer segmentation, and image and pattern recognition. Can also used to reduce the number of features in a model through the process of dimensionality reduction. Principal component analysis (PCA) and singular value decomposition (SVD) are two common approaches for this. Other algorithms used in unsupervised learning include neural networks, k-means clustering, and probabilistic clustering methods.

**Semi-supervised learning**

Semi-supervised learning offers a happy medium between supervised and unsupervised learning. During training, it uses a smaller labeled data set to guide classification and feature extraction from a larger, unlabeled data set. Semi-supervised learning can solve the problem of not having enough labeled data for a supervised learning algorithm. It also helps if it’s too costly to label enough data.

How supervised learning works: Supervised learning uses a training set to teach models to yield the desired output. This training dataset includes inputs and correct outputs, which allow the model to learn over time. The algorithm measures its accuracy through the loss function, adjusting until the error has been sufficiently minimized.

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

* **Classification** uses an algorithm to accurately assign test data into specific categories. It recognizes specific entities within the dataset and attempts to draw some conclusions on how those entities should be labeled or defined. Common classification algorithms are linear classifiers, support vector machines (SVM), decision trees, k-nearest neighbor, and random forest, which are described in more detail below.
* **Regression** is used to understand the relationship between dependent and independent variables. It is commonly used to make projections, such as for sales revenue for a given business linear regression, logistic regression and polynomial regression are popular regression algorithms.

**Challenges of supervised learning**

* Supervised learning models can require certain levels of expertise to structure accurately.
* Training supervised learning models can be very time intensive.
* Datasets can have a higher likelihood of human error, resulting in algorithms learning incorrectly.
* Unlike unsupervised learning models, supervised learning cannot cluster or classify data on its own.

Table 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**

Reinforcement machine learning is a machine learning model that is similar to supervised learning, but the algorithm isn’t trained using sample data. This model learns as it goes by using trial and error. A sequence of successful outcomes will be reinforced to develop the best recommendation or model for a given problem.

**Need of crop yield prediction**

Estimating agricultural yield prior to harvest is an important issue in agriculture, as the changes in crop yield from year to year influence international business, food supply, and global market prices. In recent times, forecasting of crop productivity at the within-field level has increased. The most influencing factor for crop productivity is weather conditions. If the weather based prediction is made more precise, then farmers can be alerted well in advance so that the major loss can be mitigated and would be helpful for economic growth. The prediction will also aid the farmers to make decisions such as the choice of alternative crops or to discard a crop at an early stage in case of critical situations. Further, predicting crop yield can facilitate the farmers to have a better vision on cultivation of seasonal crop and its scheduling. Thus, it is necessary to simulate and predict the crop yield before cultivation for efficient crop management and expected outcome. Researchers have developed different ways to track and predict crop yield, including and statistical models, remote sensed vegetation indices and machine learning approaches and so on. As there exists a non-linear relationship between crop yield and the factors influencing crop, machine learning techniques might be efficient for yield predictions. In this chapter, we are discussing various Machine Learning Models used for crop yield prediction and their importance.

**Deep Learning**

Many studies have used machine learning approaches such as regression tree, random forest, multivariate regression, association rule mining, and artificial neural networks for crop yield prediction [4,13]. Machine learning models treat the output, crop yield, as an implicit function of the input variables, such as weather components and soil conditions, which could be very complex [14]. Deep learning is broadly used in the agricultural field as it can analyze huge datasets, learn the relationships between various variables, and use nonlinear functions. These approaches can extract features for huge datasets in an unsupervised environment. When compared to traditional machine learning approaches, deep learning approaches perform better in feature extraction [15]. Since an accurate crop yield prediction relies on the factors influencing crop growth, deep learning has a strong ability to extract features from available data.

Deep neural networks have a collection of nonlinear layers that convert the untested input data into an extracted form at each layer [16]. Deep neural networks with various hidden layers are important to discover the nonlinear correlation between input and response variables [16]. Nevertheless, they are difficult to train and need recently developed hardware and optimization methodologies [9]. The vanishing gradient problem in deeper neural networks can be reduced by making use of residual skip connections for the network [9,17]. Moreover, the performance of deep learning approaches has been improved by undertaking various techniques such as stochastic gradient descent (SGD), batch normalization, and dropout. Some of the deep learning approaches are given below.

**Support Vector Machines**

Support vector machines (SVMs) were first introduced in the work of [18] on the foundation of statistical learning theory. SVM is intrinsically a binary classifier that constructs a linear separating hyperplane to classify data instances. The classification capabilities of traditional SVMs can be substantially enhanced through transformation of the original feature space into a feature space of a higher dimension by using the “kernel trick”. SVMs have been used for classification, regression, and clustering. Based on global optimization, SVMs deal with overfitting problems, which appear in high-dimensional spaces, making them appealing in various applications [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**

Decision trees (DT) are classification or regression models formulated in a tree-like architecture [24]. With DT, the dataset is progressively organized in smaller homogeneous subsets (sub-populations), while at the same time, an associated tree graph is generated. Each internal node of the tree structure represents a different pairwise comparison on a selected feature, whereas each branch represents the outcome of this comparison. Leaf nodes represent the final decision or prediction taken after following the path from root to leaf (expressed as a classification rule). A decision tree is a decision support technique that forms a tree-like structure, it consists of three components: decision nodes, leaf nodes, and a root node. It’s algorithm divides a training dataset into branches, which further segregate into other branches. This sequence continues until a leaf node is attained. The leaf node cannot be segregated further. The nodes in the decision tree represent attributes that are used for predicting the outcome. The most common learning algorithms in this category are the classification and regression trees [25], the chi-square automatic interaction detector [26], and the iterative dichotomiser [27].

**Random Forest**

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensembling,** which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. The random forest algorithm is an extension of the bagging method as it utilizes both bagging and feature randomness to create an uncorrelated forest of decision trees. Feature randomness, also known as feature bagging, generates a random subset of features, which ensures low correlation among decision trees. This is a key difference between decision trees and random forests. While decision trees consider all the possible feature splits, random forests only select a subset of those features. **The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.**

**Features of RF**

* It’s more accurate than the decision tree algorithm.
* It can also maintain accuracy when a large proportion of data is missing.
* It predicts output with high accuracy, even for the large dataset it runs efficiently.
* It can produce a reasonable prediction without hyper-parameter tuning.
* It solves the issue of overfitting in decision trees.
* In every random forest tree, a subset of features is selected randomly at the node’s splitting point.

Decision trees are the building blocks of a random forest algorithm. The main difference between the decision tree algorithm and the random forest algorithm is that establishing root nodes and segregating nodes is done randomly in the latter. The random forest employs the bagging method to generate the required prediction. Bagging involves using different samples of data (training data) rather than just one sample. A training dataset comprises observations and features that are used for making predictions. The decision trees produce different outputs, depending on the training data fed to the random forest algorithm. These outputs will be ranked, and the highest will be selected as the final output.

**Artificial Neural Networks (ANN)**

Artificial neural networks are simple neural networks that were modeled based on the human brain’s neural structure [16]. The neural network consists of nodes, which are connected to each other, where the neurons are grouped into layers. The network has three layers: the input layer, the hidden layer, and the out layer. The inputs are received by the input layer of neurons; the hidden layer with interconnected neurons performs the function and then provides the output to the output layer [28]. Moreover, to initiate the process, initial weights are assigned randomly.

**Deep Neural Networks (DNN)**

A DNN is a special kind of feed-forward neural network with many hidden layers that are fully connected. Generally, activation functions such as ReLU (Rectified linear unit) and loss functions such as L2 (Ridge regression) regularization and mean squared error are used with the hidden layers [29].

**Bayesian Neural Networks (BNN)**

BNN uses a neural network with Bayesian inference and probability distributions are used as weights in BNN. Using a Bayesian neural network can prevent the problem of overfitting without necessary validation data to evaluate the regularization parameter [30]. For better accuracy, training a BNN with a large dataset can be helpful.

**Convolution Neural Network (CNN)**

Compared to conventional neural network approaches, a CNN includes layers such as convolution layers, pooling layers, and fully connected layers, which helps in efficiently finding salient features within the data. The convolution layer consists of a convolution operation and activation function, which perform feature extraction [31]. Convolution operation includes a filter and feature map. A filter is a group of weights applied across the input and a feature map is the corresponding output for a given filter. Moreover, a pooling operation is used to perform down-sampling as it helps to detect features effectively [31]. The outputs are then passed through a nonlinear activation function as it generates nonlinearity into the output. Fully connected (FC) layers are used after convolution layers; the network has the capacity to learn the mapping between the feature and the target by increasing the FC layers [32].

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

A 2D-CNN is called a spatial method CNN whereas 3D-CNN is called a spatio-temporal method [28]. In a 2D-CNN, the input data are considered as the spatial–spectral volume, where the kernel slides along the two spatial dimensions that are across width and height. In a 3D-CNN, to the two spatial dimensions, a temporal dimension is also added. A 3D-CNN uses three-dimensional kernels, which slide along width, height, and depth and help in generating a 3D feature map [33]. The 3D-CNN approach is developed by implementing 3D convolutional layers [33].

**Faster R-CNN**

The region-based convolutional neural network (R-CNN) is predominantly used for object localization and object detection [34]. There are four different kinds of R-CNN; they are R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN. The difference in pooling methods and region proposal methods makes the R-CNNs different and their process faster.

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

LSTM is a special kind of recurrent neural network (RNN) that can learn time dependent information with an appropriate gradient-based algorithm. The LSTM comprises a chain structure, which starts with an input layer, one or more LSTM layers, and the output layer. To control the cell state and output, the LSTM uses three gates, namely the input gate, forget gate, and output gate. These gates are more likely as neural network layers, which can control the information transfer [16]. Each cell in LSTM layers has three gates; the input gate decides which information needs to be retained, the forget gate determines the amount of previous information that must be forgotten and the amount of current input that needs to be reserved, the output gate uses both the current input and previous output to decide the final output [35]. [36] proposed the ALSTM model, which had six layers, namely one input layer, one LSTM layer, one attention mechanism layer, two dropout layers, and one output layer.

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 used in crop yield prediction [41,42]. One of the advantages of this method is that mathematical analysis is relatively easier because nonlinear problems related to the input space are expressed by being matched with linear problems of high-dimension feature space [43]. In SVR, radial basis function kernel is commonly set to achieve better predictive performance [44]. Artificial Neural Networks (ANN) is the most used algorithm for crop yield prediction. Recently, deep learning, which is a sub-branch of machine learning, has provided state-of-the-art results in many different domains, such as face recognition and image classification. These Deep Neural Networks (DNN) algorithms use similar concepts of ANN algorithms; however, they include different hidden layer types such as convolutional layer and pooling layer and consist of many hidden layers instead of a single hidden layer. Deep neural networks belong to the class of representation learning models that can find the underlying representation of data without handcrafted input of features. Deep neural networks have multiple stacked non-linear layers which transform the raw input data into higher and more abstract representation at each stacked layer [16]. As the network grows deeper, more complex features are extracted which contribute to the higher accuracy of results. Given the right parameters, deep neural networks are known to be universal approximator functions, which means that they can approximate almost any function, although it may be very challenging to find the right parameters [9]. Compared with the artificial neural network models, which were shallow networks with a single hidden layer, deep neural networks with multiple hidden layers are more powerful to reveal the fundamental nonlinear relationship between input and response variables [16], but they also require more advanced hardware and optimization techniques to train. Convolutional neural networks, or CNNs, are deep learning models specialized in handling grid-like data. Such data can be images or rows of multi-column data. Deep learning refers to models composed of multiple layers. Among all DL methods, convolutional neural network (CNN) exhibits superior performance in image classification tasks [45] and regression tasks [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]. The convolution operation is the first of multiple transformations performed in a convolutional layer of CNNs. 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. In CNN, the kernel values are trained to find optimal features from data set for predicting crop yield.

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**

Different machine learning algorithms are being used for prediction of crop yield. Though different machine learning algorithms are available for use, selection of a particular algorithm is based on the nature of application and accuracy of prediction algorithm. Prediction accuracy of classifiers is validated by different metrics such as: Root mean square error (RMSE), Normalized mean square error (nRMSE), Percent Deviation, Mean Square Error (MSE).

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

This is often used measure the difference between predicted values from the model and actual observed values from the experiment that is being modelled. By this test, model performance during the calibration as well as validation period can be determined. It is also helpful in comparing individual model performance with that of other predictive 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. Normalized mean square error expressed in percentage, values close to zero indicates better model performance. nRMSE is a measure (%) of the relative difference of estimated versus observed data. The prediction is considered excellent with the nRMSE <10 %, good if 10–20 %, fair if 20–30 %, poor if >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 a growing number of applications of machine learning techniques in agriculture are required for which a large amount of data currently available from many resources can be analyzed to find the hidden knowledge. This is an advanced researched field and is expected to grow in the future. The integration of computer science with agriculture helps in forecasting agricultural crops. It is required to build on objective methodology for pre-harvest crop forecasting. Building up a suitable model will have certain merits over the traditional forecasting method. 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.