ASurvey: CROP YEID PREDICTION USING NEURAL NETWORKS & MACHINE LEARNING

[1] CH . MALLIKARJUNARAO, [2] M. HARIKA
[1.] Professor, Department of Computer Science and Engineering, Gokaraju Rangaraju Institute Of Engineering And Technology,(TS).India.
Email-: cmrao@griet.ac.in
[2.] M.Tech Student Department of Computer Science and Engineering, Gokaraju Rangaraju Institute Of Engineering And Technology,(TS).India.
Email-: harika20241d5807@grietcollege.com

Abstract
Agriculture has been identified as one of the most important sectors for the growth of nations. Agriculture is affected by the impact on the economy of the nation and on food grain statistics around the world. The challenge always lies in achieving sustainable crop production for agriculturists. Changing environmental conditions have always made it difficult for farmers to achieve optimal crop yields. The unpredictability of crop yields is mostly caused by differences in land types, the availability of resources, and changes in the weather. Consequently, scientists around the world are exploring techniques that can be used to estimate crop yield efficiently and with excellent accuracy in the months ahead. This is so that farmers may prepare for these issues in the future and act appropriately. In this article, we cover a variety of approaches to agricultural production prediction that involve machine learning and deep learning algorithms. In order to elaborate on yield predictions, a collection of data over a long period of time is used to implement various algorithms. Additionally, because we are interested in estimating the yield in light of the classification methods predicting the yield with respect to classification techniques, Unpredictable weather conditions, poor harvesting, and irrigation techniques, and mismanagement of livestock are some of the reasons for reduced food production. we have conducted a comparative analysis of which classification algorithm is the most appropriate for this purpose.

Keywords : component, formatting, style,styling.

DOI Number: 10.14704/nq.2022.20.12.NQ77036 **NeuroQuantology 2022; 20(12): 522-528**

I. INTRODUCTION

As of today, the world's population is approximately 4.84 billion. By 2050, it is expected to increase to 9.1 billion. With 70 percent growth in food requirements and urbanization accelerating, agriculture land availability will decrease dramatically. Currently, India lags behind its own food production. By 2050, it will be the most populated country on earth. Reduced food production is caused by a variety of factors, including unpredictable weather patterns, inadequate irrigation and harvesting methods, and poorly managed livestock. Climate change has caused nature to experience dramatic Re weather changes over the last few years. Climate conditions are uncertain due to the increasing average temperature of the earth. The most common challenges faced by poor farmers are droughts and heavy rain. Agricultural production is reduced by 20-25% in India because of adverse climatic conditions. The government of India annually surveys the state's economy.

In order to guarantee food security for the whole world, precision agriculture is one of the best options. A data-driven, technology-enabled, sustainable farm management system, precision agriculture is also referred to as digital agriculture. The technology consists of smart embedded devices, software, and modern information tools for agriculture decision-support. Agriculture revolutions 1 and 2 were driven largely by mechanized agriculture and the green revolution, while revolution 3 was driven largely by precision farming [1].

Global positioning system (GPS)-equipped tractors were first introduced by John Deere in 1990 to sow seeds and

spray fertilizer. Precision farming is largely focused on reducing production costs while reducing environmental effects, as well as increasing the farm's profitability. Precision agriculture has been impacted by several digital technologies such as the internet of things[1], artificial intelligence, cloud computing, analysis of data, and block chain. Sensors connected to the IOT are used in precision farming to collect information about soil nutrients, fertilizer availability, and crop growth. A variety of autonomous devices and semi-autonomous systems are being utilized for weed identification and disease detection in plants including unmanned aerial vehicles (UAV) and robots. Precision agriculture uses satellite images to monitor crops and identify diseases. The examination of the data gathered from the deployed sensors can be automated with the use of machine learning (ML) techniques. to contribute to improving the effectiveness and efficiency of farming techniques. ML algorithms are also used to predict weather and rainfall using data from sensors, climate records, and satellite imagery.Millions of farmers could be saved from suicide due to crop losses caused by weather-related uncertainty. A successful precision farming program includes intelligent livestock management.During their lifetime, it allows for monitoring their health, welfare, and productivity. Animal health is monitored by sensors and cameras, while computer vision helps us make more informed decisions to stop diseases from spreading within a community.

Modern farmers have access to modern farming solutions with autonomous tractors and automated irrigation systems. It is largely due to the availability of high-speed Internet access, advanced deep learning (DL) algorithms, and efficient computing devices that precision farming is used all over the world. The authors explained how ML could be used to improve farm supply chain performance in [2]. The value and function of digital technology in farming can be understood using this paradigm by academics and the agricultural sector. Several ML applications in agriculture were reviewed in [3], as well as how digital technologies will contribute to agriculture. A comprehensive review of the use of machine learning in precision agriculture is presented in this paper. This article proposes to provide an insight into the area of digital practices and how they are becoming more prevalent as part of the agricultural management system. Precision farming will likely be promoted throughout the world.

There is a comprehensive assessment in this paper of how every cubic centimeter can be employed for agricultural prediction management, typically for crop yield estimation. These papers provide an overview of a variety of advanced ML models and highlight a variety of options for them. Here we divide the rest of the document into sections. There is a tendency in Section II to provide a thorough assessment of various machine learning and deep learning methods used to predict crop yields for various crops. Section III describes the comparison table. Section IV focuses on our proposed methodologies for future experimentation. Section-Vgives a conclusion

II. LITERATURE REVIEW

In 2021, Dilli Paudel et al., Proposed the development of a machine learning model of yield forecasting that combined agronomic principles with machine learning based on crop modeling to help predict large-scale crop yields[4]. In this study, they employed a standard machine learning paradigm: data collection, preprocessing, splitting the test data, designing features, selecting features, training, and testing with five-fold sliding validation. Correctness, modularity, and reusability should be the hallmarks of a workflow. In order to identify predictors or features that can be explained, they focused on the following: (i) Planting Window, (ii) Pre-Planting window, (iii) Flowering Phase, (iv) Vegetative Phase, (v) Harvest Phase and (iv) Yield Formation Phase. Using supervised regression, they applied machine learning without information leakage to crop yield prediction through the use of Random Forest (RF) for model selection, Recursive Feature Elimination (RFE) with Lasso for feature elimination. Training examples that include both features and labels are used in supervised learning to develop a function to relate features to labels. There was a 70/30 split between the training and test datasets. Following that, the authors divided the training set into validation folds, in order to optimize the hyper parameters of feature selection and prediction algorithms[4]. They used the yield trend to test the hypothesis and took a time-dependent k-fold validation with the previous few years for each region included in the test set. Researchers choose a GBDT (Gradient Boost Decision Tree) prediction algorithms for normalized RMSE. Additionally remote sensing, soil data, and weather from the MCYFS datasets were incorporated into the workflow to create a reusable and modular approach to support various crops in different countries. This procedure identifies how repeatable experiments can be run using standard input data. It allows further optimizations to be applied. Five crops (Sunflower, barley, potatoes, sugar beet, wheat) and three countries the Germany, Netherlands, France are considered in the proposed case studies[4]. There was also an attempt by the authors to compare the performance of the model with a simple prediction method that predicted either a linear trend of yield or the training set average. The MCYFS department aggregated the predictions to the national level and compared them to previous forecasts. For all crops except soft wheat in Netherlands, except soft wheat in Germany, and, spring barley, soft wheat, sunflower crops from France, to find normalized RMSEs (NRMSE) were comparable for early predictions (30 days after planting)[4]. These results were compared to previous forecasts made by MCYFS. Soft wheat from Netherland had a Normalized RMSE of 7.87, for MCYFS its 6.32, sugar beet from Germany had a Normalized RMSE of 8.21 for MCYFS its 8.79 and sunflower from Francedata had a Normalized RMSE of 10.63 (10.91 for MCYFS). In contrast, predictions for Germany's soft wheat data and France's potato data and Sugar beet data were much worse than the Normalized Root Mean Square Error for Germany's soft wheat data was 16.38 (6.21 MCYFS), and for France's Sugar beet was 14.34 (MCYFS 7.42)[4].

As part of their 2021 study, Pranay Malik et al. used machine learning algorithms to analyze soil properties in order to predict fertility and yield[5]. A self-generated dataset was used for the analysis. Crop yield predictions were developed based on the KNN algorithm, Naive Bayes algorithm, and Decision Tree classifier. The authors collected a dataset of three different crops namely: tomato, Potato and chilly crops from three different soils alluvial, black and red soils. All the three soils monitored and collected the dataset of 1320 instances with four parameters to do this proposed experimentation which are moisture, pH, temperature and sunlight[5]. Authors used the data pre-processing technique on raw dataset to get rid of the redundant data and missing values.As a result of pre-processing the data, a set of machine learning algorithms was used - K nearest neighbors, Naive Bayes, and Decision Trees - to learn and analyze the example data. There are three classes that have been converted to numbers in order to make prediction easier: good, average, and poor[5]. The K nearest neighbor algorithm is based on Euclidean distance, and the prediction is made as a result of that. As part of the Naive Bayes algorithm, the Bayes classifier is used. Gini index has been used to classify data in the Decision Tree algorithm. After they divided the accuracy into 5- folds and calculated accuracy for every single fold. The final precision value was calculated by doing average of the 5 different outputs from the 5-fold.Moisture, pH, and temperature were determined to be optimal in the normal, 6-7-, and 35-40-degrees Celsius ranges, respectively[5]. To determine the average accuracy of three algorithms, the dataset was bifurcated into five folds. Based on the Euclidean distance, the KNN algorithm predicted the yield with 91% accuracy. In contrast, the Naive Bayes Algorithm predicts with an accuracy of only 76.426% using the Bayes theorem[5]. The highest prediction by Decision Trees for the Gini index was 95.361% obtained[5].

In 2020, Dhivya Elavarasan et al., An RF-RFE Wrapper-Based Feature Extraction Hybrid CFS Filter and RF-RFE Filter-Based Feature Extraction for Enhanced Agricultural Crop Yield Prediction Modeling has been proposed[6].Using the proposed hybrid feature extraction method, in the following paper they investigated a novel hybrid process of key feature extraction which is an integration of the random forest recursive feature elimination (RFRFE) and the correlation-based filter (CFS) combined framework. Using a collection of groundwater, soil, and climate characteristics, an optimal subclass of features is extracted for use in constructing crop yield predictions using machine learning. Model accuracy and correctness are calculated by estimating (i) the architecture precision based onfeatures of the dataset, (ii) the accuracy based on featureimportance's method built into the algorithm, and (iii) with the model's accuracy based on significant features based on the proposed hybrid feature extraction method[6]. Predictive accuracy, Evaluation metrics, and RFRFE feature extraction and plot analysis projected using diagnostics performance of the hybrid CFS approach are found to be profoundly satisfying compared to RF, DT, and GB algorithms. Using the proposed hybrid feature selection method, the model's predictive performance is examined using the crop dataset[6]. Efficiency of a ML model is assessed by measuring it according to various measurements or performance metrics. Some measurement or performance metrics considered for assessing developed workinclude: Mean Square error (MSE), Root mean square error (RMSE), Mean Absolute Error (MAE), Mean absolute percentage error (MAPE), and Co-efficient (R2). The accuracies and performance measures are mentioned in below table[6].

TABLE I. PERFORMANCE MEASURE AND MODEL ACCURACY OF CFS AND RF-RFE FILTER

Algorithm Name	Performance Measure with CFS Filter and RF-RFE Wrapper Feature Selection Method				
	MAE	MSE	RMSE	R^2	MAPE (%)
Random Forest	0.194	0.07	0.265	0.67	19
Decision Tree	0.341	0.182	0.426	0.55	33
Gradient Boosting	0.306	0.187	0.433	0.48	29

Algorithm Name	Model Accuracy with All Features in the Dataset (%)	Model Accuracy with Algorithm Inbuilt Feature Importance Method (%)	Model Accuracy with CFS Filter and RF-RFE Wrapper Feature Selection Method (%)
Random Forest	90.84	90.94	91.23
Decision Tree	77.05	80.75	82.58
Gradient Boosting	83.71	84.4	85.41

In 2020, N.R Prasad et al., proposed a method for crop yield prediction using random forest algorithm[7].Utilizing R package, the proposed algorithm predicts cotton yield in the state of Maharashtra in India three months before the actual harvest using the random forest (RF) algorithm. Further they took a ancillary crop data which is a major cash crop in Maharashtra state was retrieved from the data server. Later NDVI based vegetation condition index data is used and standardized precipitation index is used to get to know the crop moisture index during the growing seasons which is used to calculate the soil moisture and crop stress. The GDD ("Growing degree days")[7] is used to measure the heat accumulation in the phonological stages which is used as a predictor variable for RF model. All these combined as an input data for the RF by doing correlation analysis. To generate co-linearity of predictor variables, as well as calibrate and validate the RF model, long-term agrommet-spectral variables were extracted from multi-sensor satellites and crop yield data from 2001-17[7]. In September, December, and February the CART decision tree for recursive feature elimination with RF model functioned well in terms of predicting crop yield with 69%, 60% and 39% CART decision tree coefficient of determination (R2) in the final harvest, respectively[7]. This study demonstrated that the RF algorithm is capable of integrating and processing a wide range of inputs including satellite-derived inputs, unscaled and non-uniform information from the ground as well as expert knowledge. The RMSE and MAPE performance metrics of this model is presented in the below table[7].

TABLE II. RMSE AND MAPE PERFORMANCE METRICS[7]

Prediction model	RMSE (Kg ha^{-1})	MAPE	IA
September End	62.77	0.32	0.85
December End	65.70	0.32	0.84
February End	88.57	0.58	0.69

In 2020, Farhat Abbas et al., proposed a Crop Yield Prediction using Proximal Sensing data and ML Algorithms [5]. In addition to probing soil variables, proximal sensing can be applied to study crop yields. Using proximal sensing, the authors proposed four Machine Learning algorithms to predict potato yield, namely, linear regression (LR), support vector regression (SVR), elastic net (EN), and k-nearest neighbor(k-NN) based on soil properties and crop properties. Over two growing seasons (2017 and 2018), electrical conductivity of soil, moisture content of soil, slope value, NDVI index data, and chemical properties of soil were measured in 6 different agricultural lands in Canada, 3 different agricultural lands in Prince Edward Island (PE) and 3 different agricultural lands in New Brunswick (NB). Each growing season, four times, 30 * 30 m2 sections were measured in each field, and samples of yield were taken manually. By combining data points from three fields, four datasets representing provincial data for 2017 and 2018 were created: PE-2018, NB-2018, PE-2017, andNB-2017[8]. 80% training set and a 20% testing set were calculated based on the samples. In order to further refine the testing procedure, the k fold cross-validation method was adopted, which is used to assess the ability of machine learning algorithms to handle new, unknown data. The dataset is divided into approximately k equal-sized groups based on random sampling. Data are trained using k*1 folds, with the first fold being used as the testing set. Each dataset was examined with three folds (k = 3). Tests on one test set are less robust than tests on small datasets. Extensive testing was done in order to determine ML hyper parameters. Yield predictions were generated using models based on different statistical parameters. As noted above, SVR architecture for PE-2018, NB-2018, NB-2017,and PE-2017 datasets outperformed all other models with Root Mean Square Errors of 6.60,4.62, 6.17, and5.97 t/ha, respectively[8]. Except PE2018 dataset the rest three showed poor performance for k-NN, with RMSE of 6.916, 5.23, and.93 t/ha, respectively[8].

TABLE III. THE COMPARISION WITH ALL FOUR DATASETS AND FOUR ALGORITHMS USED[8]

Year	Algorithm	MAE (t/ha)	RMSE (t/ha)	Mean R^2	Std. Dev. (R^2)
2018	Linear Regression	3.59	4.69	0.63	0.04
	Elastic Net	3.79	4.72	0.63	0.06
	k-Nearest Neighbor	4.21	5.23	0.53	0.07
	Support vector regression	3.60	4.62	0.65	0.06
2017	Linear Regression	4.77	6.19	0.70	0.05
	Elastic Net	5.60	6.67	0.65	0.04
	k-Nearest Neighbor	5.57	6.93	0.62	0.09
	Support vector regression	4.68	5.97	0.72	0.07
2018	Linear Regression	5.01	6.24	0.53	0.09
	Elastic Net	5.27	6.54	0.49	0.11
	k-Nearest Neighbor	4.85	6.49	0.54	0.12
	Support vector regression	4.95	6.17	0.54	0.09
2017	Linear Regression	5.23	6.70	0.64	0.07
	Elastic Net	5.57	6.74	0.65	0.01
	k-Nearest Neighbor	5.62	6.91	0.64	0.05
	Support vector regression	5.18	6.60	0.65	0.06

In 2019, Lizi Wang published a paper proposing a process for predicting crop yields using CNN and RNN. The proposed CNN-RNN model, compared with RF, DFCNN, and LASSO, models used[9]. To forecast the performance of all these models' authors used soya bean and corn yield datasets which have data across the entire Corn Belt in all the thirteen states in the United States for 2016 to 2018. For crop yield prediction, a hybrid approach is proposed, combining fully connected layered models, RNNs, and CNNs. S-CNN and W-CNN systems are designed to work in conjunction with the weather and soil data in order to capture the linear and nonlinear effects, respectively[9]. In order to capture the temporal dependencies of weather data, the authors used the W-CNN model with 1D convolution to capture the spatial dependencies of soil data measured at different depths underground[9].Combining the high-level features extracted from the W-CNN and S-CNN models resulted in a fully connected layer, which also reduced the output dimension of the CNN models. This RNN model can be used to determine the time-dependence of crop yields over an extended period of time. Using information from years 't' to 't', the RNN model predicted the crop yield of a county for year 't' based on k LSTM cells[9].Based on the thesis, the CNN-RNN had an RMSE of 9% and 8%, which is substantially lower than all other methods currently used in crop prediction studies[9]. The CNN-RNN has three salient characteristics that make it a potentially useful technique for future crop yield prediction studies. Firstly, in CNN-RNN, age dependent environmental factors, as well as the genetic improvement of seed species over time can be captured without having to know a seed's genotype information[9]. Secondly, the model was successfully generalized to non-tested environments, so that the yields could be predicted with a higher degree of accuracy while performing in an untested environment. Thirdly, the model could reveal a great deal of information about weather conditions, accurate weather forecasts, soil conditions and management practices that could be associated with the variation in crop yield while combining this with the backpropagation method[9]. The accuracies of the four different models are mentioned in below table.

TABLE IV. THE COMPARISION OF ACCURACIES OF FOUR MODELS

Model	Validation Year	Training RMSE	Training Correlation Coefficient (%)	Validation RMSE	Validation Correlation Coefficient (%)
CNN-RNN	2016	13.26	93.02	16.48	85.82
	2017	12.75	93.68	15.74	88.24
	2018	11.48	94.99	17.64	87.82
RF	2016	13.38	92.74	25.48	69.52
	2017	14.31	92.39	29.40	69.03
	2018	14.40	92.39	26.02	70.55
DFNN	2016	12.34	94.43	27.23	81.91
	2017	11.21	95.09	23.88	79.57
	2018	11.54	95.25	21.37	79.85
LASSO	2016	19.88	81.81	32.58	61.90
	2017	20.62	81.83	27.06	61.18
	2018	20.81	83.63	31.30	55.95

In 2019, Hulyah Yalcin proposed a deep learning model for a relative Crop Yield predictionusing approximations onimages captured in agricultural field Using Deep Learning [7]. In this work, crop yield is estimated based on a deep learning architecture. The agricultural stations on the ground capture images every thirty minutes of the plants. Using intermediate outputs from deep learning architectures, they developed a measure for estimating a farmer's crop yield. In a deep learning architecture for training, estimates of the crop yield for agricultural parcels are used to translate high yield estimates into a relative measure of crop yield estimates. For the purposes of takingvarious phenological stages in to account by classifyingthe plants using this model, Alex-Net is usedto pre-train Convolutional Neural Network (CNN) model. In order to apply phenology recognition, the CNN model has the following layers and general structure. In order to fine-tune the CNN model, the authors did not train the initial layers because of the data accessibility of phonological stages. In order to tune the model, the first three layers are considered as frozen and only the last three layers are utilized for training. Once the top-layers have been personalized for the new dataset, the model is fine-tuned. Authors used ten times zoomed resolution images are used to train the network. A crop yield estimation measure was calculated by using the activation maps of the pre-trained CNN architecture, which was originally trained for phenology recognition. Using the following formula, authors estimated the yield of a crop[10].

$$CropYieldIndex = \frac{1}{M.N}\sum_{m=1}^{M}\sum_{n=1}^{N}\left[\begin{matrix} PhenoTypeVerification(m,n) \\ .ActivationMap(m,n)\end{matrix}\right]$$

$$PhenoTypeVerification(m,n) = \begin{cases} 1 & PhenoType(m,n) = Flowering \\ 0 & otherwise \end{cases}$$

After that, estimation is done on patches of 227 x 227 sized images with 1x resolution. A sliding window is used to scan the images, and a CNN architecture is used in order to calculate activation maps from the patches. Authors considered 0.0001 as initial learning rate to optimize the network using SGDM[10]. Maximum epoch limits are set at five. Validation data and frequency of validation are specified during training to ensure network accuracy. Afterward, the accuracy of an ensemble based on validation data is calculated using the accuracy of a network that is trained using the training data. A final assessment follows the allocation of precision as a small part of the marks that are accurately anticipated by the network. A validation set accuracy of 87.67% was achieved for this experiment[10].

In 2019, Ekaansh Khosla et al., proposed amethod for Crop yield prediction using aggregated rainfall based- modular artificial neural networks and support vector regression[11].This study aims at predicting kharif crops yields using a rainfall-based modular artificial neural network model and support vector regression that is based on the rainfall data for a particular crop area using modular artificial neural networks (MANNs). Also, for kharif crops yield prediction with support vector regression we are utilizing the rainfall datasets on an area basis and the individual crops area to predict the expected yields. With the use of the MANNs-SVR methodology, it can be made possible to formulate the right agricultural strategies that will ultimately help to improve the productivity of crops. As per the proposed model, the dataset comprises seasonal rain data collected from maritime areas of AP, along with crop production data collected from Visakhapatnam. The dataset has been cleaned, preprocessed, ANN has been used to predict rain with an entropy function, and crop production data has been added to the ANN to predict various kharif crops[11]. Modular Artificial Neural Networks (ANN) are built by first splitting the training data into several clusters, then building artificial neural networks (ANN) on top of each cluster. Fuzzy clustering based on C-Means methods is used to create a cluster in the present study[11].Fuzzy C-Means clustering is used to separate input-output data pairs based upon the magnitude of the rainfall that is low, medium, and high[11]. Each of the clusters is then fed into the Artificial Neural Network to get the final results. In this study, the multilayer perceptron is used as a means of training the network to calculate model inputs based on error back propagation. In the following analysis, six model inputs are examined in terms of root mean square error (RMSE) along with the Levenberg-Marquardt (L–M) training algorithm. For better calculations, the authors chose the LCA method since the RMSE values of these methods are very close. As soon as they completed the prediction of the kharif crop production, they used a number of regression techniques to predict it and in this situation SVR showed the best results out of all the regression techniques used (K-nearest neighbor regression, Random Forest regression, Linear Regression, and Support Vector Regression). Using their LCA methodology, the authors forecasted the rainfall for the coming years and calculated the yields of bajra, maize, rice, and ragi based on their predictions[11].

Methods	τ	m	Effective inputs	Identified ANN	RMSE
LCA	1	20	The last 12	(12-5-1)	63.74
AMI	1	12	The last 12	(12-5-1)	63.74
PMI	1	12	$X_{t-11,t-10,t-5,t}$	(4-5-1)	71.07
FNN	1	20	The last 5	(5-9-1)	79.85
SLR	1	12	Except for X_{t-4}	(11-9-1)	66.12
MOGA	1	12	$X_{t-11,t-9,t-7,t-5,t-4,t-1,t}$	(7-1-1)	76.24

In 2019 Saeed Khaki et al., proposed a Deep neural network for crop yield prediction[12]. Utilizing deep neural networks, this approach makes yield predictions based on genotype and environment data. With carefully designed deep neural networks and historical data, the artificial intelligence system was able to learn and predict the yield of new hybrid plants planted in new locations with known weather conditions based on nonlinear and complex relationships between genes, environmental conditions, and their interactions. There was a moderately significant relationship between the performance of the model and the quality of weather prediction, indicating the importance of these techniques. In the first stage the data processing in this study involved coding the genotype values as (-1, 0, 1) numeric values, which represent 'aa, 'a A, and 'AA', respectively[12]. The genotype data for about 37% of the individuals had missing values. Researchers addressed this issue by preprocessing genotype data in two steps before they were able to utilize them in neural network models. In order to consider genetic markers that had non-missing values of 97% or lower, they first used a 97% call rate to evaluate the markers. The researchers later also excluded genetic markers whose lowest common allele frequency is less than 1% or whose level of heterozygosity is below 1%, since these were less heterozygotic and as a result provide less information. Thus, in order to reduce the number of genetic markers from 19,465 to 627, we reduced the number of genes[12].It continues in the second stage of this study, where the investigators considered a 2001-15 dataset in order to predict the 2016 weather data, and they trained 72 shallow neural networks, which were used for 72 weather variables, across all locations, in order to capture the nonlinearities that exist in weather data, and the researchers learned these nonlinearities without requiring them to specify the nonlinear model beforehand, meaning that they learn these nonlinearities from their data without specifying the nonlinear model beforehand. Three deep neural network models were trained: one predicting yield and the other predicting check yield. Researchers used the difference between the two results to predict yield differences in the next stage. During the training phase, hyper parameters were used. There are 21 layers and 50 neurons in each layer in each neural network.In trials with deeper network structures, it was determined that these dimensions provided the best combination of high prediction accuracy and limited overfitting, so the weights were initialized using the Xavier method[12]. A batch of 64 SGD was later used. In the Adam optimizer, the learning rate was set at 0.03%, and every 50,000 iterations, the learning rate was divided by 2. Before activating the first hidden layer, batch normalization was used. Maximum 300,000 iterations were used to train the models. In every stacked hidden layer, residual shortcuts were used. A regularization was applied to both the hidden layer and the top layer to avoid overfitting. For the comparison of results the authors implemented three other prediction models: i) Lasso, ii) Shallow neural networks, and iii) regression trees[12]. The predictions for the yield difference were created from the differences between the outputs of each model, in order to ensure fair comparisons, and the differences of their outputs were used as a prediction for the yield difference resulting from the two models[12].

The team led by Alfini, Meftahet al., presented a system based on ANN, KNN, and SVM classifiers for enforcing ripeness grading in oil palm[13]. The external characteristics of the oil palm fruit, are taken as basis for training and testing of the system was carried out so that ripeness could be classified appropriately. A framework assessment was conducted to analyze the presentation information, the handling time, and the costs of the order capability, in order to improve the framework strategy. To implement the comparison and evaluation of supervised machine learning classifiers, we used a real-time inspection and ripeness rankingprocedure for the palm fruit types. Based on three different levels of ripeness (override ripe, ripe, and underripe) of palm samples, they employed extracted texture, color, and thorn properties to train the suggested method. A variety of feature extraction algorithms, including color histogram, mean, SD, gray-level co-occurrence matrix (GLCM), Gabor wavelet (GW), and basic grey level aura matrix, have been used in several experiments to determine the ranking procedure for palm fruit FFBs based on external fruit characteristics (BGLAM). based on experiments using the supervised machine learning methods KNN, SVM, and ANN.It was found that the algorithms were the most suitable methodology for grading the oil palm FFB ripeness. In the computer vision application, the five components of the common image processing system are executed: image acquisition, preprocessing, segmentation, object frame calculation, and classification. In terms of processing on a low-level, image acquisition and preprocessing fall under low-level processes; description,representation,and segmentation processes fall under the level of intermediate processes; and object recognition and image classification fall under high level processes. A supervised machine learning classifier is used for image classification (decision making) as the final stage of image processing for determining FFB types and ripeness for oil palm. It is a technique for developing a classifier that can deduce new instances from a set of rules that are discovered by training on preexisting examples. The real-time oil palm FFB ripeness grading system used BGLAM texture features based on ANN classifier, which resulted in the highest accuracy of 93 percent and the fastest image processing speed of 0.40 (s) when compared to other feature extraction approaches and machine learning classifiers[13].

III. Comparision Table

Year	*Algorithm*	*Dataset*	*Crop type*	*Performance metrics*	*Ref. No*
2021	Gradient boost decision tree with 5-kfold	MCYFS	Wheat(W) Barley(B) Sugar beet (SB) Sunflower (S) Potato(P)	**RMSE** W-7.86, B-6.3 SB-8.21 S-10.36 P-6.8	[4]
2021	Naïve Bayes kNN Decision tree classifier	Self obtained dataset	Tomato, Potato Chilly	KNN-91.179% NB-74.426% DT-95.36%	[5]
2020	Random Forest Decision Tree Gradient Boosting	Self obtained Indian dataset collected from Tamil Nādu state	Paddy yield	RF-91.23% DT-82.58% GB-85.41%	[6]
2020	Random Forest	VCI, GDD, SPI, LST, historical data	Cotton	R-squared 0.69(SEP) R-squared 0.60(DEC)	[7]
2020	Linear Regression, Elastic Net, k-NN Support Vector Regression	Combined datasets of six different fields in Canada, including 3 fields in Prince Edward Island and 3 fields in New Brunswick 2017-2018	Potato	RMSE LR- 4.69 EN-4.72 kNN-5.23 SVR-4.62	[8]
2019	Hybrid CNN RNN	13 US states corn belt(2016-2018)	Soyabean and corn	CNN-RNN 2016-RMSE 2016-13.26 2017-12.75 2018-11.48	[9]
2019	CNN	Self obtained dataset from 4 different agricultural stations	Sunflower	Validation accuracy – 87.67%	[10]
2019	Modular Artificial Neural network and support vector regression	Rain data from Costal andhrapradesh and Crop production data from visakhapatnam	Maize, Rice, Ragi and Bajra	Metrics in terms of yield production of yearly basis	[11]
2019	DNN LASSO SNN DT	the genotype and yield performances of 2,267 maize hybrids planted in 2,247 locations between 2008 and 2016	Maize	RMSE DNN-12.79 LASSO-21.40 SNN-18.04 DT-15.03	[12]
2018	ANN	Collected data from Multiple palm oil databases	Palm Oil	93%	[13]

IV. Conclusion

Technology enables farmers to produce optimum results with precise inputs through precise and accurate agriculture. Some of the key technological revolutions that have had an impact on the agriculture industry are smart sensors, actuators, satellite images, robotics, drones, and IoT enabled devices. There is a crucial role for these components to play as they collect data in real time and make decisions accordingly without any assistance from humans. Incorporating AI, using the data for automated behaviors that benefit our planet and make life more comfortable for humans. Our paper explores how machine learning can be applied to precision agriculture and crop yield. It discusses the impact of artificial intelligence, machine learning, and big data in smart farm management, and various researchers work with their results to manage different crops. We concluded from our research that regression algorithms are crucial for predicting crop yields, soil properties, and weather.

V. References

[1] "Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. Future generation computer systems, 29(7), 1645-1660.".

[2] "Sharma, Rohit, et al. "A systematic literature review on machine learning applications for sustainable agriculture supply chain performance." Computers & Operations Research 119 (2020): 104926.".

[3] "Liakos, Konstantinos G., et al. "Machine learning in agriculture: A review." Sensors 18.8 (2018): 2674.".

[4] "Paudel, Dilli & Boogaard, Hendrik & Wit, Allard & Janssen, Sander & Osinga, Sjoukje & Pylianidis, Christos & Athanasiadis, Ioannis. (2020). Machine learning for large-scale crop yield forecasting. Agricultural Systems. 187. 103016. 10.1016/j.agsy.2020.103".

[5] "P. Malik, S. Sengupta and J. S. Jadon, "Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine Learning Algorithms," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, p".

[6] "Elavarasan D, Vincent P M DR, Srinivasan K, Chang C-Y. A Hybrid CFS Filter and RF-RFE Wrapper-Based Feature Extraction for Enhanced Agricultural Crop Yield Prediction Modeling. Agriculture. 2020; 10(9):400. https://doi.org/10.3390/agriculture10090400".

[7] "Prasad, N R & R., Patel & Danodia, Abhishek. (2020). Crop yield prediction in cotton for regional level using random forest approach. Spatial Information Research. 29. 10.1007/s41324-020-00346-6.".

[8] "Abbas, Farhat & Afzaal, Hassan & Farooque, Aitazaz & Tang, Skylar. (2020). Crop Yield Prediction through Proximal Sensing and Machine Learning Algorithms. Agronomy. 10. 1046. 10.3390/agronomy10071046.".

[9] "Khaki, Saeed, Lizhi Wang and Sotirios V. Archontoulis. “A CNN-RNN Framework for Crop Yield Prediction.” Frontiers in Plant Science 10 (2019): n. pag.".

[10] "H. Yalcin, "An Approximation for A Relative CropYield Estimate from Field Images Using Deep Learning," 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 2019,pp. 1-6, doi: 10.1109/Agro-Geoinformatics.2019.8820693.".

[11] "Khosla, Ekaansh & Ramesh, Dharavath & Sharma, Rashmi. (2020). Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression. Environment, Development and Sustainability. 22. 10.1007/s10668-019-00445".

[12] "Khaki, S. and Wang, L., 2019. Crop yield prediction using deep neural networks. Frontiers in plant science, 10, p.621.".

[13] "Alfatni, Meftah & Mohamed Shariff, Abdul Rashid & Bejo, Siti & Saaed, Mafaz & Mustapha, Aouache. (2018). Real-time oil palm FFB ripeness grading system based on ANN, KNN and SVM classifiers. IOP Conference Series: Earth and Environmental Science. 169. 012".