Importance of machine learning to predict crop yield using weather parameters

Shreya Gupta

Division of Agricultural Physics

ICAR-Indian Agricultural Research Institute

New Delhi, India

shreya14146@gmail.com

Dipankar Barman

Division of Plant Physiology

ICAR-Indian Agricultural Research Institute

New Delhi, India

barmandipankar1993@gmail.com

Ananta Vashisth

Division of Agricultural Physics

ICAR-Indian Agricultural Research Institute

New Delhi, India

ananta.iari@gmail.com

ABSTRACT

Agriculture is a crucial sector in every country, playing a significant role in the country's gross domestic product. The success of agriculture and crop management depends heavily on the yield and market rates. Timely monitoring and accurate predictions are essential factors in determining crop yield, which are strongly influenced by weather variables. Artificial intelligence (AI) offers an automated approach to monitor crops and predict yield outcomes. This chapter focuses on the application of machine learning and hybrid techniques, incorporating weather indices and heat indices as input variables, to achieve improved and reliable crop yield predictions and monitoring. Various machine learning techniques, such as artificial neural networks, support vector machines, and hybrid machine learning, are discussed to demonstrate how AI technology contributes to enhancing crop yield. Accurate and timely crop production forecasts are crucial for making important policy decisions related to import-export, pricing, marketing, and distribution, which are issued by government agencies. Therefore, there is a need to develop statistically robust and objective crop production predictions. The advancements in computing and information storage have resulted in large volumes of data. The challenge lies in extracting meaningful insights from this raw data, leading to the development of new approaches and techniques like machine learning, which can combine data knowledge with crop yield evaluation. Innovative techniques such as artificial neural networks and support vector machines have been explored in the agricultural domain for yield monitoring and prediction. Machine learning, along with big data technologies and high-performance computing, has opened up new opportunities for data-intensive scientific research in the multidisciplinary field of agricultural technologies.

**Keywords**—yield; machine learning; hybrid machine learning; weather indices; heat indices

1. **INTRODUCTION**

Agriculture plays a vital role in the Indian economy and is crucial for the livelihoods of its people. Over 70 percent of rural livelihoods in India are dependent on agricultural activities, including crop cultivation and various agricultural services. Among the numerous crops being cultivated in India rice, wheat, maize are the most grown crops because of their high yield potential and good climate of India supporting such crops. According to Second Advance Estimates for 2022-23, total Foodgrain production in the country is estimated at record 3235.54 lakh tones which is higher by 79.38 lakh tones as compared to previous year 2021-22 production. Total production of Rice, wheat, maize and coarse grains during 2022-23 is estimated as record of 1308.37, 1121.82, 346.13 and 527.26 lakh tones respectively. It is higher by 13.65, 44.40, 8.83 and 16.25 lakh tones respectively as compared to 2021-2022. Thus, monitoring yield production is an important step which should be carried out with accuracy to enhance the policy scheme, management, storage, import, export etc. Reliable and well-timed monitoring and forecasts provide important and beneficial input for foresighted and informed planning in agriculture which is full of uncertainties.

Recent research reports have indicated that fluctuations in crop yields from year to year are often linked to climate variability and weather patterns [1]. Consequently, the impacts of climate variability and change on food and fibre production have become a major concern globally [2]. In the case of wheat production, climatic factors like rainfall, temperature, solar radiation, and relative humidity have a significant influence on yields [3]. Evapotranspiration (ET), which is the combined loss of water through evaporation from the soil surface and transpiration by plants, plays a crucial role in the crop water balance. It regulates various hydrological processes and accurate estimation of ET is essential for effective water resources planning and management throughout the crop's lifecycle [4]. Numerous empirical equations have been developed to estimate ET using meteorological data, but the Penman-Monteith FAO-56 combination equation (PMF-56) has been recommended by the Food and Agriculture Organization of the United Nations (FAO) as the standard equation for estimating ET. The PMF-56 requires several climatic parameters such as daily maximum and minimum temperature, relative humidity, solar radiation, and wind speed. Weather variability, both within and between seasons, significantly contributes to yield variability. Different weather variables impact crops differently during various stages of development. The influence of weather on crop yield depends not only on the magnitude of weather variables but also on their distribution pattern throughout the crop season.

Crop yield monitoring can be achieved through three main methods: (i) biometrical characteristics, (ii) weather variables, and (iii) agricultural inputs. These methods can be utilized individually or combined to create a composite model [5]. Numerous studies have been conducted to monitor crop yield using weather parameters. However, such monitoring studies based on statistical models need to be conducted continuously and for different agro-climatic zones due to the visible effects of changing environmental conditions and weather patterns in different locations. Crop yield monitoring plays a crucial role in ensuring food security, optimizing agricultural management practices, and promoting sustainable development. Various sources of information, such as field observations, climate data, satellite images, and crop growth simulation models, are utilized in crop yield monitoring and forecasting. Weather and climate have a significant impact on crop yield and production. Increasing intensity and frequency of climate extremes can affect crop yield, food security, and safety. Changes in climate variables can have profound effects on global socio-economic and agricultural systems. Climate variables like temperature and precipitation directly influence agriculture and changes in these variables can significantly impact crop yield and variations. Climate variables affect crop growth and development at different stages of the growth cycle.

1. **MACHINE LEARNING**

Artificial Intelligence (AI) involves the development of techniques that enable computers to learn and perform tasks. Machine learning, a subset of AI, focuses on the development of algorithms that allow machines to learn from data and perform specific activities. Machine learning methods are also named data-driven methods. It is a subtype of computer science and as it is classified as an artificial intelligence method, it bothers with the perfection of techniques which empower the computer to acquire. In simple terms evolution of algorithms which permit the machine to learn, accomplish tasks and actions. Its commonalities with statistics in many ways. Over the period of time many methodologies were developed for tasks of machine learning. It very useful may be utilized in a large area and the benefit of this technique is that a model can tackle issues which are difficult to express by calculations. The AI models discover relations among information sources and yields (output) even if the representation is not possible. Hence this permits the utilization of AI models as a rule, for instance in design acknowledgment, characterization issues spam filtering, in information mining and anticipating.

Machine learning is concerned about the development and investigation of frameworks that can gain from informational data sets, enabling system to learn without being explicitly programmed. In the predictive learning issues, the framework comprises of an irregular “output” or response” variable y and a lot of arbitrary "input" or "logical" factors. They build a model from proof acquired from a set of information which is Input and Output. The preparation stage that is basically the training phase brings about a capacity that can be applied to new Input data so as to foresee the relating Outputs. The calculations can recognize the complicated patterns in the input with connection to yield or the output by combining simple components. In rice crop yield estimation, different order procedures were applied like support vector machine, naïve bayes, bayesnet, locally weighted learning, decision tree and so forth, to get increasingly noteworthy outcome and to show signs of improvement characterization precision. Bayesnet determines better precision of forecast of rice yield in future. This reasoned the more exact rice crop yield forecast model can be built by utilizing various data mining techniques [6].Various methodologies and techniques have been developed for machine learning tasks. Machine learning algorithms can identify and learn patterns between input data and corresponding target values through training. Once trained, these algorithms can be used to predict outcomes for new, independent input data. Machine learning has great potential for predictive modelling, where all relevant data describing a situation can be presented to trained models to make predictions about agricultural systems.

1. **Types of Machine Learning (ML)**

Machine learning algorithms are often categorized as supervised or unsupervised. Supervised machine learning algorithms learn from labelled examples in training datasets to make predictions about future events. By analysing known training data, the learning algorithm produces a function that can predict output values. After sufficient training, the system can generate predictions for new input data. The learning algorithm can compare its output with the correct output, identify errors, and modify the model accordingly. Unsupervised machine learning algorithms are used when the training data is unclassified or unlabelled. In unsupervised learning, the algorithms infer hidden structures from the unlabelled data. The system explores the data and draws inferences to describe hidden structures. Unlike supervised learning, unsupervised learning does not aim to find the correct output but rather focuses on understanding the underlying patterns and structures within the data. There are various machine learning algorithms, including decision trees, artificial neural networks, support vector machines, and many others. These algorithms provide different approaches to learning and pattern recognition, enabling machines to perform specific tasks and make predictions based on the data they are trained on.

Apart from supervised and unsupervised machine earning there are other algorithms also like semi supervised, reinforcement machine learning etc. Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning, since they use both labelled and unlabelled data for training – typically a small amount of labelled data and a large amount of unlabelled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labelled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabelled data generally doesn’t require additional resources. Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behaviour within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal. Machine learning enables analysis of massive quantities of data. While it generally delivers faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. Combining machine learning with AI and cognitive technologies can make it even more effective in processing.

1. **Difference between Artificial Intelligence (AI) and Machine learning (ML)**

Artificial intelligence is a technology using which can create intelligent systems that can simulate human intelligence whereas Machine learning is about extracting knowledge from the data. Detail difference is discussed in table 1.

**Table1:** **Difference between Artificial Intelligence and Machine learning**

|  |  |
| --- | --- |
| **Artificial Intelligence** | **Machine learning** |
| Artificial intelligence (AI) is a technology that allows machines to mimic or imitate human behaviour. It provides machines with the capability to simulate and replicate various aspects of human intelligence and behaviour | Machine learning is a specific branch of AI that enables machines to learn from past data without requiring explicit programming instructions. This ability to learn and adapt autonomously sets machine learning apart from traditional programming approaches |
| The objective of AI is to create intelligent computer systems that possess human-like capabilities in order to address intricate problems. The aim is to develop smart machines that can effectively solve complex challenges, comparable to the problem-solving abilities of humans | The objective of machine learning is to enable machines to acquire knowledge from data, enabling them to provide precise outputs. The goal is to develop systems that can learn from data, identify patterns, and utilize that knowledge to make accurate predictions or generate reliable outcomes. |
| ML and deep learning are the two main subdivisions of AI | Deep learning is a main subdivision of ML |
| AI can be divided into three types, which are, Weak AI, General AI, and Strong AI. | ML classified as Supervised learning, Unsupervised learning, semi supervised and Reinforcement learning. |

1. **CONCEPT OF YIELD PREDICTION AND MONITORING**

Yield monitoring refers to the practice of measuring and analyzing crop yields to assess agricultural productivity. It involves the collection of data related to crop production, such as yield per unit area, crop quality, and other relevant parameters, to make informed decisions regarding farming practices, resource allocation, and overall farm management. In recent years, yield monitoring in India has gained significant importance due to several factors, including increasing population, changing dietary patterns, and the need for sustainable agricultural practices. Accurate and timely yield data can help farmers optimize their production processes, improve resource efficiency, and enhance profitability. Overall, yield monitoring in India is an essential component of modern farming practices, enabling farmers to make data-driven decisions, improve resource efficiency, and achieve sustainable agricultural production.

1. **Why prediction and monitoring of crop yield**

Crop yield monitoring can be used by Government, traders, policy makers, agro-based industries and agriculturists. Government use crop yield prediction in procurement, buffer-stocking, distribution, price fixation, import, export and marketing of agricultural commodities. Based on prior experience, Farmers cultivate crop, but nowadays due to the uncertainty increased in environment the accurate analysis of historic data of environment parameters should be done for successful farming. To get more harvest, analysis of previous cultivation data can be used. Estimation of crop yield mainly cereals such as rice, wheat, corn has always been an fascinating research area for agro-meteorologists, as these crops are important in national and international economic programming. Reliable crop yield forecasts which are early over large area would help policy makers as well as grain marketing agencies in planning for exports and imports business and local marketing. Any farmer is always attentive in knowing quantity of yield they are about to assume from their income generating farm.

Crop yield monitoring plays a crucial role in ensuring food security, optimizing agricultural management practices, and promoting sustainable development. Various sources of information, such as field observations, climate data, satellite images, and crop growth simulation models, are utilized in crop yield monitoring and forecasting. Weather and climate have a significant impact on crop yield and production. By monitoring crop yield, farmers can better plan and manage their agricultural operations. Yield data helps in determining optimal planting density, fertilizer and water requirements, pest and disease management strategies, and harvesting schedules. It enables farmers to allocate resources effectively and make adjustments to improve future yields. Also, it allows farmers to evaluate the performance of different crops, varieties, or management practices. By comparing yield data over time, farmers can identify which crops or practices are more successful and make informed decisions for future planting choices. Yield monitoring provides farmers with data-driven insights that support decision-making that helps in determining the profitability of specific crops, estimating market demand, negotiating contracts with buyers, and assessing the financial viability of different agricultural ventures. It helps in Detecting abnormalities and issues in the field it can be in the form of significant deviations in yield from expected levels may indicate nutrient deficiencies, pest infestations, water stress, or other problems that require immediate attention. Timely detection and intervention can help mitigate potential losses and ensure crop health. This can be very useful in Research and Development as crop yield data collected through monitoring can contribute to analysis of yield trends, study the impact of climate change on crop productivity, develop new crop varieties, and improve agricultural practices based on real-world data.

1. **Methods of Yield prediction and monitoring**

The monitoring and prediction of agricultural yield is one of the most challenging and desirable tasks for every nation and especially in country like India which is a developing country and where there is large population to fight with hunger. Prediction of yields of any crop can be done grounded on historic crop cultivation data and meteorological. There are several methods and technologies used for yield monitoring in India, including:

* Field Surveys: Traditional field surveys involve physical counting, weighing, and sampling of crops to estimate yields. This method is time-consuming and labour-intensive but can provide reasonably accurate results for small-scale farming operations.
* Remote Sensing: Satellite imagery and remote sensing technologies play a vital role in yield monitoring. These techniques use sensors mounted on satellites or aircraft to capture data on vegetation indices, crop health, and biomass. Remote sensing can provide valuable information on crop conditions and spatial variability, enabling farmers to make informed decisions about irrigation, fertilizer application, and pest control.
* Geographic Information System (GIS): GIS technology combines spatial data with yield information to create detailed maps of crop yields across different regions. These maps help identify patterns and variations in yield performance, allowing farmers to target specific areas for improvement and implement site-specific management strategies.
* Precision Agriculture: Precision agriculture techniques, such as Global Positioning System (GPS) and Geographic Information System (GIS) technologies, enable farmers to collect data at a high level of precision. Yield monitors equipped with GPS receivers and sensors can record crop yield and location data in real-time, providing accurate and detailed information for analysis and decision-making.
* Mobile Applications: The increasing use of smartphones and mobile applications has facilitated yield monitoring in India. There are various agricultural apps available that allow farmers to record and track crop yields, input usage, and other relevant data. These apps often leverage GPS technology and cloud storage to store and analyze data, providing farmers with valuable insights for yield optimization.
* Artificial intelligence and machine learning: Artificial Intelligence or AI is a branch of computer science by which can create intelligent machines behaving, thinking like a human, and able to make decisions. It replicates the human intelligence in solving Knowledge-intensive tasks thus helps in monitoring the yield of various crops by the algorithm provided and through calibration and testing.

In past, mostly yield monitoring was achieved by seeing farmer's experience on field and crop, which seems to be very cumbersome. The yield monitoring is a major issue that remains to be solved based on available data. To some extent this issue can be solved through machine learning and artificial intelligence.

1. **WEATHER PARAMETERS AND YIELD PREDICTION**

The crucial climatic factor for crop growth and production is air temperature, light, rainfall and sunshine hours [7-8]. The global mean annual temperatures towards the end of 20th century were approximately 0.70 C compared to 19th century also is probable to surge further by means of 1.8 to 6.4 0C by 2100 AD [9]. The most prominent change is inclining in the climatic temperature as a result of augmented levels of greenhouse gases in the atmosphere. Weather variables and indices like maximum temperature and minimum temperatures, rainfall and relative humidities were considered as input variables for application of Neural Networks (NNs) for crop yields forecast done at district level of Uttar Pradesh. Various weather and agricultural parameters, which includes NDVI, surface parameters (surface temperature and soil moisture) and rainfall data were used for crop yield estimation using a piecewise linear regression method with break point. [10]. The vital effect of high temperature is the augmented physiological changes, subsequently hasten maturation and low yield. The transformation of heat energy to dry matter in process photosynthesis depends on crop, sowing time and crop genetic factors [11]. The most important parameter among all the biological processes of yam production is the air temperature [12]. It affects the phenology, growth, development and yield of yam most among all the other weather parameters in the tropical wet as well as dry climate [13-14].

Dynamic simulation models of crop growth are extensively utilized for forecasting growth and harvestable yield of crops, but actually these are very comprehensive models with huge prerequisite of input data besides being highly difficult and complex in practical situation. Agro climatic models constructed on thermal indices along with weather parameters may possibly fulfil these goals. Therefore various efforts have been done in many fields to predict the crop phenology [15], growth rate [16], leaf area index (LAI) [17], yield [18], growth and crop yield by means of thermal based indices.

1. **Heat indices and yield prediction**

Thermal time is the very important independent variable to explain plant development. It can also be used as an important means for characterizing thermal responses in crops [19]. The time interval of a specific growth phase is directly associated to temperature which is used to derive the various agro climatic indices, viz., helio-thermal units, growing degree days, photo-thermal units, phenol-thermal index and heat use efficiency [20]. Information of GDD can provide an estimate of harvest date and crop development [21-23].

Growing degree days explain the thermal requirement of a crop. Nevertheless, accumulation of growing degree days and photo thermal units are comparatively constant and autonomous for crop variety and sowing date but it can modify significantly for each developmental stage in crop [24]. Growing degree days and photo thermal unit requirement differ from crop to crop. It also differs significantly from genotype to genotype. The ever-changing of sowing dates parallels to the deviations in temperature causing lengthening or shortening of growth period. Furthermore, it was also resulted that in a general the delay in sowing causes a reduction in GDD and PTU requirement of pheno-phases. Reference [25-26] witnessed that temperature has momentous impact on duration of different phonological phases and dynamics of leaf appearance. Predictive models for sweet potato (*Ipomoea batatas*) harvest in relative to growing degree days (GDD) was done using set of data of 116 planting dates using a combination of minimum CV, linear regression (LR), and several models in data mining (DM) mode to identify candidate methods of estimating relationships between GDD and harvest dates. These algorithms included regression trees, neural networks, multivariate adaptive regression splines, support vector machine, and generalized linear models. Models were based on calculated GDD and climate-related variables like maximum air temperature, mean relative humidity 20 days after transplanting (DAT), and maximum soil temperature. In this GDD method demonstrated high predictive accuracy as shown by mean square error [27]. Reference [28] Yield prediction was done in Hamedan district during time period of 2003-04 and 2004-05 for Wheat using different variables of meteorology with meteorological indices. Result was interpreted according to statistical methods which gave the best subset of meteorological indices which were selected like daily minimum temperature, growing degree days (GDD), difference of maximum and minimum temperatures (TD), sunshine hours (SH), potential evapotranspiration (PET), and water vapour pressure deficit (VPD). Yield estimation was done two months in prior of harvesting. It showed that in the statistical models 83% of yield variability was accounted for variation in meteorological indices. It can be assumed that solar radiations and temperature are the driving force of agriculture production and photothermal quotient (PTQ) use to describe effects on plant growth. Photo thermal index (PTI) is numerically equal to sum of the ratios of day length per 24 h period and multiplied by the total growing degree days [29]. Heat use efficiency (HUE) also significantly depends on genetic factors, crop type and time of sowing and has great application [30]. Helio thermal unit (HTU) helps in calculating and articulating the consequence of variable ambient temperature on the spell of phenological activities relating the crop reactions to the environmental temperature.

1. **Yield estimation Models using weather parameters**

Crop yield depends on numerous factors such as temperature and precipitation, disease, soil conditions, and anthropogenic factors like fertilizers, irrigation. Climatic data are easily available also available in most of districts of India, but some of the factors can be difficult or may be impossible to quantify thus using weather parameter can boost the prediction of yield well in advance. Weather causes direct and indirect influence on crop growth development and hence its production in terms of yield. Models are simplified illustration of the complex relationships between weather or climatic elements and the performance of the crop at different stages of growth such as biomass, and yield with the help of mathematical or statistical methods [31]. There are a many yield forecast models and generally classified in two parts first one is Statistical Models and other being Crop Simulation Models. Lately, application of Artificial Intelligence (AI), Artificial Neural Networks (ANNs), Fuzzy Systems are upcoming. The simulation models intended to forecast crop yield use detailed crop ecology and requires widespread information on soil type, plant factors information and weather data which is related to the crop growth stage. For large area yield estimation, the regression approach has been preferred over simulation because of trouble in getting all the data (for example, soil information, plant parameters, planting dates and certain agrometeorological information) which are expected to develop, test and run even a basic simulation model [32]. A problem associated with regression, illustrated by is multi-collinearity amid the independent variables. Simulation models are gaining popularity from past few decades, because it is process-based and requires intensive input data for better operation. Crop simulation models are specific and accurate thus can be useful for large spatio-temporal scales if sufficient input data is available.

 Pioneer work in crop weather relationship was done by reference [33-34]. Models developed by them require few parameters for assessind when dealing with trend of weather parameters over the crop growing period. Fisher statistical model was that impact of meteorological conditions on crop follows an orderly pattern according to mathematical law. The Fisher’s technique was modified by in 1943 by Hendricks and Scholl and later this methodology was modified by IASRI-New Delhi. According to reference [35] crop-weather models are classified into three kinds: crop-weather analysis models, crop growth simulation models and empirical statistical models. Multiple regression analysis uses one or more climate variables for its analysis. Besides from meteorological variables, soil qualities, biometric perceptions, inferred meteorological parameters, accessible soil dampness, real evapotranspiration, water stress records, temperature-based list etc. have been associated with product yields. Independent variables comprised weather variables, agro-meteorological variables, soil characteristics or some derived indices of these variables. Some popular agro-climatic indices like Thermal Interception Rate Index (TIR), Growing Degree Days (GDD) were used in models. Reference [36] used Southern Oscillation Index (SOI) with various other weather parameters or elements to forecast the crop yield. Various works for the development of crop weather yield models have been done by India Meteorological Department (IMD). They use long term yield data of district yields as dependent parameters whereas the independent variables were rainfall, temperature, relative humidity which are recorded daily, also some derived parameters like radiation, rainy days, mean temperature, evapo-transpiration, soil moisture, and yield moisture index were used. Highly correlated variable during the critical phenological stage of the crop were analyzed to establish multiple collinearities with yield.

1. **YIELD PREDICTION AND MACHINE LEARNING**

Currently crop yield forecasting is done using linear statistical models, which are not able to account nonlinear relations in the data. The cropland ecosystem being very complex and most of the processes are non-linear. This is the limitation with statistical models, such as multiple linear regression. Majority of the forecasting models in the past have been using multiple linear regressions (MLRs) for developing yield forecast [37-39]. Multiple linear regression has very big disadvantage of over-fitting, when data have number of samples less than the number of variables along with multi-collinearity, when the independent predictors are correlated [40]. To overcome the various challenges in modelling, feature selection methods in statistical analysis are included, such as Least absolute shrinkage and selection operator (LASSO), Stepwise Multiple Linear Regression (SMLR), Support Vector Machine (SVM) etc. are being developed. Neural networks which are non-linear models have been effectively used to predict crop yield by means of remotely sensed vegetation indices. Artificial neural networks to predict soybean and corn yield in the USA. So, the study attempted to use completely nonlinear models in the field of machine learning, like model-based recursive partitioning (MOB). If the design a network or data can be arranged which correctly learns and understand the relations of effective and more contributing climatic factors on crop production, this can be used to envisage the crop yield in both the long and short term and along with enough and useful data can get a ANNs model for each area. Also using ANNs can detect the most effective factors on production of crop. So, some of the factors for which the measurements are very difficult and only theoretically valid measurement are there can be ignored as there are many climatic or weather data which can be easily measured through the meteorological observatory or with the sensors. There are many other classification techniques like Naïve Bayes, Bayes Net, K Nearest Neighbour (KNN), Locally Weighted Learning (LWL) etc.

Reference [41] established weather-based model for yield predicting in turmeric intended for Coimbatore district, Tamil Nadu with coefficient of determination 0.89. Reference [42] found that FAO methodology involves developing an agro-meteorological index dependent on water surplus/water deficit in period of crop growth, having good potential for early crop yield assessment designed for rainfed crops. Reference [43] done study on wheat yield prediction based on agrometeorological data in rainfed Potohar region of Pakistan. Prediction of Rice yield in 13 districts of West Bengal using agrometeorological models was done and result obtained was a good relation between observed yield and predicted yield. Models developed have coefficient of determination ranging between 0.56 to 0.96 [44]. Reference [45] developed statistical models in coastal Karnataka for Ragi and rice using weather elements and yield data for long term. Reference [46] based on the weather data ordinal logistic model was developed for forecasting wheat yield in district Kanpur of Uttar Pradesh. They used weekly weather data on maximum temperature, minimum temperature, rainfall, relative humidity and for sixteen weeks of the crop cultivation for the period 1971-72 to 2009-10 along with the yield data of wheat crop.

1. **Steps of crop yield monitoring using Artificial intelligence (or Machine learning)**

Crop yield monitoring using machine learning involves the application of machine learning algorithms and techniques to analyze crop yield data and generate predictions or classifications related to crop productivity. Here is a general summary of the steps involved:

* Data Collection: Gather a comprehensive dataset of historical crop yield data, encompassing various factors that impact yield, such as weather conditions, soil characteristics, crop management practices, and other pertinent variables. The dataset should encompass multiple seasons and locations to capture a diverse range of conditions.
* Data Preprocessing: Clean and preprocess the collected data to ensure its quality and suitability for machine learning algorithms. Tasks include handling missing values, removing outliers, normalizing or standardizing data, and encoding categorical variables.
* Feature Selection/Engineering: Determine the most pertinent features (variables) that significantly influence crop yield. Techniques like feature selection or domain expertise can assist in identifying important variables. Feature engineering may involve creating new features through combination or transformation of existing ones to enhance model performance.
* Model Training: Select an appropriate machine learning algorithm or ensemble of algorithms based on the specific requirements of the crop yield monitoring task. Commonly used algorithms include regression models (e.g., linear regression, random forest regression), support vector machines (SVM), and neural networks. Split the dataset into training and validation sets and train the model using the training data.
* Model Evaluation and Optimization: Assess the trained model using appropriate evaluation metrics, such as mean squared error (MSE) for regression tasks or accuracy, precision, and recall for classification tasks. Optimize the model by fine-tuning hyperparameters or utilizing techniques like cross-validation to discover the best-performing configuration.
* Prediction or Classification: Once the model is trained and optimized, employ it to predict crop yield for new or upcoming seasons based on input data like weather forecasts, soil data, and crop management practices. Alternatively, the model can classify crops into yield categories (e.g., low, medium, high) using specific thresholds or criteria.
* Monitoring and Refinement: Continuously monitor the performance of the machine learning model and update it as new data becomes available. This iterative process enables ongoing refinement and improvement of the model's accuracy and predictive capabilities.

It is worth noting that the success of crop yield monitoring using machine learning depends on the availability and quality of data, appropriate feature selection, and the choice of suitable algorithms. Furthermore, domain expertise and collaboration with agricultural experts are valuable in interpreting and validating the model's outputs and making informed decisions based on the predictions or classifications.

1. **Development of weather indices**

Weather or climatic variables as such cannot be used as an input parameter for the machine learning. It has to ne first converted to the weather indices using the steps described here. Reference [47] model for distribution of weather element was modified at Indian Agricultural Statistics Research Institute (IASRI), where, the effects of changes in weather variables on yield in the wth week was expressed as second-degree polynomial in respective correlation coefficients between the yield and the weather variables [48-50]. This relationship was elucidated in much better way as weather in different weeks receives suitable weightage. Reference [51] further modified this model by considering the fact that the impact exerted by variations in weather parameters in wth week on yield is having a linear function of corresponding correlation coefficients between the yield and the weather parameters. The significant impact of trend on yield was removed while calculating correlation coefficients of yield with weather elements to be used as weights. These studies on the effects of second-degree terms of weather variables showed that: (a) the models using correlation coefficients based on adjusted yield for trend effect was found to be better than the ones using only simple correlations (b) quadratic terms of weather variables as well as second power of correlation coefficients did not make any improvement in the model.

Weather data for crop growing period have been used for generating weather indices and developing the crop yield forecast model. Weather indices used for developing crop yield forecast model using weather variables is given in Table 2.

**Table 2: Weather indices used by machine learning using simple and composite weather variables for developing yield monitoring and prediction**

|  | Simple weather indices | Weighted weather indices |  |
| --- | --- | --- | --- |
|  | Tmax | Tmin | RF | RH I | RH II | SSH | EVP | ETo | Tmax | Tmin | RF | RH I | RH II | SSH | EVP | ETo |
| Tmax | Z10 |  |  |  |  |  |  |  | Z11 |  |  |  |  |  |  |  |
| Tmin | Z120 | Z20 |  |  |  |  |  |  | Z121 | Z21 |  |  |  |  |  |  |
| Rf | Z130 | Z230 | Z30 |  |  |  |  |  | Z131 | Z231 | Z31 |  |  |  |  |  |
| RH I | Z140 | Z240 | Z340 | Z40 |  |  |  |  | Z141 | Z241 | Z341 | Z41 |  |  |  |  |
| RH II | Z150 | Z250 | Z350 | Z450 | Z50 |  |  |  | Z151 | Z251 | Z351 | Z451 | Z51 |  |  |  |
| SSH | Z160 | Z260 | Z360 | Z460 | Z560 | Z60 |  |  | Z161 | Z261 | Z361 | Z461 | Z561 | Z61 |  |  |
| EVP | Z170 | Z270 | Z370 | Z470 | Z570 | Z670 | Z70 |  | Z171 | Z271 | Z371 | Z471 | Z571 | Z671 | Z71 |  |
| ETo | Z180 | Z280 | Z380 | Z480 | Z580 | Z680 | Z780 | Z80 | Z181 | Z281 | Z381 | Z481 | Z581 | Z681 | Z781 | Z81 |

$$Y= A\_{0}+ \sum\_{i=1}^{p}\sum\_{j=0}^{1}a\_{ij}Z\_{ij}+ \sum\_{i\ne 1^{'}=1}^{p} \sum\_{j=0}^{1}a\_{ii'j} Z\_{ii'j}+ cT+e$$

$$Z\_{ij}= \sum\_{w=1}^{m}r\_{iw}^{j} X\_{iw} and Z\_{ii^{'}j }= \sum\_{w=1}^{m}r\_{ii^{'}w}^{j} X\_{iw} X\_{ii^{'}w} $$

Where,

Xw denotes the value of weather variable under study in the wth week. n is the number of weeks in the crop season and Ao, ao, a1 and a2 are model parameters. These models were extended to study combined effects of weather variables and an additional variate T representing the year for time trend. Y is yield; riw/rii'w is the correlation coefficient of yield (adjusted for trend effect) with i-th weather variable (Xiw) /product of i-th and i'-th weather (Xiw/Xi’w) variables in w-th period; m is week of forecast; p is number of weather variables used and e is error term.

In this type of method, for each weather variable, two types of weather indices were developed. First one being the simple values of weather variable during crop growing period [un-weighted index -Zi0] and the second one as weighted [weighted index Zi1]. Weights being taken as correlation coefficients between yield and weather variable in respective periods. In the same way indices were also produced for interaction of weather variables by using weekly products of weather variables taking two at a time. Combination of various weather variables for Weather indices were generated are presented in Table 1. Weather parameters viz. maximum and minimum temperature, morning and evening relative humidity, rainfall, bright sunshine hour and evaporation were used for such model. After development of weather indices empirical models along with machine learning models can be used in Machine learning algorithms.

1. **Crop Prediction using stepwise regression model and regression model**

 In the regression approach, the relationship among yield and various variables affecting it is established through statistical methods like stepwise forward and stepwise backward multivariate analysis. The efficiency of these empirical or regression models mostly depends on the past data set which are used for the model development. The cause-and-effect relationship usually implicit and are used for the estimation of yields. Empirical or regression-built crop models by using long terms yield data and weather is used to predict the crop production in terms of yield and still is being use. Simple regression techniques are widely used alternative to simulation model with long term weather data and crop yield data [52]. The application of empirical models is that it is able to give better insight about the past yield of a targeted region, and also for updating weather-based models weather interactions can be used [53-55].Reference [56] reported good accuracy of pre-harvest district wise rice yield forecast for Bihar using weather data. They also noted that remarkable growth in productivity of wheat is observed in India from the last few decades, so for proper and efficient planning and policy making, crop yield forecasting is a vital tool, which helps to manage excess production.

 Reference [57] studied the particular and combined effect of weather elements on yield of rice of eastern Uttar Pradesh. Also showed that the importance of these variables in crop yield variation. They also reported that bright sunshine hour has more effect on yield predicting tailed by wind velocity and rainfall with R² value of 67.57, 48.63 and 46.74%, respectively. The combined weather parameters such as wind velocity and rainfall, sunshine hour and rainfall, and wind velocity and sunshine hour were found to be more influencing on crop yield modelling with R² value of 82, 63, and 53.8% respectively. Reference [58] reported that 51 to 79% variability in yield of rice and 65 to 92% variability in the wheat yield can be explained through weather-based yield prediction models for the Eastern Uttar Pradesh districts. Reference [59] predicted the first season rice yield by including solar radiation as one of the predictors at Coimbatore, Tamil Nadu. Ten years data from 1988 to 1997 were collected and used for the model development with seven predictors. Stepwise regression analysis was performed for rice yield forecast by MSTAT package. The full model regression without considering variable as solar radiation (Model I) had recorded only R² value to be 0.63, compared to second model which included solar radiation have enhanced the R2 value to 0.94. In the third model they used 7 variables for stepwise regression analysis, and the developed final model which retained only four variables from the subjected input variable with an R2 value of 0.92, this model was able to predicting the Coimbatore’s rice yield.

Reference [60] developed yield prediction model for nine districts of Eastern Uttar Pradesh by using the weather and yield data of eighteen years (1991 to 2008), models were validated only for two years (2009 and 2010). Outcomes directed that model clarified 51 to 79% and 65 to 92% disparities for rice and wheat yield respectively. The percent Mean Bias Error ranged between -1.05 (Mau) to 6.17 (Mirzapur) districts for rice and from -6.56 (Mau) to 0.01 (Varanasi) in case of wheat crop. The percent Root Mean Square Error was in the range of 6.87 (Jaunpur) to 11.60 (Sant Ravidas Nagar) for rice and in case of wheat it was 5.52 (Mirzapur) to 11.11 (Mau). Thus, they reported that the models can be applied to some extent for forecasting the yield in different districts of Eastern Uttar Pradesh.

Reference [61] used discriminant function and Multiple Linear Regression (MLR) techniques for estimating wheat productivity for the district Varanasi in Uttar Pradesh. He concluded that stepwise multiple linear techniques can be used effectively for the pre-harvest wheat crop, which are more consistent in performance. Incorporating statistical indicators enhanced the precision of forecasting of wheat crop yield for both Adjusted R2 and RMSE values. Similar work has also been done by several researchers viz. reference [62] to forecast rice yield, developed pre-harvest model only after about two months of sowing. Reference [63] reported that reliable forecasting of wheat yield could be obtained when the crops were at twelve week i.e. only about two months before harvest. This study was conducted on Vindhyanchal Plateau of Madhya Pradesh for wheat yield forecast.) used discriminant function analysis intended for developing wheat yield forecasting models for Kanpur. This methodology gave very trustworthy yield forecast about two months beforehand harvest.

Multiple linear regressions are very much appropriate for the short or intermediate term forecasting, using weather variables, weighted and unweighted weather indices to generate multiple linear regression forecasting models [64-66]. Reference [67] predicted wheat yield and wheat quality using weather predictors, using regression models to explain the consequences of weather on wheat yield, protein and test weight. These variables include temperature and precipitation for growing period of wheat development stages. The forecasting efficacy of the models were boosted by adding a spatial lag effect. Wheat yield, test weight and protein showed strong correlation with the weather. Meant for developing models for the prediction of yield at different levels like district, state or national level, time series of crop yields are utilized. Trend analysis of the input-output data and Auto Regressive Integrated Moving Average (ARIMA) analysis technique is commonly used to develop model [68]. For modelling time series data gathered over the long run, ARIMA strategy is by and large utilized. One impediment of this approach is that the time arrangement under thought ought to be stationary or ought to be fit for turning out to be so by method for differencing or detrending. Reference [69] predicted the sugarcane yield using the ARIMA model.

1. **Yield estimation Models using machine learning**

Time series is a process to analyse time on parametric series data to mine significant statistics and additional features of the data to transform into an information. Time series forecasting is the model to foresee upcoming values based on formerly observed data. New idea of crop yield in average climate situations is being developed and it is applied in time series methods on the older yield data to set up a predicting model. Agricultural organization needs simple and precise estimation techniques for the prediction or estimation of rice yields in the scheduling process [70]. Compulsion was to recognize whether artificial neural network (ANN) could successfully predict rice yield for characteristic climatic circumstances of the mountainous region, reference [71] assess performance of ANN model relative to the variations of parameters and compared the efficiency of multiple linear regression with ANN models. The Generalized Regression Neural Networks (GRNN) process was used for predicting production of given crop [72]. They reported that GRNN is a suitable technique for prediction grain production. It was stated that GRNN is appropriate for multi-objectives, non-linear, and multivariate predicting. Assessment of modified k-Means clustering set of rules in crop prediction was validated and the evaluation indicated the assessment of modified k-Means over k-Means and-Means++ clustering algorithm and also stated that the modified k-Means has attained the highest number of great superiority clusters along with accurate prediction of crop and greatest precise count [73]. Model was developed for predicting yield of the sugarcane by means of fortnightly weather variable like average daily maximum temperature and minimum temperature, relative humidity in the morning and evening and total rainfall and the yield data in Coimbatore district [74]

1. **Computer software for machine learning**

Machine learning methods can be performed in MATLAB, PYTHON, R software but R is one of the most preferred. It makes statistical computing very easy also graphs are easy to plot and depict in R. Advance statistical and machine learning packages are provided in R software along with various other packages and in-built functions which makes statistical analysis very easy. It provides plots, effective data handling in huge amount and storage facility depending on or interest and use. R is very much helpful in predictive analytics, data pre-processing, statistical modelling, data visualization and deployment [75].

1. **Yield estimation Models using hybrid machine learning**

Hybrid machine learning involves use of two machine learning algorithms in combination to each other for example SVM-LASSO, this includes to two techniques in combination to each other. Reference [76] used two machine learning algorithms boosted Support Vector Machines (SVM) and Regression Trees (BRT) to predict regional winter wheat yields. The models are created through Normalized Difference Vegetation Indices (NDVI) drawn from SPOT vegetation imagery. Three types of NDVI-related forecasters were used: Single NDVI, Targeted NDVI and Incremental NDVI. BRT and SVM were initially used to select features with good significance for predicting the yield. After feature selection, BRT and SVM models were applied to the subset of selected features for yield forecasting. BRT outperform SVM. Forecast of rice yield through ARIMAX and proposed hybrid models using weather variables. Two hybrid approaches like ARIMAX-ANN and ARIMAX-SVM have been used for the rice yield along with weather variables of Aligarh in Uttar Pradesh. Based on the results obtained, performance of ARIMAX-SVM and ARIMAX-ANN models are close to each other but much superior to the conventional ARIMAX model for the considered data set. Performance of hybrid ARIMAX model was found to be quite encouraging. [77].

 Reference [78] determined yield of rice by utilizing long-term weather data with six different statistical methods. These included stepwise multiple regression alone and in alliance with principal component analysis, similarly ANN individually and in combination with principal component analysis, elastic net (ENET) and least absolute shrinkage and selection operator (LASSO) for developing regression equations for calibration of model along with validating the accuracy of the model. During the calibration R2 and root mean square error ranged between 0.22–0.98 and 24.02–607.29 kg ha-1 respectively. Independent dataset for validation resulted with the RMSE as 21.35–981.89 kg ha-1 and normalized root mean square error (nRMSE) as 0.98–36.7%. Ranking of the models depicted that LASSO (2.63) was the finest model, next best was ENET (3.07) while PCA-ANN (4.19) was the poorest model which was observed significant at p <0.001. Due to the inhibition of overfitting and reducing the magnitude of regression coefficient by penalization decreases the model complexity, LASSO showed good performance. Pair-wise multiple comparison test was showed that LASSO was the finest model alike to SMLR and ENET.

1. **Crop yield prediction model using LASSO**

LASSO is an attractive method because it improves the quality of prediction by shrinking regression coefficient, when compared to prediction models fitted through unpenalized maximum likelihood methods. Reference [79] suggested LASSO for utilizing in the crop yield forecasting technique. LASSO curtails residual sum of squares focus to the sum of the absolute value of the coefficient being less than a constant. This characteristic enables to produce some coefficient that are exactly zero and henceforth gives easier understandable models. It produces significant models like subset selection and shows the stability of even the ridge regression. The idea of LASSO is relatively general and can be practical in variety of statistical models. Majority of the forecasting models in the past few decades have been using only multiple linear regressions (MLRs) for developing crop yield prediction [80-82]. Multiple linear regression has the biggest disadvantage of over-fitting when quantity of samples is less than the quantity of variables. Also, another disadvantage is the multi-collinearity when independent predictors are correlated [83]. To combat these demerits, feature selection methods in statistical analysis like Stepwise multiple linear regression (SMLR), least absolute shrinkage and selection operator, machine learning statistical technique can be adopted.

1. **Crop yield prediction using SVM**

Support Vector Machine (SVM) first perceived in 1992 and was familiarized by Boser, Guyon, and Vapnik. SVMs are supervised learning methods applied for regression and classification. They are the type of classification of generalized linear classifiers, or in new terms it is a regression prediction and classification means that practices machine learning theory to exploit analytical accuracy while spontaneously escaping over-fitting to the data. The SVM can be used equally for grouping and regression problems and it be able to indicate as a two-layered network where in first layer the weights are non-linear while in second layer it is linear [84]. Support Vector Regression (SVR) is used in texts and in common language to describe the regression with SVM.

 Support Vector machines can be demarcated as organizations which use hypothesis space of a linear functions in a high dimensional feature space. It is trained through a learning algorithm from optimization theory that instigates a learning bias resulting from statistical learning theory. It was firstly widespread with the NIPS community and now is a dynamic part of the machine learning research everywhere in the world. SVM becomes well-known when, using pixel maps as an input gives accuracy comparable to ultra-modern neural networks with expounded features in a calligraphy recognition. Nowadays it is moreover being operated in many applications, such as face analysis, hand writing analysis, regression-based applications and pattern classification. The fundamentals of Support Vector Machines (SVM) have been elaborated by Vapnik and gained admiration due to countless promising and advanced features such as improved empirical performance, Structural Risk Minimization (SRM), which has been shown to be exclusive to traditional Empirical Risk Minimization (ERM) principle, used by conventional neural networks. SRM curtails an upper bound on the predictable risk, on the other hand ERM diminishes the faults on the training data. It is this modification which enables the SVM with a greater capability to oversimplify, which is the objective in statistical learning. SVM was developed to resolve the problem of classification, but in recent times they have been stretched to explain regression problems [85].

The generalized support vector machines (SVMs) have a two-stage neural network design. In the first stage, self-organizing feature map (SOM) is used as a clustering algorithm to divide the whole input space into several incoherent regions. A tree-structured architecture is taken in the partition to avoid the difficulty of predetermining the figure of partitioned regions. In the second stage, multiple SVM that optimum fit partitioned regions are prepared by concluding the most applicable kernel function and the optimal free parameters of SVMs. The simulation displays that it accomplishes significant upgrading in the generalization performance in appraisal with the single models It is established on the distinctive principle of the structural risk minimization principle to evaluate functions by decreasing an upper bound of the generalization error, they are made known to be very resistant to the over-fitting problem, ultimately succeeding high generalization performance in resolving various time series estimating problems.

Another crucial property of SVMs is that training SVM is alike to solving a linearly quadratic programming problem so that the solution of SVMs is always exclusive and optimal. Distinct to other networks, training necessitates non-linear optimization with the risk of getting trapped into local minima. In the modelling of time series, two of the vital problems are non-stationarity and noise. The non-stationarity infers that the time series change their dynamics between diverse regions. This will way to gradual variations in the reliance between the input and output variables. The noisy characteristic denotes the inaccessibility of comprehensive information from the historical behaviour of the time series to entirely capture the dependency between the past and the future. The noise in the data could lead to the under-fitting or over-fitting problem. It is applied to construct nonlinear nonparametric forecasting models to be used in Crop yield forecast models for spring wheat, barley and canola grown on the Canadian Prairies were developed taking vegetation indices resulting from satellite data machine learning approaches [86].

1. **Artificial neural network (ANN) for yield monitoring**

A feed-forward back-propagating ANN structure was used to develop sugarcane yield monitoring. The ANN was developed by a supervised learning procedure in R using the package neuralnet [87]. The package contains a flexible function to train feed-forward neural networks. A minimum of three layers is required in ANN: the input, hidden, and output layers (Fig. 1). The input layer contains neurons that correspond to the input variables for monitoring the yield. The output layer contains one neuron that corresponds to the value of monitoring yield. In a feed-forward neural network, the neurons in one layer are connected to the neurons in the next layer and the information flows forward from the input to the output through the hidden layers [88]. An ANN can have zero or more hidden layers. The data move between the layers across weighted connections in one direction, from the input through the hidden to the output layers.



**Figure 1: figure showing input, hidden and output layer of ANN**

 A neuron accepts data from the previous layer and calculates a weighted sum of all its inputs:

ti = $\sum\_{j=1}^{n}wij xj $

where ti is the weighted sum of all its i inputs, n is the number of inputs, w is the weight of the connection between neurons i and j, and × is the input from neuron j. A transfer function (equation given below) is applied to the weighted value to calculate the neuron output:

$$oi=f(ti)$$

where oi is the neuron output and f (ti) is a transfer function applied to the weighted value t of i inputs. The most used transfer function is the sigmoidal function for the hidden and output layers, and a linear transfer function is commonly used for the input layer. The number of hidden neurons determines the number of connections between inputs and outputs and may vary depending on the specific problem under study [89]. If too many neurons are used, then the ANN may become over-trained, causing it to memorize the training data resulting in poor result [90]. In the recent scenario of prediction of yield and weather forecast, Artificial Neural Network (ANN) getting great deal of attention. Complex problems are solved by this method even if the quality of data is less or lack of precision in data. There are numerous introductory works on ANNs has been done by many workers, reference [91] has done a well detailed study of neural network models vis-a-vis traditional statistical models. They have shown that some statistical methods including regression, principal component analysis, density function and statistical image analysis can be given neural network expressions. Reference [92] reviewed the relevant literature on neural networks, clarified the learning algorithm and made a comparison between regression and neural network models in terms of notations, terminologies and implementation. Reference [93] given the over-all summary of the work in ANN forecasting, providing the procedures for neural network modeling, general paradigm of the ANNs especially those used for forecasting. The comparative performance of neural networks is studied with traditional statistical methods, majority of the studies concluded with better performance of ANN over traditional method. For determining the quality in prediction and robustness to deviation of multilayer perceptron to linear regression, showed relative performance of Multilayer perceptron (MLP) found better than linear regression. ANNs, data driven and self-adaptive methods, are working based on prior assumptions of model. On the basis of examples, subtle functional relationships among the data are captured even if the underlying relationships are unknown or hard to describe. ANNs can identify and learn correlated patterns between input data sets and corresponding target values through training. After training, ANNs can be used to predict the outcome of new independent input data and have great capacity in predictive modelling, i.e., all the characters relating the unknown condition are uploaded to the trained ANNs, and then prediction of yield may be possible.

Artificial neural network (ANN) by name itself says that it is a network of artificial neurons which follows the concept of function as in human brain neurons. The structure of artificial neural network consists of several layers of processing units /neurons/ nodes [94]. Modeling with ANN involves two important tasks, namely, topology and learning algorithm of network. The first task topology of a networks comprises (a) fixing the number of layers, (b) the number of neurons present for each layer, (c) the node function for each neuron, (d) feedback or feedforward method, and (e) the pattern of connection between the neurons through layers. All these alterations are should be considered for improved performance of the system. In the learning phase deals with weight adjustments as well as threshold values [95]. Usually, the data is divided into three non-overlapping sets: the so-called training, validation and testing set. The training set, consisting lager portion of data, is used to teach the network in order to get the desired target function. Then the validation set is used to decide when to stop training process, to avoid over fitting, a situation where the network memorizes the training data rather than learning the law that governs them. The testing data set, which exposed to the unseen data, is used to measure performance of trained network by mean square error (MSE) or root mean square error (RMSE). Neural Network architectures were developed by using Levenberg Marquardt (LM) Algorithm as a training algorithm of weight matrix.

1. **CONCLUSION**

Weather variables affect the crop differently during different stages of development. Weather is dynamic, continuous and multi-dimensional, these unfavorable properties make yield prediction a challenging task for developing the model for crop yield forecast. Thus, there is need to develop crop yield forecast model based on weather variables so that consistent forecasts can be obtained. In recent times, there is a rising demand for applying machine learning techniques in agriculture, as there is a wealth of data available from various sources that can be analyzed to uncover valuable insights. This field of research is continuously evolving and is anticipated to expand further in the future. The convergence of computer science and agriculture plays a crucial role in predicting agricultural crop outcomes. It is essential to develop an objective approach for forecasting crop yields before harvest. By constructing an appropriate model, several advantages can be gained compared to traditional forecasting methods. Application of machine learning and hybrid machine learning for yield monitoring and prediction can be done to have the better understanding of time series data with acuuarcy and faster computational efficiency.

**REFERENCES**

1. Wassmann R, Jagadish S.V.K., Surnfleth, K., Pathak, H., Howell, G., Ismail, A,,Serraj, R., Redoña, E., Singh, R.K.and Heuer, S. (2009) Regional vulnerability of climate change impacts on Asian rice production and scope for adaptation. Adv Agron 102:99–133
2. Rimi RH, Rahman SH, Karmakar S and Hussain SG (2011) Trend analysis of climate change and investigation on its probable impacts on rice production at Satkhira, Bangladesh. Pak J Meteorol 6:37–50
3. Ji, B., Sun Y, Yang S and Wan, J. (2007) Artificial neural networks for rice yield prediction in mountainous regions. J Agric Sci 145:249–261
4. Tabari H, Kisi, O, Ezani, Hosseinzadeh, A. and Talaee, P. (2012) SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. Journal of Hydrology. 444–445:78–89
5. Agrawal, R. Jain, R.C. and Mehta, S.C. (2001). Yield forecast based on weather variables and agricultural inputs on agro-c climatic zone basis. Indian Journal of Agricultural Sciences 71(7):487-490
6. Sukhadia, K., and Chaudhari, M. B. (2019). A Survey on Rice Crop Yield Prediction in India Using Improved Classification.
7. Ekaputa, U.N. (2004). Agricultural systems. Port Harcourt: UNIK Publishers. pp. 126-134.
8. Orkwor, G.C. and Ekanayake, I.J. (1998). Growth and Development. In: Orkwor, G.C., Asiedu, R. and Ekanayake, I.J. (Eds).The progress in yam Research. Ibadan, Nigeria: IITA. pp. 39-62.
9. Kaur, R. and Rajni (2010). Impact of increasing CO2 on crop productivity. Indian Farming. June issue. 23-7.
10. Singh RS, Patel C, YadavMK and Singh KK (2014) Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. Journal of. Agrometeorology 16:199– 202
11. Sikder, S. (2009). Accumulated heat unit and phenology of wheat cultivars as influenced by late sowing heat stress condition. Journal of Agriculture & Rural Development 7(1&2).57-64.
12. Girijesh, G.K., Kumara wamy, A.S., Sridhara, S., Dinesh Kumar, M., Vageesh, T.S and Nataraju, S.P. (2011). Heat use efficiency and helio-thermal units for maize genotypes as influenced by dates of sowing under southern transitional zone of karnataka state. International Journal of Science and Nature 2(3): 529-533
13. Maduakor, H.O., Lal, R. and Opara-Nadi, O.A. (1984). Effects of methods of seedbed preparation and mulch on the growth and yield of white yam (Dioscorea rotundata L.) on an ultisol in south -east Nigeria. Field Crop Research 9: 119-130.
14. Okoh, C.A. (2004). The effect of mulching on soil physico-chemical properties and the yield of white yam. Tropical Journal of Root Tuber Crops 4(2): 24-31.
15. Hundal, S.S., Singh, R. and Dhaliwal, L.K. (1997). Agro-climatic indices for predicting phenology of wheat (Triticum aestivum) in Punjab. Indian Journal of Agricultural Sciences 67 (6): 265-8.
16. Singh, S. (1996). Growth and yield response of different crop species to low light and high temperature-humidity stress. Indian Journal of plant Physiology 2:151-155.
17. Benbi, D.K. (1994). Prediction of leaf area indices and yield of wheat. Journal of Agricultural Sciences Cambridge 122: 13-20.
18. Tripati, P., Tomar, S. K. and Singh, A.K. (1999). Crop weather models to predict the growth and yield of mustard and wheat under mustard-wheat cropping system. In: Proceedings of the National Workshop on Dynamic Crop Simulation Modelling for Agro-met Advisory Services, National Center for Medium range Forecasting, New Delhi. pp. 285
19. Dwyer, L.M. and Stewart, D.W. (1986). Leaf area development in field grown maize. Journal of Agronomy 78: 334-348.
20. Wang, J.Y. (1960). A critique of the heat units approach to plant response studies. Ecology 41: 785-90.
21. Bonhomme, R. (2000). Beware of comparing RUE values calculated from PAR vs. solar radiation or absorbed vs. intercepted radiation. Field Crops Research 68: 247-252.
22. Wurr, D.C.E., Fellows, J.R. and Phelps, K. (2002). Crop scheduling and prediction – Principles and opportunities with field vegetables. Advances in Agronomy76: 201-234.
23. Roy, S., Meena, R.L., Sharma, K.C., Kumar, V., Chattopadhyay, C., Khan, S.A. and Chakravarthy, N.V.K. (2005).Thermal requirement of oilseed Brassica cultivars at different phenological stages under varying environmental conditions. Indian J. Agric. Sci. 75(11): 717-21.
24. Phadnawis, N.B. and Saini, A.D. (1992). Yield models in wheat based on sowing time and phonological development. Annals of Plant Physiology 6: 52-59.
25. Tewari, S.K. and Singh, M. (1993).Yielding ability of wheat at different dates of sowing: a temperature development performance. Indian Journal of Agronomy 38: 204-209.
26. Slafer, G.A. and Rawson, H.M. (1994). Sensitivity of wheat phasic development to major environmental factors: A re-examination of some assumptions made by physiologists and modellers. Australian Journal of Plant Physiology 21: 393-426.
27. Villordon, A., Clark, C., Ferrin, D., and LaBonte, D. (2009). Using growing degree days, agrometeorological variables, linear regression, and data mining methods to help improve prediction of sweetpotato harvest date in Louisiana. HortTechnology, 19(1), 133-144.
28. Bazgir, S., Kamali, G., and Mortazavi, A. (2007). Wheat yield prediction through agrometeorological indices for Hamedan, Iran
29. Masle, J., Doussinault, G., Farquhar, G.D. and Sun, B. (1989). Foliar stage in wheat correlates better to photothermal time than to thermal time. Plant Cell Environ.12:235-247.
30. Rao, V.U.M., Singh, D. and Singh, R. (1999). Heat use efficiency of winter crops in Haryana. Journal of Agrometeorology 1(2): 143-8.
31. Baier, w.(1981). Crop-weather analysis models.In: Applicaions of remote sensing to agricultural production forecasting (ed breg A.), Joint research Centre, ISPRA, Italy. PP105-118.
32. Walker, G. K. (1989). Model for operational forecasting of western Canada wheat yield. Agricultural and Forest Meteorology, 44(3-4), 339-351.
33. Fisher, R.A. (1924). The influence of rainfall on the yield of London wheat at Rothamsted. Phil. Trans. Roy. Soc.,B213,89-142.
34. Hendrick,W.A.and Scholl.J.e. (1943).Technique in measuring joint relationship the joint effects of temperature and precipitation on crop yield. North Carolina Agric. Exp.Sta. Tech.Bull., 74.
35. Baier, w.(1981). Crop-weather analysis models.In: Applicaions of remote sensing to agricultural production forecasting (ed breg A.), Joint research Centre, ISPRA, Italy. PP105-118.
36. Ramakrishna, Y.S., Singh, H. P. and Nageswara, R.G. (2003). Weather based indices for forecasting food grain production in India. Journal of. Agrometeorology,5(1):1-11.
37. Dhekale, B.S., Mahdi SS, Dalvi TP, and Sawant PK (2014) . Forecast models for groundnut using meteorological variables in Kolhapur, Maharashtra. Journal of Agrometeorology 16:238–239
38. Kumar, N., Pisal, R., Shukla, S.P. and Pandye, K.K. (2014) Regression technique for South Gujarat. Mausam ,65:361–364
39. Rai, K.K., N.P.V., Bharti, B.V.S. and SA K (2013) Pre -harvest forecast models based on weather variable. Adv Biores 4:118–122
40. Verma ,U., Piepho, H.P. and Goyal, A. et al (2016) .Role of climatic variables and crop condition term for mustard yield prediction in Haryana. Int J Agric Stat Sci 12:45–51
41. Kandiannan, K., K. K. Chandaragiri, N. Sankaran, T. N. Balasubramanian and C. Kailasam (2002). Crop-weather model for turmeric yield forecasting for Coimbatore District, Tamil Nadu, India. Agricultural and Forest Meteorology, 112, 133- 137.
42. Frere, M. and Popov, G.F.(1979). Agrometeorological crop monitoring and forecasting. Plant Production and Protection Division, no. 17. Food and Agriculture Organization of the United Nations,Rome, p. 64.
43. Kazmi, D. H., and Rasul, G. (2012). Agrometeorological wheat yield prediction in rainfed Potohar region of Pakistan. Agricultural Sciences, 3(02), 170.
44. Ghosh, K., Balasubramanian, R., Bandopadhyay, S.,Chattopadhyay, N., Singh, K. K. and Rathore, L. S.(2014). Development of crop yield forecast models under FASAL – a case study of kharif rice in West Bengal. Journal of Agrometeorology, 16 (1): 1-8.
45. Rajegowda, M. B., Soumya, D. V., Padmashri, H.S., Janardhana Gowda, N. A. and Nagesha, L. (2014). Ragi and groundnut yield forecasting in Karnataka-statistical model, Journal of. Agrometeorology, 16(2): 203-206.
46. Kumari, Prity., Mishra, G. C., Pant, A. K., Shukla, G. and Kujur, S. N. 2014b. Comparison of forecasting ability of different statistical models for productivity of rice (oryza Sativa l.) in India. The Ecoscan. 8(3&4): 193 -198.
47. Hendrick,W.A.and Scholl.J.e. (1943).Technique in measuring joint relationship the joint effects of temperature and precipitation on crop yield. North Carolina Agric. Exp.Sta. Tech.Bull., 74.
48. Agrawal, R., Jain, R.C., Jha, M. P. and Singh, D. (1980): Forecasting of rice yield using climatic variables, Indian Journal of Agricultural Sciences, 50(9):680-684.
49. Agrawal, R. and Jain, R.C. (1982). Composite model for forecasting rice yield, Indian Journal of Agricultural Sciences, 52(3): 189-194
50. Jain, R.C., Agrawal, R. and Jha, M.P. (1980). Effect of climatic variables on rice yield and its forecast. Mausam,31(4):591-596
51. Agrawal, R., Jain, R.C. and Jha, M.P. (1986). Models for studying rice crop weather relationship. Mausam, 37 (1): 67-70.
52. Shi W, Tao F and Zhang Z (2013) A review on statistical models for identifying climate contributions to crop yields. J Geogr Sci 23:567–576.
53. Lobell, D. B. and Burke, M. B. (2010) .On the use of statistical models to predict crop yield responses to climate change. Agric For Meteorol 150:1443–1452.
54. Lobell DB, Schlenk er W and Costa-Roberts J (2011) Climate trends and global crop production since 1980. Science 333:616–620.
55. Basso,B., Cammarano, D. and Carfagna, E. (2013). Review of crop yield forecasting methods and early warning systems. In: Intergovernmental panel on climate change (ed) climate change 2013 - the physical science basis. Cambridge University Press, Cambridge, pp 1–30.
56. Dutta S, Patel N.K, and Srivastava, S.K. (2001) District wise yield models of rice in Bihar based on water requirement and meteorological data. Journal Indian Society Remote Sensing 29:175–181
57. Pandey, K.K., Rai, V.N., Sisodia, B.V.S. and Singh,S.K. (2015) Effect of weather variables on rice crop in eastern Uttar Pradesh, India. Plant Arch 15: 575–579.
58. Singh RS, Patel C, YadavMK and Singh KK (2014) Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. Journal of. Agrometeorology 16:199– 202
59. Kandiannan, K., K. K. Chandaragiri, N. Sankaran, T. N. Balasubramanian and C. Kailasam (2002). Crop-weather model for turmeric yield forecasting for Coimbatore District, Tamil Nadu, India. Agricultural and Forest Meteorology, 112, 133- 137.
60. Singh RS, Patel C, YadavMK and Singh KK (2014) Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. Journal of. Agrometeorology 16:199– 202
61. Garde, Y. A., Dhekale, B. S. and Singh, S. (2015). Different approaches on pre harvest forecasting of wheat yield, Journal of Applied and Natural Science ,7 (2): 839 – 843.
62. Jain, R.C., Agrawal, R. and Jha, M.P. (1980). Effect of climatic variables on rice yield and its forecast. Mausam,31(4):591-596
63. Agrawal, R. Jain, R.C. and Mehta, S.C. (2001). Yield forecast based on weather variables and agricultural inputs on agro-c climatic zone basis. Indian Journal of Agricultural Sciences 71(7):487-490
64. Kumar R., Gupta B.R.D., Athiyaman B., Singh K.K. and Shukla R.K. (1999).Stepwise regression technique to predict pigeonpea yield in Varanasi district, Journal of Agrometeorology, 1(2): 183-186.
65. Agrawal, R. and Mehta, S.C. (2007). Weather based forecasting of crop yields, pests and diseases-IASRI models. J. Indian.Society. Agricultural. Statistics., 61(2): 255-263.
66. Chauhan, V.S., Shekh, A.M., Dixit ,S.K., Mishra, A.P. and Kumar, S. (2009), Yield prediction model of rice in Bulsar district of Gujarat, Journal of. Agrometeorology, 11 (2): 162-168.
67. Lee, Byoung-Hoon, Phil Kenkelb and Wade Brorsenb, B. (2013). Pre-harvest forecasting of county wheat yield and wheat quality using weather information. Agricultural and Forest Meteorology, 168, 26– 35
68. Box, G.E.P. and Jenkins, G. (1970). Time series analysis, Forecasting and control, Holden-Day, San Francisco, CA.
69. Suresh, K.K. and Krishna Priya, S.R.(2011). Forecating Sugarcane Yield of Tamilnandu Using ARIMA Models. Sugar Tech., 13(1): 23-26.
70. Ji, B., Sun Y, Yang S and Wan, J. (2007) Artificial neural networks for rice yield prediction in mountainous regions. J Agric Sci 145:249–261
71. Dahikar, S. S., and Rode, S. V. (2014). Agricultural crop yield prediction using artificial neural network approach. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 2(1), 683-686.
72. Jin, M., and Jin, C. (2008, September). Forecasting agricultural production via generalized regression neural network. In 2008 IEEE Symposium on Advanced Management of Information for Globalized Enterprises (AMIGE) (pp. 1-3). IEEE.
73. Narkhede, U. P., and Adhiya, K. P. (2014). Evaluation of Modified K-Means Clustering Algorithm in Crop Prediction. International Journal of Advanced Computer Research, 4(3), 799.
74. Suresh, K. K., and Krishna Priya, S. R. (2009). A study on pre-harvest forecast of sugarcane yield using climatic variables. Stat. Appl. 7&8 (1&2)(New Series).
75. Team, R. C. (2013). R: A language and environment for statistical computing.
76. 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.
77. Alam, W., Ray, M., Kumar, R. R., Sinha, K., Rathod, S., and Singh, K. N. (2018). Improved ARIMAX modal based on ANN and SVM approaches for forecasting rice yield using weather variables. Indian Journal of Agricultural Sciences, 88(12), 101-105.
78. Das, B., Nair, B., Reddy,K.V. and Venkatesh, P.(2018). Evaluation of multiple linear, network and penalised regression models for prediction of rice yield based on weather parameters foe west cost of India. Int J Biometerol 1583-6
79. Tibshirani, R. (1996). Regression shrinkage and selection via lasso. J Roy StatSoc B, 58:267–288.
80. Rai, K.K., N.P.V., Bharti, B.V.S. and SA K (2013) Pre -harvest forecast models based on weather variable. Adv Biores 4:118–122
81. Dhekale, B.S., Mahdi SS, Dalvi TP, and Sawant PK (2014) . Forecast models for groundnut using meteorological variables in Kolhapur, Maharashtra. Journal of Agrometeorology 16:238–239
82. Kumar, N., Pisal, R., Shukla, S.P. and Pandye, K.K. (2014) Regression technique for South Gujarat. Mausam ,65:361–364
83. Verma ,U., Piepho, H.P. and Goyal, A. et al (2016) .Role of climatic variables and crop condition term for mustard yield prediction in Haryana. Int J Agric Stat Sci 12:45–51
84. Bray, M., and Han, D. (2004). Identification of support vector machines for runoff modelling. Journal of Hydroinformatics, 6(4), 265-280.
85. Jakkula, V. (2006). Tutorial on support vector machine (svm). School of EECS, Washington State University, 37.
86. 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.Pantazi, X. E., Moshou,
87. Günther, F., and Fritsch, S. (2010). Neuralnet: training of neural networks. R J., 2(1), 30.
88. Cherukuri, H., Perez-Bernabeu, E., Selles, M., and Schmitz, T. (2019). Machining chatter prediction using a data learning model. Journal of Manufacturing and Materials Processing, 3(2), 45.
89. Kaul, M., Hill, R. L., and Walthall, C. (2005). Artificial neural networks for corn and soybean yield prediction. Agricultural Systems, 85(1), 1-18.
90. Lawrence, J. (1994). Introduction to neural networks. Nevada City.
91. Cheng, B., and Titterington, D. M. (1994). Neural networks: A review from a statistical perspective. Statistical science, 2-30.
92. Warner, B., and Misra, M. (1996). Understanding neural networks as statistical tools. The american statistician, 50(4), 284-293.
93. Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. International journal of forecasting, 14(1), 35-62.
94. Haykin, S. 2001. “Neural Networks – A Comprehensive Foundation”. IEEE Press, New York.
95. Hagan , M. T. and Menhaj, M. (1994). Training feedforward networks with the Marquardt algorithm, IEEE Trans on Neural Networks, 5(6): 989–993.