IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# CROP PREDICTION BASED ON CHARACTERISTICS OF THE AGRICULTURAL ENVIRONMENT USING VARIOUS FEATURE SELECTION TECHNIQUES AND CLASSIFIERS

**[1]Mrs Dr. J. Sarada**
[1]Associate Professor
[1]Department of Computer Applications
[1]Chadalawada Ramanamma Engineering College College
(Autonomous),Tirupathi

**[2] Bovilla Sravani**
[2] Student
[2] Department of Computer Applications
[2] Chadalawada Ramanamma Engineering
(Autonomous), Tirupathi

## ABSTRACT

*Agriculture is a growing field of research. In particular, crop prediction in agriculture is critical and is chiefly contingent upon soil and environment conditions, including rainfall, humidity, and temperature. In the past, farmers were able to decide on the crop to be cultivated, monitor its growth, and determine when it could be harvested. Today, however, rapid changes in environmental conditions have made it difficult for the farming community to continue to do so. Consequently, in recent years, machine learning techniques have taken over the task of prediction, and this work has used several of these to determine crop yield. To ensure that a given machine learning (ML) model works at a high level of precision, it is imperative to employ efficient feature selection methods to preprocess the raw data into an easily computable Machine Learning friendly dataset. To reduce redundancies and make the ML model more accurate, only data features that have a significant degree of relevance in determining the final output of the model must be employed. Thus, optimal feature selection arises to ensure that only the most relevant features are accepted as a part of the model. Conglomerating every single feature from raw data without checking for their role in the process of making the model will unnecessarily complicate our model. Furthermore, additional features which contribute little to the ML model will increase its time and space complexity and affect the accuracy of the model's output. The results depict that an ensemble technique offers better prediction accuracy than the existing classification technique.*

***Keywords***

*Crops, Zigbee, Monitoring, Soil, Temperature sensors, Security, Data models*

## 1.INTRODUCTION

Crop prediction in agriculture is a complicated procesand multiple models have been proposed and tested to this end. The problem calls for the use of assorted datasets, given that crop cultivation depends on biotic and a biotic factors Biotic factors include those elements of the environment that occur as a result of the impact of living organisms (microorganisms, plants, animals, parasites, predators, pests), directly or indirectly, on other living organisms. This group also

includes anthropogenic factors (fertilization, plant protection, irrigation, air pollution, water pollution and soils, etc.). These factors may contribute to the occurrence of many changes in the yield of crops, cause internal defects, shape defects and changes in the chemical composition of the plant yield. The shaping of the environment as well as the growth and quality of plants is influenced by a biotic and biotic factors .A biotic factors can be divided into physical, chemical, and other. The recognized physical factors include: mechanical vibrations (vibration, noise), radiation (e.g., ionizing, electromagnetic, ultraviolet, infrared); climatic conditions (atmospheric pressure, temperature, humidity, air movements, sunlight); soil type, topography, soil rockiness, atmosphere, and water chemistry, especially salinity. The chemical factors include: priority environmental poisons, such as sulfur dioxide and derivatives, PAHs; nitrogen oxides and derivatives, fluorine, and its compounds, lead and its compounds, cadmium and its compounds, nitrogen fertilizers, pesticides, carbon monoxide. The others are: mercury, arsenic, dioxins and furans, asbestos, and a atoxins A biotic factors also include bedrock, relief, climate, and water conditions - all of which affect its properties. Soil-forming factors have a diversified effect on the formation of soils and their agricultural value.

Predicting crops yields is neither simple nor easy. The methodology for predicting the area under cultivation is, according to Myers and Muriithi a set of statistical and mathematical techniques useful in an evolving and improving optimization process. It also has important uses in design, development, and formulation new as well as improving existing products. Presentation or performance of statistical analysis requires the possession of numerical data. Based on them, conclusions are drawn as to various phenomena and further, on this basis, binding economic decisions can be made. According to Muriithi [6], the better you describe certain phenomena in terms of numbers, the more you can say about them, and with increasing data accuracy you can also obtain more accurate information and make more accurate decisions.

The biggest problem in the temperate climate zone is assessment of agro climatic factors in terms of shaping the yield of winter plant species, mainly cereals. The key factor influencing wintering yield, which provides access to days with a temperature over of 5_ C, their number and frequency, and the number of days in the wintering period with temperatures above 0_C and 5_C. A number of these can be estimated on the basis of public statistics and yield regression statistics in years. Developed models for checking the situation that assess whether they want to be a probation of state policy in the field of intervention in the cereal market. Efficient forecasting of productivity requires forecasting of agro meteorological factors.

Aspects related to the variability of these factors may pose a particular problem.Many researchers have dealt with this issue with varying degrees of success .

Grabowska predicted narrow-leaf lupine yields for 2050-2060 using weather models and three climate change scenarios for Central Europe: E-GISS model, HadCM3 and GFDL. The _t of the models was assessed by means of the determination coefficient R2, corrected coefficient of determination R2adj, standard error of estimation and the coefficient of determination R2pred calculated using the Cross Validation procedure. The selected equation was used to forecast lupine yield under the conditions of doubling the CO2 content in the atmosphere. These authors stated that the influence of meteorological factors on the yield of narrow-leaved lupine varied depending on the location of the station. The temperature (maximum, average, minimum) at the beginning of the growing season, as well as rainfall during the _owering - technical maturity period, most often had a significant influence on the yield. It has been shown that the predicted climate changes will have a positive effect on the lupine yield. The simulated protability was higher than observed in 1990-2008, and HadCM3 was the most favorable scenario.

D¡browska-Zieli«ska assessed the usefulness of plant biophysical parameters, calculated from the ranges of rejected electromagnetic radiation recorded by the new generation satellites Sentinel-2 and Proba-V, for forecasting crop yields in Poland. In 2016-2018, ground measurements were carried out in arable fields in the area included in the global crop monitoring network GEO Joint Experiment of Crop Assessment and Monitoring JECAM. Classification of crops was performed using optical and radar images Sentinel-1 and RadarSat-2. The prototypical model of Biomass and Evapo transpiration PRO was used to simulate the growth of winter wheat cultivation, to forecast its biomass size. Got high accuracy of 94% of the size of biomass modeled with real biomass.

Li found that accurate, high-resolution yield maps are needed to identify spatial patterns of yield variability, to identify key factors in_uencing yield variability, and to provide detailed management information in precision farming. Varietal differences may significantly affect the forecasting of potato tuber yields with the use of remote sensing technologies. These authors argue that improving potato crop forecasting with remote sensing of unmanned aerial vehicles (UAVs) by incorporating varietal information into machine learning methods has the best chance at present .There are different challenges in this research area. Currently, crop prediction models generate actual results that are satisfactory, though they could perform better. This paper attempts to propose an enhanced crop prediction model that addresses these issues. The prediction process depends on the two fundamental techniques of feature selection [FS] and classification. Prior to the application of FS techniques, sampling techniques are applied to balance an imbalanced dataset.

## 2. LITERATURE SURVEY

### 2.1 DIFFERENT AUTHORS

Li found that accurate, high-resolution yield maps are needed to identify spatial patterns of yield variability, to identify key factors in_uencing yield variability, and to provide detailed management information in precision farming. Varietal differences may significantly affect the forecasting of potato tuber yields with the

use of remote sensing technologies. These authors argue that improving potato crop forecasting with remote sensing of unmanned

aerial vehicles (UAVs) by incorporating varietal information into machine learning methods has the best chance at present

There are different challenges in this research area. Currently, crop prediction models generate actual results that are satisfactory, though they could perform better. This paper attempts to propose an enhanced crop prediction model that addresses these issues. The prediction process depends on the two fundamental techniques of feature selection [FS] and classification. Prior to the application of FS techniques, sampling techniques are applied to balance an imbalanced dataset.

### 2.2 DOMAIN DESCRIPTION

Agriculture is a growing field of research. In particular, crop prediction in agriculture is critical and is chiefly contingent upon soil and environment conditions, including rainfall, humidity, and temperature. In the past, farmers were able to decide on the crop to be cultivated, monitor its growth, and determine when it could be harvested. Today, however, rapid changes in environmental conditions have made it difcult for the farming community to continue to do so. Consequently, in recent yeas, machine learning techniques have taken over the task of prediction, and this work has used several of these to determine crop yield. To ensure that a given machine learning (ML) model works at a high level of precision, it is imperative to employ efficient feature selection methods to preprocess the raw data into an easily computable Machine Learning friendly dataset.

## 3. PROBLEM STATEMENT

### 3.1 EXISTING SYSTEM

You posited an adaptable and precise technique to anticipate yields by employing openly accessible remote sensing data. The methodology enhances existing procedures in three different ways. To begin with, a remote detecting network is applied to propose a working methodology. Next, a novel dimensionality reduction procedure is presented that uses a convolutional neural network (CNN) alongside long-term memory. Finally, a Gaussian process is used to investigate and examine the spatio-transient structure of the data and enhance its accuracy. Anantha *et al.* [16] implemented a recommendation system using an associate ensemble model with majority voting. The random tree, Chi-square Automatic Interaction Detection (CHAID), kNN, and Naive Bayes (NB) are used as learners to help determine the most appropriate crop, taking into consideration soil parameters, with the results showing high accuracy and potency. The classified image generated by these techniques consists of ground truth-applied mathematics information Further, it incorporates such data as the parameters of the square measure in terms of the weather and crop yield, as well as state and district-wise crop produce.

All of the above are employed to predict specific crop yields in a given set of circumstances. developed a forecasting model which uses the default settings along with RF regression for crop yield production.

Fernando studied data on annual coconut production from 1971 to 2001 in a particular region and assessed its economic impact. The research revealed that the loss sustained by the economy in crop shortage terms was

around US $50 million. Ji advanced an estimation technique to predict rice yields. The study attempted to determine the effectiveness of artificial neural networks (ANN) in predicting rice yield in mountainous regions. It assessed the efficacy of the ANN, relative to biological parametric variations, and compared the efficiency of multiple bilinear regression models with the ANN model. Boryan proposed a decision tree-based technique to depict openly accessible state-level crop cover groups, in accordance with guidelines laid down by the Cropland Data Layer (CDL) and National Agricultural Statistics Service (NASS), and utilizing ground truth collected during the June Agricultural Survey. The proposed work outlines theNASS CDL program.

It presents information dealing with handling strategies, order and approval, precision evaluation, and CDL item particulars, and product cost estimation procedure. Hansen and Loveland proposed the use of Landsat to acquire satellite imagery that facilitates remote sensing of the environment.

### 3.2 DISADVANTAGE OF EXISTINTG SYSTEM

1) The system is not implemented RECURSIVE FEATURE ELIMINATION (RFE).
2) The system is not implemented Sampling techniques which are applied during preprocessing to balance the dataset and maximize the prediction performance.

## 4.PROPOSED SYSTEM

### 4.1 PROPOSED SYSTEM

Boruta is a random forest-based classification algorithm [38] that involves the voting of versatile unbiased indistinct classifiers in decision trees. The importance of a characteristic is estimated by calculating the loss of classification exactness caused by the random permutation of attributes within objects. The average and standard deviation of the loss of accuracy are calculated, and the average loss is divided by the standard deviation to obtain the Z score to measure average fluctuations in mean accuracy loss among crops.

A `shadow' attribute is made for each tree by randomly rearranging the values of the initial attributes across objects. The importance of every attribute is determined by analyzing all the attributes in the system. Given the random nature of the fluctuations, the shadow attributes are used as a reference to point to the most important ones. As is to be expected, the degree of accuracy depends greatly on the shadow attributes. Consequently, the values will be re-shuffled constantly to obtain optimal results.

The Boruta algorithm comprises the following steps: 1. The data system, which is extended by affixing copies of all the shadow attributes, is always prolonged by 5 shadow attributes. The added attributes are shuffled with the original attribute to remove any correlation with the response. The Z score is computed by running a random forest algorithm on the widespread information system. The Maximum Z Score Attributes (MZSA) are calculated and any attribute with a value higher than the MZSA is assigned a ``hit''. . For attributes with undetermined importance, a two-sided test of equality with the MZSA is carried out. Attributes with importance significantly lower than the MZSA are identified as `unimportant' and permanently eliminated from the information system. Attributes with importance significantly higher than the MZSA are marked `important'. Shadow attributes are thus eliminated from the information system. The

process is repeated until all attributes are marked with a level of importance

### 4.2 ADVANTAGE OF PROPOSED SYSTEM

1) The RFE technique is a wrapper feature selection technique that starts with the entire dataset. The ranking method crucial to the RFE technique orders the dataset from the best to the worst, based on which salient features are selected.
2) The main advantage the RFE has over other methods is that it categorically verifies every feature's role in processing the output of the model and eliminates features only based on their performance.

## 5.IMPLEMENTATION

### 5.1 Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Attack Status, View Attack Status Ratio, Download Trained Data Sets, View Attack Status Ratio Results, View All Remote Users.

### 5.2 View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

### 5.3 Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT ATTACK STATUS TYPE, VIEW YOUR PROFILE.

## 6.CONCLUSION

Predicting crops for cultivation in agriculture is a difficult task. This paper has used a range of feature selection and classification techniques to predict yield size of plant cultivations. The results depict that an ensemble technique offers better prediction accuracy than the existing classification technique. Forecasting the area of cereals, potatoes and other energy crops can be used to plan the structure of their sowing, both on the farm and country scale. The use of modern forecasting techniques can bring measurable financial benefits.

## 7. FUTURE ENHANCEMENT

A `shadow' attribute is made for each tree by randomly rearranging the values of the initial attributes across objects. The importance of every attribute is determined by analyzing all the attributes in the system. Given the random nature of the fluctuations, the shadow attributes are used as a reference to point to the most

important ones. As is to be expected, the degree of accuracy depends greatly on the shadow attributes. Consequently, the values will be re-shuffled constantly to obtain optimal results.

## 8.REFERENCE

[1] R. Jahan, ``Applying naive Bayes classi_cation technique for classi_cation of improved agricultural land soils," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 6, no. 5, pp. 189_193, May 2018.

[2] B. B. Sawicka and B. Krochmal-Marczak, ``Biotic components in_uencing the yield and quality of potato tubers," *Herbalism*, vol. 1, no. 3, pp. 125_136, 2017.

[3] B. Sawicka, A. H. Noaema, and A. Gáowacka, ``The predicting the size of the potato acreage as a raw material for bioethanol production," in *Alternative Energy Sources*, B. Zdunek, M. Olszáwka, Eds. Lublin, Poland: Wydawnictwo Naukowe TYGIEL, 2016, pp. 158_172. B. Sawicka, A. H. Noaema, T. S. Hameed, and B. Krochmal-Marczak, ``Biotic and abiotic factors in_uencing on the environment and growth of plants," (in Polish), in *Proc. Bioró»norodno±¢ .rodowiska* Znaczenie, Problemy, Wyzwania. Materiaªy Konferencyjne, Puªawy, May 2017.

Available: https://bookcrossing.pl/ksiazka/321192 R. H. Myers, D. C. Montgomery, G. G. Vining, C. M. Borror, and S. M. Kowalski, ``Response surface methodology: A retrospective and literature survey," *J. Qual. Technol.*, vol. 36, no. 1, pp. 53_77, Jan. 2004.

[4] D. K. Muriithi, ``Application of response surface methodology for optimization of potato tuber yield," *Amer. J. Theor. Appl. Statist.*, vol. 4, no. 4, pp. 300_304, 2015, doi:.M. Marenych, O. Verevska, A. Kalinichenko, and M. Dacko, ``Assessment of the impact of weather conditions on the yield of winter wheat in Ukraine in terms of regional," *Assoc. Agricult. Agribusiness Econ. Ann. Sci.*, vol. 16, no. 2, pp. 183_188, 2014.

[5] J. R. Ol¦dzki, ``The report on the state of remotesensing in Poland in 2011_2014," (in Polish), *Remote Sens. Environ.*, vol. 53, no. 2, pp. 113_174, 2015.

[6] K. Grabowska, A. Dymerska, K. Poáarska, and J. Grabowski, ``Predicting of blue lupine yields based on the selected climate change scenarios," *Acta Agroph.*, vol. 23, no. 3, pp. 363_380, 2016.

[7] D. Li, Y. Miao, S. K. Gupta, C. J. Rosen, F. Yuan, C. Wang, L. Wang, and Y. Huang, ``Improving potato yield prediction by combining cultivar information and UAV remote sensing data using machine learning," *Remote Sens.*, vol. 13, no. 16, p. 3322, Aug. 2021, doi: 10.3390/rs13163322.

[8] N. Chanamarn, K. Tamee, and P. Sittidech, ``Stacking technique for academic achievement prediction," in *Proc. Int. Workshop Smart Info-Media Syst.*, 2016, pp. 14_17.

[9] W. Paja, K. Pancerz, and P. Grochowalski, ``Generational feature elimination and some other ranking feature selection methods," in *Advances in Feature Selection for Data and Pattern Recognition*, vol. 138. Cham, Switzerland: Springer, 2018, pp. 97_112.

[10] D. C. Duro, S. E. Franklin, and M. G. Dubé, ``A comparison of pixelbased and object-based image analysis with selected machine learning algorithms for the classi_cation of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sens. Environ.*, vol. 118, pp. 259_272, Mar. 2012.

[11] S. K. Honawad, S. S. Chinchali, K. Pawar, and P. Deshpande, ``Soil classi_cation and suitable crop prediction," in *Proc. Nat. Conf. Comput. Biol., Commun., Data Anal.* 2017, pp. 25_29.

[12] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, ``Deep Gaussian process for crop yield prediction based on remote sensing data," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 4559_4565.

[13] D. A. Reddy, B. Dadore, and A. Watekar, ``Crop recommendation system to maximize crop yield in ramtek region using machine learning," *Int. J. Sci. Res. Sci. Technol.*, vol. 6, no. 1, pp. 485_489, Feb. 2019.

[14] N. Rale, R. Solanki, D. Bein, J. Andro-Vasko, andW. Bein, ``Prediction of Crop Cultivation," in *Proc. 19th Annu. Comput. Commun.Workshop Conf. (CCWC)*, Las Vegas, NV, USA, 2019, pp. 227_232.

[15] J. Jones, G. Hoogenboom, C. Porter, K. Boote, W. Batchelor, L. Hunt, P. Wilkens, U. Singh, A. Gijsman, and J. Ritchie, ``The DSSAT cropping system model," *Eur. J. Agronomy*, vol. 18, nos. 3_4, pp. 235_265, 2003.

[16] M. T. N. Fernando, L. Zubair, T. S. G. Peiris, C. S. Ranasinghe, and J. Ratnasiri, ``Economic value of climate variability impact on coconut production in Sri Lanka," in *Proc. AIACC Working Papers*, vol. 45, 2007, pp. 1_7.

[17] B. Ji, Y. Sun, S. Yang, and J. Wan, ``Arti_cial neural networks for rice yield prediction in mountainous regions," *J. Agricult. Sci.*, vol. 145, no. 3, pp. 249_261, Jun. 2007.

[18] C. Boryan, Z. Yang, R. Mueller, and M. Craig, ``Monitoring U.S. agriculture: The U.S. department of agriculture, national agricultural statistics service, cropland data layer program," *Geocarto Int.*, vol. 26, no. 5, pp. 341_358, 2011.

[19] M. C. Hansen and T. R. Loveland, ``A review of large area monitoring of land cover change using Landsat data," *Remote Sens. Environ.*, vol. 122, pp. 66_74, Jul. 2012.

[20] D. K. Bolton and M. A. Friedl, ``Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics," *Agricult. Forest Meteorol.*, vol. 173, pp. 74_84, May 2013.

[21] J. Dempewolf, B. Adusei, I. Becker-Reshef, M. Hansen, P. Potapov, A. Khan, and B. Barker, ``Wheat yield forecasting for Punjab province from vegetation index time series and historic crop statistics," *Remote Sens.*, vol. 6, no. 10, pp. 9653_9675, Oct. 2014.

[22] H. D. Shannon and P. M. Raymond, ``Managing weather and climate risk to agriculture in North America," *Central Amer. Caribbean*, vol. 10,pp. 50_56, Dec. 2015.