A New Approach of Machine Learning and Deep Learning Algorithms based Crop Yield Prediction

[1]S. Shanthi, [2]K Venkata Sai Vaishnavi, [3]B Supraja, [4]B Vijitha and [5]B Vineela Reddy
[1,2,3,4,5]Department of CST, Madanapalle Institute of Technology and Science, Andhra Pradesh – 517325.
[1]shan.sece@gmail.com, [2]19691A28H4@mits.ac.in, [3]19691A28F1@mits.ac.in, [4]19691A28H6@mits.ac.in,
[5]19691A28H8@mits.ac.in

Article Info
A. Haldorai et al.(eds.), *2nd International Conference on Materials Science and Sustainable Manufacturing Technology,* Advances in Computational Intelligence in Materials Science.
Doi: https://doi.org/10.53759/acims/978-9914-9946-9-8_18

Abstract – The science and skill of nurturing plants and wildlife are referred to as agriculture. India ranks second in the world for farming, which takes up 60.45% of the country's territory. The economy of India is primarily supporting agricultural, agro-industrial sectors. Crop rotation, the consistency of the soil, air and surface temperatures, precipitation, and other elements all have an impact on how well crops are grown. Further crucial are soil constituents including nitrogen, phosphate, and potassium. The corpus of work currently being done in this field includes a crop choice model that makes use of ML methods (Random Forest, Decision Tree, ANN). In this paper, recommended model enhanced using Deep Learning techniques, in addition to crop prediction, precise data on the amounts of necessary soil components and their individual prices are attained. Compared to the present model, it provides a better degree of accuracy. In order to help farmers to predict a profitable crop, analyses the available data. Variables related to the soil and climate taken into consideration to anticipate an acceptable yield. This objective show's that Python-Based System using cunning strategies for predicting, bountiful harvest possible while using the least amount of resources. In this work, the SVM machine learning algorithm is combined with the LSTM and RNN deep learning algorithms.

Keywords – Random Forest, Decision Tree, ANN, Crop, Yield, Prediction, SVM.

I. INTRODUCTION

Cultivation is one of the main professions in the nation. The nation's financial status greatly improves by investing in a variety of agricultural enterprises. It is thus regarded as the most adaptable method of generating income. India uses 60.45% of its land agriculture. It results, satisfaction of almost 1.2 billion people's needs. In present age, agriculture business modernization is an extensive process. As a result, farmers are shifting in their favour and making more money while spending less [1]. Data Scientific (DA), together with the guidance of certain software and framework, analyses the informative indexes in order to draw conclusions about the data they include. In the past, the yield was predicted based on the rancher's knowledge of a particular plot of land and harvest [2].

Farmers concentrate on accumulating an ever-increasing number of harvests as circumstances change gradually. Given the present circumstances, many farmers want further information on increased yields. They have no idea about what will earn from the crop [3]. A thorough understanding and evaluation of crop performance under natural conditions may also improve farm profitability. The suggested structure in this research calls for information on the client's geographic region. The region provides the soil's nutrients, including nitrogen, phosphorus, and potassium [4]. The two extra dataset, feature and crop data set collected from kaggle.com website, is took into consideration in the considering as section and contain unique information that is considered static information. Construction of the data and harvest about different yields determined from various government websites are represented by the static information. The suggested structure makes predictions in Artificial Intelligence (AI), such as Multiple Linear Regression, in order to identify the pattern within the data. Moreover, it is treated using the requirements mentioned. As a result, it will provide the finest possible reaps given the biological circumstances. Hence, this system just requires the customer's location to recommend potential advantageous yields. It allows the farmer to choose which harvest to cultivate.

II. LITERATURE SURVEY

Sheenoy et al. provided a paper lays forth a solution for the decline in transportation costs. The IOT-based technique is utilized to reduce the number of agents and intermediary hops between customers and ranchers, which helps the rancher even more. The study implements integrated processes and offers a prediction-based system that recommends crops that will provide the most profit. [5].

Monali et al. said, Crops are analyzed and categorized so that predictions about their yields can be made. Data mining methods are used to conduct classification [6]. This work addresses a number of categorization rules, including

KNN and Naive-Bayes. These guidelines been researched are acknowledged as being precise for the data set utilized in this study. [7].

Hemageetha et al. provided few data mining methods, including association rule mining, classification, clustering, market-based analysis, and decision trees. The concept of data mining is extensively covered. Several Data Mining methods, such as K-Means, Naive-Bayes classifier, and J48, are described in this study [8]. Similar to how soil classification uses association rule mining, naive bayes, and genetic algorithms. The clustering in the soil database is finally handled. It helped us understand and analyse different data mining methods. It becomes really beneficial when the job of this study work is being built up. It facilitates mining of datasets obtained from remotely placed sensors. [9].

Nagini et al. This report presents exploratory investigation and provides an explanation of how many prediction models were created. For a sample dataset, the different regression methods are used in order to distinguish and analyse their individual characteristics. Linear, nonlinear, multiple-linear, polynomial, ridge, and logistic regression are the techniques covered in the study [10].

Awanit et al. In this study, a design for projecting agricultural output for the current year is suggested. To determine crop yield, data mining technique K Means is used. In this, fuzzy logic is also used as a tool to anticipate the harvest. On a certain part of land, a set of rules are employed for cultivation, precipitation, and agricultural production. A rule-based predicted logic, often known fuzzy logic, is what is referred to as it. This publication [11] explains how to use K-Means to analyse the datasets. As a set of fuzzy logic rules, these principles are once again applied to the prediction of the crop that shall maximize profit based on the cost of crops that are in previous years and current soil and weather information.[12].In recent years, much classification has been utilized for brain tumors such as machine learning algorithms and other optimization techniques. Lot of researchers had developed feature selection for classification process. The stochastic diffusion search a technique for effectively selects the features[13,14].Shanthi et al.[15] had developed cancer prediction using radiometric feature selection and machine learning algorithm. Here, they mainly focused on feature selection process. This method was effectively reduce the computation complexity and increase the classification accuracy.

III. PROPOSED MODEL

In provided design, machine learning, deep learning techniques used to deliver best crop yield. The proposed model experiments with a crop data set. When the weather criterion is taken into account, the crop selected is based on the existing environment, the soil, as well as its contents. Deep learning is used to do a variety of effective computations, like selecting the best crop when there are many possibilities. This method allows for precise crop prediction. As indicated in **Fig 1**, the **SVM** method is used for Machine learning, while the RNN, LSTM algorithms are taken for Deep learning.

Architecture of formulated model

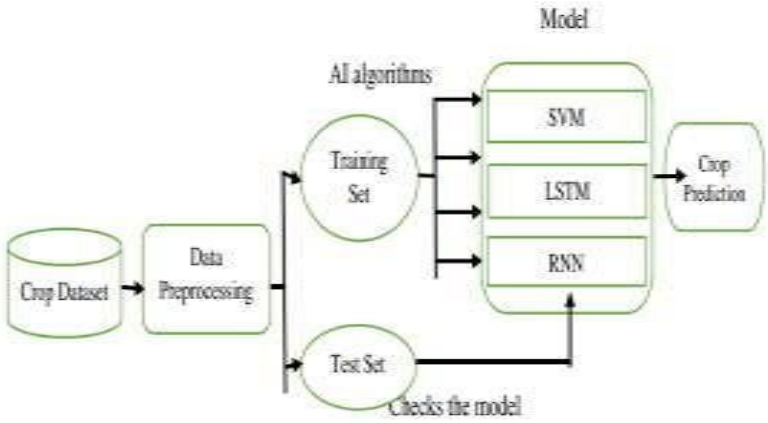


Fig 1. Crop Prediction is used in the Built Model's Architecture.

Algorithms Used

Support vector machine (SVM)

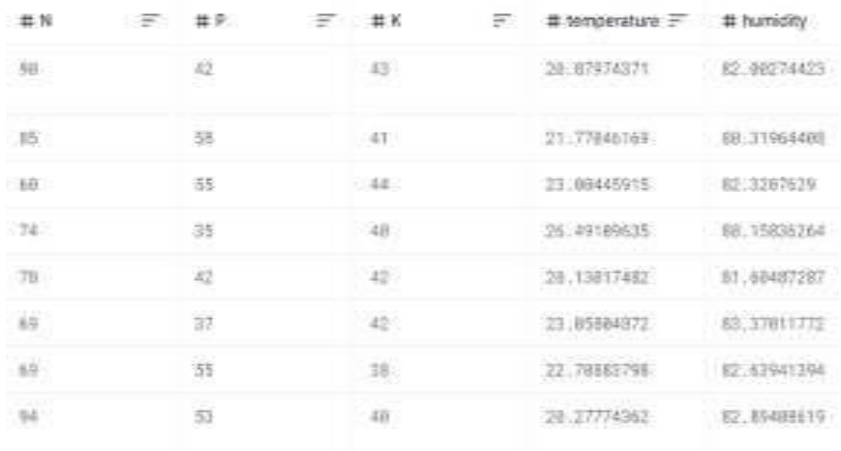
step-1:Impact the necessary packages.
step-2: Load the input data.
step-3:Selection from the data set of the necessary number of features.
step-4: Plot SVM boundaries by using original data.
step-5: Set the regularization parameter's value.
step-6:Lastly,SVM classifies object is constructed.

Long-short term memory (LSTM)
step-1: In **Keras,** define a Neural Network as a series of layers.
step-2: Compile the network that needs the numerous parameters that are supplied.
step-3: Fit the network that needs an input patterns matrix X and an output patterns array Y that are identical, as well as the required training data.
step-4:Utilize the training data to evaluate the network. Rarely model can be evaluated on a test or validation set.
step-5:Make the necessary predictions in the format that the network's output layer specifies.

Recurrent neural network (RNN)
step-1: One time step of input is sent to the network.
step-2: Calculate the current state using the previous state and the current input.
step-3: Current state is ht and it turns out to be ht-1 for the next time step.
step-4:Depending on the issue, any number of time steps can be taken. Data is combined from all the prior states.
step-5: Final current state is used to compute the outcome once all time steps have been completed.
step-6: In order to update the weights, error propagated backwards toward network. **RNN** is subsequently prepared.

IV. DATASET AND PERFORMANCE METRICS.

Dataset Description
In this study, A dataset with set of agricultural parameters utilized to evaluate the model. In addition, another dataset is utilized for feature purpose. The two datasets are taken from Kaggle repository which holes a size of **7841 kb** with record of 6 rows and features of 250 columns. this datasets prediction includes temperature, rainfall, pH level, and relative humidity. Suppose, if the crop is wheat, the prediction parameters may be adjusted to any value from the range of data included in the dataset. That is similar for all the crops which are present in dataset.



# N	# P	# K	# temperature	# humidity
90	42	43	20.87974371	82.00274423
85	58	41	21.77846169	80.31964408
60	55	44	23.004459[illegible]	82.3207629
74	35	40	26.491096[illegible]	80.15836264
78	42	42	20.13017482	81.60487287
69	37	42	23.058049[illegible]	83.37011772
69	55	38	22.708838[illegible]	82.63941394
94	53	40	20.27774362	82.89409[illegible]

Fig 2. Dataset for Crop.

Fig 2 depicts a section of crop dataset.This crop is wheat; the soil has a pH of 5.5; the temperature is 13; the amount of rainfall and relative humidity vary depending on the climate conditions of different days; and each crop given a goal value to help identify it.

Performance Metrics
The performance of suggested models can be calculated in order to obtain the result. To ensure the correctness of the output, certain equations are used. The following are these formulas:

Accuracy = (TP+TN)/(TP+FP+TN+FN)
Precision = TP/(TP+FP)
Recall = TP/(TP+FN)

TP is described as occurring when situations are projected to be positive and turn out to be positive. When situations are expected to be positive but turn out to be negative, this is known as **FP**. TN occurs when instances are projected to be negative and turn out to be negative. When instances are expected to be negative but turn out to be positive, this is known as FN.

V. EXPERIMENTATION & RESULT ANALYSIS

The dataset of harvested crops is loaded to begin the execution of the study task. The relevant libraries and packages were imported, and process of preprocessing the data resumed. The data is divided into training data and test data. The necessary AI algorithms are used in a model that has been created to choose the best crop to grow on a particular plot of land.

Result Analysis

The study project is put into practise utilising a crop dataset that was obtained from website kaggle.com. It has a variety of crops, including sugarcane, rice, wheat, millets, green gramme, maize, peas, pigeon peas, and so on. A few prediction parameters are incorporated in it, including pH value, rainfall, temperature, area and relative humidity. Two sets of data are required by machine learning, deep learning algorithms, known as the Training set and Test set, in order to create a prediction model. The survey data that has been gathered from previous occurrences makes up the trained data. while the data from the most recent poll is test data. When the code has been executed, a visual representation of the answer is shown for each of the original parameter values.**Fig 3** shows usage of fertilizer vs yield and **Fig 4** shows scattered plot of pesticides along with yields. These variables include the use of pesticides, fertilizers, water, area, and the sun. Based on the data for these characteristics, the yield is anticipated:

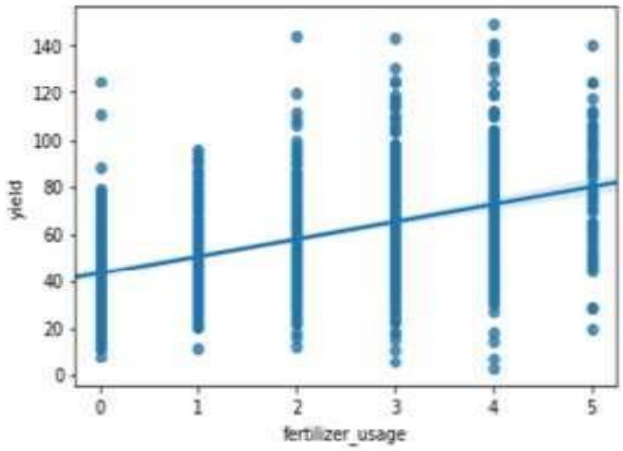


Fig 3. Usage of Fertilizer Vs Yield.

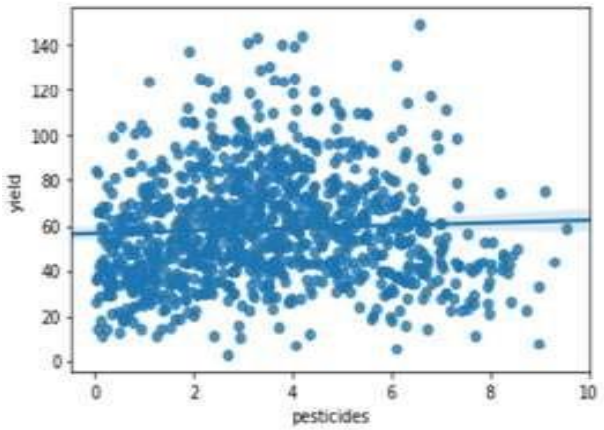


Fig 4. Scattered plot of Pesticides along with Yields.

Performance Evaluation

A set of parameters is used to analyse the dataset. The following are examples of these variables: area, humidity, soil type, precipitation, and temperature. These are the characteristics that are used to assess the dataset and forecast a higher yield. On a certain plot of land, a certain crop is expected to produce. Wheat, rice, maize, peas, pigeon peas, green gramme, potatoes, sugarcane, soya beans, and other crops are among them.

Table 1. Performance Analysis

Algorithms	Features	Crops	Accuracy
DT,ANN,RF (Used in existing paper)	Temperature, Rainfall, Location, Soil	Wheat, Rice, Soyabean	93%
LTSM,RNN,SVM (Used in proposed work)	Temperature, Rainfall, pH value,	Wheat,Pea, Rice, Maize, Millets, Pigeon Pea, Greengram, sugarcane, Soyabean	97%

Table 1 demonstrates that the dataset is evaluated across set of crops based on a number of characteristics.Accuracy is 93% when using Random Forest and Artificial neural network (ANN) methods. Whereas the accuracy is estimated to be 97% when using the Support vector machine (SVM) Long-short term memory (LSTM), and Recurrent neural network (RNN)methods. Hence, it is evident that Deep learning algorithms, in addition to Machine learning algorithms, are essential for more accurately forecasting yield.

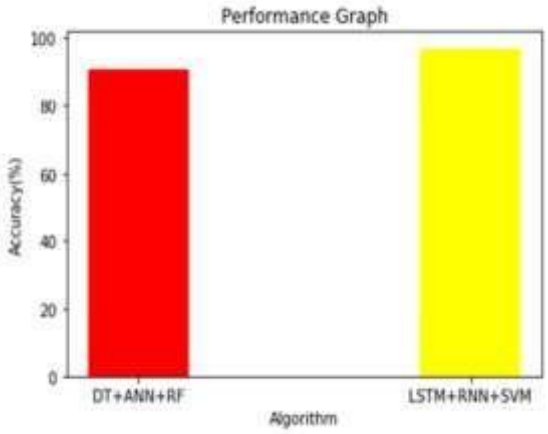


Fig 5. Graph Shows the Accuracy Level.

The accuracy for both techniques is visually shown in **Fig 5**. While using ANN and The first, red bar, of the Random Forest algorithm's accuracy was 93%. The second bar, which is represented by a yellow bar, shows that accuracy was 97% when the RNN,LSTM and SVM algorithms were combined. As a result, the second result is more accurate.

VI. CONCLUSION AND FUTURE SCOPE

The implemented model was developed by using Artificial intelligence algorithms to solve problem of farmers losing money on their farms since they don't know how to cultivate in different types of soil and weather. Deep learning (LSTM, RNN) and machine learning (SVM) methods are used to build the model. Out of all the crops that are now available, the model offers the best crops that should be grown on land with the lowest costs after reviewing the prediction parameters. There are no other studies that employ the same methods for crop prediction, according to research. As a consequence, it may be argued that research endeavour is more accurate than earlier studies that are used other crop forecast approaches. The computed accuracy is 97%. It has a sizable potential for growth and be integrated and inter faced with a flexible, multifaceted application. Due to their need for education, the farmers would get clear information on the maximum agricultural yield on their mobile devices. By doing this, the work may be completed at that precise time without costing the rancher any money, even if he is at home. There will be rapid expansion in agribusiness sector, will be helpful for farmers produce more food.

References

[1]. Introduction September 2021 Situation Assessment of Agricultural Households and Land and Livestock Holdings of Households in Rural India PIB
[2]. Sonal Agarwal and Sandhya Tarar 2021 A Hybrid Approach for Crop Yield Prediction using Machine Learning and Deep Learning Algorithms IOP
[3]. Introduction 2017 The Future of Food and Agriculture FAO
[4]. Mudra Verma, Kaushal Vyas, Kunal Sahjwani and Mahendra Patil April 2022 Soil Profile Based Crop Prediction System IRJET
[5]. Sheenoy et al and RJ Reddy July 2022 Prediction of Suitable Crop Using Machine Learning IJRPR

[6]. Monali Paul, Santosh K. Vishwakarma, Ashok Verma. Analysis of soil behaviour and prediction of crop yield using data mining approach. Computational Intelligence and Communication Networks (CICN). 2015; 766-771.
[7]. Kusuma Lata, Sajidullah S. Khan October 2019 Experimental Analysis of Machine Learning Algorithms Based on Agricultural Dataset for Improving Crop Yield Prediction IJEAT
[8]. N. Hemageetha, G.M.Nasira March 2012 Vegetable Price Prediction using Data Mining Classification Technique ICPRIME
[9]. P K Arunesh May 2016 Analysing Soil Data using Data Mining Classification Techniques IJST
[10]. S. Nagini, T. V. R. Kanth and B. V. Kiranmayee, "Agriculture yield prediction using predictive analytic techniques," 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), Noida, 2016, pp. 783-788, doi: 10.1109/IC3I.2016.7918789.
[11]. Awanit Kumar,Shiv Kumar August 2015 Prediction of Production of Crops using K-mean & Fuzzy Logic IJCSMC
[12]. V Latha Jothi ,Neelambigal A, Nithish Sabari S and Santhosh K 2020 Crop Yield Prediction using KNN Model IJERT
[13]. Shanthi, S., & Rajkumar, N. “Lung cancer prediction using stochastic diffusion search(SDS) based feature selection and machine learning methods”, Neural Processing Letters, 53(4),26172630,2021.
[14]. Shanthi, S., Akshaya, V.S., Smitha, J.A., & Bommy, M.(2022).Hybrid TABU search with SDS based feature selection for lung cancer prediction. International Journal of Intelligent Networks,2022
[15]. Shanthi, S., & Rajkumar, N., “Nonsmall-cell lung cancer prediction using radiomic features and machine learning methods”, International Journal of Computers and Applications,1-9,2019.