Crop Yield Prediction by Hybrid Technique with Crop Datasets

Kusum Lata[1], Sajidullah Khan[2]

[1]School of Computer Science and Engineering, Sandip University
Nashik, Maharashtra, India
kusumrao444@gmail.com

[2] School of Computer Science and Engineering, Sandip University
Nashik, Maharashtra, India
sajidullah.khan@sandipuniversity.edu.in

Abstract*: *Agriculture is one of the intense domains across the globe which has greater impact on the development of a country. There are various tools and techniques developed for the farmers and they are taking advantages of it by collaborating with leading government and public and private organizations. Also, the power of artificial intelligence is realized in agriculture field with the application of machine learning and deep learning algorithms. Numerous models have been proposed using the conventional algorithms, but still it is needed to improve the prediction accuracy. Therefore, in this proposed model a hybrid technique is designed by combining the Machine learning, deep learning algorithms and optimization with particle swarm optimization PSO methods to improve the prediction accuracy. In the proposed model, SVM is used as Machine leaning algorithm and RNN-LSTM is used as deep learning algorithm. The crop data sets of Maharashtra for previous years are used as input to the model and prediction is done for 10 years. The proposed model has potential in improving the yield prediction for various crops like onion, grapes, cotton etc. produced in the Maharashtra State of India.

***Keywords*: Hybrid, SVM, RNN-LSTM, PSO, accuracy, agriculture**

(Article history: Received: 30th April 2021 and accepted 20th December 2021)

I. INTRODUCTION

India is the largest producer of various crops and other agricultural items. The agricultural output has a main contribution to the Indian economy. The crop yield is one of the most influential parameters in the agricultural sector. According to this crop yield parameter the various other related activities like crop insurance, storage, market development is planned and monitored. This field requires most attention as the population of the country is increasing day by day and food security is a big challenge [1]. The crop yield mainly projected on the weather (Climatic) and non-weather (soil, irrigation, seed, fertilizers, etc.) factors.

In this area of study many authors, institutes and agencies are working with the aid of artificial intelligence, machine learning, deep learning algorithms. The concepts of machine learning and deep learning belongs to artificial intelligence domain. Various techniques and models have been proposed in the favor of farmers with the help of machine learning and deep learning algorithms. These techniques have enhanced performance over the conventional statistical technique. Machine learning is the innovative approach to solve the complex task in order to obtain the optimal preferred results. Machine learning practices the concept of statistics to build models, the key motive is to find out the interpretations from a sample dataset. After a machine learning model is finalized, its illustration and algorithmic explanation for understanding is very important. The machine learning is also measured in terms of classification accuracy in various applications. Machine learning is used in numerous areas, including credit card fraud detection, biomedical data analysis, forecasting as a recommendation support system. Generally, many complex issues that involve decision-making can also be considered problems designed for machine learning. These scenarios can be tackled with learning from past experiences, observations to find out the solutions for the problems. Machine learning is widely applied in everyday problems which are impossible to solve with the conventional techniques. The main applications of machine learning are listed as below:

- Credit Card fraud detection
- Online Product recommendations
- Sentiment analysis
- Weather forecasting for agriculture
- Stock market prediction
- Customer categorization
- Breakdown analysis and findings
- Image processing
- Healthcare patient data analysis
- Anomaly detection and solution
- Virtual assistants
- Business intelligence

In the digital era and cloud environment a large amount data is accessible to study and transform it to beneficial results. This study of this data is possible by applying machine leaning algorithms to find out the relations between the variables. The main goals of machine learning are to develop computer algorithms & models to deal with real world scenarios. Machine learning has a prime role in all innovative applications created for humans. The machine learning is basically classified into supervised, unsupervised and semi supervised techniques. Further supervised learning is classified into two types namely classification and regression. The classification is suitable where we need to predict a category or class based on values. The best example of classification is mail account where mails are classified into social, spam etc. This type of supervised learning is not suitable for predictions having the values like distance, weight etc. The second type of supervised technique i.e. regression is suitable where we need to predict the continuous values like age of a person, salary of employee etc. On the other hand, unsupervised is classified into three types namely clustering, association and dimension reduction (Generalization) [2]. The main concept behind unsupervised is that only input data variables are known and output variables are not known. Clustering type of machine learning divides the data into some clusters having similar patterns. Association type of machine learning applied to large datasets to find out the relations between various variables. Dimensionality reduction can be seen as a data preparation method applied on data prior to modeling. It can be performed after data cleaning before training a model. Dimensionality reduction is a technique to reduce the no. of input variables in a dataset. Semi-supervised learning or reinforcement learning uses a small amount of labeled data reinforcing a larger set of unlabeled data. The main concept behind reinforcement learning is that there is no training labeled datasets and the agent learns from past experience.

The traditional machine learning techniques are not capable to handle the complex task and not calculating the accurate predictions, hence evolution of Deep Learning (DL) ideas come in the artificial intelligence space. The deep learning is a subcategory of machine learning in artificial intelligence domain [3]. It has potential of learning from unstructured or unlabeled data i.e., unsupervised. Nowadays DL approaches are very much required because of effective learning, precision, and accuracy. Deep learning models are having enhanced performance as compared traditional models in the field of prediction of various real-life things. DL algorithms are deep neural networks, recurrent neural networks and convolution neural networks etc. The various authors applied machine learning and deep learning algorithms in the area of crop yield prediction across the globe but the prediction accuracy is not optimal [11].

Hence in this paper, we will discuss the new technique known as hybrid technique. In the conventional yield Prediction techniques only one machine learning algorithm is used to build a model. On the other hand, in the hybrid technique various algorithms can be combined in order to get the optimal results. The detailed illustration of the both the techniques is shown in Fig1 and Fig 2. In this paper, a hybrid machine learning technique/model is created using the SVM approach as a traditional machine learning and RNN-LSTM as a deep learning [15][16]. Further the output is obtained by PSO optimization technique is done Testing and training of both the algorithms are done independently. This hybrid technique will enable the farmers and governments to get the preferred output which will further boost the Indian economy [23].

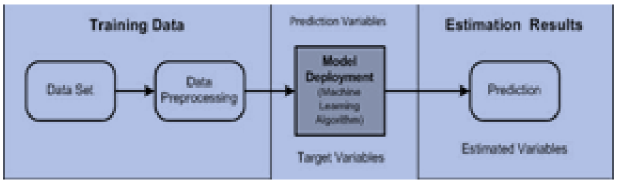

Fig 1. Conventional machine learning workflow

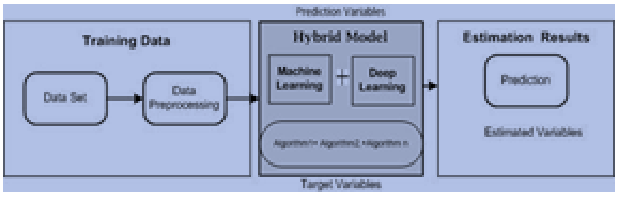

Fig 2. Hybrid technique workflow

II. RELATED WORKS

In this section related work done on crop yield prediction using hybrid machine learning algorithms is discussed. The researchers tried to enhance the accuracy by with different combinations of machine learning algorithms to construct the hybrid algorithms for crop yield prediction.

Sonal Agarwal, Sandhya Tarar [1] 2021 proposed a hybrid approach for crop yield prediction using machine learning and deep learning algorithms to predict the best crop production. This approach will analyze the given data and help the farmers in forecasting a crop to increase the revenues. In this research, the authors have done the experiments on given crop dataset like current atmosphere, soil with its constituents, climate and soil parameters. The authors have collected the data from kaggle.com and this dataset contains the various parameters such as temperature, rainfall, pH value, relative humidity and also an area. In this dataset, a number of crops are taken like wheat, rice, maize, millet, pea, pigeon pea, sugarcane and green gram etc. In the machine learning, Support vector machine (SVM) algorithm is used while in the deep learning long short-term memory (LSTM) and recurrent neural network (RNN) is used for execution of the dataset. This proposed model estimates and assesses the different parameters for the available crops that should be grown on the land with less expanses. The various techniques studied in this model to get the best accuracy. Firstly, the authors have applied artificial neural network (ANN) and random forest (RF) algorithms for a set of crops and find out the accuracy i.e., 93%. Secondly, they have applied long short-term memory (LSTM), support vector machine (SVM) and recurrent neural network (RNN) and find out the accuracy i.e., 97%. Therefore, the authors have concluded that the use of both machine learning algorithm and deep learning algorithm plays an essential role in predicting the improved crop yield with upgraded accuracy.

U. Muthaiah, S. Markkandeyan, Y. Seetha [2] 2019 proposed a classification model and hybrid feature selection method to improve crop performance on Mango and Maize dataset. The selection of important features is done with the Particle Swarm Optimization–Support Vector Machine

(PSO-SVM) classification algorithm for the Mango and Maize datasets. In Particle Swarm Optimization, a particle is considered as each solution of the optimization problems and for the selection of significant features authors have made use of PSO-SVM. The authors have exhibited the several experiments on these datasets and it's also giving more generated rules and features selection with PSO-SVM algorithm and fuzzy decision tree. In this research, the authors concluded that the proposed methodology achieves the great accuracy for the classification using Maize and Mango datasets with less error and high positive rate. Although the methodology used here brings efficient outcomes as compared to the other existing techniques.

Saeed Nosratabadi, Karoly Szell, Bertalan Beszedes, Felde Imre, Sina Ardabili, Amir Mosavi [3] proposed a novel hybrid machine learning model for crop yield prediction. In this research, the performance of artificial neural networks-imperialist competitive algorithm (ANN-ICA) and artificial neural networks-gray wolf optimizer (ANN-GWO) models are estimates for the crop yield prediction. The research study is done in a large irrigated area in Kerman, Iran. The dataset consists of the numerous attributes like rainfall, solar radiation, temperatures and agricultural products like wheat, barley, potato and sugar beets are taken from 1998-2006 in this research study. This dataset is divided into two parts i.e., training and testing. For the years 1998-2005 is taken for training phase whereas the dataset for 2005-2006 is taken for testing phase. The system was evaluated using different performance criteria such as RMSE, MEA, R matrices are applied to check the accuracy for the ANN-ICA and ANN-GWO in the crop yield perdition. The accuracy and performance of ANN-GWO is better than ANN-ICA. In the meantime, a different set of features affect the performance of the model, it is suggested that future research perceive a different set of features and do the comparison for suitable outcomes. In future, the improvement of various hybrid and ensemble machine learning algorithms are done using deep learning models.

Shivi Sharma, Geetanjali Rathee, Hemraj Saini [4] 2018 presents big data analytics for crop prediction mode using optimization technique. The data comprises of the soil and environment features i.e., average temperature, average humidity, total rainfall and production yield are used in predicting the two classes. The classes can be termed as good yield and bad yield. The proposed approach is divided into three segments like the data pre-processing, grey wolf optimization-based feature selection and SVM_GWO with the support vector machine (SVM) and also a hybrid classifier model is used to enhancing the features. The data set used here comprises of historical data, agricultural equipment and sensor data, social and web-based data, streamed data and business, industries and external data. The authors tried their best to attain the best accuracy of prediction models via different parameters for the future precision agriculture. A combinational approach is used to improve the classification accuracy, precision, recall and F-measure by selecting the optimal constraints. The result shows that the proposed SVM_GWO method is better as compare to the SVMs classification algorithm. In the future, one can apply the various machine learning techniques for the different kind of challenges like artificial neural network (ANN), random forest (RF) and use the hybrid algorithms to select the best features.

Ms. Shreya V. Bhosale, Ms. Ruchita A. Thombare, Mr. Prasanna G. Dhemey, Ms. Anagha N. Chaudhari [5] done the analysis on the crop yield prediction using data analytics and hybrid approach. The authors have made use of various techniques like K-means clustering, Apriori algorithm, Naive Bayes algorithm etc. This research work will consist of the various attributes such as district, land area, soil type, season, crop name, production and rainfall. The database is used for analysis and after that pre-processing takes place. The authors have studied the various problems faced by the farmers in India and help the farmers to increase the yield for their crops. Therefore, authors have created a system which will predict the name of crop and the yield for particular farm [5].

III. SYSTEM MODEL

Crop yield prediction very important for farmers and governments in order to plan the various agriculture related activities. Farmers are trusting on the traditional way of crop management and crop yield prediction based on belief, physical visits and reports. Many techniques have proposed for crop yield prediction using machine learning but the accurate predictions are not achieved. With the motive to help the farmer's community it is entrusted to develop some model to optimize the predictions. This section covers the system model required to realize the hybrid technique for crop yield prediction. The algorithms used in this technique are described as below:

A. *Machine learning and deep learning algorithms*

1) SVM

A support vector machine (SVM) is a supervised machine learning algorithm that uses classification algorithms for two-group classification problems. The SVM classifier is used to recognize the classes, which are nearly associated to the known classes. Support vectors are nothing but they are the points that is actually passing through the marginal plane that we have actually created in parallel to the hyperplane. It may be having number of points like two, three and any number points that are passing through this particular marginal plane. We'll be considering those as support vectors. So, this helps us to determine the maximum distance of the marginal plane. If the hyperplane is having the massive distance to nearest features of any type of class, then it'll be considered as the good separation. Here, SVM splits the data into decision surface and decision surface further divides that data into hyperplane of two classes. So, the ultimate goal of SVM is to increase the margins between the hyperplane of two classes [7]. Primarily, SVM can only solve the problems related to binary classification but now they can also solve the multi class problems. SVM (Support Vector Machine) is a machine learning algorithm that comes under the supervised category and is used for binary classifications problems. The objective of this algorithm is to plot a hyper plane in an N-dimensional space, where N is the number of features that are going to be in a dataset, that distinctly classify the data points [13].

Maximal Margin Classifier: The distance between the line and closest data points is called a margin whereas the best line that can separate the two classes as the largest margin is called as Maximal-Margin hyperplane. This is a hypothetical classifier used to describe how SVM works. Like, if we have two input variables then it'll form a two-dimensional space and hyperplane is a line used to splits the input variable space. SVM consists of the two classes such as class 0 and class 1. So here to select the best separate points for the input variable space by their class one can create the hyperplane. This can be signified as in equation 1:

$$B0 + (B1 * X1) + (B2 * X2) = 0 \quad (1)$$

Now, B1 and B2 regulate the slope of a line and B0 is intercept. X1 and X2 are two input variables. By using this line one can make the classifications. We can make the use of these input variables in the line equation and estimate whether a new point is above or below the line.

Soft Margin Classifier: In real life problems, the two class datasets are only rarely linearly separable. The data in the real life is smudged and this data cannot be easily separated with the hyperplane. There are two types of deviations: An instance may lie on the wrong side of the hyperplane and be misclassified. Secondly, an instance may be on the right side but may lie in the margin i.e., not sufficiently away from the hyperplane.

Support Vector Machines Kernels: SVM algorithms make the use of set of mathematical functions called as kernel. The purpose of kernel is to take data as input and convert it into the required form. SVM algorithms are implemented using a kernel. In linear SVM, the learning of linear hyperplane is done by transforming the problems through the linear algebra. Linear SVM can be restated using the inner product of any two given observations. This is calculated as shown in equation2:

$$f(x) = B0 + \sum_{i=1}^{n}(a_i * (x * x_i)) \quad (2)$$

This is an equation used to calculate the inner product of a new input vector (x) with all support vectors in the given training data. B0 is the coefficient and a_i for each input values must be assessed after training data by learning algorithms.

Linear Kernel SVM: Linear kernel is used where the data is linearly separable and can be separated by using a single line. This is the one of the most useful kernels used in SVM. This is used where it'll comprise of large number of features for a certain dataset. In this linear kernel, the training of SVM is faster as compare to other kernels. This can be shown as equation 3:

$$K(x, x_i) = \sum(x * x_i) \quad (3)$$

Polynomial Kernel SVM: Polynomial kernel signifies the similarity of vectors for the training set of data in a feature space over polynomials of the original variables used in kernels. The degree of polynomial is given to the learning algorithms. This is same as linear kernel. This polynomial kernel is used for the curved lines in the input space.

Radial Kernel SVM: This is commonly used in support vector machine classifications. This kernel is a type of function which is used to approximate the other function as shown in equation 4:

$$K(x, x_i) = e^{-gamma * \sum(x - x_i^2)} \quad (4)$$

Here, gamma is used as parameter that specify the learning algorithm and the default value for gamma is 0.1 i.e., 0<gamma<1.

2) RNN-LSTM

Recurrent neural networks are the type of neural network which is designed to captures the information about the sequences or the time series data. They can take variable size inputs and gives us variable size outputs and these RNN works really well with time series data. RNN were designed to work with when the input data and output data is in the form of sequences like text of natural languages. Recurrent neural network has input layer, one or more recurrent layers and output layer. This recurrent layer depends upon the previous layer's output [6]. RNN includes some examples of sequence prediction problems like One-to-One, One-to-Many, Many-to-One and Many-to-Many etc. The RNN is used to assess the current feature on the basis of previous features.

Long short-term memory (LSTM) is special type of deep RNN comprises of multiple layers. Traditionally, the recurrent neural networks are difficult to train. The Long Short-Term Memory (LSTM), network is the most successful RNN because it overcomes the problems of training a recurrent neural network and in turn has been used as a wide range of applications. LSTM remember the past data in memory. The problem of vanishing gradient of RNN is fixed in the version [16]. Back-propagation is used to train the model. The LSTM network has three gates i.e., input gate, forward gate and output gate.

3) PSO(Particle Swarm Optimization)

PSO is a population based stochastic algorithm inspired by social behavior of bird flocking or fish schooling. Dr. Kennedy and Dr. Eberhart introduced the PSO algorithm in 1995. Particle is nothing but a small localized object which has some property like physical and chemical. Swarm is the collection of something that moves somewhere in large numbers and optimization is the action of making the best or most effective use of a situation/resource. We can consider it as possible solution to a given problem. PSO can solve the computationally hard optimization problems. This is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to given measure of quality. It solves a problem by having population of candidate solutions. There are two common swarm inspired approaches in computational intelligence areas i.e., Ant colony optimization (ACO), Particle swarm optimization (PSO)

PSO gives optimal solutions from a set of given solutions where each participating element represents a solution itself. Each participant has its own position and velocity vector associated with it. In PSO algorithm, we need to calculate the fitness function to maximize or minimize the positing of particles [23]. This is an iterative algorithm for each particle calculates the fitness and update if fitness function is better than the previous one.

B. Data Set

Dataset is obtained from the website https://data.gov.in and other govt. of Maharashtra websites. In these websites the agricultural datasets are uploaded by government agencies containing data for previous years. The data set file has the huge records of various crops of districts of Maharashtra state of India. The dataset is having variables like temperature, humidity, rainfall, crop yield, and soil etc. Generally, the data sets used in prediction of crop yield may be Historical datasets, Remote sensing datasets, GPS based datasets, social datasets.

C. Proposed Model

The proposed model is fundamentally a combination of three machine learning techniques i.e., SVM, RNN-LSTM and PSO. Initially the data set is divided into two parts in order to train and test the model. The proposed model is simulated as below[8]:

Step-1: Data Pre-processing Phase- Prepare the data to test the model.
Step-2: Train and test SVM-Input the data to SVM model
Step-3: Train and test RNN-LSTM-Input the data to RNN LSTM2
Step-4: Combine the output of SVM and RNN-LSTM using Hybrid function.
Step-5: Apply PSO algorithm for final yield prediction
Step-6: Recommendation to farmers based on model output

The complete illustration of the proposed model is shown in Fig3.

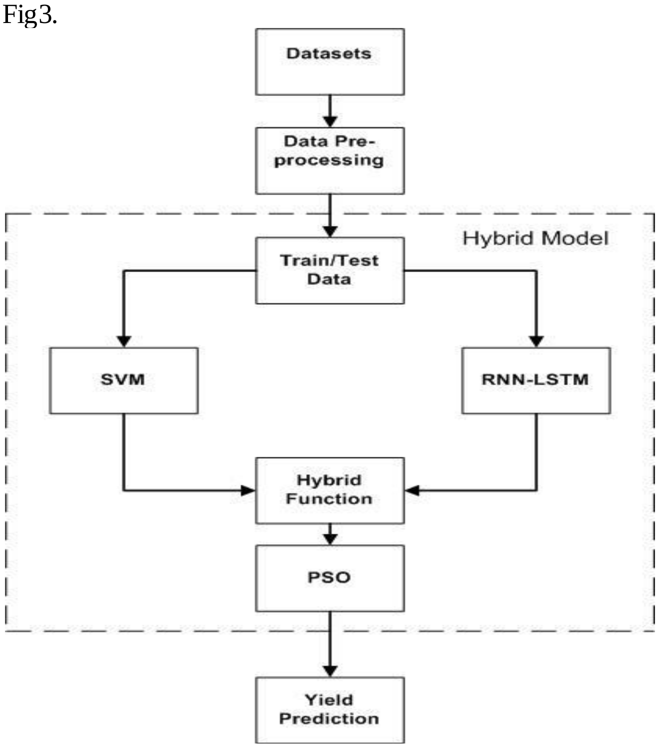

Fig 3. Design of basic hybrid model

The SVM Model is trained with three kernel functions. The kernel function may be linear, polynomial and radial basis functions. The outcomes of the functions are treated as outputs. Similarly, the RNN-LSTM model is trained with the same datasets. After the training of both the models separately, the output of both the models is average by hybrid function as below:

$$f(x) = \sum_i w_i \int_i x$$

Here, i represent the index for an algorithm and w represents the weight of f(x). The weighted average output is calculated by this hybrid function in order to make the accurate predictions. This output will be the input to the PSO algorithm. The PSO will declare the final prediction.

The main objective of this model is to predict the yield for the forthcoming years to facilitate the farmers and other stake holders in the agriculture domain. Also, the importance of the variables like temperature, rainfall etc. can be formulated for various crops.

IV. Experimental Results

The proposed model is realized on dataset obtained from various government agencies. After cleaning the last 10 (2010-2020) year's data of crop yield of Maharashtra State of India. The data which we have collected consists of the crop details and it contains the 10,000 rows and the data is stored in the csv format as shown in table 1. In order to analyze the data it is necessary to clean the data because the raw data is unstructured data, noisy and also consists of missing data. So, data preprocessing is done as the first step in machine learning models. With the preprocessed data, we can easily find out the dependent and independent factors for our model.in this model production factor is dependent factor and it changes with the various available independent factors like temperature, crop name, season and area.

STATE	DISTRICT	ZONE	CROP	YEAR	SEASON	AREA	PRODUCT	YIELD	MIN TEM	MAX TEM	RH	RAIN	SOIL
Maharashtra	Raigad	Konkan	Rice	2010	Kharif	124600	294800	2.37	32	21.4	64	3357.5	Laterite
Maharashtra	Ratnagiri	Konkan	Rice	2010	Kharif	78000	223100	2.86	30.2	21.8	53	3357.5	Laterite
Maharashtra	Sindhudurg	Konkan	Rice	2010	Kharif	73900	197000	2.67	32.1	24.2	68	3357.5	Laterite
Maharashtra	Raigad	Konkan	Gram	2010	Rabi	1800	1600	0.89	34.3	26.5	59	3357.5	Laterite
Maharashtra	Raigad	Konkan	Tur	2010	Kharif	1300	600	0.46	32	21.4	64	3357.5	Laterite
Maharashtra	Raigad	Konkan	Groundnut	2010	Kharif	100	100	1	30.2	21.8	53	3357.5	Laterite
Maharashtra	Raigad	Konkan	Moong	2010	Kharif	300	200	0.67	32.1	24.2	68	3357.5	Laterite
Maharashtra	Raigad	Konkan	Niger seed	2010	Kharif	400	100	0.25	32	21.4	64	3357.5	Laterite
Maharashtra	Raigad	Konkan	Other kharif pulse	2010	Kharif	300	100	0.33	30.2	21.8	53	3357.5	Laterite
Maharashtra	Raigad	Konkan	Ragi	2010	Kharif	10800	8300	0.77	32.1	24.2	68	3357.5	Laterite
Maharashtra	Raigad	Konkan	Sesamum	2010	Kharif	200	100	0.5	34.3	26.5	59	3357.5	Laterite
Maharashtra	Raigad	Konkan	Small Millets	2010	Kharif	3800	1700	0.45	32	21.4	64	3357.5	Laterite
Maharashtra	Raigad	Konkan	Urad	2010	Kharif	600	300	0.5	30.2	21.8	53	3357.5	Laterite
Maharashtra	Ratnagiri	Konkan	Tur	2010	Kharif	700	300	0.43	32.1	24.2	68	3357.5	Laterite
Maharashtra	Ratnagiri	Konkan	Groundnut	2010	Kharif	200	200	1	32	21.4	64	3357.5	Laterite
Maharashtra	Ratnagiri	Konkan	Moong	2010	Kharif	200	200	1	30.2	21.8	53	3357.5	Laterite
Maharashtra	Ratnagiri	Konkan	Niger seed	2010	Kharif	2200	300	0.14	32.1	24.2	68	3357.5	Laterite
Maharashtra	Ratnagiri	Konkan	Other kharif pulse	2010	Kharif	800	400	0.5	34.3	26.5	59	3357.5	Laterite
Maharashtra	Ratnagiri	Konkan	Ragi	2010	Kharif	17000	20700	1.22	32	21.4	64	3357.5	Laterite
Maharashtra	Ratnagiri	Konkan	Sesamum	2010	Kharif	100	50	0.5	30.2	21.8	53	3357.5	Laterite
Maharashtra	Ratnagiri	Konkan	Small Millets	2010	Kharif	7000	4300	0.61	32.1	24.2	68	3357.5	Laterite

Table 1. Crop Dataset

A. Yield Prediction using Test Data:

The main objective is to predict the yield for various crops cultivated across the Maharashtra, India. For the processing of the given model, the 70% of data is used for training and 30% is used for testing dataset. The predicted values for the test data for the years 2010 to 2020 is shown in table and graph in table 2:

Table 2: Percentage error for Kharif crops using Hybrid Model

Crop Year	Actual Yield (AY)	Predicted Yield (PY)	Difference (DF)	Percentage Error
2010	314.64	320.21	5.57	1.77
2011	297.16	312.86	15.7	5.28
2012	251.44	276.66	25.22	10.03
2013	300.53	314.45	13.92	4.63
2014	186.51	190.32	3.81	2.04
2015	150.44	167.98	17.54	11.66
2016	302.86	325.23	22.37	7.39
2017	243.83	260.34	16.51	6.77
2018	200.31	226.83	26.52	13.24
2019	45.02	49.67	4.65	10.33
2020	28.95	32.75	3.8	13.13

Based on above table it is found that the average percentage error is 7.84 for kharif crop using hybrid model. The year wise actual and predicted yield is shown in fig.4.

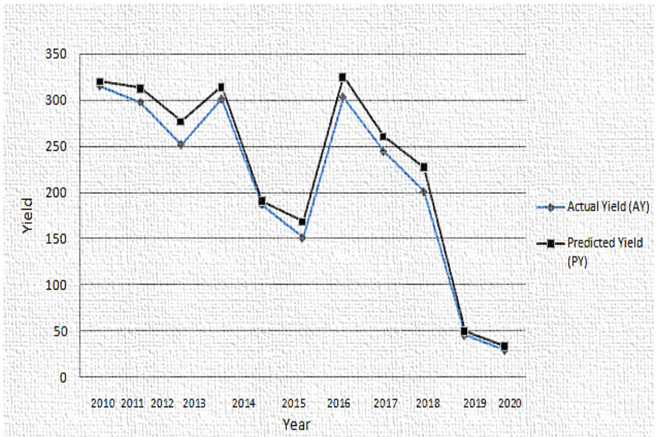

Fig.4. Percentage error for Kharif crops using Hybrid Model

Similarly, for rabi crops predictions are made using the hybrid model is described in Table 3. as shown below:

Table 3: Percentage error for Rabi crops using Hybrid Model

Crop Year	Actual Yield (AY)	Predicted Yield (PY)	Difference (DF)	Percentage Error
2010	253.50	261.65	8.15	3.21
2011	208.86	214.45	5.59	2.68
2012	157.02	163.11	6.09	3.88
2013	179.24	184.18	4.94	2.76
2014	154.74	159.46	4.72	3.05
2015	149.78	155.23	5.45	3.64
2016	798.38	815.23	16.85	2.11
2017	216.58	224.78	8.2	3.79
2018	156.11	159.67	3.56	2.28
2019	68.97	75.11	6.14	8.90
2020	46.14	50.34	4.2	9.10

Based on above table it is found that the average percentage error is 4.13 for rabi crops using hybrid model. The year wise actual and predicted yield is shown in fig. 5.

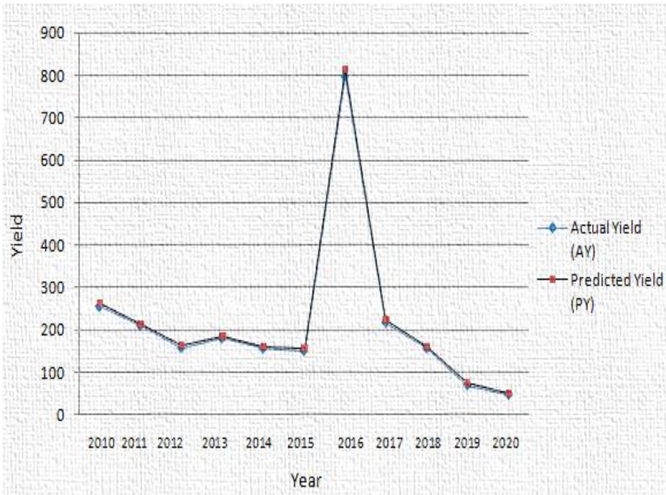

Fig 5. Percentage error for Rabi crops using Hybrid Model

V. Data Analysis and Discussion

The outcome of the proposed model is evaluated with the traditional machine learning algorithms like SVM, KNN and RNN-LSTM in order to check the effectiveness. The metrics used to measure the performance of the model are stated in Table 4. These matrices envisage about the accuracy and strength of the model.

Table 4: Performance Metrices

S.No.	Matrices	Equations
1.	Mean Error (ME)	$\frac{\sum_{i=1}^{n}(P_i - A_i)}{n}$
2.	Mean Percentage Error (MPE)	$\frac{1}{n}\sum_{i=1}^{n}\frac{P_i - A_i}{P_i}$
3.	Mean Absolute Error (MAE)	$\frac{1}{n}\sum_{i=1}^{n}\lvert P_i - A_i\rvert$
4.	Mean Absolute Percentage Error (MAPE)	$\frac{1}{n}\sum_{i=1}^{n}\frac{\lvert P_i - A_i\rvert}{P_i}$
5.	Coefficient of Determination (R^2)	$\left(\frac{\sum_{i=1}^{n}(P_i-\hat{P}_i)\sum_{i=1}^{n}(A_i-\hat{A}_i)}{\sqrt{\sum_{i=1}^{n}(P_i-\hat{P}_i)2\sum_{i=1}^{n}(A_i-\hat{A}_i)2}}\right)$
6.	Root Mean Square Error (RMSE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - A_i)^2}$

7.	Accuracy	$\frac{TN+TP}{TN+TP+FN+FP}$

Where, n denotes the length of testing datasets, P_i is the predicted data and A_i is the actual data. $\hat{P}_i$ and $\hat{A}i$ are the mean values for the predicted and actual datasets. The hybrid model always outperforms than the machine learning model and the same way this proposed model will have great accuracy as compared to other models. The overall accuracy is shown in below table 5 and figure 6.

Table 5. Accuracy of Machine Learning Models

Machine Learning Model	Overall Accuracy
SVM	84
KNN	75
RNN-LSTM	82
HYBRID (SVM+RNN-LSTM)	89

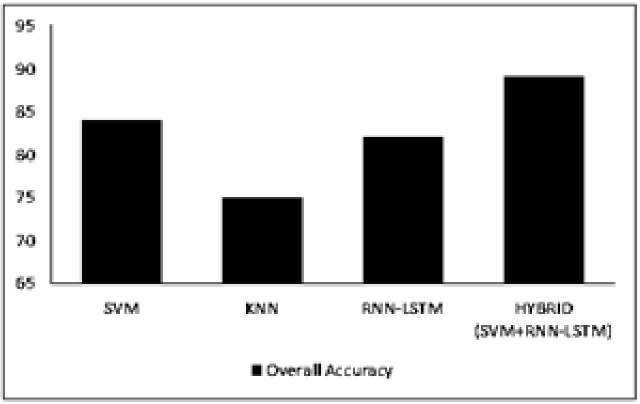

Fig. 6 Accuracy of Machine Learning Models

VI. CONCLUSION

The early and accurate crop yield prediction is having great importance for various stakeholders like farmers, governments, agencies etc. to plan the activities associated with yield i.e., market planning, insurance, storage, food security etc. The two-machine learning and deep learning model are combined and further optimize with PSO to visualize the hybrid model. It is entrusted that the hybrid model is more powerful than the individual models. The algorithms used in this study are having innumerable advantages, thus performance is excellent. The motive of this research is that the efficient algorithms and multidimensional dataset can entrust farmers with early crop yield predictions and recommendations. The various combinations of model can be combined and tested for complex prediction problems. We will implement these algorithms in future for crop yield predictions to strengthen the farmers by creating a simple GUI.

ACKNOWLEDGMENT

Sincere gratitude is given to my mentor Dr. Sajidullah Khan, School of Computer Sciences & Engineering, Sandip University, Nasik for their motivation and support for this work. We thank the all-support staff of Sandip University, Nashik for providing the necessary infrastructure for the research work.

REFERENCES

[1] Sonal Agarwal and Sandhya Tarar," A Hybrid Approach for Crop Yield Prediction using Machine Learning and Deep Learning Algorithms: Journal of Physics: Conference Series 1714 (2021) 012012.

[2] U. Muthaiah, S. Markkandeyan, Y. Seetha," Classification Models and Hybrid Feature Selection Method to Improve Crop Performance", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-11S2, September 2019.

[3] Saeed Nosratabadi, Karoly Szell, Bertalan Beszedes, Felde Imre, Sina Ardabili, Amir Mosavi, "Hybrid Machine Learning Models for Crop Yield Prediction".

[4] Shivi Sharma, Geetanjali Rathee, Hemraj Saini, "Big Data Analytics for Crop Prediction Mode Using Optimization Technique", 5th IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC-2018), 20-22 Dec, 2018, Solan, India.

[5] Ms. Shreya V. Bhosale, Ms. Ruchita A. Thombare, Mr. Prasanna G. Dhemey, Ms. Anagha N. Chaudhari," Crop Yield Prediction Using Data Analytics and Hybrid Approach", 2018 Fourth International Conference on Computing Communication Control and Automation.

[6] Sun Jie., Di Liping., Sun Z., Shen Y., Lai Z. County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model. Sensors. 2019;19:4363. doi: 10.3390/s19204363.

[7] Sujatha, R., Isakki, P., "A study on crop yield forecasting using classification techniques", International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE), pp.1-4, 2016.

[8] Rakesh Kumar, M.P. Singh, Prabhat Kumar, J.P. Singh, "Crop Selection Method to maximize crop yield rate using machine learning technique",2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM),27 August 2015.

[9] J. Liu, C. E. Goering, Lei Tian, 2001. "A neural network for setting target corn yields". Transactions of the American Society of Agricultural Engineers44 (3):705-713.

[10] Rossana MC, L. D. (2013). "A Prediction Model Framework for Crop Yield Prediction". Asia Pacific Industrial Engineering and Management Society Conference Proceedings Cebu, Philippines, 185.

[11] Naveen L., Mohan H S., 2019. Analyzing Impact of Weather Forecasting Through Deep Learning in Agricultural Crop Model Predictions International Journal of Applied Engineering Research, p. 4379-4386.

[12] Dhivya B H., Manjula R., Siva Bharathi S., 2017. Madhumathi R4. A Survey on Crop Yield Prediction based on Agricultural Data, p.4177-4183.

[13] S. Ying-xue, X. Huan, Y. Li-jiao, Support Vector Machine-Based Open Crop Model (SBOCM): Case of Rice Production in China, Saudi Journal of Biological Sciences (2017), doi: http://dx.doi.org/10.1016/j.sjbs.2017.01.024.

[14] Saeed Khaki, Lizhi Wang, "Crop Yield Prediction Using Deep Neural Networks", Front. Plant Sci., 22 May 2019 | https://doi.org/10.3389/fpls.2019.00621.

[15] Saeed Khaki, Lizhi Wang, Sotirios V. Archontoulis, "A CNN-RNN Framework for Crop Yield Prediction", Front. Plant Sci., 24 January 2020 | https://doi.org/10.3389/fpls.2019.01750.

[16] Sagarika Sharma, Sujit Rai, Narayanan C. Krishnan, "Wheat Crop Yield Prediction Using Deep LSTM Model", 3 Nov. 2020.

[17] Maya Gopal P. S. & Bhargavi R. (2019): Performance Evaluation of Best Feature Subsets for Crop Yield Prediction Using Machine Learning Algorithms, Applied Artificial Intelligence, DOI: 10.1080/08839514.2019.1592343.

[18] Preetam Tamsekar, Nilesh Deshmukh, Parag Bhalchandra, Govind Kulkarni, Kailas Hambarde and Shaikh Husen, "Comparative Analysis of Supervised Machine Learning Algorithms for GIS-Based Crop Selection Prediction Model", Computing and Network Sustainability, Lecture Notes in Networks and Systems 75, https://doi.org/10.1007/978-981-13-7150-9_33.

[19] K.A. Shastry and H.A. Sanjay, Hybrid prediction strategy to predict agricultural information, Applied Soft Computing Journal (2020), doi: https://doi.org/10.1016/j.asoc.2020.106811.

[20] Suresh Kumar N, Sarumathi R, Subiksha A, Yamini K, Prasanna Vasudevan, "An advanced prediction of crop yield using SVM classifier", International Journal of Advanced Science and Technology. Vol. 29, No. 7, (2020), pp. 10483-10488.

[21] Nischitha K, Dhanush Vishwakarma, Mahendra N, Ashwini, Manjuraju M.R, "Crop Prediction using Machine Learning Approaches", International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV9IS080029 Published by: www.ijert.org Vol. 9 Issue 08, August-2020.

[22] Kiran Kumar Paidipati, Christophe Chesneau, B. M. Nayana, Kolla Rohith Kumar, Kalpana Polisetty and Chinnarao Kurangi, "Prediction of Rice Cultivation in India- Support Vector Regression Approach with Various Kernels for Non-Linear Patterns". Agri Engineering 2021, 3, 182–198. https://doi.org/10.3390/ agriengineering3020012.

[23] Jinglin Du, Yayun Liu, Yanan Yu and Weilan Yan A, "Prediction of Precipitation Data Based on Support Vector Machine and Particle Swarm Optimization (PSO-SVM) Algorithms", Algorithms 2017, 10, 57; doi:10.3390/a10020057.

[24] S. Manimekalai, K. Nandhini, "An Optimized Feature Selection and Classification of Soil Parameters for Crop Yield Prediction", International Journal for Research in Engineering Application & Management (IJREAM) ISSN: 2454-9150 Vol-04, Issue-05, Aug 2018.

AUTHOR PROFILE

Kusum Lata is a Research Scholar in the Department of Computer Science and Engineering, Sandip University, Nashik. Her broad research area is machine learning.

Dr. Sajidullah S. Khan, Associate Professor, Department of Computing Sciences and Engineering, Sandip University, Nasik.