

## RESEARCH ARTICLE

# Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning

PRIYANKA SHARMA<sup>1</sup>, PANKAJ DADHEECH<sup>1</sup>, (Member, IEEE), NAGENDER ANEJA<sup>2</sup>,  
AND SANDHYA ANEJA<sup>3</sup>, (Member, IEEE)

<sup>1</sup>Department of Computer Science and Engineering, Swami Keshvanand Institute of Technology, Management and Gramothan, Jaipur, Rajasthan 302017, India

<sup>2</sup>School of Digital Science, Universiti Brunei Darussalam, Gadong BE1410, Brunei Darussalam

<sup>3</sup>School of Computer Science and Mathematics, Marist College, Poughkeepsie, NY 12601, USA

Corresponding author: Nagender Aneja (nagender.aneja@ubd.edu.bn)

**ABSTRACT** Agriculture contributes a significant amount to the economy of India due to the dependence on human beings for their survival. The main obstacle to food security is population expansion leading to rising demand for food. Farmers must produce more on the same land to boost the supply. Through crop yield prediction, technology can assist farmers in producing more. This paper's primary goal is to predict crop yield utilizing the variables of rainfall, crop, meteorological conditions, area, production, and yield that have posed a serious threat to the long-term viability of agriculture. Crop yield prediction is a decision-support tool that uses machine learning and deep learning that can be used to make decisions about which crops to produce and what to do in the crop's growing season. It can decide which crops to produce and what to do in the crop's growing season. Regardless of the distracting environment, machine learning and deep learning algorithms are utilized in crop selection to reduce agricultural yield output losses. To estimate the agricultural yield, machine learning techniques: decision tree, random forest, and XGBoost regression; deep learning techniques - convolutional neural network and long-short term memory network have been used. Accuracy, root mean square error, mean square error, mean absolute error, standard deviation, and losses are compared. Other machine learning and deep learning methods fall short compared to the random forest and convolutional neural network. The random forest has a maximum accuracy of 98.96%, mean absolute error of 1.97, root mean square error of 2.45, and standard deviation of 1.23. The convolutional neural network has been evaluated with a minimum loss of 0.00060. Consequently, a model is developed that, compared to other algorithms, predicts the yield quite well. The findings are then analyzed using the root mean square error metric to understand better how the model's errors compare to those of the other methods.

**INDEX TERMS** Agriculture, crop yield prediction, decision tree, machine learning, deep learning.

## I. INTRODUCTION

Agriculture is vital to humanity, not only for food but also for employment and the economy. Crops or arable land are relatively "new," even though people have been eating grains and plants for over 1,00,000 years. Around 11,000 years ago, during the Neolithic era, often referred to as the New Stone Age, people first actively managed the land and its vegetation. In India, agriculture provides a considerable portion of the country's economic support and the majority of the country's food needs. Due to India's

rapid population growth and important climatic changes, the demand chain and food supply must be maintained. Several scientific approaches have been included in agriculture to preserve the harmony between the supply and demand of food. The significant climatic variance makes it difficult for farmers to choose how to be more flexible and sustainable [1]. With modern technology and innovative farming techniques, agriculture must produce more with fewer inputs. Therefore, crop production estimation is crucial in identifying issues with food security. In India, Agriculture uses 70 percent of the water worldwide [2].

More than 50% of Indian workers were employed in agriculture in 2018, which contributed 17%-18% of the

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

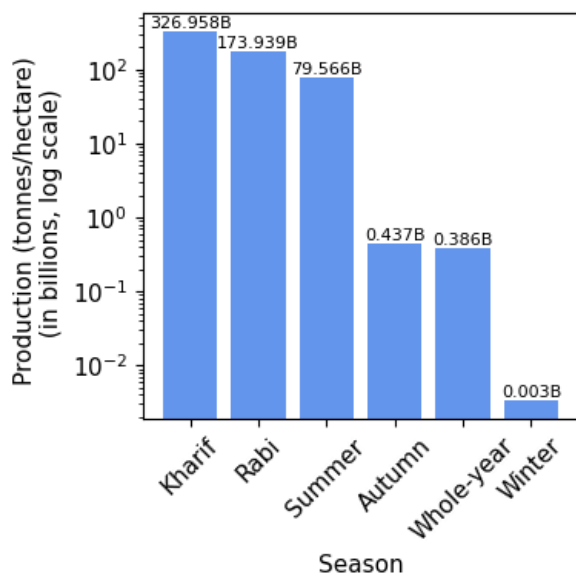


FIGURE 1. Season-wise crop production.

country's GDP. India's national area held steady at about 3,28,726 thousand hectares between 1971 and 2020 [3]. By 2025, India's agricultural industry is expected to grow to USD 24 billion. With 70 percent of sales coming from retail, India has the world's 6th largest food and grocery market. According to preliminary estimates for the fiscal year 2022-2023, the nation will produce 149.92 million tons of food grains altogether (Kharif alone). India's rapid population growth primarily drives the sector.

Additional evidence favoring these results from increasing income levels in urban and rural locations, boosting the demand for agricultural products across the country. As a result, in addition to the introduction of numerous e-farming applications, the increased utilization of cutting-edge technologies is driving sectors such as geographic information systems (GIS), artificial intelligence (AI), blockchain and remote sensing technologies, and drones.

Figure 1 shows the season-wise analysis of crop production for Kharif, Rabi, Summer, Autumn, and Winter seasons and the year. The crop production is maximum in the Kharif season and minimum in winter.

Farmers can boost production under favorable conditions and decrease production loss under unfavorable conditions by using crop yield estimation. Numerous variables, such as farmer practices, choices, pesticides, fertilizers, weather, and market prices, impact crop yield positive forecasts. It is possible to estimate crop yield using statistical information from yields of previous years, together with climate, area-wise output, and rainfall. Recently, machine learning has advanced in various industries, including agriculture.

Numerous machine learning approaches, such as decision trees, artificial neural networks, support vector machines, and

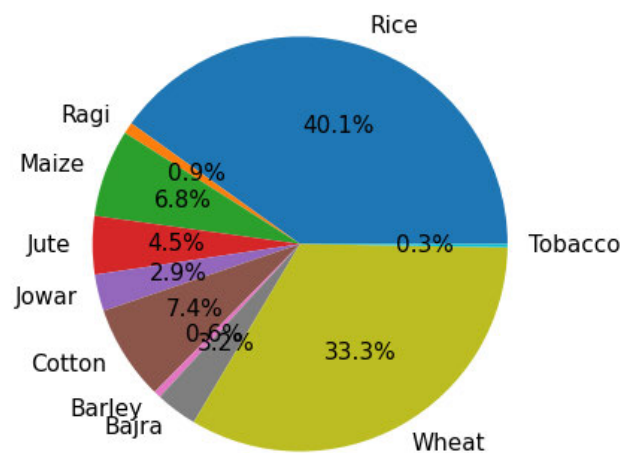


FIGURE 2. Average yearly crop production 1997-2020 [4].

deep learning [5], have been applied to calculate agricultural production.

The crop production for India from 1997 to 2020 is shown in Figure 2. From the analysis shown, it can be observed that the majority of the land in the country was planted with wheat and rice, which produced more than 73% of the nation's staple grains.

India accounts for around 40 percent of global rice (Basmati and Non-Basmati) trade and exports to more than 150 countries. Trade ministry data showed that exports rose 11 percent to 2.16 million tons [6] in the first half of 2022-23.

The illustration shows that rice has the highest production and area allocation. The agricultural industry had strong growth in terms of exports over the past year. Exports of rice, basmati and non-basmati totaled USD 6.12 billion in FY22 (through December 2021), as represented in Figure 3.

As farming developed throughout the centuries, people ceased utilizing the stars and animal sacrifices to manage their crops in favor of ever-more-scientific techniques. Mechanization, mathematics, and scales were significant as farming expanded throughout the industrial revolution. Modern approaches were being used more and more. By the 1900s, farming was using regression analysis to gauge agricultural output. A statistical method known as regression analysis combines many statistical methods to determine the correlations between a dependent and independent variable. Regression models for crop forecasting frequently rely on inputs like weather, soil characteristics, and past crop yields. For almost a century, farming has now employed these models. Regression approaches became more popular as relevant data and processing power became more widely accessible. Classification and fruit recognition have also been emerging areas in image classification and computer vision in the agriculture sector [7].

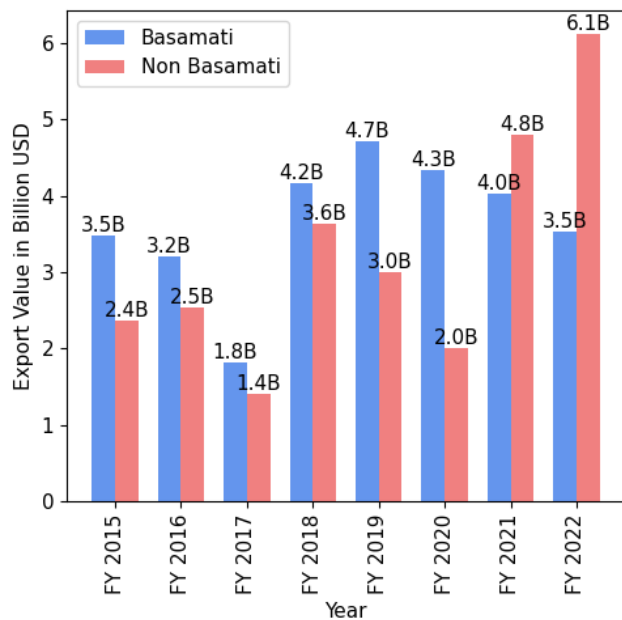


FIGURE 3. Rice exports from India in 2015-2022.

### A. RESEARCH CONTRIBUTIONS

India is a predominantly agrarian nation with around 48.9% of the population is employed in agriculture in some capacity. One of the most pressing issues is farmers' suicides. Farmers are killing themselves, which is a national calamity. According to a report of the Indian National Crime Records Bureau (NCRB), more than 2 lakh Indian farmers committed suicide between 1995 and 2020. This serious issue tends to come from farmers' failure to repay loans typically obtained from banks and private landlords. This is because many losses used to occur due to inadequate knowledge about the crops, such as which crop should be sown in which season. The same insight served as the impetus for this study. By utilizing historical agricultural production statistical data for India, a robust mechanism is needed to estimate yield production at the initial stage. Early foresight can assist farmers to avoid committing suicide, suffering financial losses from their crops, and ultimately affecting the nation's GDP. Forecasting agricultural yields isn't a simple task as several factors like rainfall, wind speed, soil characteristics, climate, humidity, temperature, etc., affect agricultural production, and no single dataset is available for the same; this data needs to be collected from multiple sources. Although the same concern has been the subject of numerous studies, improved performance is still preferred. Unquestionably, the most captivating scientific frontiers that inspire academics to investigate further aspects of prediction are combined with machine learning and deep learning. The primary contributions of this research are to propose a reliable agricultural yield prediction method by analyzing the data gathered from official government websites and then applying various ML and DL models to improve accuracy

metrics and test losses while offering a standard for contrast in this area of research. This study will help the researchers to understand the problem and solution domains effectively and will enhance their knowledge.

### B. ARTICLE ORGANIZATION

The objective of this study is to predict agriculture yields. For a thorough understanding of the literature, Section II looks closely at the theories and related research about crop yield prediction according to the investigation's goals and the research gaps found in the literature. Section III describes the study region, data sources, technological perspective, research methodology, and approaches for agriculture yield prediction, along with the relationships among various dataset features. Section IV shows the model performance of various approaches (i.e., results). Section V concludes the entire work and also suggests additional work.

## II. LITERATURE REVIEW

By using machine learning approaches, crop yield estimation can be achieved. The dataset, which contains the total area under cultivation, the canal length, the average maximum temperature, and the water sources (tanks and wells) for irrigation, were used to forecast the crop output. According to a study, the computational model developed was superior to those produced using Regression Tree, Lasso, Deep Neural Network, and Shallow Neural Network approach. RMSE is fifty percent of the standard deviation and twelve percent of the average yield for the dataset validation using projected weather data [8].

From 1998 to 2002, during the Kharif season, the accuracy was 97.5 percent using parameters: min/max/average temperatures, area, rainfall, production, and yield [9]. The study focused on estimating crop yields for the Kharif season in the Vishakhapatnam district of Andhra Pradesh. Rainfall significantly influences Kharif crops' production; hence, the scientists used modular artificial neural networks to forecast rainfall first before utilizing support vector regression to estimate crop output using rainfall and area data. To boost crop yield, these two approaches were used.

The research focused on the following four goals: investigating the Artificial Neural Network (ANN) model to forecast soybean and corn yields under unfavorable weather conditions; examining the model's capacity for regional, state, and local estimation; assessing the performance of the ANN concerning variation of the parameter; and evaluating the evolved ANN model in comparison to additional models of multivariate linear regression [10]. The study used artificial neural networks to assess rice output in various cities in Maharashtra, India. Maharashtra's 27 districts' information was gathered from the open records of the Indian government.

The study estimates superior crop yield using ML methods such as Support Vector Regression (SVR), Random Forest, Artificial Neural Network (ANN), and K-Nearest Neighbor (KNN). Seven hundred forty-five examples comprise the research's data set, with 70 percent of those cases being

randomly assigned for model training and the remaining 30 percent for testing and performance evaluation. According to the final analysis, Random Forest achieves the highest level of accuracy [11]. Utilizing Long-Short Term Memory (LSTM), satellite data in southern Brazil, the research suggests a unique model to forecast soybean yield.

The primary aim of the research is to compare the effectiveness of LSTM neural networks, random forest, and multivariate OLS linear regression [12]. Rainfall, Land surface temperature, and Vegetation indices are used as independent variables in the forecasting process for soybean data, and step two is to determine how soon the model can reasonably anticipate the yield. For all forecasts except DOY 16, Long Short Term Memory outperforms all other algorithms. Multivariate OLS linear regression outperforms all other algorithms for DOY 16 [13]. The results of using a Sequential Minimal Optimization Classifier are analyzed in this research. The experiment was conducted using the WEKA tool and data from 27 districts in Maharashtra, India.

According to the experiment's results on the same dataset, other strategies outperform Sequential minimum optimization. Sequential minimum optimization demonstrated the minimum accuracy and bad quality, whereas Multilayer Perceptron and BayesNet demonstrated the best accuracy and improved quality [14].

Crop productivity estimation is proposed using a Deep Belief Network (DBN) and Parallel Layer Regression (PLR) approach. Here, five major crops flourishing in Karnataka: pulse, ragi, and rice, are the subject of a DBN approach. The suggested methodology predicts each location in the relevant database to grow one of the five crops. Last but not least, the experimental findings demonstrate that the approach has a significant potential for accurate prediction of agricultural productivity in terms of accuracy, specificity, and sensitivity that its effectiveness has been validated using real-time data and interactions with people [15].

Approaches for machine learning to estimate agricultural yield: linear/ lasso/ridge regression and decision trees have been used. Some machine learning methods were inferior to the Decision tree [16].

A Crop Yield Prediction System (CYPS) is implemented using a KNN algorithm. For a farmer, however, yield predictions must be based on various factors that could influence crop production and quality. Season, crop type, and production area are three aspects that influence yield production; therefore, authors use specific fields like year, crop, area, region, and season to predict yield production. Making decisions related to agricultural risk management requires accurate knowledge of crop yield history [17].

KNN, Random Forest, and Decision Tree Classifier were examined by Rao et al. [18] using two distinct metrics: Gini and Entropy. Findings show that Random Forest has given the most accurate results.

Good performances of 91.35% and 91.17% have been obtained for VGG\_19 and VGG\_16 based on feature vectors [19]. Vanipriya et al. [20] have proposed using Hydroponics

to deal with low agricultural production issues in India as it has high efficiency. Additionally, it offers soil cultivation a greener alternative. Food production depends on the economy and also on the yield of agricultural production [21].

### III. PROPOSED WORK

#### A. STUDY AREA

India is the subject of this study because of its diverse climate, which ranges from humid to arid in the south to temperate alpine in the north. It spans 3,28,726 thousand hectares [22] from the Himalayan peaks capped in the snow to the rain forests in southern central tropical, out of which 2,97,319 thousand hectares (between 1971 and 2020) is for agricultural usage [3]. To conduct this study, ten key Indian crops (majorly grown crops)- rice, jowar, maize, bajra, tobacco, jute, barley, ragi, cotton, and wheat have been chosen.

#### B. DATA SOURCES

For creating the dataset, the information was gathered from the publicly available official websites <https://data.gov.in> and <https://aps.dac.gov.in> (i.e. multiple sources) [23], [24]. The dataset generated contains the details for the years 1997 to 2020: State name, district name, crop year, seasons, crop type, rainfall, wind speed, humidity, area under irrigation, area, production, and yield. Figure 2 depicts the crop's yield for the years mentioned. To forecast the crop yield for India, this study uses a Decision Tree, Random Forest, XG Boost Regression, Convolutional Neural Network, and Long-short Term Memory Network. Furthermore, accuracy, standard deviation, root mean squared error and test loss are used for validation.

The flow of the work done is demonstrated in Figure 4. Here, the dataset is partitioned into training and testing data. 70 percent of the entire data is selected for training purposes, and the remaining 30 percent is used for validation. On the training dataset, k-fold cross-validation is employed. The number of folds taken is four, and training data is randomly assigned to each fold. The trained model is then evaluated.

#### C. METHODS

##### 1) K-FOLD CROSS VALIDATION

In k-FOLD Cross Validation, k-1 folds are used as training data to train the model, and rest of the data is used to validate the resulting model.

Accuracy, the performance metric, is computed at each step while this process is performed k times. k-Fold cross validation guarantees that in both the training and test sets, every observation from the original dataset has a chance to appear. If we just have a small amount of input data, this is one of the finest approaches. The flow chart for the proposed model is shown in Figure 5.

##### 2) DECISION TREE

Classification and regression issues are resolved using a decision tree that performs two essential functions: firstly,

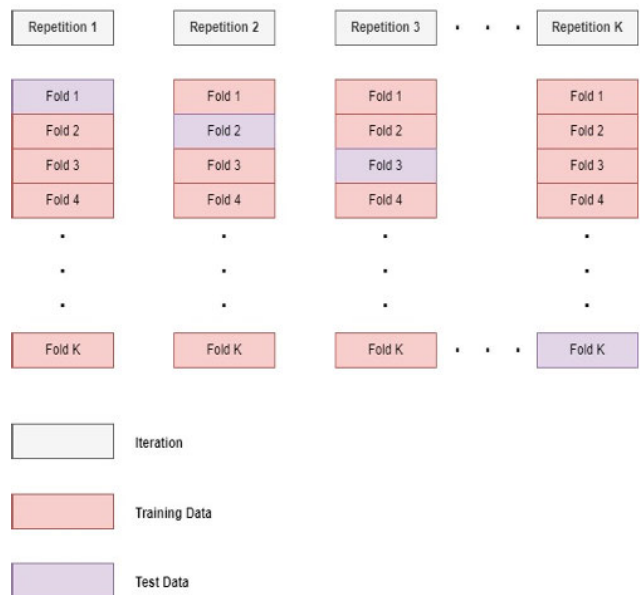


FIGURE 4. The k-fold cross-validation.

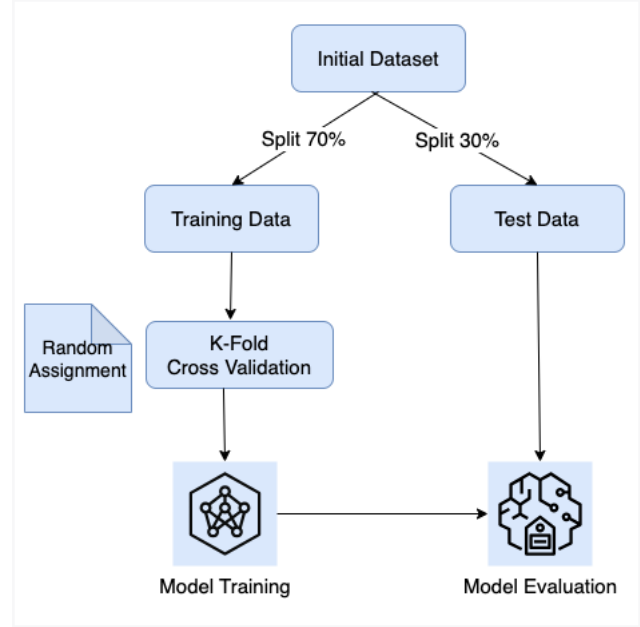


FIGURE 5. Flow chart of the approach used.

it categorizes the features that are pertinent to each decision, and secondly, it determines the best course of action based on the selected features. The plausible choice is given a probability distribution by the Decision Tree algorithm [25]. Here, every node represents a feature, the branch devotes to a selection, and the leaf node denotes the outcome. One characteristic should be the decision tree's root node to begin tree production. Data splitting is then necessary to finish the decision tree.

### 3) RANDOM FOREST

An uncorrelated forest of decision trees is built by amalgamating the bagging approach and feature randomness,

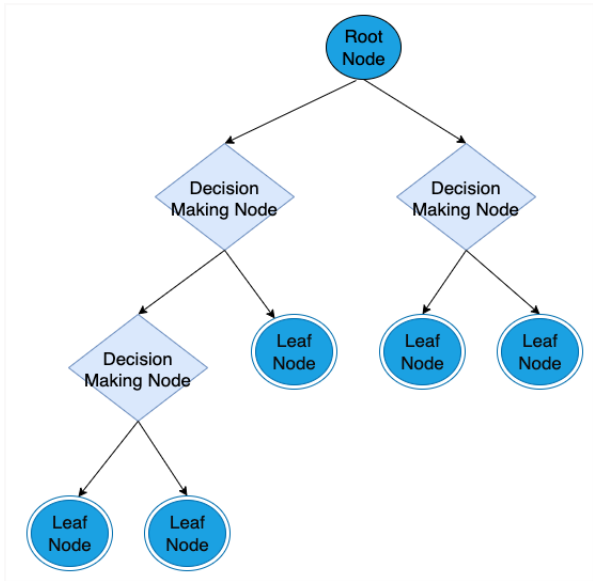


FIGURE 6. Decision tree.

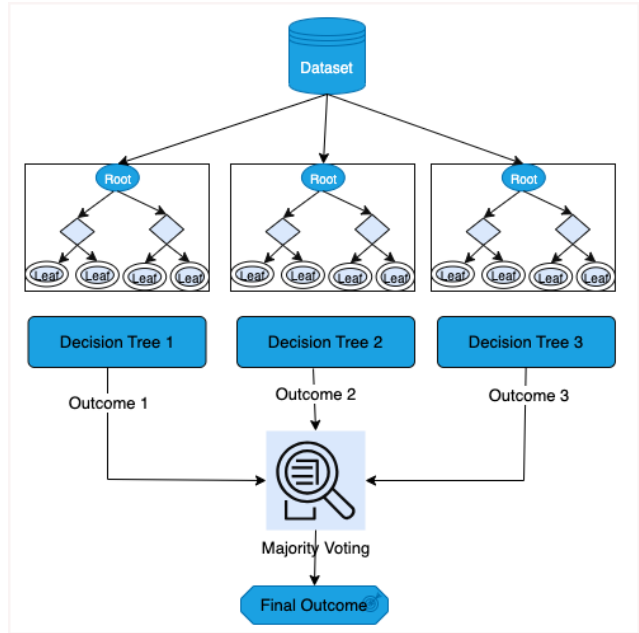


FIGURE 7. Random forest.

i.e., the bagging approach's extension is RF [26]. Low correlation across decision trees is ensured via feature bagging or feature randomization, or the random subspace technique, which provides a random selection of features. RF only chooses a portion of the feature splits that decision trees can take into account [27].

Each decision tree in the ensemble is built using the bootstrap sample, a data sample obtained from a training set. The rest 1/3rd of the training sample is used as test data.

### 4) XGBoost

A popular and successful open-source alternative of the gradient-boosted trees method is called Extreme Gradient Boosting, i.e., XGBoost. Regression trees [28] act as weak



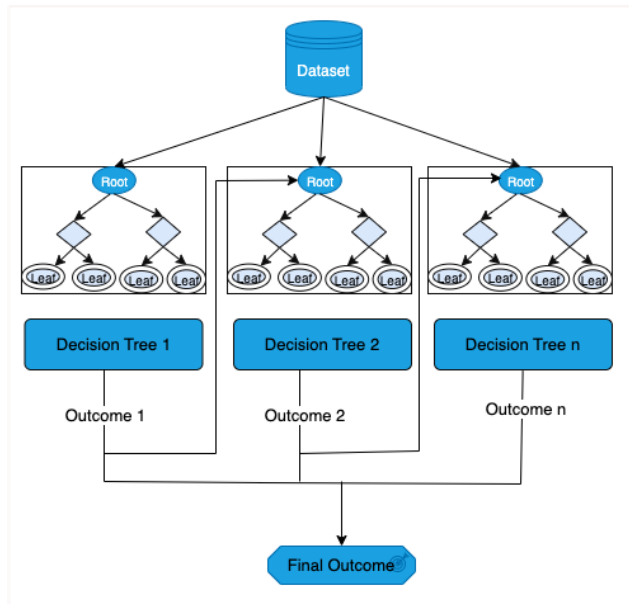


FIGURE 8. XGBoost.

learners when gradient boosting is used for regression. With the use of a continuous score, a connection is established between every input data point and a leaf. XGBoost's convex loss function takes a penalty term for model complexity to minimize a regularized (L1 and L2) objective function/regression tree functions [29].

Repeatedly, new trees are added to the training process for predicting the errors/residuals of previous trees, which are then amalgamated with preceding trees to make the ultimate forecast. Since the technique reduces the loss while adding new models, it is frequently called "gradient boosting" [30].

Decision tree, random forest, and XGBoost are each represented graphically in Figures 6, 7, and 8, respectively.

## 5) CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolution Neural Networks use some mathematical operations in between their layers called Convolution. This implementation of CNN has a total of 7 layers. The first layer is a Conv1d layer with 64 filters of the size (3\*3), each with a kernel size of 3. Layer 2 (MaxPooling1D layer) has a pool size of two. Dropout is visible on layer 3. The output from layer three is flattened by layer four, which uses the ReLU activation function, and this output is then sent to layer 5. A neural network's hidden layer 6 has 330 neurons in it. The output layer, layer 7, uses the SoftMax function and has 11 neurons for 11 kinds of output. CNN architecture of the model used is shown in Figure 9.

## 6) LONG SHORT-TERM NETWORK (LSTM)

A type of recurrent neural network (RNN) called a Long Short-Term Memory (LSTM) network, shown in Figure 10, specifically handles sequential data, such as text, speech, and time series. Sequential data can be used to teach LSTM

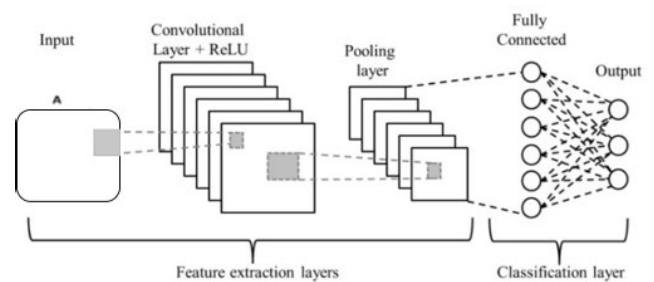


FIGURE 9. CNN architecture.

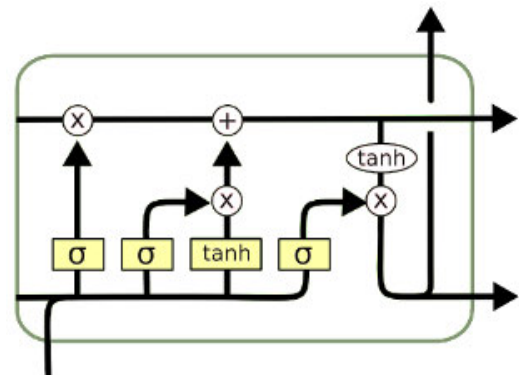


FIGURE 10. Working of LSTM.

networks long-term dependencies. For both models, i.e., CNN and LSTM, the optimizer used is ADAM optimizer that is used to change or tune the attributes of a neural network [31], [32] such as layer weights, learning rate, etc., and it is a combination of "RMSP" and "gradient descent with momentum" algorithm. The loss is calculated using MSE (i.e., mean squared error). For training the model, the number of epochs taken is 50, the batch size is 32, and the validation split is 0.2. LSTM architecture of the model used is shown in Figure 10.

## IV. RESULTS

The features of the dataset are related to one another. The authors identified the crop as a significant feature and plotted the link. Figure 11 shows the production count for each of the 10 crops taken consideration in the study. Figure 12 shows the historical correlations between Area and Crop Type. It can be observed that the maximum area is allotted to wheat, followed by Rice.

Figures 11 and 12 show the historical correlations between the dataset's features, such as Crop type, Area, and Production.

Table 1 presents the forecast outcomes. The study concludes that Random Forest outperforms other machine learning algorithms regarding accuracy.

Using statistical data, Random Forest produces India's most accurate crop production estimation with accuracy = 98.96 %, MAE = 1.97, RMSE = 2.45, and SD = 1.23

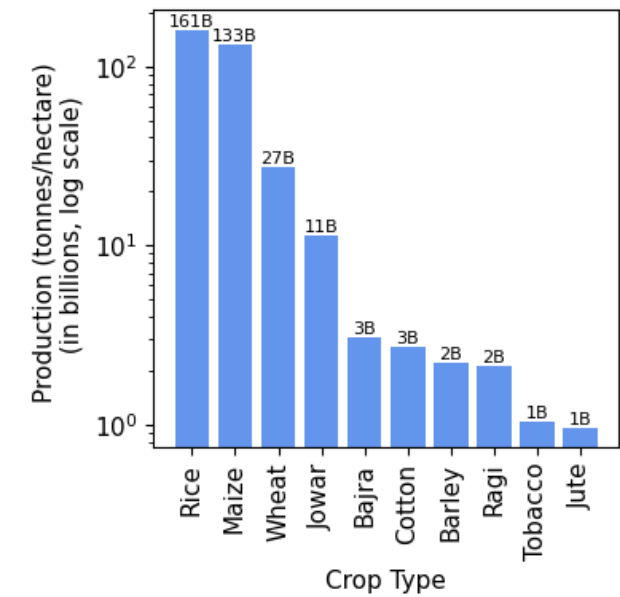


FIGURE 11. Relationship between crop and Production.

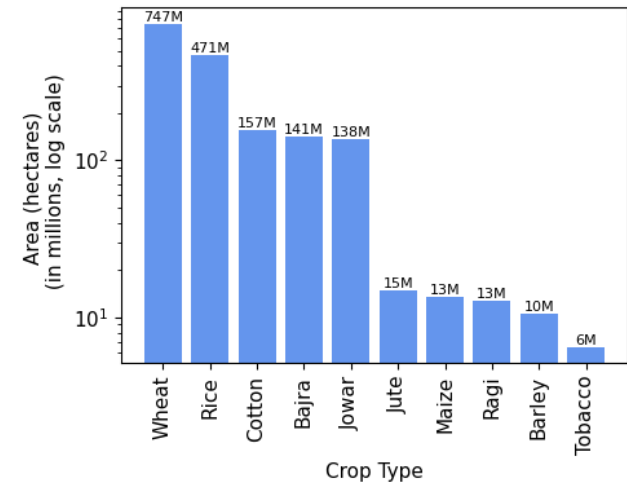


FIGURE 12. Area allotted to crops.

TABLE 1. Simulating model performance with area and production as inputs.

Model	Accuracy	Mean Absolute Error (MAE)	Root Mean Square Error (RMSE)	Standard Deviation (SD)
Random Forest	98.96	1.97	2.45	1.23
Decision Tree	89.78	4.58	5.86	2.75
XGBoost	86.46	6.31	7.89	3.54

(Table 1). For the Decision tree and XGBoost, the accuracy, mean absolute error, root mean square error, and standard deviation are 89.78%, 4.58, 5.86, 2.75 and 86.46%, 6.31, 7.89, 3.54. Figure 13, Figure 14, and Figure 15 show the

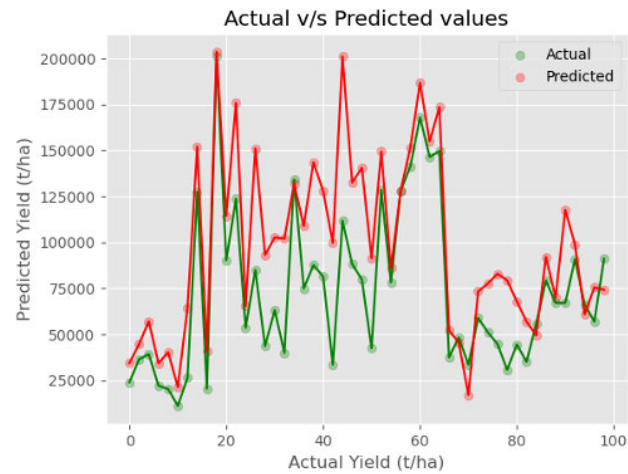


FIGURE 13. Model performance of decision tree.

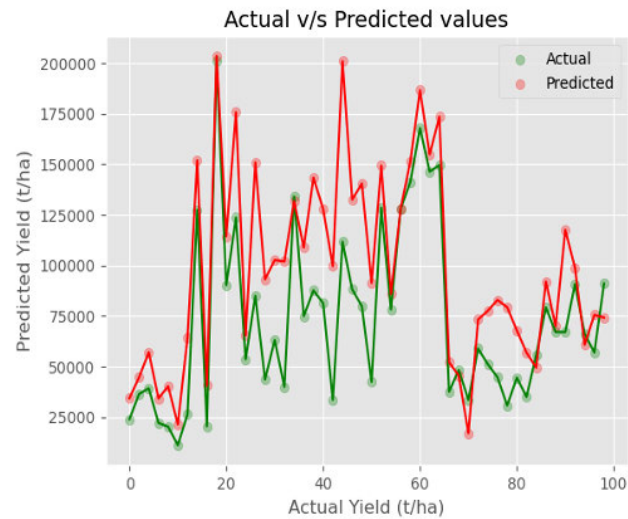


FIGURE 14. Model performance of random forest.

TABLE 2. Test results of CNN and LSTM.

Model	Test Loss
CNN	0.00060
LSTM	0.00063

model performance of Decision Tree, Random Forest, and XGBoost, respectively.

With accuracy ratings of 89.78 and 86.46, respectively, Decision Tree and XGBoost Regression performs less compared to Random Forest. Machine learning is a “black box” technique because it doesn’t interpret much. Random Forest outperformed the other three regression approaches in this study’s application of machine learning techniques to forecast the estimated crop output for India.

Figure 16 details the layers used in CNN and LSTM. Table 2 shows the Test loss for CNN and LSTM, evaluated as 0.00060 and 0.00063 respectively.

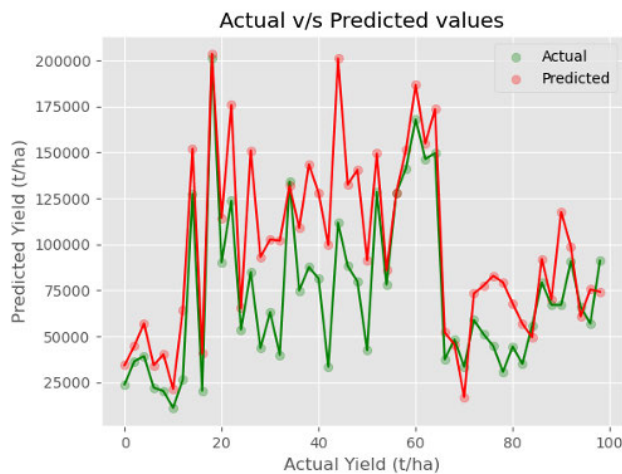


FIGURE 15. Model performance of XGBoost regression.

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 2, 64)	448
max_pooling1d_1 (MaxPooling 1D)	(None, 1, 64)	0
lstm_1 (LSTM)	(None, 50)	23000
dense_3 (Dense)	(None, 1)	51
=====		
Total params: 23,499		
Trainable params: 23,499		
Non-trainable params: 0		

FIGURE 16. Layer description of CNN and LSTM.

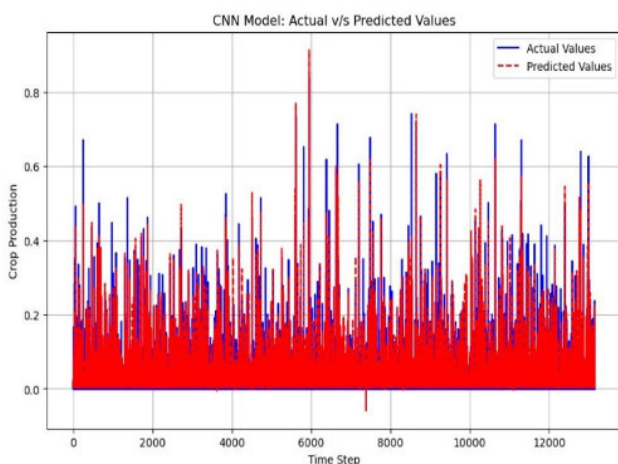


FIGURE 17. Model performance of CNN.

Figure 17 and 18 shows the model performance of CNN and LSTM respectively.

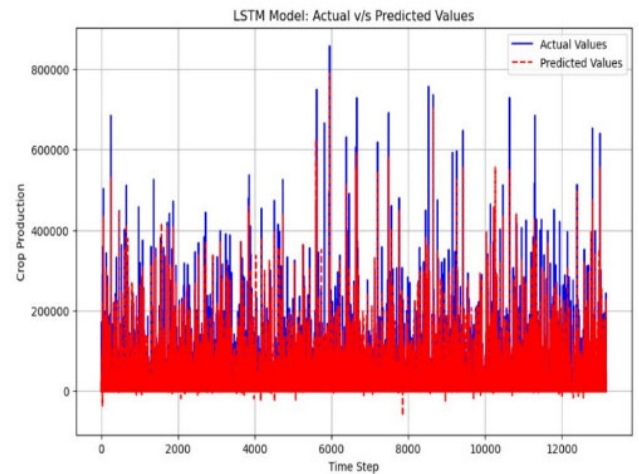


FIGURE 18. Model performance of LSTM.

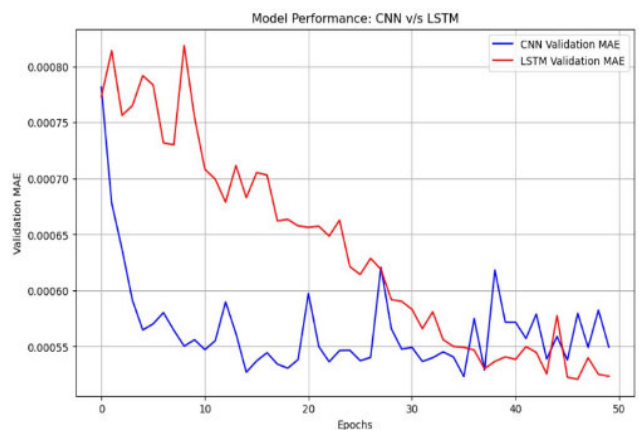


FIGURE 19. CNN v/s LSTM.

Figure 19 illustrates a comparison of CNN and LSTM. Result show that a change in the number of epochs leads to a significant change in mean absolute error. Here, it can be observed that CNN performs better than LSTM as the loss is less with CNN.

## V. CONCLUSION

The demand and supply for food have grown more difficult to manage as the population grows. To assist farmers, experts have worked hard over the past few years to anticipate agricultural yield production. In order to forecast India's crop yield, this study uses various machine learning and deep learning approaches. The study underlines the advantages of cutting-edge procedures. It is beneficial for small-scale ranchers, as they may use the predictions to estimate crop production for upcoming years and plant it appropriately. Five machine learning and deep learning algorithms, Decision Tree, Random Forest, XGBoost regression, Convolutional Neural Network, and Long-Short Term Memory Networks are applied to the dataset taken into consideration. When data is analyzed at the country level, Random Forest



(with accuracy-98.96%, mean absolute error-1.97, RMSE-2.45 and standard deviation-1.23) and CNN (with minimum loss-0.00060) perform better according to the current prediction. Experimental findings demonstrate that the approach has a great potential for precise crop productivity prediction and its effectiveness has been validated using real-time data and interactions with people. More data for each crop year having more historically precise information about the climate and environment is needed. More deep learning models need to be applied to the dataset to identify the method that performs the best. To increase the model's accuracy in crop production prediction, remote sensing data could be amalgamated with statistical data of districts. The prediction can be more accurate using satellite imagery land cover or satellite image classification.

## REFERENCES

- [1] N. Gandhi, L. J. Armstrong, O. Petkar, and A. K. Tripathy, "Rice crop yield prediction in India using support vector machines," in *Proc. 13th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, Jul. 2016, pp. 1–5.
- [2] M. Khan and S. Noor, "Irrigation runoff volume prediction using machine learning algorithms," *Eur. Int. J. Sci. Technol.*, vol. 8, pp. 1–22, Jan. 2019.
- [3] NIC IN. (2023). *Annual Report 2020–21*. Accessed: Jan. 20, 2023. [Online]. Available: <https://agricoop.nic.in/Documents/annual-report-2020-21>
- [4] Govt India. (2023). *Crop Production Statistics Information System*. [Online]. Available: <https://aps.dac.gov.in/APY/Index.htm>
- [5] M. Khan and S. Noor, "Performance analysis of regression-machine learning algorithms for predication of runoff time," *Agrotechnology*, vol. 8, no. 1, pp. 1–12, 2019.
- [6] IBEF. (2023). *Agriculture in India: Industry Overview, Market Size, Role in Development*. Accessed: Jan. 20, 2023. India. [Online]. Available: <https://www.india.gov.in/>
- [7] H. S. Gill, G. Murugesan, B. S. Khehra, G. S. Sajja, G. Gupta, and A. Bhatt, "Fruit recognition from images using deep learning applications," *Multimedia Tools Appl.*, vol. 81, no. 23, pp. 33269–33290, Sep. 2022.
- [8] S. Khaki and L. Wang, "Crop yield prediction using deep neural networks," *Frontiers Plant Sci.*, vol. 10, p. 621, May 2019.
- [9] E. Khosla, R. Dharavath, and R. Priya, "Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression," *Environ., Develop. Sustainability*, vol. 22, no. 6, pp. 5687–5708, Aug. 2020.
- [10] P. S. Maya Gopal and R. Bhargavi, "Optimum feature subset for optimizing crop yield prediction using filter and wrapper approaches," *Appl. Eng. Agricult.*, vol. 35, no. 1, pp. 9–14, 2019.
- [11] P. S. M. Gopal and R. Bhargavi, "Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms," *Appl. Artif. Intell.*, vol. 33, no. 7, pp. 621–642, Jun. 2019.
- [12] The Food and Agriculture Organization (FAO). (2023). *FAOSTAT*. Accessed: Jan. 20, 2023. [Online]. Available: [https://www.fao.org/faostat/en/#rankings/countries%20\\_by\\_com%20modity\\_exports](https://www.fao.org/faostat/en/#rankings/countries%20_by_com%20modity_exports)
- [13] F. Raimundo, A. Glória, and P. Sebastião, "Prediction of weather forecast for smart agriculture supported by machine learning," in *Proc. IEEE World AI IoT Congr. (AllIoT)*, May 2021, pp. 160–164.
- [14] A. Suruliandi, G. Mariammal, and S. P. Raja, "Crop prediction based on soil and environmental characteristics using feature selection techniques," *Math. Comput. Model. Dyn. Syst.*, vol. 27, no. 1, pp. 117–140, Jan. 2021.
- [15] P. Mohan and K. K. Patil, "Crop production rate estimation using parallel layer regression with deep belief network," in *Proc. Int. Conf. Electr., Electron., Commun., Comput., Optim. Techn. (ICEECCOT)*, Dec. 2017, pp. 168–173.
- [16] M. Kavita and P. Mathur, "Crop yield estimation in India using machine learning," in *Proc. IEEE 5th Int. Conf. Comput. Commun. Autom. (ICCCA)*, Oct. 2020, pp. 220–224.
- [17] S. Kunchakuri, S. Pallerla, S. Kande, and N. R. Sirisala, "An efficient crop yield prediction system using machine learning algorithm," in *Proc. 4th Smart Cities Symp. (SCS)*, vol. 2021, Nov. 2021, pp. 120–125.
- [18] M. S. Rao, A. Singh, N. V. S. Reddy, and D. U. Acharya, "Crop prediction using machine learning," *J. Phys., Conf. Ser.*, vol. 2161, no. 1, 2022, Art. no. 012033.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [20] C. H. Vanipriya, Maruyi, S. Malladi, and G. Gupta, "Artificial intelligence enabled plant emotion expresser in the development hydroponics system," *Mater. Today, Proc.*, vol. 45, pp. 5034–5040, Jan. 2021.
- [21] A. Tomar, G. Gupta, W. Salehi, C. H. Vanipriya, N. Kumar, and B. Sharma, "A review on leaf-based plant disease detection systems using machine learning," in *Proc. ICRIC*, vol. 1, 2022, pp. 297–303.
- [22] Govt India. (2023). *Profile*. Accessed: Jan. 20, 2023. [Online]. Available: <https://www.india.gov.in/india-glance/profile>
- [23] Govt India. (2023). *Data*. Accessed: Jan. 20, 2023. [Online]. Available: <https://data.gov.in>
- [24] Govt India. (2023). *Crop Production Statistics Information System*. Accessed: Jan. 20, 2023. [Online]. Available: <https://aps.dac.gov.in/APY/Index.htm>
- [25] D. J. Reddy and M. R. Kumar, "Crop yield prediction using machine learning algorithm," in *Proc. 5th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2021, pp. 1466–1470.
- [26] S. Bhansali, P. Shah, J. Shah, P. Vyas, and P. Thakre, "Healthy harvest: Crop prediction and disease detection system," in *Proc. IEEE 7th Int. Conf. Conver. Technol. (I2CT)*, Apr. 2022, pp. 1–5.
- [27] S. Agarwal and S. Tarar, "A hybrid approach for crop yield prediction using machine learning and deep learning algorithms," *J. Phys., Conf. Ser.*, vol. 1714, no. 1, Jan. 2021, Art. no. 012012.
- [28] R. Gupta, A. K. Sharma, O. Garg, K. Modi, S. Kasim, Z. Baharum, H. Mahdin, and S. A. Mostafa, "WB-CPI: Weather based crop prediction in India using big data analytics," *IEEE Access*, vol. 9, pp. 137869–137885, 2021.
- [29] N. Bali and A. Singla, "Emerging trends in machine learning to predict crop yield and study its influential factors: A survey," *Arch. Comput. Eng.*, vol. 29, pp. 1–18, Mar. 2022.
- [30] S. Iniyar, V. A. Varma, and C. T. Naidu, "Crop yield prediction using machine learning techniques," *Adv. Eng. Softw.*, vol. 175, Jan. 2023, Art. no. 103326.
- [31] Z. Guo, Q. Liu, and J. Wang, "A one-layer recurrent neural network for pseudoconvex optimization subject to linear equality constraints," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 1892–1900, Dec. 2011.
- [32] A. V. S. Kumar, P. Sharma, S. Kaur, O. S. Saleh, H. R. Chennamma, and A. Chaturvedi, "AI-equipped IoT applications in high-tech agriculture using machine learning," in *Handbook of Research on AI-Equipped IoT Applications in High-Tech Agriculture*. Hershey, PA, USA: IGI Global, 2023, pp. 38–64.



**PRIYANKA SHARMA** was born in Jaipur, Rajasthan, India, in 1991. She received the B.Tech. degree in information technology and the M.Tech. degree in computer science from Rajasthan Technical University, Kota, Rajasthan, in 2012 and 2016, respectively. She is currently an Assistant Professor with the Department of Computer Science and Engineering (NBA Accredited), Swami Keshvanand Institute of Technology, Management and Gramothan (SKIT), Jaipur, Rajasthan (Accredited by NAAC A++ Grade). She was also appointed as a Faculty Advisor of the Google Developer Students Club, SKIT. She has guided various M.Tech. and UG research scholars. She has published more than 15 research papers in national and international journals/conferences and book chapters. Also, she has taken part in forums hosted by Infosys, TCS, Oracle, and IBM. She has received accolades a number of times in a variety of fields. Her research interest includes investigating the advancements in machine learning and deep learning applications in various problem domains. She is a member of ACM and various professional societies. She was also given recognition by Infosys, India, for her remarkable performance in the Campus Connect Program.



**PANKAJ DADHEECH** (Member, IEEE) is currently a Professor and the Deputy Head of the Department of Computer Science and Engineering (NBA Accredited), Swami Keshvanand Institute of Technology, Management and Gramothan (SKIT), Jaipur, Rajasthan, India (Accredited by NAAC A++ Grade). He has more than 18 years of experience in teaching. He has been appointed as a Ph.D. Research Supervisor with the Department of Computer Science and Engineering, SKIT (Recognized Research Centre of Rajasthan Technical University, Kota). He has also guided various M.Tech. research scholars. He has published 22 patents and granted six patents at Intellectual Property India, Office of the Controller General of Patents, Design and Trade Marks, Department of Industrial Policy and Promotion, Ministry of Commerce and Industry, Government of India. He has also published and granted five Australian patents at the Commissioner of Patents, Intellectual Property Australia, Australian Government. He has also published and granted one German patent, one South African patent, and one USA patent. He has also registered and granted two research copyrights at the Registrar of Copyrights, Copyright Office, Department for Promotion of Industry and Internal Trade, Ministry of Commerce and Industry, Government of India. He has presented 60 papers at various national and international conferences. He has 67 publications in various international and national journals. He has published seven books and 32 book chapters. His research interests include high-performance computing, cloud computing, information security, big data analytics, intellectual property rights, and the Internet of Things. He is a member of many professional organizations, such as the CSI, ACM, IAENG, and ISTE. He has chaired technical sessions at various international conferences and contributed as a resource person in various FDPs, workshops, STPs, and conferences. He is also acting as the guest editor of various reputed journal publishing houses and conference proceedings and the Bentham Ambassador of Bentham Science Publisher.



**NAGENDER ANEJA** received the M.E. degree in computer technology and applications from the Delhi College of Engineering and the Ph.D. degree in computer engineering from the J. C. Bose University of Science and Technology, YMCA, Faridabad, India. He was an Associate IP Lead for Microsoft Patent Research Services with CPA Global India for five years. He is currently an Assistant Professor with the School of Digital Science, Universiti Brunei Darussalam, Brunei Darussalam. He is also the Founder of <https://researchid.co/>, a platform allowing faculty members and researchers to create profiles for their research. Several process innovations have been developed by him, including the development of automation tools for patent analysis with CPA Global,

as well as the development of an expert directory for Universiti Brunei Darussalam. He has also held teaching and research positions at various universities in India and Brunei Darussalam. He is a highly experienced and accomplished researcher who has published over 35 papers in top academic journals and conferences. He has more than 20 years of experience in computer science. He has two patents from the U.S. patent and Trademark Office. His research interests include deep learning, computer vision, and natural language processing. He also received several awards, including the Brunei ICT Award, in 2016, for a mobile app that creates a social network with nearby users by matching interests in airplane mode. He also received the Research Award 2020–2021 from the J. C. Bose University of Science and Technology, YMCA, for several SCIE-indexed publications during the Ph.D. degree. His website, <http://naneja.github.io>, provides further information about his work and research interests.



**SANDHYA ANEJA** (Member, IEEE) is currently a Researcher, an Inventor, and a Computer Scientist with a strong passion for teaching. She is also an Assistant Professor with the School of Computer Science and Mathematics, Marist College, Poughkeepsie, NY, USA. She is also a Visiting Research Scholar with the Department of Computer Science, Purdue University. She has over 15 years of experience teaching computer science to undergraduate and graduate students at the University of Delhi and Universiti Brunei Darussalam. As a Researcher, she contributed to developing a mobile application to facilitate the matching of interests on available mobile devices and allow exchanging of messages and files. The application allows broadcasting names and a limited number of keywords representing users' interests without any connection in a nearby region. The broadcasting region creates a mobile wireless network limited by the Wi-Fi region that is around 200 meters. She also received a U.S. patent on this technology. As a Computer Scientist, she has received project funding from the University of Delhi as a PI and Universiti Brunei Darussalam as a co-PI. She has extensively worked on Brunei government-funded projects with IBM researchers. She is also a Contributor to Sandia and DARPA-funded projects with Purdue University.

• • •