Acta Geophysica (2024) 72:4613–4632
https://doi.org/10.1007/s11600-024-01312-8

RESEARCH ARTICLE - ATMOSPHERIC & SPACE SCIENCES

Prediction of crop yield in India using machine learning and hybrid deep learning models

Krithikha Sanju Saravanan[1] · Velammal Bhagavathiappan[1]

Received: 19 July 2023 / Accepted: 23 January 2024 / Published online: 19 March 2024

Abstract

Crop yield prediction is one of the burgeoning research areas in the agriculture domain. The crop yield forecasting models are developed to enhance productivity with improved decision-making strategies. The highly efficient crop yield forecasting model assists farmers in determining when, what and how much to plant on their cultivable land. The main objective of the proposed research work is to build a high efficacious crop yield prediction model based on the data available for the period of 21 years from 1997 to 2017 using machine learning and hybrid deep learning approaches. Two prediction models have been proposed in this research work to predict the crop yield accurately. The first model is a machine learning-based model which uses the CatBoost regression model and its hyperparameters are tuned which improves the performance of the yield prediction using the Optuna framework. The second model is the hybrid deep learning model which uses spatio-temporal attention-based convolutional neural network (STACNN) for extracting the features and the bidirectional long short-term memory (BiLSTM) model for predicting the crop yield effectively. The proposed models are evaluated using the error metrics and compared with the latest contemporary models. From the evaluation results, it is shown that the proposed models significantly outperform all other existing models and CatBoost regression model slightly performs better than the STACNN-BiLSTM model, with the R-squared value of 0.99.

Keywords CatBoost regression · Bayesian optimization · Spatial attention mechanism · Temporal attention mechanism · Convolutional neural network · Bidirectional long short-term memory

Introduction

Agriculture plays a key role in the entire socio-economic growth of the nation. Agriculture not only helps the people with their livelihood, but it also supports people and the nation by providing industrial developments (Zambon et al. 2019; Kumar et al. 2021), financial managements (Calicioglu et al. 2019; Belhadi et al. 2021) and international trade operations (Green et al. 2019; Kastner et al. 2021). Therefore, poverty in the country can be reduced or eliminated with the development fostered by the agriculture. Agriculture is the main alimentation for about 70% of the rural people in Tamil Nadu (Kumar and Venugopal 2016). In the Agricultural sector, any small intervention results in major impact in the economic development of the State. The nation is much focused on the upliftment of farmers duly considering the challenges in Agriculture. The challenges in agriculture may be categorized as the longstanding problems and the emerging issues depend on the prevailing global climate, rainfall, high cost of farm inputs, food security, groundwater depletion and soil fertility along with usage of latest technologies.

In India, during the last Agricultural Budget (2022–2023), the Chief Minister of Tamil Nadu has announced a visionary strategy to increase 60% of net cultivated area to 75%. Zero poverty, sustainable consumption, no famine and production plans are the four significant goals of the agriculture department of Tamil Nadu. Similarly, each state in India has different goals and strategies for improving crop yield. By enhancing agriculture in India, the yield obtained can be increased which in turn increases the socio-economic growth

Edited by Dr. Ahmad Sharafati (ASSOCIATE EDITOR) / Prof. Theodore Karacostas (CO-EDITOR-IN-CHIEF).

Krithikha Sanju Saravanan
krithikhasanju3008@gmail.com

Velammal Bhagavathiappan
velammalkarthik@gmail.com; velammalbl@annauniv.edu

[1] Department of Computer Science and Engineering, College of Engineering Guindy, Anna University, Chennai, India

of the nation (Pawlak and Kolodziejczak 2020; Fukase and Martin 2020). The crop yield prediction is a milestone in agriculture (Shakoor et al. 2019). For sustainable development of the nation, it is important to improve potential yield to fight against famine and starvation (Gil et al. 2019; Tian et al. 2022). A country's policymaker depends upon precise forecast of crop yield, to make appropriate export and import assessments to reinforce national food security. Cultivators and farmers are privileged by the accurate yield forecasts to make financial and management decisions.

Crop yield prediction is the main task of the decision-makers at regional, national and international levels for whirlwind decision-making with respect to the agricultural yield. A precise crop yield prediction model can assist farmers to decide on what to grow and when to grow (Van Klompenburg et al. 2020). It helps in making import and export decisions, evaluating performance of yield, strategies to be carried out by farmers for better results. Crop yield prediction is most important for global production of food and development (Ristaino et al. 2021). Based on previous yield experiences, the farmers normally predict their yield in olden days. Earlier statistical models have been used to predict the crop yield (Inoue et al. 1998) but resulted in substandard performance and it is found to be tedious because it consumes more time. Now at present, machine learning models (Palanivel and Surianarayanan 2019, Paudel et al. 2021, Burdett and Wellen 2022, Vance et al. 2022, Kuradusenge 2023) and deep learning models (Sharma et al. 2020; Oikonomidis et al. 2022; Saravanan and Bhagavathiappan 2022) are used to predict the crop yield for improving the performances of the prediction.

The crop yield mainly depends upon climatic conditions, soil quality, pest infestations, landscapes, water quality and availability, planning of harvest activity, etc. The crop yield forecast processes and strategies vary with time and they are profoundly nonlinear in nature (Whetten et al. 2017). Machine learning and deep learning models enable better yield decisions, improved efficiencies and it also helps farmers to cultivate according to their requirements. These approaches resolve non-linear or linear-based agricultural systems with remarkable forecasting ability. Further, machine learning resembles an umbrella that holds various significant strategies and methodologies (Murdoch et al. 2019). Deep learning is a subgroup of machine learning that can determine the yield outcomes from varying arrangements of raw data (Gupta and Nahar 2022). On observing the most prominent models in agriculture for predicting the crop yield, the artificial and deep neural networks are used. Despite the advancements in deep learning and machine learning models for forecasting the yield, there still exists a potential to enhance the accuracy of these predictions and minimize the error rates. Hence, the proposed machine learning-based crop yield prediction work aims at improving the prediction accuracy, automatic feature handling and enhances the performance with optimized hyperparameters selection. Further, the proposed hybrid deep learning model aims to enrich the prediction accuracy with modeling spatio-temporal dependencies and to boost the performance with very minimal error values.

In this research work, two crop yield prediction models are proposed with machine learning and deep learning approaches. The machine learning model of this research work is the CatBoost regression model with tuned hyperparameters using Optuna framework. The novel hybrid deep learning model is developed by combining a spatio-temporal attention-based CNN model with the BiLSTM network. The hybrid deep learning model's component utilizes the Gaussian error linear units (GELU) activation function, a critical element that substantially boosts the model's predictive prowess. The proposed work handles the requirement of predicting the crop yield effectively and can aid in the production of the crops in future. Section "Literature survey" of the research article describes the thorough literature review that was conducted for the crop yield prediction research work. Section "Materials and methods" provides a detailed design and explanation of the overall methodologies of the proposed crop yield prediction systems. Section "Results" discussed about the step-by-step implementation results of this research work. Section "Conclusion" is the conclusion and Section 6 provides references for this research work.

Literature survey

This section provides the literature survey that has been conducted for the crop yield prediction research work. A deep neural network model has been proposed by Crane-Droesch (2018) to determine the crop yield. The proposed method is a semi-parametric type which uses both complicated and parametric structures at the same time. This approach works better than statistical methods and neural networks for forecasting the corn yield in the US. Different climatic models were used and it has been seen that there are negative impacts on the change in climate. However, the impacts are not projected very well with statistical techniques. Chlingaryan et al. (2018) have reviewed the scientific advancement in the last 15 years on machine-based learning techniques for precise crop yield prediction and assessment of nitrogen status. As a result, the researcher concludes that the rapid developments in sensing technology and machine learning approaches would offer cost-effective and detailed solutions for effective estimation and decision-making of the crop yield. In future, a more robust framework for precision agriculture will incorporate artificial intelligence methodologies fusing different sensor methods and hybrid systems that integrate various machine learning and signal processing techniques.

Crop Yield prediction has been performed by Khaki and Wang (2019) using a deep neural network (DNN) consisting of 21 hidden layers each with 50 neurons. This work was submitted for Syngenta Crop Challenge and superior performance was observed. Large corn hybrid datasets were used for the implementation and uses the data of 2,267 maize hybrid plants from various locations in the USA. The proposed model had superior accuracy with root mean square error (RMSE) of 12.41. Feature selection has also been performed using DNN which successfully reduced the size of the training model. The work has been compared with other conventional methods like Lasso regression and regression tree in order to prove the effectiveness of the DNN algorithm. However, the error rate can still be decreased with a better technique. Gopal and Bhargavi (2019) have examined the intrinsic relationship between multiple linear regression (MLR) and artificial neural network (ANN). The researcher has proposed a hybrid MLR–ANN model for efficient prediction of crop yields. The hybrid model is designed to evaluate the prediction efficiency when MLR interrupt and correlations are added to initialize weights and bias input layer of the ANN. The back propagation learning algorithm with feed forward artificial neural network was utilized to predict the exact yield of paddy crops. Rather than random bias and initialization of weights, this hybrid model initializes the input layer bias and weights by utilizing the MLR coefficients and bias. The predictive accuracy of the hybrid model is contrasted with the models support vector regression (SVR), MLR, ANN, random forest (RF), and k-nearest neighbors (KNN) using performance metrics. The computational time was estimated for both the conventional ANN and hybrid MLR–ANN. As a result, the proposed MLR–ANN hybrid model provides greater precision than the traditional models.

A convolutional neural network model has been proposed by Nevavuori et al. (2019) for predicting the crop yield. The data are based on RGB and NDVI data, various aspects like training algorithms, regularization, hyperparameters tuning and depth of the network are identified. The MAPE and MAE have been calculated where the MAE for the time period till June 2017 is 484.3 kg/ha and the MAPE is 8.8. For the later part of the crop growth, i.e., after June 2017, the MAE is 624.3 kg/ha and MAPE of 12.6. It is observed that the proposed method worked better for RGB data than NDVI data. Li et al. (2019) has proposed statistical crop models for rainfed corn in the Midwest USA and resolves the existing issues via an extensive diagnostic analysis. The approach is robust sufficient to absorb new data either from atmosphere or satellite sources. These models could be extended to crop yield forecasting within season and for the evaluation of climatic change effects. As a limitation, this model has sample extrapolation problems as collinearity between predictors. This method opens the way for future growth and implementation of statistical yield models.

Sharma et al. (2020) introduced a model for predicting the crop yield using the satellite imagery. The model uses the combination of CNN and LSTM to process the satellite images. The model predicts the yield for the selected states in India and the error values calculated for those states are high. The hybrid method for predicting the crop yield using machine learning and deep learning models was proposed by Agarwal and Tarar (2021). The soil and climatic conditions were used for the models to predict the crop yield. Initially, SVM-based machine learning method was used then the RNN and LSTM models were used representing deep learning models. All the three models were combined to predict the yield and the overall accuracy obtained is 97%. Nayana et al. (2022) proposed a crop yield prediction framework for wheat in India. The proposed framework uses PCA for feature extraction and multivariate adaptive regression splines as prediction technique. This work focuses only on wheat yield and the proposed hybrid model has low RMSE value of 23 and MAE value of 18.925. Mallikarjuna et al. (2022) proposed a crop yield prediction model using data containing the external factors that are important for crop growth. The proposed method used two techniques, SVM and XGBoost model for predicting the crop yield. A comprehensive approach for predicting the crop yield with hybrid machine learning models was proposed by Saravanan and Bhagavathiappan (2022). Three hybrid models were proposed namely PCA-AdaBoost, PCA-XGBoost and LSTM-based stacked auto-encoder-DNN. The hybrid deep learning model performs best in all calculated error metrics compared to the other two models and has low RMSE value for the crop yield prediction model. From the explored research works for predicting the crop yield, it is seen that the error rates are high and can be reduced further by constructing a novel prediction system.

Challenges

- The data for crop yield are often insufficient because only very few benchmark databases are available and that too for limited regions across the globe.
- The crop yield prediction models using either machine learning or deep learning algorithms require more time for processing the data. So, a minimal time-consuming model must be constructed which is a demanding work in the prediction systems.
- Prediction models using regression algorithms cause the problem of overfitting in many cases.
- At times, the regression model used for prediction analysis fails or yields poor performances when the input dataset size is very large.
- The crop yield varies with space and time. So, it is important to maintain both spatial and temporal granularity in the prediction system.

Contributions

The proposed crop yield prediction work consists of the following distinct contributions,

- A crop yield prediction model is constructed using the CatBoost algorithm for the regression process. It is a machine learning based novel gradient boosting method for automatically handling the categorical features of the crop yield dataset.

- Tuning of hyperparameters is done using the Optuna framework for Bayesian optimization. The Optuna framework selects the accurate values for the hyperparameters of the CatBoost regressor. Thus, by using the Optuna framework for tuning the hyperparameters the performance of the prediction system is improved and the error values are minimized.
- A hybrid deep learning model for crop yield prediction is established using spatio-temporal attention based CNN and BiLSTM model. The CNN model uses GELU activation function which is the key feature of the proposed deep learning model for improving the performance of the prediction system.

Materials and methods

The research work proposed two separate models for predicting the crop yield. The first model is proposed based on the machine learning algorithms and the second model is proposed using the hybrid deep learning models. The main objective of these two models is to predict the crop yield with less error rates and high precision. Figure 1 represents the proposed frameworks for predicting the crop yield using machine learning and deep learning models. The proposed machine learning based crop yield prediction model comprises of the following steps,

I. Data collection and preparation.
II. Data preprocessing using data imputation.
III. Constructing CatBoost regression model with hyperparameter tuning using optuna framework based on Bayesian optimization for predicting the crop yield.
IV. Evaluating the constructed CatBoost model by calculating the losses using performance measures.

The proposed hybrid deep learning based crop yield prediction model comprises of the following steps,

I. Data collection and preparation.
II. Data preprocessing using data imputation and one hot encoding.
III. Feature extraction using Spatio-temporal attention-based convolutional neural network.
IV. BiLSTM model for predicting the crop yield.
V. Evaluating the constructed hybrid deep learning model by calculating the losses using performance measures.

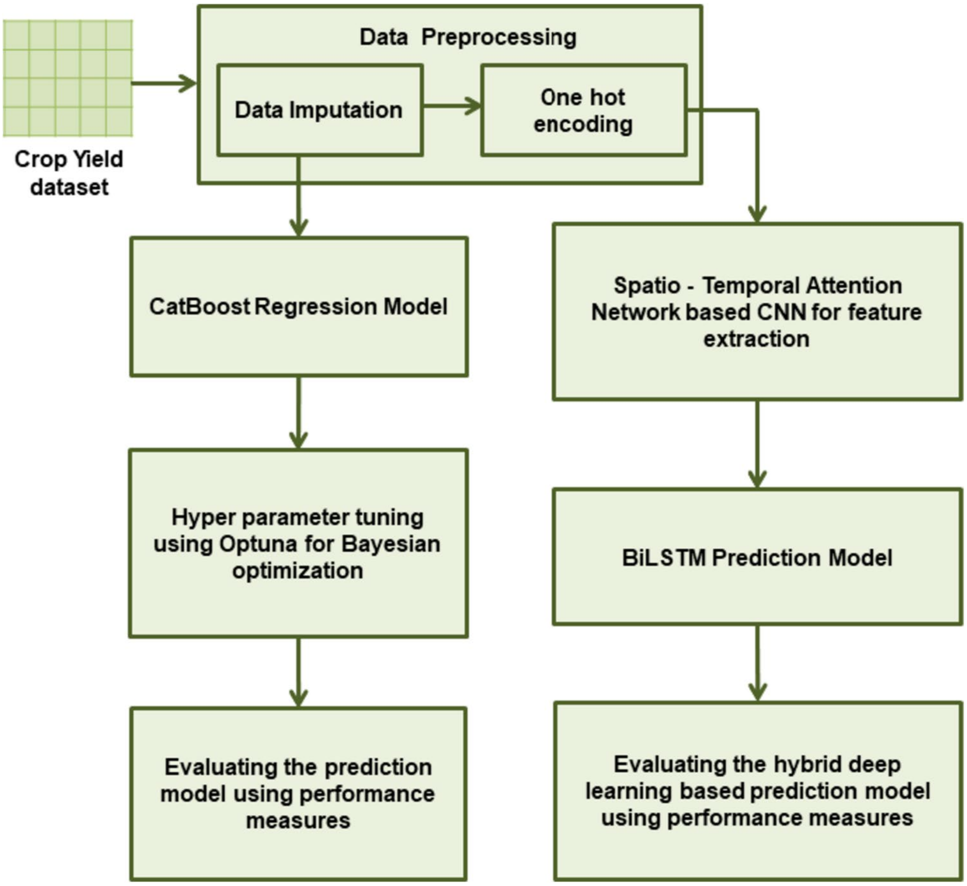

Crop Yield dataset
Data Preprocessing
Data Imputation
One hot encoding
CatBoost Regression Model
Hyper parameter tuning using Optuna for Bayesian optimization
Evaluating the prediction model using performance measures
Spatio - Temporal Attention Network based CNN for feature extraction
BiLSTM Prediction Model
Evaluating the hybrid deep learning based prediction model using performance measures

Fig. 1 The proposed frameworks for predicting the crop yield

Dataset description

The dataset for crop yield is collected from Kaggle website (https://www.kaggle.com/datasets/abhinand05/crop-production-in-india/data or https://www.kaggle.com/code/anjali21/indian-production-analysis-and-prediction/ data) and Tata-Cornell Institute (TCI) website (http://data.icrisat.org/dld/src/crops.html). The Kaggle website crop yield data are collected from government websites for 646 districts of 33 Indian states and consists of historical information on crop yields between the years 1997–2015. Seven instances are present in the Kaggle crop yield dataset namely state name, district name, crop year, season, crop name, area and production with 246,091 different attributes. The Kaggle crop yield dataset consists of 124 types of crops grown all over in India. The crop yield dataset for district level Indian agriculture from TCI website is created by International Crops Research Institute for the Semi-Arid Tropics and TCI. The crop yield TCI dataset is collected for the years 2016 and 2017 which has 18,009 different attributes. Both the datasets are used for the proposed work by merging the data according to the common instances. The collected historical data are validated by cross checking the values randomly with authorized government data with the help of domain experts and also by interpreting the results obtained from current research works with the datasets.

Data preprocessing

The data preprocessing is an essential step because it is common to encounter missing values in the collected datasets while trying to understand and analyze the data. The collected crop yield dataset can have some missing values so it is important to handle those missing values before processing the data with proposed models. Data imputation method is used for handling the missing values of the crop yield datasets. Data imputation is required because the missing data can create the following issues,

(1) Huge volume of missing data can cause distortion in the dataset.
(2) The dataset becomes incompatible for most of the machine learning-based Python libraries.
(3) Affects the performance of the model.

The data imputation preprocessing method is the technique for replacing the missing values in the dataset with substituted data value (Farhangfar et al. 2007; Ramli et al. 2013). There are three basic imputation methods namely mean, median and mode (Geng et al. 2021). For the proposed research work, mean imputation method is used to handle the missing values. Thus, the missing data in the crop yield dataset are substituted by the data obtained from the mean imputation method.

CatBoost regression method

The Yandex researchers developed a gradient boosting machine learning-based ensemble algorithm and named it as CatBoost algorithm in the year 2017 (Shyam et al. 2020; Agrawal et al. 2021). It works on the principle of gradient boosting technique on decision trees and effectively handles the categorical features, hence it got its name as CatBoost from Categorical Boosting. The predominant features of CatBoost algorithm are:

(1) Supports categorical features without encoding.
(2) Accuracy is improved compared to XGBoost and LightGBM algorithms.
(3) Open-source algorithm and easy for implementing with packages in R and Python.
(4) Robust model thus improves the performance.
(5) Faster training and prediction model.
(6) Simpler process for tuning the hyperparameters.

CatBoost can directly and automatically handle the categorical features without encoding by converting the data into CatBoost's special pool data type using pool() class. The names of the categorical features have to be specified in the cat_features parameter. If the data is only numeric then pool data type conversion is not required. For building the CatBoost regression model, CatBoostRegressor() class is used with its required hyperparameter values (Yasir et al. 2022).

Two critical algorithmic advances are introduced by CatBoost algorithm,

(1) Ordered boosting implementation.
(2) The permutation driven innovative model for processing categorical features.

The random permutations of training samples are used for both advanced techniques to fight against prediction shift caused by a unique kind of target leakage existing in all the present implementations of gradient boosting algorithms. The one hot encoding by CatBoost is used for all the features with one_hot_max_size unique values and the default value is 2. The one hot encoding for preprocessing in CatBoost makes the system to work slower and less efficient. The CatBoost algorithm uses a unique method called minimal variance sampling (MVS) (Kim et al. 2022). The MVS is a weighted sampling method for regularization of stochastic gradient boosting models. Using MVS, the number of samples required for each iteration of boosting is reduced with the improvement in the quality of the model. The accuracy of the split scoring is maximized by the way of sampling the features for each boosting tree. The significant change compared with the other gradient boosting algorithms is the process of the tree construction which is known as leaf growth. The CatBoost algorithm uses symmetric trees by default. Figure 2 represents the basic architecture of the prediction model using

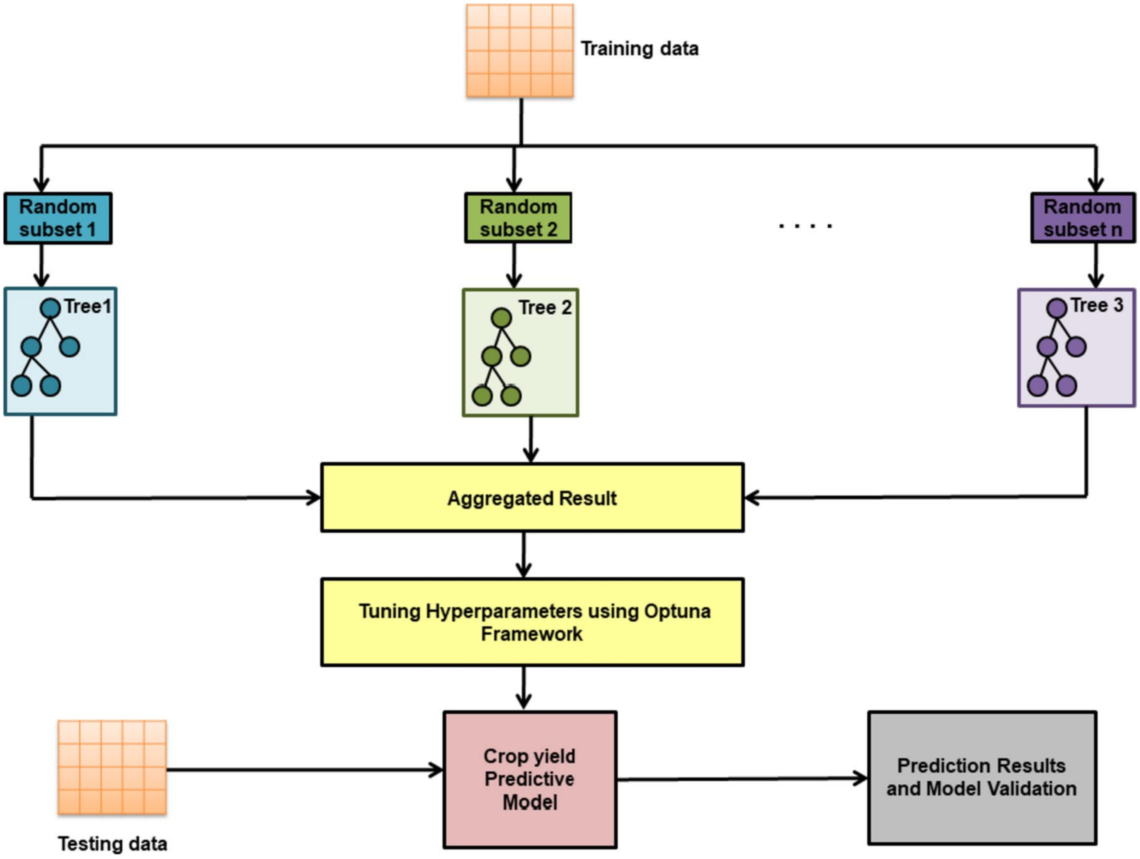

Training data
Random subset 1
Random subset 2
. . . .
Random subset n
Tree1
Tree 2
Tree 3
Aggregated Result
Tuning Hyperparameters using Optuna Framework
Crop yield Predictive Model
Testing data
Prediction Results and Model Validation

Fig. 2 Basic architecture of the prediction model using CatBoost algorithm with tuned hyperparameters by Optuna framework

CatBoost algorithm with tuned hyperparameters by Optuna framework. Figure 3 shows the implementation flow of ordered boosting algorithm.

The ordered boosting algorithm unfolds through the following sequence of steps,

I. Set of feature vectors derived from training data where each vector is associated with its corresponding crop yield and number of boosting iterations are given as the input.

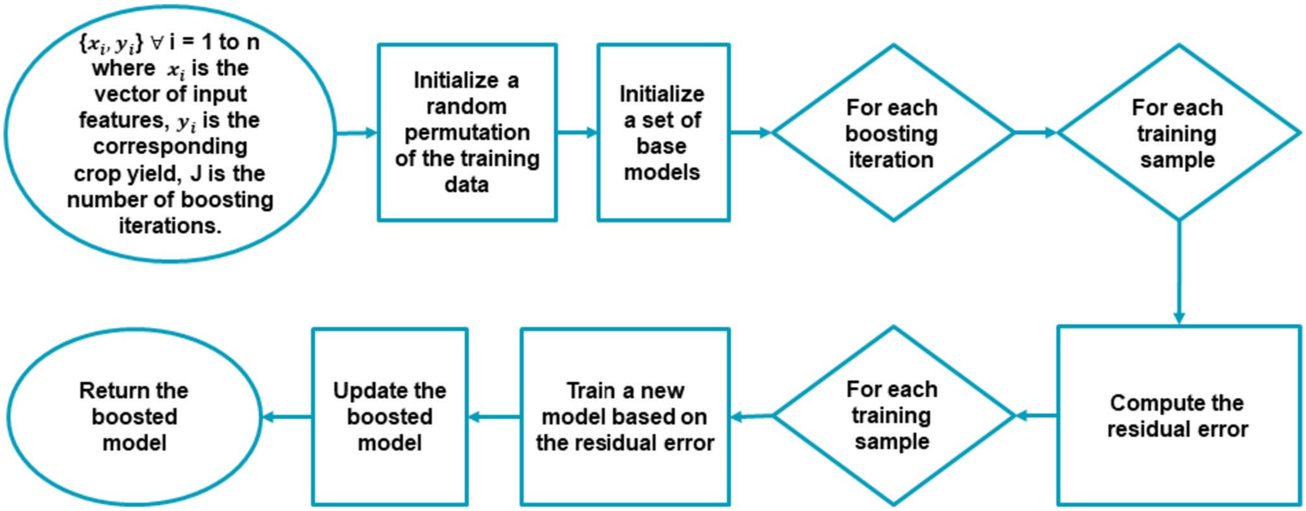

$\{x_i, y_i\}$ ∀ i = 1 to n where x_i is the vector of input features, y_i is the corresponding crop yield, J is the number of boosting iterations.
Initialize a random permutation of the training data
Initialize a set of base models
For each boosting iteration
For each training sample
Compute the residual error
For each training sample
Train a new model based on the residual error
Update the boosted model
Return the boosted model

Fig. 3 Implementation flow of ordered boosting algorithm

II. A random permutation of indices from 1 to n, denoted as σ is created. Here, with the crop yield data, it is used for shuffling the training data to ensure randomness into the boosting process.

III. Set of base models, denoted as B_j, where each base model is initially set to 0. These base models will be iteratively updated during the boosting process.

IV. For each boosting iteration (t) from 1 to J and for each training sample (j) in the permuted order, the algorithm computes the residual error, denoted as r_j. This error represents the difference between the actual crop yield y_j and the prediction based on the current boosted model $B_{\sigma(j)-1}$ for the corresponding input features x_j.

V. For each training sample (j) in the permuted order, train a new base model ΔB using the input features x_k and the corresponding residual error r_k for samples up to the current index (j).

VI. Following the training of the new base model, it is added to the existing boosted model B_j for the current training sample. This addition is performed to correct errors and enhance the overall predictive performance of the model.

VII. After completing a total of J boosting iterations, the algorithm returns the final boosted model B_n. This model embodies the collective knowledge acquired during the iterative process and serves as a robust predictor for crop yield.

Algorithm 1 Algorithm for ordered boosting

```
Input:
$(x_i, y_i) = \{(x_1, y_1), (x_2, y_2), (x_3, y_3) \dots (x_n, y_n)\}$ // Set of n input-output pairs where $x_i$ is the vector of input
features and $y_i$ is the corresponding crop yield
J = Number of boosting iterations
Output:
$B_n$ = Final boosted model for predicting crop yield.
$\sigma$ = random permutation [1, n]  // Initialize random permutation for the training data
$B_j$ = 0 for j = 1 to n              // Initialize set of base models
for t = 1 to J do                     // For each boosting iteration
  for j = 1 to n do                   // For each training sample
    $r_j = y_j - B_{\sigma(j)-1}(x_j)$          // Compute the residual error
  end for
end for
for j = 1 to n do                     // For each training sample
  $\Delta B$ = Learn_Model($(x_k, r_k) : \sigma(i) \le j$)  // Train a new base model based on the residual error
  $B_j = B_j + \Delta B$                       // After learning the new model is added to the existing boosting model
end for
return $B_n$        // Return the final boosting model after completing all boosting iterations.
```

The CatBoost algorithm uses an effective method of encoding for the categorical columns that have a unique number of categories larger than one_hot_max_size. This encoding method reduces overfitting and is similar to mean encoding. The encoding method used by the CatBoost algorithm is an ordered target encoding method for preprocessing the categorical features (Prokhorenkova et al. 2018). The random permutation of the dataset is performed and encoding the target is done on each sample using only the data objects that are placed before the current data object. CatBoost algorithm can construct a new categorical feature by combining the existing features.

Steps involved in converting categorical features to numeric features in CatBoost algorithm,

(1) Permutation of training data in random order.
(2) Quantization–Conversion of target value as integer from floating point.
(3) Encoding the values of categorical features.

$$\text{Average\_target} = \frac{\text{Count\_In\_Class} + \text{Prior}}{\text{Total\_count} + 1}, \tag{1}$$

where Count_In_Class denotes the number of times the label value is equal to 1 for data objects with current categorical feature, Prior denotes the constant number defined by the starting parameter, Total_Count denotes the total number of data objects with most frequent feature value. For composite features, the classes and the average target are swapped with the mathematical representation as $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ be the permutation (Ishfaque et al. 2022), so $x_{\sigma_{j,k}}$ is substituted with:

$$x_{\sigma_{j,k}} = \frac{\sum_{i=1}^{j=1}\left[x_{\sigma_{i,k}} = x_{\sigma_{j,k}}\right] y_{\sigma_i} + a.P}{\sum_{i=1}^{j=1}\left[x_{\sigma_{i,k}} = x_{\sigma_{j,k}}\right] + a}, \tag{2}$$

where $\left[x_{\sigma_{i,k}} = x_{\sigma_{j,k}}\right]$ will be 1 when the required criteria are fulfilled, P denotes prior value and a is the hyperparameter which denotes the weight of the prior value. The average of the entire dataset is used for performing regression and computes prior probability. Though CatBoost model has many advantages, there are few major weaknesses while working with the CatBoost regression model. The main weaknesses are,

(1) It is a computationally complex model and training a CatBoost model can be time-consuming, especially for large datasets. While CatBoost model provides some interpretability features such as feature importance scores, it is difficult to understand the inner workings of the model and how it makes its predictions. This limits its applicability in situations where interpretability is crucial.
(2) Like other decision tree-based algorithms, CatBoost model is prone to overfitting when the model learns the training data too well and fails to generalize to new data leading to poor performance on unseen data.
(3) CatBoost model is very sensitive to its hyperparameters and tuning hyperparameters can be a time-consuming and challenging process. If the hyperparameters are not tuned properly, then it leads to inefficient learning process of the model.

The random search (Bergstra and Bengio 2012) and grid search are the two conventional methods of hyperparameter optimization (Zahedi et al. 2021; Hossain and Timmer 2021). These two conventional methods require more space and time as the volume of the dataset increases. Hence, the Optuna framework is used for tuning the hyperparameters with Bayesian optimization in the proposed prediction model. The Optuna framework is an open source and automatic optimization software framework (Akiba et al. 2019; Sandha et al. 2021). The most important hyperparameters of the CatBoost algorithm are learning rate, number of trees, tree depth, L2 regularization, random strength, bagging temperature, border count, internal dataset order and tree growing policy. Bayesian optimization technique using Optuna framework finds the best values for the CatBoost algorithm's hyperparameters (Shekhar et al. 2021; Pravin et al. 2022) which improves and maximizes the performance of the prediction system. The Bayesian optimization technique works on the principle of Bayes theorem that finds global optimization through iteratively constructing the probabilistic model of functions mapping the objective function from the values of the hyperparameters. The function optimization is the technique intriguing to find maximum or minimum of an objective function. In this system, the objective function is to maximize the performance of the prediction and it is given as:

$$p^* = \text{argmaxf(p);p}\epsilon\text{P}. \tag{3}$$

Here, f(p) represents an objective score to maximize the performance of the model, p^* is the set of hyperparameters that provides the lowest value to the objective score and p can have any value in the domain P. The proposed CatBoost model consists of tuned hyperparameters, namely iterations, colsample_bylevel, depth, L2_leaf_reg, learning rate and subsample. An iteration refers to a single round of building trees in the ensemble model of the CatBoost model. During each iteration, the algorithm selects a subset of features in the crop yield dataset and a split point to create a new tree. The leaf values of the tree are then updated based on the training data. This process is repeated until the required number of trees has been built. A higher number of iterations will generally lead to a more complex model that can fit the training data better. However, too many iterations can also lead to overfitting which means that the model will perform poorly on new data. Here, the iterations value in the range 100 to 1000 is given as the hyperparameter range for the Optuna framework. The colsample_bylevel parameter manages the fraction of columns to be randomly sampled for each tree. A lower value will result in more diversity among the trees but may also lead to overfitting. A higher value will result in less diversity but may also lead to underfitting. Here, the colsample_bylevel value in the range 0.5–1.0 is given as the input to the Optuna framework. The depth parameter controls the maximum depth of the trees. A deeper tree can learn more complex patterns in the crop yield dataset but may also lead to overfitting. A shallower tree will be less likely to overfit but may not be able to learn complex patterns. Here, the depth value in the range 3–10 is given as the input to the Optuna framework. The l2_leaf_reg parameter controls the L2 regularization of the leaf weights. L2 regularization penalizes large weights which can help to prevent overfitting. A higher value of l2_leaf_reg will result in more regularization and a lower value will result in less regularization. Here, the l2_leaf_reg value in the range 0.1 to 10 is given as the input to the Optuna framework. The learning rate parameter regulates the step size of the gradient descent algorithm. A higher value of learning rate will result in faster training but may also lead to overfitting. A lower value of learning rate will result in slower training but may also lead to underfitting. Here, the learning rate value in the range 0.001 to 0.1 is given as the input to the Optuna framework. The subsample parameter governs the fraction of training samples that are used to build each tree in the ensemble model. A lower value of subsample will result in less data being used to build each tree which can lead to more diversity among the trees. This can be helpful in preventing overfitting, especially when working with large datasets. However, a lower value of subsample can also lead to underfitting. Here, the subsample value in the range 0.6–0.9 is given as the input to the Optuna framework. Finally, the iteration of 1000, colsample_bylevel of 0.9, depth of 5, l2_leaf_reg of 6, learning rate of 0.1, subsample of 0.8 is selected for the proposed system using Optuna framework with Bayesian optimization. Early stopping criteria are used as the stopping criteria in this proposed CatBoost model. It is a widely used technique that monitors the model's performance on a validation set during training. If the validation performance stops improving for a ten number of iterations, then the training is halted to prevent overfitting. This technique is particularly effective in the Optuna framework as it allows the optimization algorithm to focus on hyperparameter combinations that lead to better generalization performance. The lead time for prediction using the proposed CatBoost model can be for a short duration of 2 years.

Steps involved in the construction of CatBoost regression model for predicting the crop yield.

(1) Import the required libraries in Python.
(2) Load the crop yield dataset and perform mean-based data imputation method.
(3) Split the dataset for training and testing.
(4) The CatBoost model locates and learns about the categorical features.
(5) Train the CatBoost regression model with the training dataset using ordered boosting method with permutation-based MVS.
(6) Tune the hyperparameters of CatBoost model with Bayesian optimization using Optuna framework.
(7) Test the CatBoost regression model with testing dataset.
(8) Validate the regression model with error metrics.

One hot encoding

One hot encoding is a preprocessing step which is used for processing the data variables by categorizing them inorder to get a better prediction result using machine learning or deep learning models (Li et al. 2018; Okada et al. 2019; Dahouda and Joe 2021). By using one hot encoding, each categorical value is converted into a new categorical column and allocates a binary value either 0 or 1 to the rows of the categorical column. The binary vector is the representation for each integer value. The CNN cannot directly work with the categorical data, so these categorical data are transformed to binary numeric data using one hot encoding. After encoding the input data, the encoded data are given as the input to the proposed CNN model for predicting the crop yield.

Feature extraction using spatio-temporal attention-based convolutional neural network model

The encoded result is given as the input to the CNN model and it starts learning the encoded data to extract the features. The CNN used for extracting the features (Jang et al. 2020) of crop yield dataset consists of four convolutional layers with activation function, four batch normalization layer each followed by max pooling layer and finally a fully connected layer. The activation function used for the proposed model is GELU (Hendrycks and Gimpel 2016; Nguyen et al. 2021) because it is suited for the model by producing less error rates with high precision in predicting the crop yield. Figure 4a represents the proposed hybrid

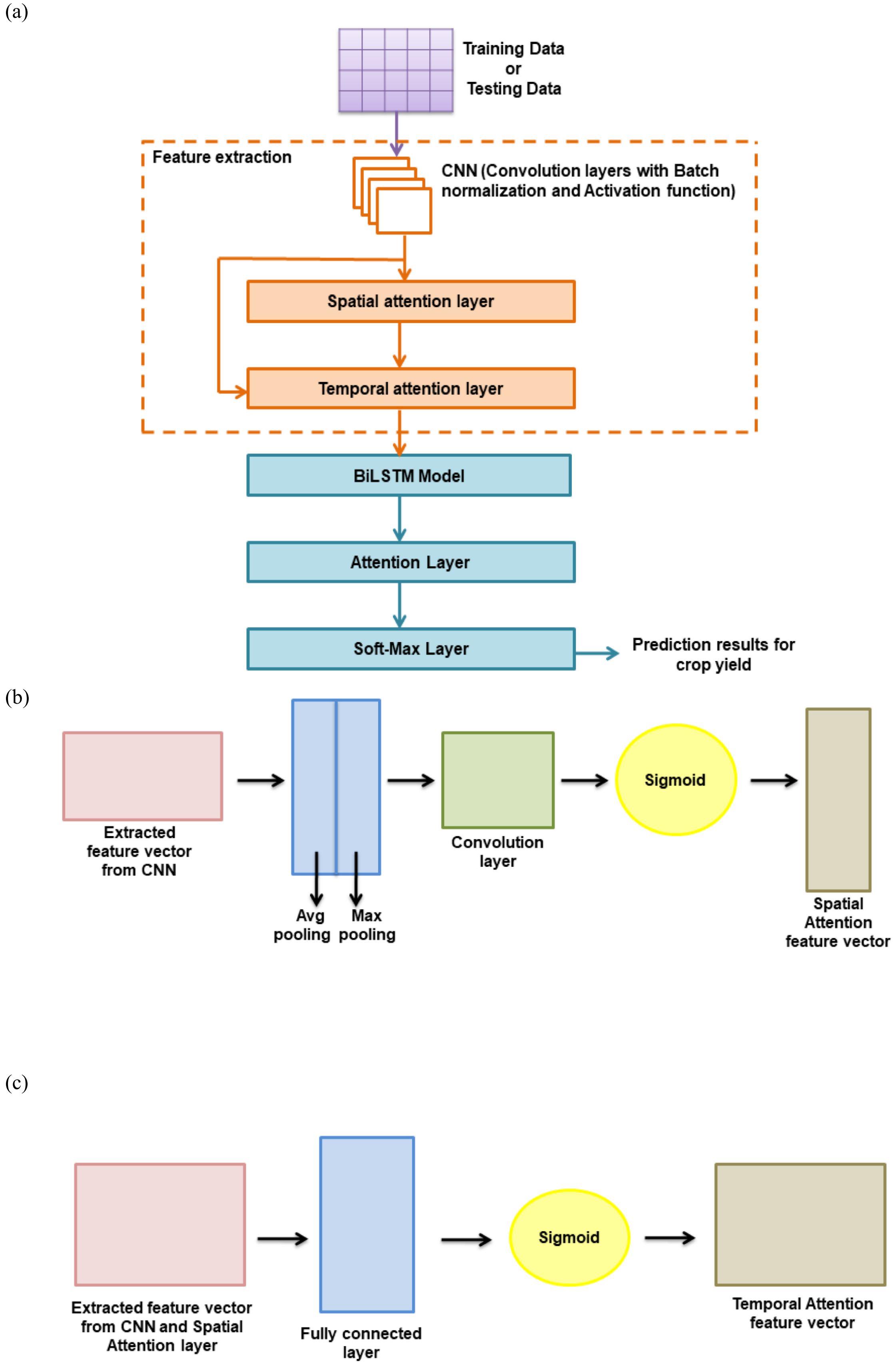

(a)
Training Data
or
Testing Data
Feature extraction
CNN (Convolution layers with Batch normalization and Activation function)
Spatial attention layer
Temporal attention layer
BiLSTM Model
Attention Layer
Soft-Max Layer
Prediction results for crop yield
(b)
Extracted feature vector from CNN
Avg pooling
Max pooling
Convolution layer
Sigmoid
Spatial Attention feature vector
(c)
Extracted feature vector from CNN and Spatial Attention layer
Fully connected layer
Sigmoid
Temporal Attention feature vector

Fig. 4 **a** Proposed hybrid deep learning-based crop yield prediction model. **b** Spatial attention module. **c** Temporal attention module deep learning-based crop yield prediction model. Mathematically, the GELU activation function is represented as:

$$\mathrm{GELU}\left(z\right) = 0.5z(1 + \tanh\left(\frac{\sqrt{2}}{\pi}\left(z + 0.044715z^3\right)\right)\right). \quad (4)$$

The feature vectors of CNN are given as the input to spatial attention network. This network consists of an average pooling layer with maximum pooling layer followed by the convolutional layer and the sigmoid function for extracting the spatial attention features (Pu et al. 2021). The spatial attention network for extracting the high-level features that are spatial is mathematically represented as:

$$A_s(\mathrm{X}) = \sigma\left(x\left[X^s_{\max};X^s_{\mathrm{avg}}\right]\right), \quad (5)$$

where $X^s_{\max}$ represents the maximum pooling operation on the CNN extracted features, X^s_{avg} represents the average pooling operation on the CNN extracted features, σ is the sigmoid function and x denotes the convolutional operation on $X^s_{\max}$ and X^s_{avg}. Next the feature vectors of CNN with spatial attention network are given as the input temporal attention network and the network consists of a fully connected layer and a sigmoid function for extracting the temporal attention features (Tang et al. 2021). Figure 4b and c represents the spatial attention module and temporal attention module. Finally, all the collected features are combined forming the features of crop yield data which is given as the input to the prediction model.

The architecture of STACNN model consists of an input layer with the number of training samples (80% from the total data), 7 features, 1 (number of channels). The convolutional layer 1 follows the input layer with filter size 32, 3 × 3 kernel size and GELU activation function. A batch normalization layer and a max pooling layer with 2 × 2 pool size and stride 2 follows the convolutional layer 1. Next is the convolutional layer 2 with filter size 64, 3 × 3 kernel size and GELU activation function. A batch normalization layer and a max pooling layer with 2 × 2 pool size and stride 2 follows the convolutional layer 2. Next is the convolutional layer 3 with filter size 128, 3 × 3 kernel size and GELU activation function. A batch normalization layer and a max pooling layer with 2 × 2 pool size and stride 2 follows the convolutional layer 3. Next is the convolutional layer 4 with filter size 256, 3 × 3 kernel size and GELU activation function. A batch normalization layer and a max pooling layer with 2 × 2 pool size and stride 2 follows the convolutional layer 4. The subsequent layer is the fully connected layer with 256 units and GELU activation function. Then, the spatial network is built using the average pooling layer, max pooling layer and a convolutional layer. Both the pooling layers in spatial network are with 2 × 2 pool size and stride 2. The convolutional layer of the spatial network is with filter size 1, 1 × 1 kernel size and Sigmoid activation function. After this the temporal network is built using a fully connected layer with 256 units and Sigmoid activation function. A dropout layer with 0.2 is followed by a batch normalization layer with fully connected layer of 128 units and GELU activation function forms the overall feature extraction model.

BiLSTM model for predicting the crop yield

The combined features are given as the input to the BiLSTM model. The BiLSTM is the neural network containing the information in both directions (Hameed and Garcia 2020; Ghasemlounia et al. 2021). Here the extracted combined features flow in two directions i.e., forward and backward directions to preserve the past information and future information (Abduljabbar et al. 2021). This feature of BiLSTM is the main difference from the LSTM model. The BiLSTM model in this research work consists of two LSTMs to process the features for constructing the prediction model. The forward layer output and backward layer output are denoted by $\overrightarrow{h}$ and $\overleftarrow{h}$. BiLSTM network generates an output vector Y and it is mathematically represented as:

$$\mathrm{Y} = \sigma\left(\overrightarrow{h},\overleftarrow{h}\right), \quad (6)$$

where σ is used to combine both the outputs. Similar to LSTM model, the final output of BiLSTM is also represented as vector. The output vector of BiLSTM is given as the input to attention mechanism. The attention mechanism concentrates only on the important information and computes the weight function. Using the attention layer and SoftMax layer, the effective prediction model for crop yield has been constructed. The model is constructed with 30 epochs, 1000 batch size and 0.01 learning rate gives best prediction results.

The architecture of proposed BiLSTM model consists of a two BiLSTM layer with 128 units and 64 units. Next is the dense layer 1 with 64 units and GELU activation function. A dropout of 0.2 and batch normalization are placed after the dense layer 1. Next is the dense layer 2 with 32 units and GELU activation function. A dropout of 0.2 and batch normalization are placed after the dense layer 2. Now the attention layer is used to focus on the most relevant parts of the input sequence when making the yield prediction. Finally, a SoftMax layer is employed with dense units 1 and sigmoid activation function. The model is compiled using Adam optimizer (Bock and Weiß, 2019; Chandriah and Naraganahalli 2021) which gives the best performance for forecasting the crop yield. Early stopping criteria is used as the stopping criteria for 10 number of iterations. The lead time for prediction using the proposed hybrid STACNN-BiLSTM model can be for a duration of 6 to 7 years depending on the dataset.

Steps involved in the construction of STACNN-BiLSTM regression model for predicting the crop yield.

(1) Import the required libraries in Python.
(2) Load the crop yield dataset and perform mean-based data imputation method and one hot encoding.
(3) Split the dataset for training and testing.
(4) Extract the features using spatio-temporal attention-based convolutional neural network model for the training dataset.
(5) From the extracted features, train the BiLSTM prediction model.
(6) Test the BiLSTM model for the testing dataset.
(7) Validate the prediction model with error metrics.

Loss calculation

The loss is calculated for the regression-based prediction model and deep learning-based prediction model using the evaluation metrics. In this research work, five techniques such as mean squared error (MSE) (Prasad and Rao 1990; Wang and Bovik 2009), mean absolute error (MAE) (Willmatt and Matsuura 2005), root mean squared error (RMSE) (Yuan 2022), mean absolute percentage error (MAPE) (De Myttenaere et al. 2016) and R-squared (Cameron and Windmeijer 1997) are used to find the error percentages of the predicting models.

Mean absolute error (MAE)

A mean absolute error metric is used for calculating the accuracy of the continuous variable (Kim and Kim 2016; Reich et al. 2016). The MAE is used to compute the similarity of the prediction for the possible results. The summation of all absolute errors is the mean absolute error (MAE) and it is also known as absolute accuracy error. The absolute error is the quantity of error during observation. It is the summation of absolute dissimilarity between the predicted values and actual values. The expression for the mean absolute error (MAE) is represented as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{m=1}^{n}\left|p_m - a_m\right|, \tag{7}$$

where n = the number of observed errors, Σ = summation (which means "sum of all"), a_m = actual value, p_m = predicted value and $|p_m - a_m|$ = absolute errors.

Mean squared error (MSE)

The MSE is sensitive to outliers. So, even if there are few outliers for a well fitted model it has a very high error value (Chen and Liu 1993). It is the average squared dissimilarity values between the actual values and predicted values. The expression for the MSE is given below:

$$\mathrm{MSE} = \sum_{m=1}^{n}\frac{(p_m - a_m)^2}{n}. \tag{8}$$

Root mean squared error (RMSE)

The RMSE is the square root of mean square error and it is the square root of MSE (Kamble and Deshmukh 2017). The expression for the RMSE is given below:

$$\mathrm{RMSE} = \sqrt{\sum_{m=1}^{n}\frac{(p_m - a_m)^2}{n}}. \tag{9}$$

Mean absolute percentage error (MAPE)

MAPE is an evaluation measure that is used to calculate how accurate the prediction system is (Al-Khowarizmi et al. 2021). It calculates the prediction accuracy as a percentage. The expression for the mean absolute percentage error (MAPE) is represented as:

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{m=1}^{n}\frac{\left|a_m - p_m\right|}{\left|a_m\right|}. \tag{10}$$

R-squared (R^2)

It is the statistical measure of the regression model that calculates the proportion of variance in the resultant that is explained by the predictor variables (Coxe et al. 2009; Osgood 2017). In multiple regression models, R-squared corresponds to the squared correlation between the actual values and the predicted values by the model. The R-squared is measured by sum of squares of residuals from the regression model divided by total sum of squares of errors from the average model and then it is subtracted from 1. It takes any value between 0 and 1. When the R-squared value is high, i.e., 1, then it is the best model. It is formulated as:

$$R^2 1 - \left[\frac{RSS}{TSS}\right] = 1 - \frac{\left(a_m - p_m\right)^2}{\left(a_m - \bar{P}_m\right)^2}, \tag{11}$$

where RSS denotes the sum of squares of residuals, TSS denotes the total sum of squares and $\overline{P}_m$ denotes the mean value of m.

Results

The collected crop yield datasets are divided into two in the ratio 4:1, i.e., 80% of the data is used for training and 20% of the data is used testing. By testing the models with existing data, the prediction models can be evaluated. The proposed two prediction models can also predict the crop yield for the upcoming years. Table 1 denotes the sample of the crop yield dataset used for the prediction models and Table 2 represents the state wise total crop yield in India between 1997 and 2015 according to the Kaggle crop yield dataset. The summation of errors in deep learning models is denoted as loss of the model. Accuracy of the deep learning models is the process of calculating the prediction efficiency of the model. When the model's accuracy is less, it is due to the presence of large number of errors, i.e., high loss values. When the model's accuracy is high, then it is due to low loss values, i.e., small number of errors. So, it is important for any deep learning model to calculate its loss and accuracy for each epochs during training and testing. The STACNN-BiLSTM model has been trained multiple times. According to the results, the prediction models have been run ten times for predicting the crop yield. The models have been trained on different subsets of the data using k-fold cross-validation technique (Anguita et al. 2012; Jung 2018) to evaluate the performances of the model and to avoid the overfitting. The training loss with accuracy and testing loss with accuracy are plotted by calculating the average of all the ten runs. Figures 5 and 6 denote the accuracy and loss values plots for both training and testing of the proposed hybrid deep learning model.

Based on the plot from Figs. 5 and 6, the model has a high accuracy on both the training and testing crop yield

Table 2 State wise total crop yield in India between 1997 and 2015 according to the Kaggle crop yield dataset

State_Name	Production (Ton)
Andaman and Nicobar Islands	1,057,814.64
Andhra Pradesh	17,324,590,296
Arunachal Pradesh	6,823,912.6
Assam	2,111,751,759
Bihar	366,483,596.7
Chandigarh	101,015,864.5
Dadra and Nagar Haveli	1,847,871
Goa	505,755,757.6
Gujarat	524,291,337
Haryana	381,273,890
Himachal Pradesh	17,805,168.6
Jammu and Kashmir	13,291,015.7
Jharkhand	10,777,741.75
Karnataka	863,429,811.7
Kerala	97,880,045,376
Madhya Pradesh	448,840,738.7
Maharashtra	1,263,640,606
Manipur	5,230,917
Meghalaya	12,112,496
Mizoram	1,661,539.83
Nagaland	12,765,950
Odisha	160,904,070.1
Puducherry	384,724,502
Punjab	586,385,001
Rajasthan	281,320,270.5
Sikkim	2,435,735
Tamil Nadu	12,076,443,049
Telangana	335,147,930
Tripura	12,522,917
Uttar Pradesh	3,234,492,663
Uttarakhand	132,177,355
West Bengal	1,397,904,390

Table 1 Sample of the crop yield dataset

State_name	District_name	Crop_year	Season	Crop	Area (hectare)	Production (ton)
Andaman and Nicobar Islands	Nicobars	2001	Whole year	Coconut	18,190	64,430,000
Andhra Pradesh	Anantapur	1998	Kharif	Dry chillies	4000	10,000
Bihar	Gopalganj	2006	Rabi	Maize	5626	11,134
Chandigarh	Chandigarh	2008	Rabi	Wheat	600	2700
Haryana	Kaithal	2012	Whole year	Sugarcane	3511	349,000
Karnataka	Belgaum	2013	Kharif	Bajra	12,444	8204
Tamil Nadu	Villupuram	2013	Whole year	Turmeric	2252	5040
West Bengal	Purulia	2014	Rabi	Urad	220	113
Punjab	Amritsar	2016	Kharif	Rice	366,000	1,188,000

Fig. 5 Accuracy plot for STACNN-BiLSTM model

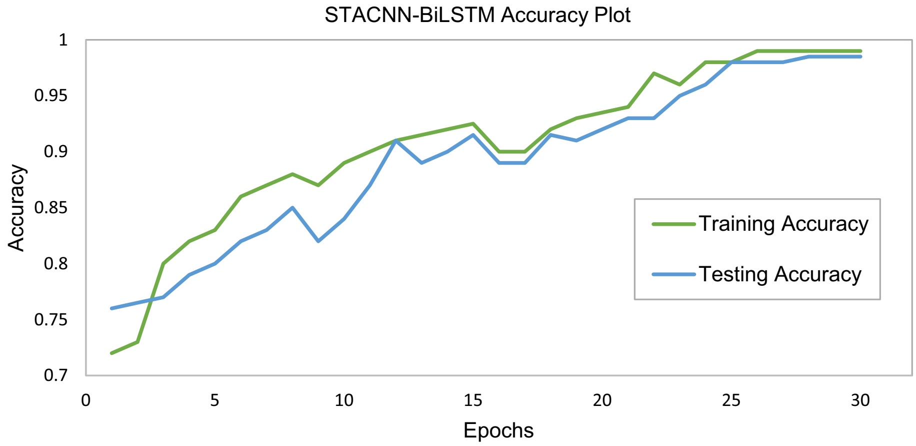
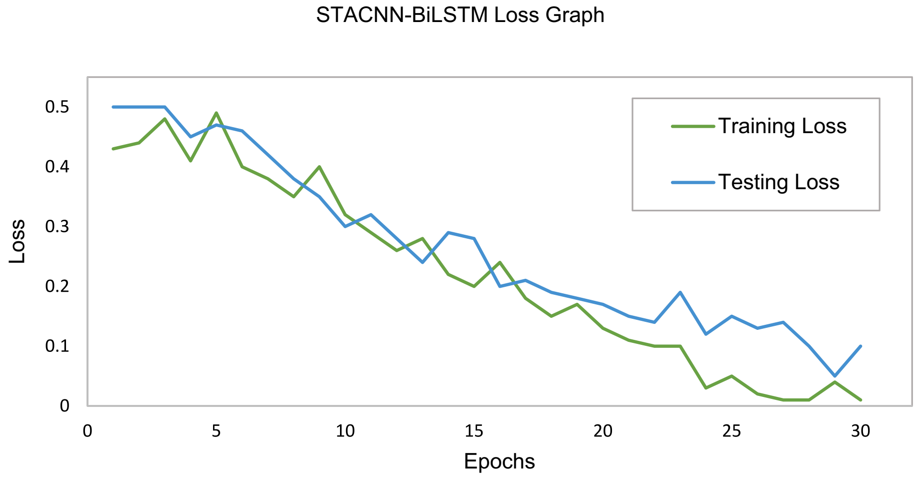


Fig. 6 Loss plot for STACNN-BiLSTM model



data which implies that it can predict the crop yield well based on the input features. It is seen that the accuracy increases as the number of epochs increases indicating that the model learns from the data and improves its predictions over time. The model has a low loss on both the training and testing data which means that it minimizes the differences between the predicted and actual crop yield. The loss decreases as the number of epochs increases indicating that the model converges to a good solution and reduces its uncertainty over time. The model does not show any signs of overfitting or underfitting which are common problems in deep learning models that affects the performances of the model. The proposed STACNN-BiLSTM model avoids these problems by having high accuracy and minimal loss on both the training and testing sets indicating that it has a good balance between complexity and simplicity. The model reaches a plateau after around 25 epochs which means that it does not improve much after that point. This indicates that the model has reached its optimal performance, so the number of epochs is 30. It also implies that the model has enough data and features to make accurate predictions.

The proposed two models, predict the crop yield by the process of regression. In the regression plot, the actual value of the yield is compared with the predicted value of the yield. Then, the performance of the regression model is calculated by comparing the predicted yield values for the given input with the actual yield values. The regression graph can be plotted using the predicted and actual values of the yield. Figures 7 and 8 denote the regression graph for the actual and predicted values of the yield using CatBoost regression model and STACNN-BiLSTM model. By comparing both the regression plots, it is seen that the points of the yield values are more evenly distributed around the regression line for the proposed CatBoost model than the proposed hybrid deep learning model. Hence, this proves the effectiveness of both prediction models for forecasting the yield.

The proposed prediction models are evaluated using the error metrics. From the evaluation, it is seen that the CatBoost prediction model has the minimal error compared to the hybrid deep learning model for all error metrics. Table 3 shows the evaluation results of the proposed prediction models using the error metrics. Inferring from Table 3, it is seen that the differences in the RMSE, MSE and MAE values

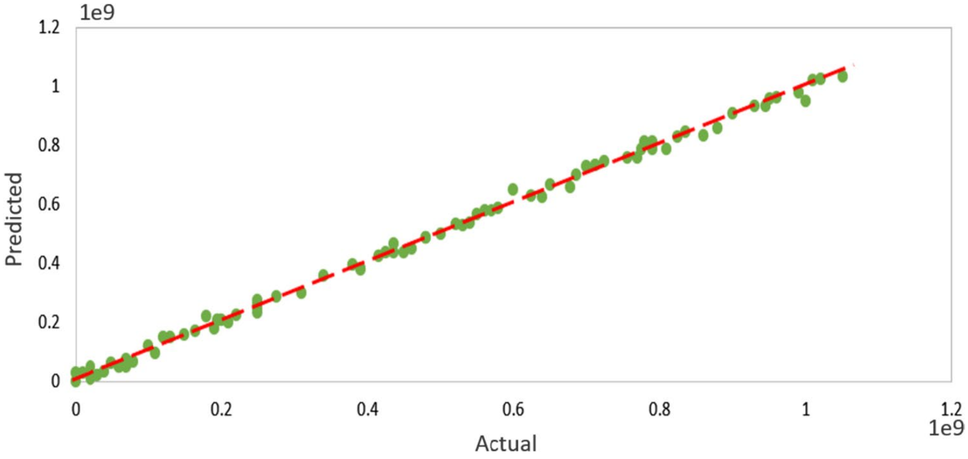


Fig. 7 Regression graph for the actual and predicted values of the yield using CatBoost model

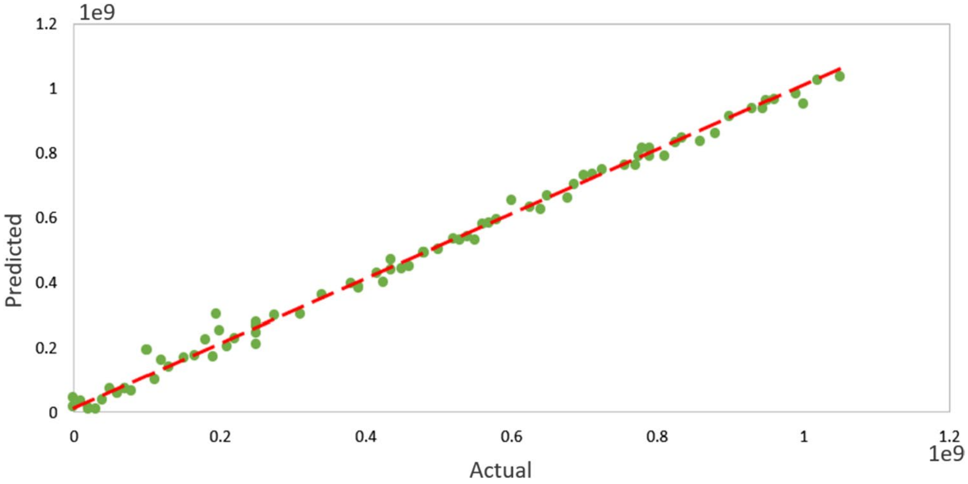


Fig. 8 Regression graph for the actual and predicted values of the yield using STACNN-BiLSTM model

Table 3 Performance comparison of the proposed prediction models

Evaluation metric	CatBoost regression model	STACNN-BiLSTM prediction model
RMSE	0.228	0.233
MSE	0.052	0.054
MAE	0.136	0.158
MAPE	0.182	0.245
R–Squared (R^2)	0.99	0.98

Table 4 RMSE values calculated for all the combination of the selected high performing activation functions and optimizer methods

Activation function	Optimization Method			
	SGD	RMSPROP	ADAGRAD	ADAM
RELU	0.982	0.927	0.778	0.418
LEAKY RELU	0.935	0.869	0.710	0.394
PARAMETRIC RELU	0.886	0.808	0.653	0.362
GELU	0.691	0.634	0.457	0.233

between the two models 0.005, 0.002 and 0.022, respectively. The differences in RMSE, MSE and MAE are relatively very small. Thus, indicates both the prediction models are making fairly accurate predictions. However, the difference in the MAPE value is 0.063 which is slightly high. This proves that the CatBoost regression model is making more accurate predictions relative to the magnitude of the actual yield values. The R-squared values of 0.99 for the CatBoost model and 0.98 for the STACNN model indicate that the CatBoost model is able to explain 99% of the variance in the yield data while the STACNN model is able to explain 98% of the variance in the yield data. Both the models explain a very high proportion of the variance in the yield data. This means that the models are able to predict the yield accurately based on the input features. However, the CatBoost model has a slightly higher R-squared value than the STACNN model which suggests that it is slightly a better model for predicting the yield than the STACNN-BiLSTM model.

The RMSE values are calculated for all the combination of the selected high performing activation functions and optimizer methods and it is shown in Table 4. The activation function and optimization method are chosen for the STACNN-BiLSTM model based on its performance for predicting the crop yield. The optimization methods have a larger impact on the RMSE than the activation functions. The RMSE values increase from Adam to AdaGrad to RMSProp to SGD regardless of the activation function. This suggests that Adam is the most effective optimization method for predicting the crop yield, while SGD is the least effective. The activation functions have a smaller impact on the RMSE than the optimization methods. The RMSE values decrease from RELU to Leaky RELU to parametric RELU to GELU regardless of the optimization method. This suggests that GELU is the most suitable activation function for predicting the crop yield, while RELU is the least suitable.

Based on Table 4, the optimization method that yields the lowest RMSE value of 0.233 for predicting crop yield is Adam paired with the GELU activation function. Hence, the best combination of optimization method and activation function is Adam and GELU with the lowest RMSE. It means that this combination of the model has the highest accuracy and the least error among all other combinations with the models. The worst combination of optimization method and activation function is SGD and RELU which has the highest RMSE of 0.982. It means that this combination of the model has the lowest accuracy and the most error among all other combinations with the models.

Discussion

The proposed CatBoost model has RMSE of 0.228, MSE of 0.052, MAE of 0.136, MAPE of 0.182 and R^2 of 0.99. The proposed hybrid deep learning model has RMSE of 0.233, MSE of 0.054, MAE of 0.182, MAPE of 0.245 and R^2 of 0.98. The performances of the proposed prediction models are compared with latest contemporary models and it is seen that both the proposed models outperform all other existing models for predicting the crop yield. Table 5 shows the comparison of the performances of the proposed models with the latest existing models and the evaluated error metrics values are indicating the proposed two models are best with respect to the existing systems.

Earlier the research work done by Khaki and Wang (2019) proposed a DNN model for predicting the yield of the hybrid corn plant. Two thousand two hundred and sixty-seven different maize hybrid plants grown in different locations of the USA between the years 2008–2016 were used to train the DNN model. The DNN model forecasts yield using genotype and environmental data, achieving an RMSE of 12.41. In contrast, the CatBoost and STACNN-BiLSTM models exhibit significantly lower RMSE values, measuring 0.228 and 0.233, respectively, along with other essential performance metrics. The DNN model suffers from black box property, i.e., the DNN model fails to provide insights that makes difficulty in production of testable hypotheses, whereas the CatBoost model is interpretable providing insights into the relative importance of different features in the predictions. Furthermore, the CatBoost model facilitates the acquisition of valuable knowledge about the factors influencing the yield, potentially leading to the development of new hypotheses. STACNN-BiLSTM is partially interpretable where the CNN part can provide insights into spatial patterns and the BiLSTM part can be challenging to interpret due to its sequential nature.

The sugarcane yield forecasting using data mining with crop simulation methods namely RF, gradient boosting machine, SVM and agroecological zone model was proposed by Hammer et al., (2020) between the years 2011–2015 for 18 sugar mills database with three seasons. Observations indicate that when predicting sugarcane yield, models such as RF, gradient boosting machine, SVM and agroecological zone model yield RMSE and MAE values ranging from 19.702 to 33.368 and 14.928 to 23.698. In stark contrast, the CatBoost and STACNN-BiLSTM models demonstrate significantly lower RMSE and MAE values, registering at 0.228 and 0.136 for CatBoost model, 0.233 and 0.182 for STACNN-BiLSTM model, respectively. Despite the utilization of sugarcane data spanning only five years and

Table 5 Comparison of the performances of the proposed models with the latest existing models

S. No	Proposed by	Year	RMSE	MSE	MAE	MAPE	R^2
1	Khaki and Wang	2019	12.41	–	–	–	–
2	Hammer et al.	2020	19.7	–	14.831	–	–
3	Oikonomidis et al.	2022	0.266	0.071	0.199	–	0.87
4	Burdett et al.	2022	27.31	746	10.07	–	0.93
5	Vance et al.	2022	–	–	0.240	–	0.98
6	Saravanan and Bhagavathiappan	2022	6.5	42.25	4.2	3	–
7	Nayana et al.	2022	23	–	18.92	–	0.99
8	Kuradusenge et al.	2023	320.39	–	257.44	–	0.85
9	Proposed work–CatBoost Regression model		0.228	0.052	0.136	0.182	0.99
	Proposed work—STACNN—BiLSTM model		0.233	0.054	0.182	0.245	0.98

encompassing three seasons across 18 mills, the error metrics indicate a slightly elevated level. This suggests that the predictive performance of these models may face challenges when extrapolated to larger datasets. In this research work, a dataset spanning 20 years and covering six seasons across 33 states in India was employed by demonstrating the importance of considering a more extensive and diverse dataset for robust model evaluation.

Oikonomidis et al., (2022) in their experimental study for predicting the soybean yield in 9 states of the US for the period from 1980 to 2018 using ML and DL models namely XGBoost, CNN-DNN, CNN-XGBoost, CNN-RNN and CNN-LSTM. It is stated that out of five models assessed, the CNN-DNN stands out as the top performer, boasting lower values for RMSE (0.266), MSE (0.071), MAE (0.199) and higher R^2 (0.87). However, it is noteworthy that all these error metric values, although indicating superior performance, are surpassed by the two models introduced in this research work. Oikonomidis et al. models were exclusively utilized for soybean crops within the corn belt in the US and did not provide information on the selection process for features or details about how and which features were chosen. In contrast, in this research, a broader scope was undertaken incorporating 124 different crops and encompassing spatio-temporal features.

Burdett et al. (2022) undergone project study regarding corn and soybean yield forecasting in Southwestern Ontario, Canada, for 1,45,500 observations of 17 fields. They have concluded that RF using cross-validation method was found to be better with 27.31 RMSE, 10.07 MAE, 746 MSE and 0.93 R^2 when compared to multiple linear regression, ANN and decision trees. The project by Burdett et al. was focused for only one year; the performance measures for many years of yield data cannot be ascertained. The two proposed models in this research work demonstrate even lower error metrics than RF with notable differences between their values. Additionally, this research work spans 20 years of data allowing for the prediction of performance over multiple years if sufficient data are available.

The yield prediction of Alfalfa variety for four spatial regions, namely, Kentucky, Georgia, Wisconsin and Mississippi by Vance et al. (2022) suggest that the RF method exhibits good results with MAE (0.240) and R^2 (0.982) while comparing with decision trees, KNN, SVM, Bayesian ridge regression and linear regression. Whereas the proposed two models of this work applied to 33 spatial regions containing 124 varieties of crops exhibit significantly lower MAE and higher R^2 values. Specifically, the CatBoost model records 0.136 for MAE and 0.99 for R^2, while the STACNN-BiLSTM model achieves 0.182 for MAE and 0.98 for R^2. Saravanan and Bhagavthiappan (2022) conducted crop yield prediction for 124 varieties of crops in India for the period 1997 to 2015. Among three proposed methodologies, LSTM-based stacked autoencoder with DNN gives superior results with 6.5 RMSE, 4.2 MAE and 3 MAPE when compared with PCA-XGBoost and PCA-AdaBoost. The research work by Saravanan and Bhagavthiappan used only the kaggle crop yield dataset, whereas the proposed work utilized kaggle dataset and TCI data for improving the crop yield predictions.

Nayana et al., (2022) performed wheat yield forecasting for four states (Haryana, Punjab, Rajasthan, Uttar Pradesh) in India during the period 1962 to 2018 and explored that PCA combined with multivariate adaptive regression splines (MARS) shows encouraging results with 23 RMSE, 18.92 MAE and 0.99 R^2. However, these values for RMSE and MAE were found to be very higher when compared to the performance of the proposed two models in this work. Later, several research works identified that the hybrid PCA-based MARS model did not result in accurate prediction for many features and vary based on the context. In contrast, the proposed models of this research work demonstrate the capability to yield accurate predictions for numerous features. The yield prediction for Irish potatoes and maize for Musanze, district of Rwanda was conducted by Kuradusenge et al. (2023) using RF, polynomial regression and SV regression techniques. Among three, RF highlighted better outcomes with 320.39 RMSE, 257.44 MAE and 0.85 R^2. The limitation of Kuradusenge et al. research work using the RF model has higher values of RMSE and MAE with lower value of R^2 for smaller dataset. In contrast, the proposed models of this research work have lower values of RMSE and MAE with higher values of R^2 for larger dataset.

From the comparison of proposed models with the existing models, it is inferred that each existing model is designed for a limited number of crops, spatial regions and years. Additionally, just a small number of evaluation metrics are used to assess the current models, while all relevant evaluation metrics are used to assess the proposed models. From the evaluated values of performance measures for the proposed models, it is inferred that the error metrics values are very low and R^2 value is high that resulted in more accurate prediction of crop yield. On exploring all the above existing research works, the proposed CatBoost model and hybrid deep learning model have been identified as outperforming in predicting crop yield based on the assessed performance measures. Among the two proposed prediction models, the CatBoost model has demonstrated superior performance, substantiated by its evaluation using error metrics.

Conclusion

The proposed machine learning-based CatBoost regression model and the hybrid deep learning-based crop yield prediction model are successful in modeling the crop yield data for 21 years from 1997 to 2017. Among two proposed models, the CatBoost regression model performs imperceptibly better than the SPACNN–BiLSTM model for predicting the crop yield. The prime strategies behind remarkable performance of the proposed CatBoost regression model are the ability to directly handle the categorical features of the crop yield data and tuning of hyperparameters with Optuna framework. While comparing with the latest contemporary models, both the proposed models outperform all other latest existing models by accurately predicting the crop yield with very low error values. Environmental impacts like natural disasters, global warming and pollution may influence the performance of the proposed models. Future work can be focused on collecting the crop yield historical and real-time data for most of the countries and for many years because still the agriculture domain mainly lacks in the research areas due to its insufficient benchmark data availability across the web.

Author contributions KSS contributed to the conceptualization, methodology, resources, data curation, writing—original draft preparation, writing—review and editing, investigation and validation. VB was involved in the methodology, resources, supervision and validation.

Funding No funding has been claimed for this research work.

Availability of data and materials The dataset for crop yield is collected from Kaggle website (The source of the dataset used in this research work is https://www.kaggle.com/datasets/abhinand05/crop-production-in-india/data or https://www.kaggle.com/code/ anjali21/indian-production-analysis-and-prediction/data) and Tata-Cornell Institute (TCI) website. The Kaggle website crop yield data are 646 districts of 33 Indian states and consists of historical information of crop yields between the years 1997–2015. Seven instances are present in the Kaggle crop yield dataset namely State name, District name, Crop year, Season, Crop name, Area and Production with 246091different attributes. The Kaggle crop yield dataset consists of nearly 124 types of crops grown all over in India. The crop yield dataset for district level Indian agriculture from TCI website is created by International Crops Research Institute for the Semi-Arid Tropics and TCI. The crop yield TCI dataset is collected for the years 2016 and 2017 which has 18,009 different attributes. Both the datasets are used for the proposed work by merging the data according to the common instances.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval Since this research work deals with text data, ethical approval is not applicable.

References

Abduljabbar RL, Dia H, Tsai PW (2021) Unidirectional and bidirectional LSTM models for short-term traffic prediction. J Adv Transp 2021:1–16. https://doi.org/10.1155/2021/5589075

Agarwal S, Tarar S (2021) A hybrid approach for crop yield prediction using machine learning and deep learning algorithms. J Phys Conf Ser 1714(1):012012. https://doi.org/10.1088/1742-6596/1714/1/012012

Agrawal D, Minocha S, Goel AK (2021) Gradient boosting based classification of ion channels. In: 2021 International conference on computing, communication, and intelligent systems (ICCCIS), pp 102–107. IEEE. https://doi.org/10.1109/ICCCIS51004.2021.9397161

Akiba T, Sano S, Yanase T, Ohta T, Koyama M (2019) Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp 2623–2631. https://doi.org/10.1145/3292500.3330701

Al-Khowarizmi RS, Nasution MK, Elveny M (2021) Sensitivity of MAPE using detection rate for big data forecasting crude palm oil on k-nearest neighbor. Int J Electr Comput Eng (IJECE) 11(3):2696–2703. https://doi.org/10.11591/ijece.v11i3.pp2696-2703

Anguita D, Ghelardoni L, Ghio A, Oneto L, Ridella S (2012) The 'K' in K-fold cross validation. In: ESANN, pp 441–446.

Belhadi A, Kamble SS, Mani V, Benkhati I, Touriki FE (2021) An ensemble machine learning approach for forecasting credit risk of agricultural SMEs' investments in agriculture 4.0 through supply chain finance. Ann Oper Res. https://doi.org/10.1007/s10479-021-04366-9

Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. J Mach Learn Res 13(2):281–305

Bock S, Weiß M (2019) A proof of local convergence for the Adam optimizer. In: 2019 International joint conference on neural networks (IJCNN), pp 1–8. IEEE. https://doi.org/10.1109/IJCNN.2019.8852239

Burdett H, Wellen C (2022) Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. Precis Agric. https://doi.org/10.1007/s11119-022-09897-0

Calicioglu O, Flammini A, Bracco S, Bellù L, Sims R (2019) The future challenges of food and agriculture: an integrated analysis of trends and solutions. Sustainability 11(1):222. https://doi.org/10.3390/su11010222

Cameron AC, Windmeijer FA (1997) An R-squared measure of goodness of fit for some common nonlinear regression models. J Econom 77(2):329–342. https://doi.org/10.1016/S0304-4076(96)01818-0

Chandriah KK, Naraganahalli RV (2021) RNN/LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting. Multimed Tools Appl 80(17):26145–26159. https://doi.org/10.1007/s11042-021-10913-0

Chen C, Liu LM (1993) Forecasting time series with outliers. J Forecast 12(1):13–35. https://doi.org/10.1002/for.3980120103

Chlingaryan A, Sukkarieh S, Whelan B (2018) Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. Comput Electron Agric 151:61–69. https://doi.org/10.1016/j.compag.2018.05.012

Coxe S, West SG, Aiken LS (2009) The analysis of count data: a gentle introduction to Poisson regression and its alternatives. J Pers Assess 91:121–136. https://doi.org/10.1080/00223890802634175

Crane-Droesch A (2018) Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environ Res Lett 13:114003. https://doi.org/10.1088/1748-9326/aae159

Dahouda MK, Joe I (2021) A deep-learned embedding technique for categorical features encoding. IEEE Access 9:114381–114391. https://doi.org/10.1109/ACCESS.2021.3104357

De Myttenaere A, Golden B, Le Grand B, Rossi F (2016) Mean absolute percentage error for regression models. Neurocomputing 192:38–48. https://doi.org/10.1016/j.neucom.2015.12.114

Farhangfar A, Kurgan LA, Pedrycz W (2007) A novel framework for imputation of missing values in databases. IEEE Trans Syst Man Cybern A 37:692–709. https://doi.org/10.1109/TSMCA.2007.902631

Fukase E, Martin W (2020) Economic growth, convergence, and world food demand and supply. World Dev 132:104954. https://doi.org/10.1016/j.worlddev.2020.104954

Geng R, Li M, Sun M, Wang Y (2021) Comparing methods of imputation for time series missing values. In: IoT and big data technologies for health care (pp. 333–340). Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-94182-6_24

Ghasemlounia R, Gharehbaghi A, Ahmadi F, Saadatnejadgharahassanlou H (2021) Developing a novel framework for forecasting groundwater level fluctuations using Bi-directional long short-term Memory (BiLSTM) deep neural network. Comput Electron Agric 191:106568. https://doi.org/10.1016/j.compag.2021.106568

Gil JDB, Reidsma P, Giller K, Todman L, Whitmore A, van Ittersum M (2019) Sustainable development goal 2: Improved targets and indicators for agriculture and food security. Ambio 48:685–698. https://doi.org/10.1007/s13280-018-1101-4

Gopal PM, Bhargavi R (2019) A novel approach for efficient crop yield prediction. Comput Electron Agric 165:104968. https://doi.org/10.1016/j.compag.2019.104968

Green JM, Croft SA, Durán AP, Balmford AP, Burgess ND, Fick S, Gardner TA, Godar J, Suavet C, Virah-Sawmy M, Young LE (2019) Linking global drivers of agricultural trade to on-the-ground impacts on biodiversity. Proc Natl Acad Sci USA 116:23202–23208

Gupta A, Nahar P (2022) Classification and yield prediction in smart agriculture system using IoT. J Ambient Intell Humaniz Comput, pp.1–10. https://doi.org/10.1007/s12652-021-03685-w

Hameed Z, Garcia-Zapirain B (2020) Sentiment classification using a single-layered BiLSTM model. IEEE Access 8:73992–74001. https://doi.org/10.1109/ACCESS.2020.2988550

Hammer RG, Sentelhas PC, Mariano JC (2020) Sugarcane yield prediction through data mining and crop simulation models. Sugar Tech 22:216–225. https://doi.org/10.1007/s12355-019-00776-z

Hendrycks D, Gimpel K (2016) Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415. https://doi.org/10.48550/arXiv.1606.08415

Hossain MR, Timmer D (2021) Machine learning model optimization with hyper parameter tuning approach. Glob J Comput Sci Technol D Neural Artif Intell 21(2).

Inoue Y, Moran MS, Horie T (1998) Analysis of spectral measurements in paddy field for predicting rice growth and yield based on a simple crop simulation model. Plant Prod Sci 1(4):269–279. https://doi.org/10.1626/pps.1.269

Ishfaque M, Salman S, Jadoon KZ, Danish AAK, Bangash KU, Qianwei D (2022) Understanding the effect of hydro-climatological parameters on dam seepage using shapley additive explanation (SHAP): a case study of earth-fill tarbela dam, Pakistan. Water 14(17):2598. https://doi.org/10.3390/w14172598

Jang B, Kim M, Harerimana G, Kang SU, Kim JW (2020) Bi-LSTM model to increase accuracy in text classification: combining Word2vec CNN and attention mechanism. Appl Sci 10(17):5841. https://doi.org/10.3390/app10175841

Jung Y (2018) Multiple predicting K-fold cross-validation for model selection. J Nonparametr Stat 30(1):197–215. https://doi.org/10.1080/10485252.2017.1404598

Kamble VB, Deshmukh SN (2017) Comparision between accuracy and MSE, RMSE by using proposed method with imputation technique. Orient J Comput Sci Technol 10(4):773–779. https://doi.org/10.13009/ojcst/10.04.11

Kastner T, Chaudhary A, Gingrich S, Marques A, Persson UM, Bidoglio G, Le Provost G, Schwarzmüller F (2021) Global agricultural trade and land system sustainability: implications for ecosystem carbon storage, biodiversity, and human nutrition. One Earth 4(10):1425–1443

Khaki S, Wang L (2019) Crop yield prediction using deep neural networks. Front Plant Sci 10:621. https://doi.org/10.3389/fpls.2019.00621

Kim S, Kim H (2016) A new metric of absolute percentage error for intermittent demand forecasts. Int J Forecast 32(3):669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003

Kim B, Lee DE, Hu G, Natarajan Y, Preethaa S, Rathinakumar AP (2022) Ensemble machine learning-based approach for predicting of FRP–concrete interfacial bonding. Math 10(2):231. https://doi.org/10.3390/math10020231

Kumar MV, Venugopal P (2016) E-Agriculture and rural development. J Chem Pharm Sci 9(4):3356–3362

Kumar S, Raut RD, Nayal K, Kraus S, Yadav VS, Narkhede BE (2021) To identify industry 4.0 and circular economy adoption barriers in the agriculture supply chain by using ISM-ANP. J Clean Prod 293:126023. https://doi.org/10.1016/j.jclepro.2021.126023

Kuradusenge M, Hitimana E, Hanyurwimfura D, Rukundo P, Mtonga K, Mukasine A, Uwitonze C, Ngabonziza J, Uwamahoro A (2023) Crop yield prediction using machine learning models: case of irish potato and maize. Agric 13(1):225. https://doi.org/10.3390/agriculture13010225

Li J, Si Y, Xu T, Jiang S (2018) Deep convolutional neural network based ECG classification system using information fusion and one-hot encoding techniques. Math Probl Eng 2018:1–10. https://doi.org/10.1155/2018/7354081

Li Y, Guan K, Yu A, Peng B, Zhao L, Li B, Peng J (2019) Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the US. Field Crop Res 234:55–65. https://doi.org/10.1016/j.fcr.2019.02.005

Mallikarjuna Rao GS, Dangeti S, Amiripalli SS (2022) An Efficient modeling based on XGBoost and SVM algorithms to predict crop yield. In: Advances in data science and management (pp. 565–574). Springer, Singapore. https://doi.org/10.1007/978-981-16-5685-9_55

Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B (2019) Definitions, methods, and applications in interpretable machine learning. Proc Natl Acad Sci USA 116:22071–22080. https://doi.org/10.1073/pnas.1900654116

Nayana BM, Kumar KR, Chesneau C (2022) Wheat yield prediction in India using principal component analysis-multivariate adaptive regression splines (PCA-MARS). AgriEng 4:461–474. https://doi.org/10.3390/agriengineering4020030

Nevavuori P, Narra N, Lipping T (2019) Crop yield prediction with deep convolutional neural networks. Comput Electron Agric 163:104859. https://doi.org/10.1016/j.compag.2019.104859

Nguyen A, Pham K, Ngo D, Ngo T, Pham L (2021) An analysis of state-of-the-art activation functions for supervised deep neural network. In: 2021 International conference on system science and engineering (ICSSE) (pp. 215–220). IEEE. https://doi.org/10.1109/ICSSE52999.2021.9538437

Oikonomidis A, Catal C, Kassahun A (2022) Hybrid deep learning-based models for crop yield prediction. Appl Artif Intell. https://doi.org/10.1080/08839514.2022.2031823

Okada S, Ohzeki M, Taguchi S (2019) Efficient partition of integer optimization problems with one-hot encoding. Sci Rep 9:13036. https://doi.org/10.1038/s41598-019-49539-6

Osgood DW (2017) Poisson-based regression analysis of aggregate crime rates. In: Quantitative methods in criminology (pp. 577–599). Routledge.

Palanivel K, Surianarayanan C (2019) An approach for prediction of crop yield using machine learning and big data techniques. Int J Comput Eng Technol 10:110–118

Paudel D, Boogaard H, de Wit A, Janssen S, Osinga S, Pylianidis C, Athanasiadis IN (2021) Machine learning for large-scale crop yield forecasting. Agric Syst 187:103016. https://doi.org/10.1016/j.agsy.2020.103016

Pawlak K, Kołodziejczak M (2020) The role of agriculture in ensuring food security in developing countries: Considerations in the context of the problem of sustainable food production. Sustainability 12:5488. https://doi.org/10.3390/su12135488

Prasad NN, Rao JN (1990) The estimation of the mean squared error of small-area estimators. J Am Stat Assoc 85:163–171. https://doi.org/10.1080/01621459.1990.10475320

Pravin PS, Tan JZM, Yap KS, Wu Z (2022) Hyperparameter optimization strategies for machine learning-based stochastic energy efficient scheduling in cyber-physical production systems. Digit Chem Eng 4:100047. https://doi.org/10.1016/j.dche.2022.100047

Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2018) CatBoost: unbiased boosting with categorical features. Adv Neural Inf Process Syst 31.

Pu C, Huang H, Yang L (2021) An attention-driven convolutional neural network-based multi-level spectral–spatial feature learning for hyperspectral image classification. Expert Syst Appl 185:115663. https://doi.org/10.1016/j.eswa.2021.115663

Ramli MN, Yahaya AS, Ramli NA, Yusof NFFM, Abdullah MMA (2013) Roles of imputation methods for filling the missing values: a review. Adv Environ Biol 7:3861–3870

Reich NG, Lessler J, Sakrejda K, Lauer SA, Iamsirithaworn S, Cummings DA (2016) Case study in evaluating time series prediction models using the relative mean absolute error. Am Stat 70:285–292. https://doi.org/10.1080/00031305.2016.1148631

Ristaino JB, Anderson PK, Bebber DP, Brauman KA, Cunniffe NJ, Fedoroff NV, Finegold C, Garrett KA, Gilligan CA, Jones CM, Martin MD (2021) The persistent threat of emerging plant disease pandemics to global food security. Proc Natl Acad Sci 118(23):e2022239118. https://doi.org/10.1073/pnas.2022239118

Sandha SS, Aggarwal M, Saha SS, Srivastava M (2021) Enabling hyperparameter tuning of machine learning classifiers in production. In: 2021 IEEE third international conference on cognitive machine intelligence (CogMI), pp 262–271. https://doi.org/10.1109/CogMI52975.2021.00041

Saravanan KS, Bhagavathiappan V (2022) A comprehensive approach on predicting the crop yield using hybrid machine learning algorithms. J Agrometeorol 24(2):179–185. https://doi.org/10.54386/jam.v24i2.1561

Shakoor N, Northrup D, Murray S, Mockler TC (2019) Big data driven agriculture: big data analytics in plant breeding, genomics, and the use of remote sensing technologies to advance crop productivity. Plant Phenome J 2(1):1–8. https://doi.org/10.2135/tppj2018.12.0009

Sharma S, Rai S, Krishnan NC (2020) Wheat crop yield prediction using deep LSTM model. arXiv preprint arXiv:2011.01498. https://doi.org/10.48550/arXiv.2011.01498

Shekhar S, Bansode A, Salim A (2021) A comparative study of hyperparameter optimization tools. In: 2021 IEEE asia-pacific conference on computer science and data engineering (CSDE), pp 1–6. https://doi.org/10.1109/CSDE53843.2021.9718485

Shyam R, Ayachit SS, Patil V, Singh A (2020) Competitive analysis of the top gradient boosting machine learning algorithms. In: 2020 2nd international conference on advances in computing, communication control and networking (ICACCCN), pp 191–196. https://doi.org/10.1109/ICACCCN51052.2020.9362840

Tang P, Du P, Xia J, Zhang P, Zhang W (2021) Channel attention-based temporal convolutional network for satellite image time series classification. IEEE Geosci Remote Sens Lett 19:1–5. https://doi.org/10.1109/LGRS.2021.3095505

Tian F, Wu B, Zeng H, Watmough GR, Zhang M, Li Y (2022) Detecting the linkage between arable land use and poverty using machine learning methods at global perspective. Geogr Sustain 3(1):7–20. https://doi.org/10.1016/j.geosus.2022.01.001

Van Klompenburg T, Kassahun A, Catal C (2020) Crop yield prediction using machine learning: a systematic literature review. Comput Electron Agric 177:105709. https://doi.org/10.1016/j.compag.2020.105709

Vance J, Rasheed K, Missaoui A, Maier F, Adkins C, Whitmire C (2022) Comparing machine learning techniques for alfalfa biomass yield prediction. arXiv preprint arXiv:2210.11226. https://doi.org/10.48550/arXiv.2210.11226

Wang Z, Bovik AC (2009) Mean squared error: love it or leave it? A new look at signal fidelity measures. IEEE Signal Process Mag 26(1):98–117. https://doi.org/10.1109/MSP.2008.930649

Whetton R, Zhao Y, Shaddad S, Mouazen AM (2017) Nonlinear parametric modelling to study how soil properties affect crop yields and NDVI. Comput Electron Agric 138:127–136. https://doi.org/10.1016/j.compag.2017.04.016

Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 30(1):79–82. https://doi.org/10.3354/cr030079

Yasir M, Karim AM, Malik SK, Bajaffer AA, Azhar EI (2022) Prediction of antimicrobial minimal inhibitory concentrations for Neisseria gonorrhoeae using machine learning models. Saudi J Biol Sci 29(5):3687–3693. https://doi.org/10.1016/j.sjbs.2022.02.047

Yuan S (2022) Review of root-mean-square error calculation methods for large deployable mesh reflectors. Int J Aerosp Eng. https://doi.org/10.1155/2022/5352146

Zahedi L, Mohammadi FG, Rezapour S, Ohland MW, Amini MH (2021) Search algorithms for automated hyper-parameter tuning. arXiv preprint arXiv:2104.14677. https://doi.org/10.48550/arXiv.2104.14677

Zambon I, Cecchini M, Egidi G, Saporito MG, Colantoni A (2019) Revolution 4.0: industry vs. agriculture in a future development for SMEs. Processes 7(1):36. https://doi.org/10.3390/pr7010036