International Journal of Advances in Applied Sciences (IJAAS)
Vol. 15, No. 1, March 2026, pp. 405~415
ISSN: 2252-8814, DOI: 10.11591/ijaas.v15.i1.pp405-415

Crop prediction in Tamil Nadu according to environmental and soil factors using hybrid machine learning architecture

Sundaraj Kannan Susee[1], Shenbagaramasubramanian Shenbaga Vadivu[2], Murugesan Senthil Kumar[3]
[1]Department of Electronics and Communication Engineering, Mohamed Sathak A J College of Engineering, Chennai, India
[2]Department of Information Technology, SRM Valliammai Engineering College, Chennai, India
[3]Department of Computer Science and Engineering, Karpaga Vinayaga College of Engineering and Technology, Chengalpattu, India

Article Info

Article history:

Received May 14, 2025
Revised Nov 26, 2025
Accepted Feb 17, 2026

Keywords:

Bidirectional long short-term memory
Convolutional LSTM
Crop prediction
Environmental factors
Linear support vector machine
Soil parameters

ABSTRACT

Mathuranthagam, Tamil Nadu, India is the site of this research initiative that employs state-of-the-art hybrid machine learning (ML) architectures to forecast crop suitability in relation to environmental and soil characteristics. The model takes advantage of the strengths of linear support vector machine (SVM) classifier, bidirectional long short-term memory (BiLSTM), and convolutional LSTM (ConvLSTM) networks, and the data to capture complicated temporal and spatial correlations. To prepare the dataset for model training, it is normalized using min-max scaling and then feature selected using a Jaya optimization technique. The dataset contains variables such as humidity, rainfall, temperature, and pH. Both the BiLSTM and the ConvLSTM improve the model's comprehension of context from both previous and subsequent time steps. The ConvLSTM also records spatial dependencies. A powerful decision-making tool for differentiating across crop varieties is the linear SVM classifier. Comparing the hybrid model's performance to that of traditional LSTM approaches using measures such as recall, accuracy, precision, and F1-score shows that it performs much better. Using this approach can see how deep learning (DL) can supplement more conventional ML methods and see how important local environmental data is for agricultural policy and planning.

Corresponding Author:

Sundaraj Kannan Susee
Department of Electronics and Communication Engineering, Mohamed Sathak A J College of Engineering
Chennai, India
Email: suseekannan07@gmail.com

1. INTRODUCTION

Agriculture is a fundamental component of India’s economy, with crop yield significantly influenced by soil properties and environmental factors. In areas include Mathuranthagam, Tamil Nadu, India erratic weather patterns, soil heterogeneity, and climate change provide considerable obstacles to conventional crop planning and management. Traditional methods, dependent on historical data or expert heuristics, often neglect the complex spatio-temporal interactions among environmental and soil variables, resulting in inadequate crop selection and reduced yields. Precise crop suitability prediction is essential for farmers, insurers, and regulators, facilitating informed choices about crop planning, risk management, and resource distribution. This research introduces a hybrid machine learning (ML) framework that combines convolutional long short-term memory (ConvLSTM), bidirectional LSTM (BiLSTM), and linear support vector machine (SVM) classifiers to predict crop suitability for 22 principal crops, such as rice, maize, chickpea, pulses, fruits, and commercial commodities. The model utilises ConvLSTM to capture spatial dependencies in soil and climatic data, BiLSTM to analyse temporal trends, and SVM to enhance classification performance. The input features are temperature, rainfall, humidity, soil pH, NPK concentration, and texture, which are normalised and optimised by feature selection using the Jaya algorithm. The primary aims of this study are to i) improve prediction accuracy compared to individual classifiers or conventional models, ii) provide seasonal and geographical crop suitability maps, and iii) deliver actionable insights for agricultural planning. The primary contributions are: i) a region-specific dataset that captures the soil and climate variability of Mathuranthagam, ii) a hybrid model exhibiting exceptional performance with an accuracy, and iii) visualisations comprising feature importance rankings, seasonal suitability charts, and geographic information systems (GIS)-based prediction maps to facilitate decision-making. Crop forecasting in Mathuranthagam, Chengalpattu, Tamil Nadu, India is crucial for food security, agriculture, and farmer livelihoods. Weather, rainfall, and soil variables, including pH, organic content, and nutrient levels, complicate crop prediction.

An IoT-based climate-adaptive crop recommendation method that uses deep ensemble learning and data detected by IoT sensors to accurately propose the best crops to grow in specific locations at certain times [1]. This research presents a neural network mathematical optimisation hybrid model for environmentally friendly food supply chains [2]. It deals with issues including manufacturing expenses, water shortages, pollution, and industrial waste. This research used the Food and Agriculture Organization (FAO)-agro-climate approach in conjunction with ML algorithms to wheat yield (WY) production in Southwest Iran by analysing soil and environmental parameters [3]. Continuous WY mapping was accomplished using artificial neural networks (ANN) and random forests (RF) [4]. This research compared the effects of climatic variables on the transmission of anthrax infections in India with those of more conventional parameters, such as the distribution of animal populations and the soil-moisture management, soil enrichment, and sustainable agriculture. The study proposes a smart composting approach that uses IoT and gradient boosting algorithms [5]. Predicting agricultural yields, mapping soil fertility, assessing food grain quality, and predicting pest and disease outbreaks are all being transformed by AI applications [6]. The goal of this research is to create a crop forecast system that uses weather, soil pH, and nutrient data collected in real-time and integrated into the IoT [7]. The technology uses deep learning (DL) and ML to determine the local environment and then proposes crops that will thrive there. The pressing problem of crop prediction is addressed by incorporating genetic algorithms into the predictive model and making use of state-of-the-art machine-learning approaches [8].

The IoT is a tool for optimising agricultural output via the use of advanced sensor technologies, protocol connections, ML techniques, and real-time monitoring [9]. An RF model is used to provide precise forecasts based on data sent over long-range wide area network (LoRaWAN) by sensors that evaluate soil nutrients, moisture, and weather conditions. Through the integration of IoT and GIS technologies, this research study presents a novel approach to real-time monitoring of soil health and nutrient status in agricultural regions [10]. Used IoT data and ML algorithms to predict and advise farmers on crops in real time [11]. Ensemble techniques have become popular for crop prediction, mixing numerous models. Bagging and boosting reduce overfitting and improve crop prediction model generalization [12]. This research evaluates the efficacy of several ML models for soil classification, including decision trees, k-nearest neighbors (k-NN), ANN, and SVM [13]. Farmers use soil and environmental variables to forecast their crops. By including pertinent information such as soil composition, temperature, and humidity, ML models using filter, wrapper, and embedding approaches decrease redundancy and temporal complexity [14]. An efficient sequential pattern discovery approach evaluates the PrefixSpan algorithm that recursively makes frequent patterns from prefixes, thus minimizing the search space and improving efficiency [15]. Improved food safety, economic steadiness, and resource efficiency are all outcomes of this study's use of ML to forecast agricultural production [16]. The proposed method predicts the need for expensive and time-consuming traditional soil testing by predicting crop compatibility for soil based on pH, moisture, temperature, and humidity measurements [17]. Utilising decision trees, RF, naive Bayes, k-NN, and SVM proposes an ML strategy for crop type prediction utilising soil and environmental variables [18].

By using environmental parameters to forecast crop results, ML is transforming agriculture. For better yield prediction and better farming methods, this research analyses meteorological conditions, soil qualities, and crop traits to create ML models [19]. Through the integration of data input and IoT sensors, a web-based application predicts and recommends agricultural yields using ML algorithms [20]. The system finds the best methods for prediction works by analysing models such as SVM, RF, and gradient boosting. Aiming to revolutionise crop management using an ML method, this project is motivated by the pressing need for modern agricultural techniques [21]. Crop recommendation methods that take soil and environmental data into account are developed using DL models such as DENSE, recurrent neural network (RNN), and LSTM [22]. Smart agriculture yield and fertilizer optimization system (SAYFOS), a new method for optimizing agricultural yields and fertilisers. With SAYFOS, users can monitor the crops and soil conditions in real time due to its superior data analytics, IoT technology, and ML algorithms [23]. Using ML methods for input characteristics, the research demonstrates two strong ML architectures for regression and classification for the crop recommendation dataset [24]. Reduced agricultural yield loss in India is achieved using algorithms such as linear regression, logistic regression, and SVM to determine the optimal crop for a given soil and environment [25]. Sustainable agriculture benefits from this model's ability to maximize efficiency while reducing negative effects on the environment. By rotating crops, farmers may improve soil fertility, structure, and biodiversity [26].

i) Problem domain: the complex and ever-changing soil and environmental conditions of Mathuranthagam, Tamil Nadu, India pose substantial challenges to the agriculture sector. Improving agricultural methods, increasing crop yields, and providing sustainable food production all depend on accurate crop prediction, as shown in Figure 1. This research presents a hybrid ConvLSTM-BiLSTM-SVM model for crop prediction in Mathuranthagam, Tamil Nadu, India using soil and environmental data. Significant contributions include a region-specific dataset, feature optimisation, spatio-temporal modelling, enhanced performance relative to baselines, and decision-support visualisations for agriculturalists, insurance strategizing, and policy development.
ii) Design: Mathuranthagam, Tamil Nadu, India soil and environmental data are used in the architectural design to improve crop forecasts. Weather parameters, including temperature, humidity, pH, and rainfall, are collected initially. Split the data into training and test datasets after min-max scaling and missing values. Jaya optimization minimizes data dimensionality and picks key features. It improves model performance. SVM classifiers and powerful DL algorithms, such as BiLSTM and ConvLSTM networks, are used to create Figure 2 models.

Figure 1. Region of Maduranthagam map

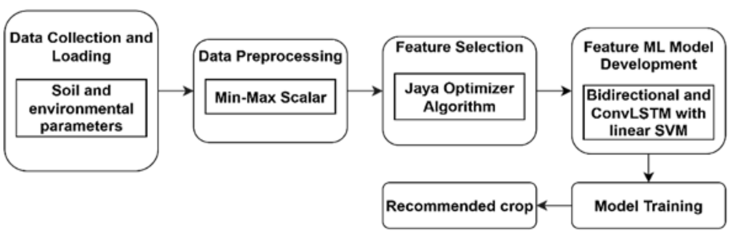

Figure 2. Architecture diagram

2. PROPOSED METHOD

This study's novelty is in its hybrid architecture, which uniquely combines linear SVM, BiLSTM, and ConvLSTM to effectively harness temporal, spatial, and classification capabilities. Furthermore, it employs Jaya-based feature selection for optimization and integrates both soil and climatic variables from localized Mathuranthagam data, providing the model methodologically creative and contextually pertinent for crop prediction in Tamil Nadu. The research uses BiLSTM and ConvLSTM for temporal crop forecasting, since LSTMs proficiently capture long-term relationships and seasonal variations in soil and climatic data. Their integration with ConvLSTM models captures spatial correlations, yielding great accuracy and interpretability, while alternatives such as transformer or gated recurrent unit (GRU), despite their potential, increase complexity. Weather and soil characteristics, including humidity, temperature, pH, and precipitation in record of Mathuranthagam's distinct. Regularizing data via min-max scaling improves model performance and consistency. The min-max scaler sets a per-feature minimum and maximum value during fitting. During transformation, the (1) converts all feature values to 0-1.

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where X represents the initial feature. X_{min} and X_{max} are the feature minimum and maximum values.

i) Properness: the scale sets upper and lower boundaries for every training dataset feature. Next, translate feature values to a 0-1 scale. Scaling is needed for all features to affect model learning equally.
ii) Normalization: continuous variables, including temperature, rainfall, humidity, and soil nutrient values (NPK and pH), are standardized by min-max normalization to ensure uniformity across all features.
iii) Categorical encoding: crop categories are represented by one-hot encoding, facilitating the model's efficient processing of categorical data.
iv) Missing data management: incomplete or missing values are addressed by mean imputation for numerical attributes and mode imputation for categorical variables, therefore preserving dataset integrity. The code imports the data into a NumPy array, normalizes the features with the min-max scaler, and then outputs the scaled data to test. The proposed hybrid architecture integrates linear SVM, BiLSTM, and ConvLSTM networks to exploit their synergistic advantages for crop prediction.
v) ConvLSTM: captures spatial relationships within the dataset, including differences in soil and environmental characteristics over Mathuranthagam.
vi) LSTM: analysis temporal patterns by processing sequences in both directions, therefore capturing interdependence across several cropping seasons.
vii) Linear SVM: functions as a conclusive classifier, delivering effective decision-making for crop appropriateness by distinguishing crop classes according to the acquired feature representations.
viii) Fusion strategy: the outputs from BiLSTM and ConvLSTM are concatenated and input into the linear SVM using a stacked ensemble methodology. This enables the model to include spatial, temporal, and classification functionalities, improving forecast precision.

2.1. Linear support vector machine classifier

The SVM is a supervised ML algorithm primarily used for classification tasks. It works by finding a hyperplane that best divides the data into classes. Linear SVM is particularly useful when the data is linearly separable, meaning that a straight line (or hyperplane in higher dimensions) can separate the different classes. In the context of crop suitability, the linear SVM classifier would be used to predict which type is most suitable based on environmental and soil features such as humidity, rainfall, temperature, and pH. SVMs are robust against overfitting, especially in high-dimensional spaces, and they can efficiently perform non-linear classification using the kernel trick. In this case, the linear SVM helps differentiate between crop varieties effectively, providing powerful decision-making.

2.2. Bidirectional long short-term memory

A BiLSTM network is a type of RNN designed to process sequential data. Bidirectionality: unlike traditional LSTMs that process data sequentially from the past (forward), BiLSTMs process data in both directions (past-to-future and future-to-past). This is especially beneficial when predicting crop suitability because it allows the model to capture both past and future temporal dependencies. For example, weather conditions such as rainfall or temperature at any given time might influence crop growth not only in the past but also in the future. LSTMs are a type of RNN that are good at capturing long-range dependencies in sequences, making them particularly suited for time-series forecasting, such as predicting the suitability of crops over time based on weather and soil conditions.

2.3. Convolutional long short-term memory

The ConvLSTM is another variant of the LSTM that incorporates convolutional operations into the LSTM framework, making it suitable for data with spatial-temporal dependencies. Spatial-temporal dependencies: ConvLSTM is particularly useful when the data has both temporal and spatial components. In the context of crop forecasting, spatial data can include geographic locations, soil quality variations, or different environmental conditions across regions. By using convolutional operations in conjunction with LSTM, the ConvLSTM captures spatial relationships between different regions of the dataset, enhancing the model's ability to learn both the temporal and spatial features of the data. The ConvLSTM will capture dependencies not only across time but also across spatial dimensions.

2.4. Min-max scaling

Before applying the ML models, the dataset is normalized using min-max scaling, a feature scaling technique that rescales all feature values to a common range. The main reason for applying min-max scaling is to make sure that all features have equal weight in the learning process. This is particularly important for models such as SVMs, neural networks, and others that are sensitive to the scale of input features. Without scaling, features with larger values could dominate the learning process, making the model inefficient.

2.5. Jaya optimization technique for feature selection

Jaya optimization is an optimization algorithm that works on finding the best possible solution by iteratively improving the candidate solutions. Feature selection: in your methodology, Jaya optimization is used to select the most relevant features from the dataset. In crop suitability forecasting, features such as humidity, rainfall, and pH are used, but not all features may be equally important. Jaya optimization helps in automatically identifying and selecting the best features that contribute most to predictive accuracy. Jaya optimization works by iterating through candidate solutions and adjusting them based on the best-found solution and the worst-found solution.

2.6. Performance metrics

Using metrics for both classification and regression, evaluate the hybrid model's performance. When false positives are expensive, accuracy is more important than precision in determining total correctness because precision assesses the accuracy of positive predictions. An all-encompassing performance measure, the F1-score strikes a good balance between recall and accuracy by identifying real positive events.

3. RESULTS AND DISCUSSION

This study's dataset was collected from Mathuranthagam, Tamil Nadu, India and it was designed to assist reliable crop prediction using the hybrid ML architecture. The dataset included soil, crops, and environmental data. Providing geographical resolution down to the block level, the data set captures the localized agro-climatic conditions in Mathuranthagam and spans numerous cropping seasons. Rice, maize, and chickpea are some of the most important crops grown in the area, and this information contains their cultivation patterns and practices. Soil characteristics such as pH, nitrogen (N), phosphorus (P), potassium (K), and soil texture are essential for determining crop compatibility and fertility. Temperature, precipitation, and humidity are examples of environmental factors that mirror climatic and seasonal changes that impact agricultural yields. Feature selection utilizing Jaya optimization and min-max normalization are two of the data preparation steps used to keep the most important variables for training the model. Complex spatio-temporal relationships may be captured by the hybrid architecture, which consists of ConvLSTM for spatial dependencies, BiLSTM for temporal patterns, and linear SVM for classification. The dataset integrates both soil and climatic information. The proposed hybrid architecture for crop prediction in Mathuranthagam, Tamil Nadu, India is assessed using a thorough technique to provide robustness and dependability. Hyperparameter tuning is conducted for the ConvLSTM and BiLSTM networks, optimizing the number of layers, hidden units, dropout rates, and learning rates using grid search and Bayesian optimization, while the regularization parameter of the linear SVM is modified for optimum classification. Jaya-based feature selection significantly improves model efficiency. The model is validated using multi-seasonal records of rice, maize, and chickpea to measure performance stability and ensure accurate forecasts despite climatic fluctuation. Performance measures such as accuracy, precision, recall, and F1-score are used to quantify both predictive capability and mistake rates. This assessment approach demonstrates that the hybrid model provides consistent, high-accuracy, and context-specific predictions of crop suitability, making it appropriate for practical applications such as farmer advisories, crop insurance planning, and regional agricultural policy formulation.

Figure 3 shows the monthly trends of temperature, humidity, soil pH, and rainfall at Mathuranthagam. Data was gathered over a span of 100 consecutive days in Mathuranthagam, Tamil Nadu, India a representative agricultural area. Measurements included daily temperature and humidity recorded by weather stations or IoT sensors, soil pH assessed using a digital meter, and precipitation obtained from meteorological stations. The dataset concentrated on primary crops: rice, maize, and chickpea, documenting their distinct microclimates. Documented parameters included daily mean temperature, relative humidity, interpolated soil pH, and daily precipitation. This data aims to predict temporal changes in crop adaptability, functioning as input for a hybrid ML architecture (ConvLSTM-BiLSTM-SVM) to improve agricultural planning. Figure 3(a) shows the temperature graph, which likely represents the variation in temperature over a given period. Temperature is a key factor influencing crop growth, and this graph could be showing how temperature fluctuations throughout the year or during specific seasons impact crop productivity. Figure 3(b) shows a humidity graph showing the changes in relative humidity over time. High humidity levels can promote fungal growth and reduce crop yields, while low humidity can lead to water stress for crops. Figure 3(c) denotes the pH graph displaying the soil's acidity or alkalinity over time, measured on a scale from 0 to 14, with 7 being neutral. Figure 3(d) shows a rainfall graph that represents the amount of precipitation received over a given period, typically shown in millimeters or inches. The dataset used for crop classification consists of 2,200 samples uniformly allocated among 22 crop classes, with each class including 100 samples [27]. The dataset was divided to achieve balanced representation, including 70% for training (1,540 samples) and 30% for testing (660 samples). Every crop category provides 70 training samples and 30 testing samples, ensuring the dataset remains impartial across classifications, including coconut, rice, papaya, maize, chickpea, pigeon peas, pomegranate, moth beans, mung beans, black beans, banana, lentil, apple, mango, grapes, muskmelon, jute, orange, watermelon, kidney beans, cotton, coffee. Figure 4 illustrates the ranks of feature relevance, highlighting rainfall, temperature, and soil pH as the predominant contributors. Figure 5 depicts fluctuations in seasonal suitability, highlighting temporal changes in agricultural potential. Figure 6 shows prediction maps that illustrate spatial suitability zones.

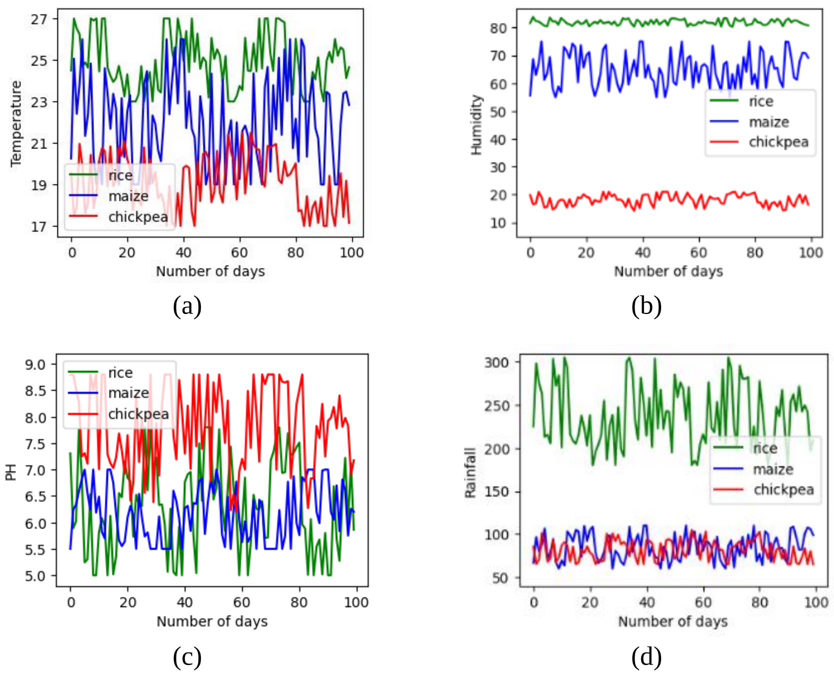

Figure 3. Seasonal variation of key environmental factors (a) temperature, (b) humidity, (c) pH, and (d) rainfall

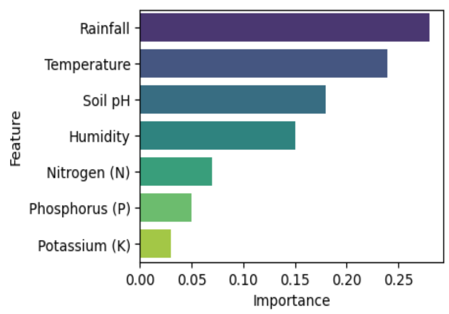

Figure 4. Feature importance

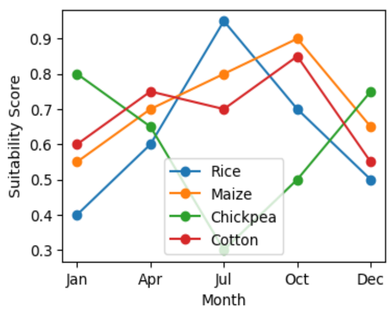

Figure 5. Monthly suitability trend

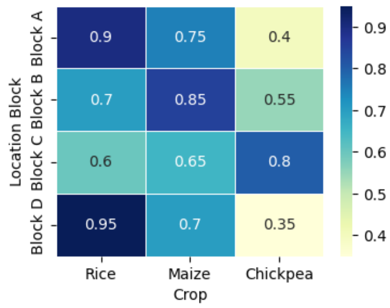

Figure 6. Spatial distribution

This equitable division facilitates rigorous training, impartial assessment, and reduces overfitting, hence improving the model's generalization for various crop prediction tasks. The ConvLSTM confusion matrix in Figure 7 shows a high level of accuracy across all crop classes, as shown by the mostly strong diagonal entries. Showing strong performance with minimal uncertainty across comparable crops, small off-diagonal values indicate rare misclassifications. Figure 8 shows the performance evaluation of the classification report.

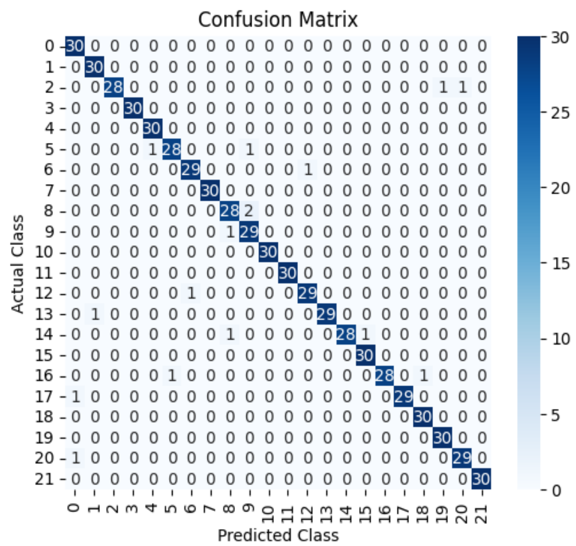

Figure 7. Confusion matrix analysis for crop classes

Accuracy: 0.9758
Precision: 0.9764
Recall: 0.9758
F1 Score: 0.9757

Classification Report:

	precision	recall	f1-score	support
0	0.9375	1.0000	0.9677	30
1	0.9677	1.0000	0.9836	30
2	1.0000	0.9333	0.9655	30
3	1.0000	1.0000	1.0000	30
4	0.9677	1.0000	0.9836	30
5	0.9655	0.9333	0.9492	30
6	0.9667	0.9667	0.9667	30
7	1.0000	1.0000	1.0000	30
8	0.9333	0.9333	0.9333	30
9	0.9062	0.9667	0.9355	30
10	1.0000	1.0000	1.0000	30
11	1.0000	1.0000	1.0000	30
12	0.9667	0.9667	0.9667	30
13	1.0000	0.9667	0.9831	30
14	1.0000	0.9333	0.9655	30
15	0.9677	1.0000	0.9836	30
16	1.0000	0.9333	0.9655	30
17	1.0000	0.9667	0.9831	30
18	0.9677	1.0000	0.9836	30
19	0.9677	1.0000	0.9836	30
20	0.9667	0.9667	0.9667	30
21	1.0000	1.0000	1.0000	30
accuracy			0.9758	660
macro avg	0.9764	0.9758	0.9757	660
weighted avg	0.9764	0.9758	0.9757	660

Figure 8. Model performance evaluation

It may include metrics such as the following:

i) Accuracy: the proportion of correct predictions (both true positives and true negatives) out of the total predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

ii) Precision: the ratio of correctly predicted positive observations to the total predicted positives.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

iii) Recall: the ratio of correctly predicted positive observations to all actual positives.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

iv) F1-score: the weighted average of precision and recall, used to balance both metrics, particularly in cases of imbalanced classes.

$$F1 - Score = 2 \times \frac{(Precision * Recall)}{(Precision + Recall)} \tag{5}$$

Missing or interference in environmental data can bias outcomes, whilst insufficient dataset sizes may lead to overfitting. Class imbalance may result in deceptive accuracy for unusual crops unless suitable safeguards are implemented. Inadequate control of temporal and spatial correlations may distort model efficacy. Ultimately, using training data for testing or inadequate cross-validation may artificially enhance performance metrics. This research improves agricultural decision-making by forecasting crop suitability in Mathuranthagam, Tamil Nadu, India using a hybrid model that combines linear SVM, BiLSTM, and ConvLSTM. It incorporates climatic variables such as temperature and precipitation, assisting farmers in optimum crop selection and facilitating regional agricultural planning. The models include climate variability, which facilitates the adoption of resilient farming practices and methods, enhancing food security and promoting sustainable agriculture in the face of changing climatic conditions.

3.1. Accuracy

Accuracy refers to the percentage of correctly predicted instances (i.e., the number of correct crop predictions divided by the total number of instances). The model likely shows high accuracy because the hybrid nature of the model, using linear SVM, BiLSTM, and ConvLSTM, enables it to capture both spatial and temporal dependencies in the data, leading to better predictions.

i) Precision: precision measures the proportion of correct positive predictions.
ii) Recall measures the model's ability to correctly identify all the relevant instances. A high recall value indicates that the model is good at detecting all possible crop types, reducing the likelihood of missing an appropriate crop recommendation.
iii) F1-score: the F1-score is the harmonic mean of precision and recall, providing a balanced measure of the model's performance.

3.2. Hybrid model vs. traditional LSTM models

When comparing the hybrid model (linear SVM+BiLSTM+ConvLSTM) with traditional LSTM models, the hybrid approach is expected to outperform in several ways, such as higher generalization. The inclusion of linear SVM alongside DL models ensures better generalization over a variety of crop types and environmental conditions. Traditional LSTM models may struggle to capture complex spatial relationships, whereas ConvLSTM excels in this domain. Table 1 demonstrates that the hybrid ConvLSTM-BiLSTM-SVM model outperforms its individual models, achieving superior accuracy, precision, recall, and F1-score via the successful integration of spatial, temporal, and classification functionalities. Table 2 hybrid design demonstrates high performance across all parameters, significantly improving individual classifiers, regression models, and rule-based systems by adeptly capturing spatial, temporal, and categorisation patterns.

Table 1. Performance comparison of hybrid and individual models

Model component	Accuracy	Precision	Recall	F1-score
Linear SVM	0.9421	0.9440	0.9421	0.9420
BiLSTM	0.9615	0.9623	0.9615	0.9612
ConvLSTM	0.9678	0.9682	0.9678	0.9677
Hybrid (SVM+BiLSTM+ConvLSTM)	0.9758	0.9764	0.9758	0.9757

Table 2. Comparative analysis of crop classification techniques

Model type	Accuracy	Precision	Recall	F1-score
Rule-based system	0.7820	0.7900	0.7800	0.7850
Statistical regression	0.8700	0.8750	0.8700	0.8720
Linear SVM	0.9421	0.9440	0.9421	0.9420
BiLSTM	0.9615	0.9623	0.9615	0.9612
ConvLSTM	0.9678	0.9682	0.9678	0.9677
Hybrid (SVM+BiLSTM+ConvLSTM)	0.9758	0.9764	0.9758	0.9757

This study's findings indicate that a hybrid ML architecture integrating linear SVM, BiLSTM, and ConvLSTM significantly enhances crop prediction accuracy (0.9758) and F1-score (0.9757) relative to standalone models. This demonstrates that the integration of temporal (BiLSTM) and spatial (ConvLSTM) dependencies with a robust classifier (SVM) more successfully captures complicated environmental-soil interactions than traditional methods. The superior prediction performance proposes that local environmental and soil data may be dependably used to guide agricultural planning, enhance crop selection, and minimize resource waste. The proposed hybrid ML architecture combines ConvLSTM, BiLSTM, and linear SVM. ConvLSTM identifies spatial dependencies in soil and environmental characteristics, BiLSTM analysis temporal patterns, and SVM executes the final crop classification using the outputs from stacked LSTM, thus harnessing both spatio-temporal and classification advantages. The information, gathered from Mathuranthagam, Tamil Nadu, India has 2,200 samples spanning 22 crops, including attributes such as temperature, humidity, rainfall, soil pH, and NPK levels. A 70:30 ratio was used for training (1,540 samples) and testing (660 samples). The hybrid model had an accuracy of 0.9758, outperforming the performance of separate models. Hyperparameters for LSTM layers and SVM were calibrated by cross-validation and parameter change studies to enhance performance.

A hybrid model that combines ConvLSTM, BiLSTM, and linear SVM within a complementary framework. ConvLSTM captures spatial relationships, BiLSTM models temporal trends, and linear SVM performs classification based on features retrieved by LSTM. This layered methodology allows an integrated predictive system that improves the accuracy and reliability of crop predictions, outperforming individual models and demonstrating the efficacy of integrating spatio-temporal learning with traditional categorization techniques. The proposed hybrid SVM-BiLSTM-ConvLSTM model with Jaya feature selection incorporates Mathuranthagam's climatic and soil data. This ensures that farming decisions in Tamil Nadu are localized, data-driven, and sustainable, given the state's specific environmental conditions. Connecting the model to agricultural extension networks, GIS platforms, and mobile applications allows for data-driven decision assistance in agriculture, as well as real-time crop alerts, geographic visualization, and farmer outreach. Despite its robust performance, the proposed hybrid ConvLSTM-BiLSTM-SVM architecture has several limitations. Initially, although the dataset is balanced for experimental purposes, actual agricultural data often exhibit imbalances, potentially impacting performance for under-cultivated crops.

The model is trained on soil and climatic data particular to Mathuranthagam, perhaps restricting its applicability to other places in Tamil Nadu without localized calibration. While forecasts exhibit high accuracy, practical implementation in real-time encounters obstacles, including the need for constant data streams from meteorological stations, IoT sensors, and dependable internet access in remote regions. Rectifying these deficiencies is crucial for the scalable, agriculturist-friendly implementation. Improving the hybrid system's scalability, accuracy, and reliability for more widespread agricultural uses may involve future extensions such as integrating dynamic weather for real-time adaptability, fusing satellite imagery to capture large-scale spatial variability, and federated learning frameworks to guarantee privacy-preserving crop prediction while enabling interaction across regions. In the future, these results may facilitate the creation of decision-support systems for farmers and policymakers, allow for real-time crop recommendations using IoT sensors, and inform adaptive tactics in response to changing climatic conditions.

4. CONCLUSIONS

This research effectively addressed the issue of predicting crop suitability in Mathuranthagam, Tamil Nadu, India by using a hybrid ML framework that combines linear SVM, BiLSTM, and ConvLSTM networks. The hybrid model enhanced prediction performance by effectively capturing both temporal and spatial dependencies, achieving an accuracy of 0.9758 and an F1-score of 0.9757, improving linear SVM (accuracy 0.9421), BiLSTM (0.9615), and ConvLSTM (0.9678). Min-max normalization and Jaya optimization efficiently used environmental and soil variables, such as humidity, rainfall, temperature, and pH, for model training. The quantitative findings validate that the integration of DL with traditional ML methodologies offers a formidable solution for crop forecasting and underscores the importance of localized environmental data in agricultural strategy and decision-making. Future works may concentrate on augmenting the dataset, integrating real-time sensor data, and introducing supplementary environmental characteristics to improve model generalization and flexibility, thereby facilitating precision agriculture across various geographies.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sundaraj Kannan Susee		✓	✓	✓		✓	✓	✓	✓					✓
Shenbagaramasubramanian Shenbaga Vadivu		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
Murugesan Senthil Kumar	✓		✓	✓	✓		✓	✓		✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author, [SKS], on request.

REFERENCES

[1] A. A. M. A. Riazulhameed, C. S. Ranganathan, P. Pandey, B. Sakthisaravanan, K. Sasikala, and S. Murugan, “Smart composting solutions for organic waste management and soil enrichment in agriculture with IoT and gradient boosting,” *4th International Conference on Sustainable Expert Systems, ICSES 2024-Proceedings*, pp. 339–345, 2024, doi: 10.1109/ICSES63445.2024.10763259.

[2] M. Dhanaraju, P. Chenniappan, K. Ramalingam, S. Pazhanivelan, and R. Kaliaperumal, “Smart farming: internet of things (IoT)-based sustainable agriculture,” *Agriculture*, vol. 12, no. 10, 2022, doi: 10.3390/agriculture12101745.

[3] M. M. Mamoudan, A. Jafari, Z. Mohammadnazari, M. M. Nasiri, and M. Yazdani, “Hybrid machine learning-metaheuristic model for sustainable agri-food production and supply chain planning under water scarcity,” *Resources, Environment and Sustainability*, vol. 14, Dec. 2023, doi: 10.1016/j.resenv.2023.100133.

[4] S. R. Mousavi, V. A. J. Mahjenabadi, B. Khoshru, and M. Rezaei, “Spatial prediction of winter wheat yield gap: agro-climatic model and machine learning approaches,” *Frontiers in Plant Science*, vol. 14, 2023, doi: 10.3389/fpls.2023.1309171.

[5] C. C. S. Basavaraddi, C. S. Ranganathan, K. Sindhuja, N. Mohankumar, M. Jagadeeswaran, and S. Murugan, “Real-time incident detection and response with isolation forest algorithm and cloud infrastructure,” *8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2024 - Proceedings*, pp. 1024–1029, 2024, doi: 10.1109/I-SMAC61858.2024.10714593.

[6] C. S. Ranganathan, A. A. M. A. Riazulhameed, G. Brindha, N. Mohankumar, S. Yuvaraj, and S. Murugan, “Cloud-integrated wireless sensor networks for dynamic household air pollution management,” *2024 Asia Pacific Conference on Innovation in Technology (APCIT)*, 2024, doi: 10.1109/APCIT62007.2024.10673715.

[7] B. R. Babu, M. A. Haile, D. T. Haile, and D. Zerihun, “Real-time sensor data analytics and visualization in cloud-based systems for forest environment monitoring,” *International Journal of Advances in Signal and Image Sciences*, vol. 9, no. 1, pp. 29–39, 2023, doi: 10.29284/ijasis.9.1.2023.29-39.

[8] S. Sharafat *et al.*, “An IoT-enabled AI system for real-time crop prediction using soil and weather data in precision agriculture,” *Smart Agricultural Technology*, vol. 12, 2025, doi: 10.1016/j.atech.2025.101263.

[9] T. Mahmud *et al.*, “An approach for crop prediction in agriculture: integrating genetic algorithms and machine learning,” *IEEE Access*, vol. 12, pp. 173583–173598, 2024, doi: 10.1109/ACCESS.2024.3478739.

[10] Z. Alsalami, G. Mohammed, and T. Srinivas, “Enhancing crop yield prediction using IoT-based soil moisture and nutrient sensors,” *SHS Web of Conferences*, vol. 216, 2025, doi: 10.1051/shsconf/202521601029.

[11] V. G. Sivakumar, V. V. Baskar, M. Vadivel, S. P. Vimal, and S. Murugan, “IoT and GIS integration for real-time monitoring of soil health and nutrient status,” *International Conference on Self Sustainable Artificial Intelligence Systems, ICSSAS 2023 - Proceedings*, pp. 1265–1270, 2023, doi: 10.1109/ICSSAS57918.2023.10331694.

[12] S. K. S. Durai and M. D. Shamili, “Smart farming using machine learning and deep learning techniques,” *Decision Analytics Journal*, vol. 3, Jun. 2022, doi: 10.1016/j.dajour.2022.100041.

[13] D. Tripathi, S. K. Biswas, and B. Purkayastha, “A comparative analysis of ensemble learning techniques for crop yield prediction: CYPELA,” in *2023 4th International Conference on Computing and Communication Systems (I3CS)*, Mar. 2023, pp. 1–6, doi: 10.1109/I3CS58314.2023.10127268.

[14] S. K. Modi, Chahat, A. L. Yadav, P. S. Rana, and Kalpana, “Innovative deep learning methods for soil classification and crop yield prediction,” *Proceedings - 2024 International Conference on Computational Intelligence and Computing Applications, ICCICA*, pp. 502–508, 2024, doi: 10.1109/ICCICA60014.2024.10585003.

[15] B. Shadaksharappa, “Efficient sequential pattern discovery with PrefixSpan for large-scale data analytics,” *International Journal of Modern Computation, Information and Communication Technology*, vol. 8, no. 1, pp. 33–40, 2025, doi: 10.65000/s2h9vg39.

[16] P. S. Ramapraba *et al.*, “Implementing cloud computing in drug discovery and telemedicine for quantitative structure-activity relationship analysis,” *International Journal of Electrical and Computer Engineering*, vol. 15, no. 1, pp. 1132–1141, 2025, doi: 10.11591/ijece.v15i1.pp1132-1141.

[17] G. S. Pravallika, L. Kundana, K. S. Thanvi, G. Sirisha, and C. Rupa, “Proficient smart soil based IoT system for crop prediction,” *Proceedings of the 2nd International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 752–757, 2020, doi: 10.1109/ICIRCA48905.2020.9183054.

[18] A. K. Nasreen, M. Shenbagapriya, S. K. Seeni, P. Veda, B. Meenakshi, and S. Murugan, "Robotic restroom hygiene solutions with IoT and recurrent neural networks for clean facilities," *7th International Conference on Inventive Computation Technologies (ICICT)*, pp. 1842–1847, 2024, doi: 10.1109/ICICT60155.2024.10544900.

[19] M. Srikanth, V. Sharma, S. Patil, and R. V. Pujar, "Natural disaster prediction with cloud integration for enhancing validation using gradient boosting algorithm," *International Journal of Industrial Engineering*, vol. 7, no. 2, pp. 57–64, 2023.

[20] M. Jaswal, A. Kumari, K. Sangha, and D. Arora, "Smart agriculture: combining crop recommendation, yield prediction, and environmental analyzers for optimized farming," *Proceedings of International Conference on Emerging Technologies and Innovation for Sustainability (EmergIN)*, pp. 245–251, 2024, doi: 10.1109/EmergIN63207.2024.10961623.

[21] H. Mavi, S. K. Upadhyay, N. Srivastava, R. Sharma, and R. Bhargava, "Crop recommendation system based on soil quality and environmental factors using machine learning," *Proceedings - 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP)*, pp. 507–512, 2024, doi: 10.1109/INNOCOMP63224.2024.00089.

[22] S. Bag, N. Padhy, and R. Panigrahi, "Enhancing crop recommendation systems using deep learning techniques on soil & environmental data," in *2025 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, Feb. 2025, pp. 601–606, doi: 10.1109/ESIC64052.2025.10962589.

[23] N. M. Basavaraju, U. B. Mahadevaswamy, and S. Mallikarjunaswamy, "Design and implementation of crop yield prediction and fertilizer utilization using IoT and machine learning in smart agriculture systems," *2nd IEEE International Conference on Networks, Multimedia and Information Technology (NMITCON)*, 2024, doi: 10.1109/NMITCON62075.2024.10699184.

[24] A. Badshah, B. Y. Alkazemi, F. Din, K. Z. Zamli, and M. Haris, "Crop classification and yield prediction using robust machine learning models for agricultural sustainability," *IEEE Access*, vol. 12, pp. 162799–162813, 2024, doi: 10.1109/ACCESS.2024.3486653.

[25] C. S. Unnathi and S. K. Devi, "Sustainable farming: crop prediction and detection of crop deficiency using various algorithms," *4th International Conference on Sustainable Expert Systems, ICSES 2024 - Proceedings*, pp. 1451–1457, 2024, doi: 10.1109/ICSES63445.2024.10763179.

[26] L. Sartore, C. Boryan, A. Rosales, and A. Nagle, "Crop prediction uncertainty maps," *International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4735–4739, 2024, doi: 10.1109/IGARSS53475.2024.10642026.

[27] A. Ingle, "Crop recommendation dataset," kaggle.com. [Online]. Available: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset/data

BIOGRAPHIES OF AUTHORS

Dr. Sundaraj Kannan Susee received her M.E degree from Anna University in 2007 and Ph.D. degree from Anna University, Chennai, India in 2019. She has more than 11 years of teaching experience. She is currently serving as an assistant professor in the Department of Electronics and Communication Engineering at Mohamed Sathak A J College of Engineering, Chennai, India. Her research interests are in medical image processing, computer networking, and wireless communications. She can be contacted at email: suseekannan07@gmail.com.

Ms. Shenbagaramasubramanian Shenbaga Vadivu is working as an assistant professor (Sel.G) in the Department of Information Technology, SRM Valliammai Engineering College, Tamil Nadu. She has 20 years of teaching experience, and her areas of interest include AI, machine learning, IoT, data science, and data analytics. The domain of her research work is smart irrigation, particularly selecting the crop suitable for the soil. Her research emphasizes the practical application of cutting-edge technologies and challenges in agriculture, particularly soil taxonomy and crop suitability for soil. Her thirst for knowledge makes her dynamic in the fields of IT and agriculture. She can be contacted at email: shenbagavadivus.it@srmvalliammai.ac.in.

Dr. Murugesan Senthil Kumar received the M.E degree from Anna University in 2006 and the Ph.D. degree from Anna University, Chennai, India, in 2015. He has 23 years of teaching experience. He is currently serving as a professor and associate dean in the Department of Computer Science and Engineering, Karpaga Vinayaga College of Engineering and Technology, Chengalpattu, Tamil Nadu, India. His research interests are in medical image processing, software engineering, cybersecurity, and cloud computing. He is a reviewer of the International Journal of Computer Science Engineering and Information Technology Research, and many other journals. He can be contacted at email: msen1982@gmail.com.