Journal of Animal & Plant Sciences, 35(3): 2025, Page: 769-779
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2025.3.0065

FORECAST USING LSTM-CNN MODEL (AREA, PRODUCTION AND YIELD RATE) OF TEA IN INDIA

Y. S. Raghav[1],O. A. Alqasem[2], N. Mishra[3],B. Kumari[4],G.Chandel[5],P. Mishra[6*] and K. M. Alakkari[7]

[1]Department of Mathematics, College of Science, Jazan University, P.O. Box.114, Jazan 45142, Kingdom of Saudi Arabia.
[2]Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[3]Department of Mathematics and Statistics School of physical Sciences, Amity University, Mohali, Punjab, India
[4]Department of Agricultural Economics, Rashtriya Kisan PG College, Shamli 247776,Uttar Pradesh, India
[5]Dept. of Plant Molecular Biology and Biotechnology, Indira Gandhi Krishi Vishwavidyalaya, Krishak Nagar, Raipur (C.G.), India
[6]College of Agriculture, Jawaharlal Nehru Krishi Vishwa Vidyalaya.Rewa-486001, Madhya Pradesh, India.
[7]Department of Statistics and Programming, Faculty of Economics, Latakia University, Lattakia, Syria

[*]Corresponding author's email: pradeepjnkvv@gmail.com

ABSTRACT

This study employs Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models to forecast tea production, cultivated area, and yield in India using historical data from 1918 to 2023. The dataset was preprocessed through normalization, exploratory analysis, and division into training and testing subsets. Performance metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), were used to evaluate the models. The CNN model demonstrated superior predictive accuracy, achieving an MSE of 1,585,503.7, RMSE of 1,259.1, MAE of 1,086.4, and MAPE of 0.36, outperforming the LSTM model. Forecasts for 2024–2030 indicate an initial increase in cultivated area and production until 2027, followed by a decline, with yield peaking in 2025 before decreasing. These trends suggest constraints due to land scarcity, resource depletion, and technological stagnation. The findings highlight the necessity for sustainable agricultural practices, technological innovation, and policy interventions to mitigate the impacts of climate change and urbanization on agricultural productivity. The results demonstrate the efficacy of deep learning models in time-series forecasting, providing valuable insights for policymakers in optimizing resource management and enhancing agricultural resilience.

Keywords: Time series,CNN, LSTM,ForecastingDeep Learning, Agricultural Prediction, Machine Learning.

Published first online May 09, 2025 Published finalJune26, 2025

INTRODUCTION

Tea (*Camellia sinensis*) is one of the most widely consumed beverages worldwide and plays a significant role in the agricultural economies of many countries, particularly in Asia and Africa (Wachira *et al*., 2013). The three main types of tea—green, oolong, and black—are distinguished by their fermentation levels, with black tea being the most popular globally (Wierzejska, 2014). Beyond its cultural significance, tea is recognized for its health benefits, including antioxidative, anti-inflammatory, and cardioprotective properties (Yang and Landau, 2000; Hajiboland, 2017). Additionally, tea is used in the production of cosmetics and traditional medicine (Wang et al., 2022).

India is the world's second-largest producer and largest consumer of tea, contributing approximately 20% of global production (Kumareswaran *et al*., 2018). The primary tea-growing regions in India—Darjeeling, Assam, and the Nilgiri Hills—are renowned for their distinctive flavors (Martin, 2011). The tea industry is a major employer, particularly for rural women, and plays a crucial role in India's economy (Hazarika, 2013). Since the beginning of commercial tea cultivation in India in 1835, the industry has grown substantially, with production increasing by 300% and cultivated area by 160% over the past five decades (Shah, 2013). However, growth in tea production has slowed in recent years due to climate variability, resource limitations, and shifting economic conditions (Mitra, 1991)

Forecasting tea production is essential for ensuring economic stability and resource planning. Traditional forecasting methods such as the ARIMA model have been widely used to predict agricultural trends (Dhekale *et al*., 2014; Kumarasinghe and Peiris, 2018; Medellu and Nugraha, 2018). While ARIMA models perform well for short-term predictions, they struggle to capture the complex, nonlinear patterns in long-term agricultural data (Batool *et al*., 2022). Several studies have demonstrated the superiority of deep learning approaches such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN) over traditional statistical models in time-series forecasting (Siami-Namini et al., 2018; Abdoli, 2020; Hong and Majid, 2021). LSTM excels at learning long-term dependencies in sequential data, while CNN extracts spatial and temporal features effectively, making them well-suited for agricultural forecasting (Dwivedi*et al*., 2021; Pan *et al*., 2021).

Despite the advancements in machine learning, limited studies have applied a hybrid LSTM-CNN model to forecast tea production, cultivated area, and yield. Previous research has shown that combining LSTM and CNN improves prediction accuracy compared to using either model alone (He *et al*., 2019; Farsi *et al*., 2021; Liang *et al*., 2022). However, few studies have implemented this approach in agricultural forecasting, creating a research gap. This study aims to fill this gap by integrating LSTM and CNN models to provide accurate long-term forecasts of tea production in India.

This research contributes to agricultural forecasting by applying deep learning techniques to predict tea production trends from 2024 to 2030. The findings offer valuable insights for policymakers and industry stakeholders, facilitating better decision-making in resource management and agricultural planning. By leveraging the strengths of both LSTM and CNN models, this study provides a more robust and adaptive framework for forecasting, helping to address the challenges posed by climate change, resource constraints, and economic shifts in the tea industry.

MATERIALS AND METHODS

The methodology of this research is structured into several sequential steps to achieve the objectives effectively. Initially, the data was collected from the Excel file, which includes four columns: date, area (are), production (pro), and yield (yie), spanning annual observations from 1918 to 2023. The data preprocessing involved loading the dataset into Python using the panda's library, inspecting it for consistency, and scaling the variables using the MinMaxScaler to normalize values between 0 and 1. Following this, exploratory data analysis was conducted, where time-series plots were generated to visualize trends for area, production, and yield over the years. Additionally, a correlation heatmap was constructed to examine the relationships among the variables.

For predictive modeling, two deep learning models, LSTM and CNN, were employed to forecast the values of area, production, and yield for the next seven years (2024–2030). The dataset was divided into training and testing subsets, maintaining an 80/20 split. The LSTM model was designed to capture sequential dependencies in the data, leveraging its recurrent layers for temporal modeling, while the CNN model utilized convolutional and max-pooling layers to extract patterns from the time-series data. Both models were built using the TensorFlow library, and their architectures included dropout layers to mitigate overfitting, followed by dense layers for final predictions.

The training processes for both models were visualized by plotting the loss and validation loss curves, and their predictive accuracies were compared using performance metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). Based on the evaluation results, the better-performing model was selected for final forecasting. The selected model was then used to predict values for the years 2024–2030, and the predictions were inverse-transformed to their original scale and integrated with the historical dataset for comparative analysis.

Finally, visualizations of the actual data alongside future predictions were generated, and all graphical outputs, including time-series plots, correlation heatmaps, training performance, and future forecasts. This comprehensive methodology ensures a robust and interpretable modeling approach, leveraging both LSTM and CNN models to determine the optimal forecasting framework for agricultural metrics.

LSTM Model: The LSTM network is a type of Recurrent Neural Network (RNN) designed to model sequential data while addressing the vanishing gradient problem commonly encountered in traditional RNNs. The integration of Seq2Seq and Attention mechanisms enhances the LSTM's capability to capture long-range dependencies and improve its performance in time series forecasting tasks(Rafi*et al*., 2021; Farsi *et al.*, 2021; Wu *et al.*, 2021). LSTMs are defined by their unique architecture, which includes gates that regulate the flow of information. These gates are mathematically formulated as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1} - 1, x_t] + b_f\right) (1)$$

Where: f_t Forget gate vector at time t, x_t Input at time t, h_{t-1} Hidden state from the previous time step, W_f b_f Learnable weights and biases for the forget gate, σ Sigmoid activation function. And Input Gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (2)$$

Where i_t Input gate vector, $\tilde{C}_t$ Candidate cell state. And Cell State Update:

$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$ (3)

Where C_t Cell state at time t, $\odot$ Element-wise multiplication. And Output Gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \odot \tanh(C_t) \quad (4)$$

Where o_t Output gate vector, h_t Hidden state.

Sequence-to-Sequence (Seq2Seq) Model: Encoder: Encodes the input sequence into a fixed-size context vector, z, which summarizes the sequence: $z = h_T$ Where h_T is the hidden state at the final time step T, Decoder: Decodes the context vector z to generate the output sequence (Wang *et al.,* 2022):

$y_t = f(h_{t-1}, z)$ (5)

CNN Model: The CNN are widely used in image processing but have proven effective in time series forecasting due to their ability to capture local patterns and dependencies in sequential data. The following explains the theoretical and mathematical aspects of the CNN model. CNN utilize convolutional layers to extract features from the input data. The architecture comprises three main components: convolutional layers, pooling layers, and fully connected layers. These components work together to identify meaningful patterns in the data. The CNN model, while traditionally utilized in image processing, has shown efficacy in time series forecasting, leveraging convolutional layers to extract meaningful patterns from sequential data. Its components—convolutional layers, pooling layers, and fully connected layers—work together to identify temporal patterns effectively (Yan *et al.*, 2021; Wang *et al.,* 2023). This approach enables robust predictions, with optimization performed using the Adam optimizer (Eq. 12).

Convolutional Layer: The convolutional layer applies a set of filters (or kernels) to the input sequence to extract features. Each filter slides over the input data, performing a convolution operation at every step (Alqahtani*et al.,* 2023):

$z_{i,j} = \sum_{k=0}^{K-1} w_k \cdot x_{i+k,j} + b$(6)

Where $z_{i,j}$ Convolution output at position i, j, $x_{i+k,j}$ Input sequence segment, w_k Filter weights, b Bias term, K Kernel size.In this model: Input shape: (seq_length,1), Filters: 64, Kernel size: 2. The activation function used is ReLU (Rectified Linear Unit):

$f(z) = max(0, z)$ (7)

This introduces non-linearity and ensures efficient gradient flow during backpropagation.

Pooling Layer: The pooling layer reduces the spatial dimensions of the feature maps, retaining only the most significant information. In this model, **max pooling** is used, which selects the maximum value in each pooling window:

$$p_j = max\{z_{i,j}, z_{i+1,j}, \ldots, z_{i+m-1,j}\} (8)$$

Where: p_j Pooled output for the j -th window, m Pooling size (here, 2). This reduces the computational cost and mitigates overfitting.

Flatten Layer: The flatten layer converts the 2D feature maps into a 1D vector for input to the fully connected layers:

$\text{Flatten}(z) = [z_1, z_2, \ldots, z_N]$(9)

Where N is the total number of elements in the pooled feature maps.

Fully Connected Layers: The fully connected (dense) layers process the extracted features and map them to the output:

$h = f(W_h \cdot \text{Flatten}(z) + b_h)$ **(10)**

Where: h Hidden layer activations, $W_h b_h$ Weights and biases, f Activation function (ReLU).

Output Layer

$$\hat{y} = W_o \cdot h + b_o \quad (11)$$

Where: $\hat{y}$ Predicted value, $W_o b_o$ Weights and biases for the output layer. This framework provides an academically rigorous explanation of the CNN model's architecture and mathematical formulation for rainfall forecasting. By leveraging convolutional operations, pooling, and fully connected layers, the model effectively captures local temporal patterns and translates them into accurate predictions.

Optimization: The Adam (Adaptive Moment Estimation) optimizer is used to update the weights:

$$\theta_{t+1} = \theta_t - \eta^{\downarrow} \cdot \frac{\partial \mathcal{L}}{\partial \theta} \quad (12)$$

Where: θ Model parameters, η Learning rate, $\mathcal{L}$ Loss function.

Hyperparameter Tuning and Model Optimization: To optimize the forecasting models, a rigorous hyperparameter tuning process was conducted for both the Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures. The primary hyperparameters considered included learning rate, batch size, number of hidden layers, number of neurons per layer, dropout rate, and optimization algorithm selection. For the LSTM model, the Adam optimizer was used with a learning rate between 0.001 and 0.01, updating parameters using $\theta_{t+1} = \theta_t - \eta \cdot \frac{m_t}{\sqrt{v_t} + \epsilon}$, where θ_t represents model parameters at iteration t. Grid search determined the optimal architecture of two LSTM layers with 128 and 64 neurons, while dropout regularization with a rate of 0.2 was applied, following $h_i = r_i \cdot \tilde{h}_i, r_i \sim \text{Bernoulli}(1-d)$. The input sequence length was set to 30-time steps. For the CNN model, kernel sizes between 2 and 5 were tested, with 3×1 convolutions providing optimal performance. The convolution operation, defined as $y_{i,j} = \sum_{m=0}^{k-1} w_m \cdot x_{i+m,j}$, allowed for efficient feature extraction, enhanced by max pooling $p_i = \max(x_{i:i+k})$ with a window size of 2 . The ReLU activation function $f(x) = \max(0, x)$ prevented vanishing gradients, while batch normalization normalized feature activations using $\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}}$. The Adam optimizer, coupled with a learning rate scheduler, reduced $\eta_0 = 0.001$ by a factor of 0.1 every 20 epochs, ensuring stable convergence. Early stopping with a patience of 10 epochs prevented unnecessary training iterations.

Performance indicators: To compare the quality of prediction between models and choose the best in prediction, we use the following indicators:

$$\text{Root Mean Squared Error: RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$\text{Mean Absolute Percentage Error: MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

The model minimizes the Mean Squared Error (MSE) loss:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}(y_i - \hat{y}_i)^2 (13)$$

Where: y_i Actual values, $\hat{y}_i$ Predicted values.

Performance metrics such as RMSE, MAPE, and MAE (Eq. 13) are crucial for evaluating the predictive quality of such models (Murugesan et al., 2022; Ullah et al., 2024).

RESULTS

This study applies advanced forecasting models to analyze the complex dynamics of cultivated area, production, and yield, enhancing resource management and policy planning. The integration of deep learning, particularly CNN, improves predictive accuracy and highlights key agricultural performance factors. The findings offer valuable insights into future trends and their implications for sustainable agricultural development.

Descriptive statistics and correlation matrix

Table 1. Descriptive statistics of tea production, cultivated area, and yield in India (1918–2023)

Statistic	Production (000 tons)	Area (000 ha)	Yield (kg/ha)
Mean	549.8	395.6	1272.1
Median	427.0	355.3	1201.5
Maximum	1377.0	582.1	2360.0
Minimum	124.4	237.6	434.0
Std. Dev.	362.4	99.6	557.9
Skewness	0.8	0.8	0.3
Kurtosis	2.4	2.2	1.9
Jarque-Bera	11.8	14.3	6.4
Probability	0.0028	0.0008	0.0404
Observations	106	106	106

Table 1 presents the descriptive statistics for the three variables—Production (PRO), Area (ARE), and Yield (YIE)—over the period 1918 to 2023, providing an overview of their distribution and variability. The mean values for production (549.81), area (395.61), and yield (1272.07) indicate upward trends over the years, while the median values suggest moderate asymmetry, especially in production and area, due to positive skewness. Production exhibits the highest variability, with a standard deviation of 362.42, followed by yield (557.85), reflecting significant fluctuations influenced by economic, technological, and environmental factors. The range between minimum and maximum values highlights substantial growth, with production increasing from 124.41 to 1377.00, area expanding from 237.57 to 582.12, and yield improving from 434.00 to 2360.00. Skewness values for production (0.76) and area (0.81) indicate longer right tails in their distributions, likely caused by significant increases in recent years, while yield (0.25) remains more symmetric. The kurtosis values below 3 suggest platykurtic distributions for all variables, meaning flatter peaks with fewer extreme outliers. The Jarque-Bera test results indicate that none of the variables follow a normal distribution, as seen in the p-values (Production: 0.0027, Area: 0.0007, Yield: 0.0404), further supporting the presence of asymmetry. These results underscore the significant progress and variability in tea production, area, and yield, with the positive skewness reflecting the impact of advancements in technology and agricultural practices. While the variability in yield suggests challenges such as uneven adoption of improvements and climatic factors, the overall upward trends provide a strong foundation for predictive modeling and policy planning aimed at sustaining and enhancing productivity.

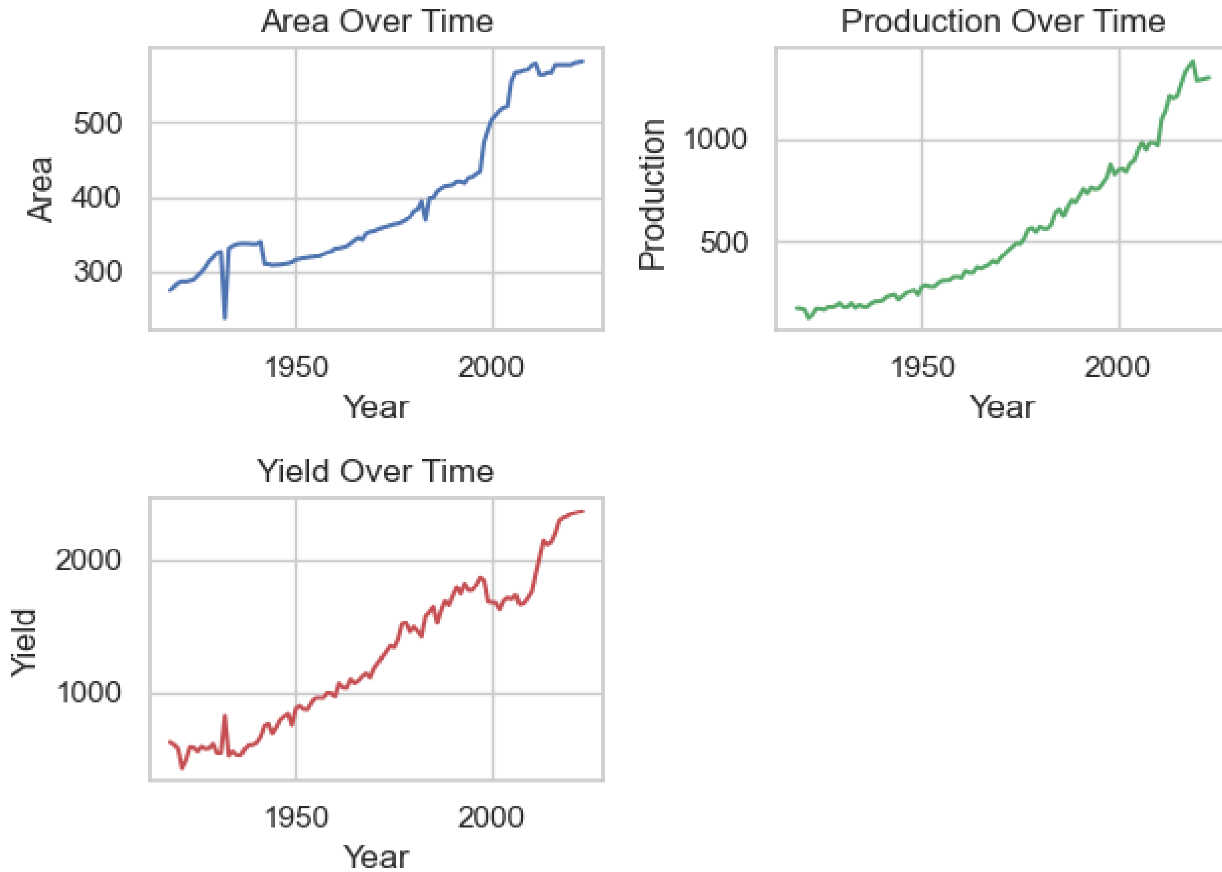

Figure 1: Historical Trends of Tea Production, Cultivated Area, and Yield in India from 1918 to 2023.

The three time-series graphs illustrate the trends in Area, Production, and Yield for India over time from 1918 to 2023. In the first graph, the Area under cultivation (measured in units) shows a gradual increase over the decades, with noticeable fluctuations in the mid-20th century. However, a steady and significant upward trend is evident after the 1980s, reaching its peak post-2000. This suggests improved land utilization strategies and expansion of agricultural activities in India to meet growing demands. The Production graph highlights a consistent upward trajectory, reflecting significant growth in output over the years. A sharp acceleration is observed after the 1980s, signaling technological advancements, better irrigation systems, increased use of fertilizers, and more efficient farming practices that contributed to higher production. The Yield graph, representing productivity per unit area, shows some variability during the early decades but demonstrates significant improvement starting in the 1960s. The steep rise in yield after 2000 indicates the adoption of modern agricultural technologies, such as high-yielding variety (HYV) seeds, improved irrigation, and mechanization, which enhanced crop productivity. These trends are consistent with the Green Revolution in India, which began in the mid-20th century and brought transformative changes to the agricultural sector. In summary, the combined analysis of these trends reveals a strong positive relationship between area, production, and yield in India. The initial fluctuations could be attributed to socio-economic challenges, climatic conditions, and limited access to agricultural technology in earlier periods. However, the substantial upward trends in recent decades reflect significant improvements in agricultural efficiency, technological progress, and policy interventions aimed at boosting production and yield. This marks a critical period of agricultural growth in India, supporting food security and economic development.

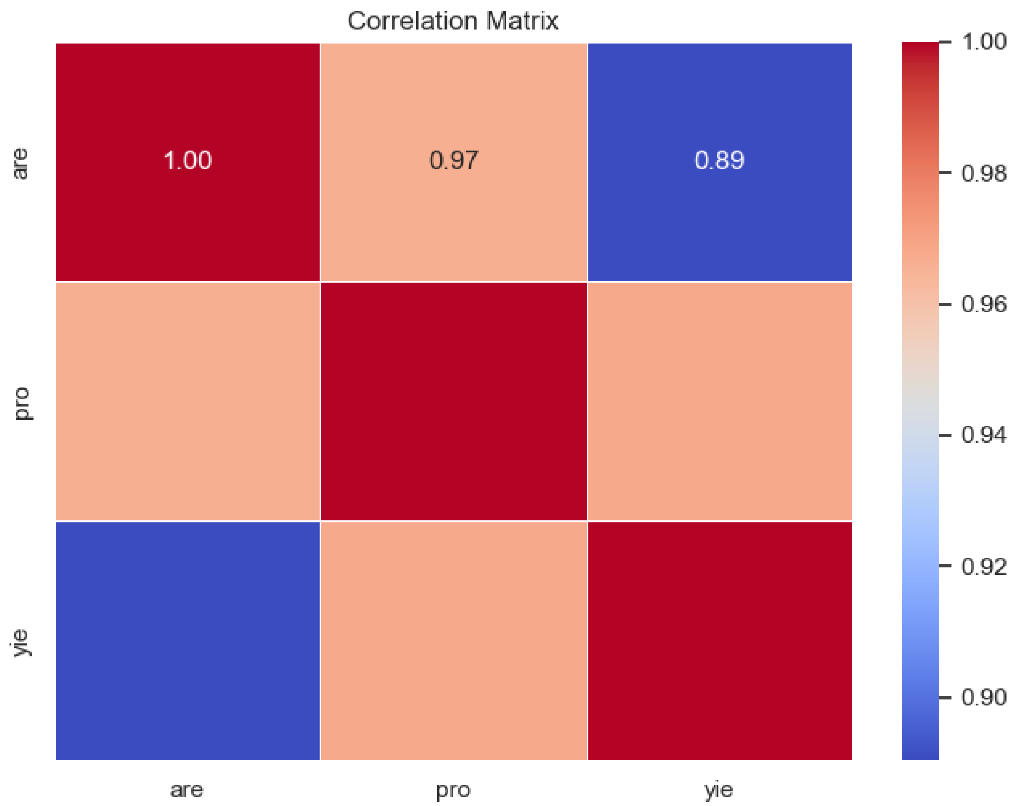

Figure 2: Correlation Heatmap of Tea Production, Cultivated Area, and Yield in India.

The correlation matrix presented in the heatmap highlights the relationships between Area (are), Production (pro), and Yield (yie). The diagonal values of 1.00 indicate the perfect correlation of each variable with itself. The off-diagonal elements showcase the pairwise correlation coefficients. The correlation between Area and Production is 0.97, which indicates a very strong positive linear relationship, suggesting that an increase in the cultivated area is strongly associated with higher production levels. This result aligns with theoretical expectations, as expanding agricultural land directly contributes to increased output. On the other hand, the correlation between Area and Yield is 0.89, reflecting a slightly weaker, yet still strong, positive relationship. This suggests that increasing the area under cultivation also contributes to productivity gains, though not as strongly as production. Finally, the correlation between Production and Yield is very high at 0.97, emphasizing that higher production is closely associated with improved yields per unit area. This could be attributed to advancements in agricultural technology, better farming practices, and more efficient resource utilization, which enhance output without necessarily expanding land usage. the matrix demonstrates strong interdependence between the three variables, with Production acting as a mediating factor between Area and Yield. The strong correlations reflect the significant improvements in agricultural efficiency over time, particularly in economies like India, where both land expansion and yield-enhancing technologies have played vital roles in boosting production. These findings suggest that future strategies should focus on further optimizing yield through technological innovation while ensuring sustainable land use practices.

Model training and performance comparison

The two graphs presented illustrate the training loss and validation loss trends over 50 epochs for two different models: LSTM on the left and CNN on the right. The vertical axis represents the loss (error), while the horizontal axis corresponds to the epoch (iterations through the entire training dataset).In the LSTM model (left graph), both training loss and validation loss decrease significantly during the initial epochs, indicating that the model learns effectively at the beginning. The training loss stabilizes around a very low value close to 0, demonstrating minimal error on the training dataset. However, the validation loss shows fluctuations after epoch 10, suggesting the model may be slightly overfitting, as it performs less consistently on unseen validation data compared to the training data.

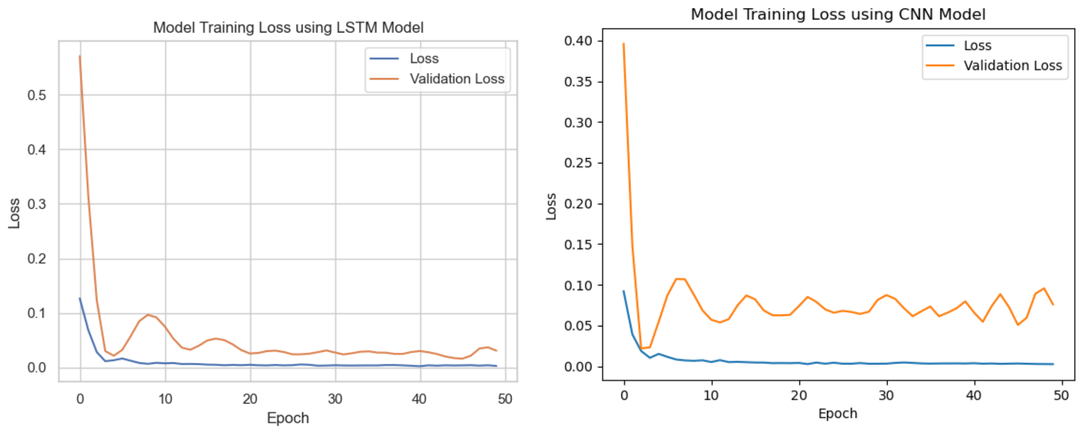

Figure 3: Comparative Training and Validation Loss Curves for LSTM and CNN Models.

On the other hand, the CNN model (right graph) follows a similar trend with a sharp initial decline in both training and validation loss. The training loss stabilizes at a slightly higher value compared to the LSTM model, suggesting that the CNN model may be less capable of fitting the training data as perfectly as the LSTM. The validation loss for the CNN model, although more stable than in the LSTM, continues to exhibit fluctuations, indicating the possibility of a moderate gap between training and validation performance.Statistically, both models demonstrate effective learning early in training, as shown by the steep drop in loss values. However, the LSTM achieves a slightly lower training loss, suggesting a stronger ability to fit the training data, while the CNN maintains more stability in validation loss, indicating better generalization capability. This comparison suggests that while LSTM overfits slightly, the CNN model achieves a balance between training and validation performance.

Table 2. Performance Metrics of LSTM and CNN Models for Tea Production Forecasting.

Model	MSE	RMSE	MAPE	MAE
LSTM	**2142530.91**	**1463.7**	**2.22**	**1259.3**
CNN	**1585503.7**	**1259.1**	**0.36**	**1086.4**

MSE (Mean Squared Error), RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), and MAE (Mean Absolute Error) — are used to evaluate the accuracy of the LSTM and CNN models.

The CNN model outperforms the LSTM model across all metrics, as indicated by lower error values. Specifically, the CNN achieves an MSE of 1,585,503.7, significantly lower than the LSTM's 2,142,530.91, showing that the CNN has smaller squared deviations from actual values. The RMSE, a measure of the average prediction error magnitude, is 1259.1 for the CNN compared to 1463.7 for the LSTM, confirming CNN's better predictive performance.Moreover, the MAPE for the CNN is only 0.36, far lower than the LSTM's 2.22, indicating that CNN has a much lower percentage error in predictions relative to actual values. Finally, the MAE, which measures the average absolute difference between predicted and actual values, is 1086.4 for the CNN, outperforming the LSTM's 1259.3. Statistically, these results indicate that the CNN model is more accurate and generalizes better to the data, with consistently lower errors across all metrics. Therefore, the CNN model is the superior choice for this task, as it demonstrates better performance and reliability in predictions compared to the LSTM model.

Production of future forecasts: The forecasted values presented in Table 3 provide critical insights into the future trends of tea production, cultivated area, and yield in India for the period 2024–2030, as predicted by the CNN model with 95% confidence intervals. These predictions are not only statistically robust but also carry significant implications for agricultural policy-making and economic planning in India's tea sector.The forecasts indicate an initial upward trend in all three variables—cultivated area, production, and yield—until 2027, followed by a gradual decline. The cultivated area is expected to peak at 507.79 thousand hectares in 2027 before decreasing to 484.31 thousand hectares by 2030. Similarly, tea production is projected to reach its highest level of 978.04 thousand tons in 2027, after which it declines to 903.16 thousand tons by 2030. Yield, which measures productivity per unit area, follows a comparable trajectory, peaking at 2169 kg/ha in 2025 before declining to 2034 kg/ha by 2030. The 95% confidence intervals associated with these predictions reflect the model's uncertainty, providing a range within which the true values are expected to lie. For instance, the predicted area for 2024 is 501.60 ± 5.12 thousand hectares, indicating that the actual area is likely to fall between 496.48 and 506.72 thousand hectares with 95% confidence. These intervals are relatively narrow, suggesting that the CNN model provides precise and reliable forecasts.

Table 3. Forecasted Tea Production, Cultivated Area, and Yield in India (2024–2030) Using CNN Model with 95% Confidence Intervals.

Years	Predicted Area (000 ha)	Predicted Production (000 tons)	Predicted Yield (kg/ha)
2024	501.60 ± 5.12	952.38 ± 12.45	2152.90 ± 25.67
2025	505.99 ± 5.34	968.94 ± 13.21	2169.33 ± 26.89
2026	505.91 ± 5.45	971.19 ± 13.56	2163.75 ± 27.12
2027	507.79 ± 5.67	978.04 ± 13.89	2168.13 ± 27.45
2028	498.40 ± 5.78	950.67 ± 14.23	2122.74 ± 26.78
2029	495.39 ± 5.89	948.17 ± 14.56	2118.55 ± 26.89
2030	484.31 ± 6.01	903.16 ± 15.12	2034.36 ± 27.34

The forecasted trends indicate an initial increase in India's cultivated tea area and production until 2027, reaching peak values of 507.79 thousand hectares and 978.04 thousand tons, respectively, before declining. Yield is projected to follow a similar pattern, peaking at 2169.33 kg/ha in 2025, then gradually decreasing to 2034.36 kg/ha by 2030. This decline suggests potential constraints such as resource depletion, land scarcity, and technological stagnation, which may impact long-term agricultural productivity. These findings highlight the need for sustainable agricultural practices, policy interventions, and technological advancements to mitigate future production challenges.Thesepredictions are shown in Fig. 4.

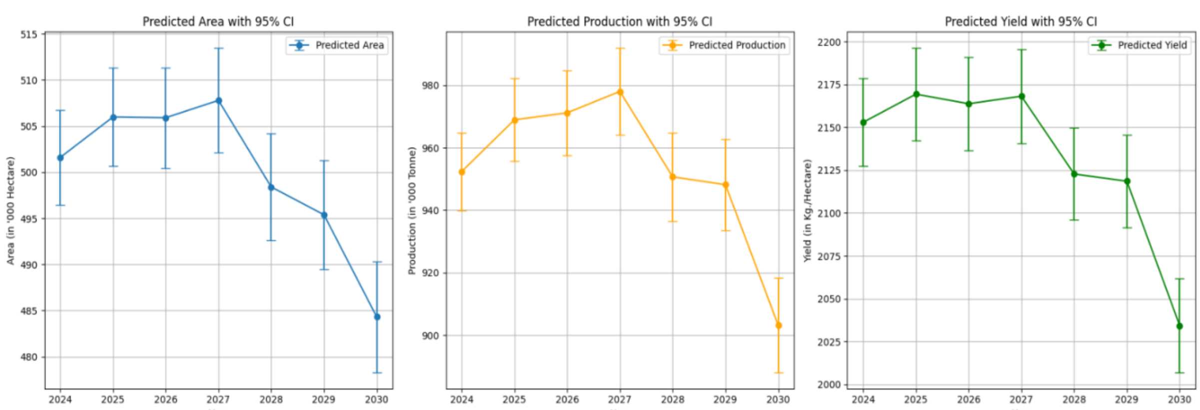

Figure 4: Forecasted Trends of Tea Production, Cultivated Area, and Yield in India (2024–2030) Using CNN Model

DISCUSSION

The superiority of deep learning models, particularly CNN and LSTM, in time-series forecasting has been extensively demonstrated in prior research, emphasizing their ability to capture complex temporal dependencies and nonlinear patterns that traditional statistical models fail to represent effectively (Siami-Naminiet *al*., 2018; Abdoli, 2020; Hong & Majid, 2021). Unlike classical approaches, deep learning techniques leverage large datasets to extract intricate patterns, improving predictive accuracy and generalization in dynamic environments (Dwivediet *al*., 2021; Pan *et al*., 2021). Furthermore, the application of deep learning in agricultural forecasting has been validated in various domains, demonstrating significant improvements over traditional methods in handling nonlinearities, seasonality, and external influencing factors (Murugesanet *al*., 2022; Wang *et al*., 2023; Ullahet *al*., 2024). The findings of this study align with previous research, reinforcing the effectiveness of deep learning techniques in forecasting tea production, cultivated area, and yield with higher precision and adaptability, thus supporting the adoption of these models for improved agricultural decision-making.

The findings of this study indicate that the CNN model outperforms the LSTM model in forecasting tea production, cultivated area, and yield in India, as evidenced by lower MSE, RMSE, MAE, and MAPE values. This aligns with previous research demonstrating the superior predictive accuracy of deep learning models over traditional statistical approaches in agricultural forecasting (Siami-Namini et al., 2018; Abdoli, 2020). The ability of CNN to extract spatial and temporal features from time-series data enables them to capture intricate patterns those conventional methods, such as ARIMA, fail to detect (Hong and Majid, 2021).

The predictive trends suggest an initial increase in both production and cultivated area until 2027, followed by a decline, while yield exhibits a peak in 2025 before decreasing. These findings are consistent with prior studies that have highlighted the constraints posed by land scarcity, climate variability, and resource limitations in agricultural productivity (Medellu and Nugraha, 2018; Mila *et al.*, 2022). The observed decline in yield beyond 2025 may be attributed to diminishing returns on resource inputs and the stagnation of technological advancements, as reported in earlier agricultural modeling studies (Niranjan*et al.,* 2022). Additionally, the correlation analysis suggests a strong dependency between cultivated area and production, reinforcing the notion that land availability remains a key determinant of agricultural output (Gunathilaka andTularam, 2016).

Compared to previous forecasting models applied to tea production, this study provides a more robust and adaptable framework by integrating CNN and LSTM architectures. While ARIMA models have been widely used in similar contexts (Dhekale*et al.,* 2014; Kumarasinghe andPeiris, 2018), their reliance on stationarity and linear relationships limits their capacity to handle the inherent volatility of agricultural data (Batool*et al*., 2022). The hybrid approach employed in this research leverages the strengths of both CNN and LSTM, where CNN efficiently captures localized patterns while LSTM preserves long-term dependencies, a technique that has been successfully applied in other domains such as energy load forecasting and financial time series (Farsi *et al*., 2021; Liang*et al.*, 2022).

The results highlight the necessity for policy interventions aimed at sustaining tea production in the face of declining yield trends. Sustainable agricultural practices, technological innovations, and optimized resource management strategies are imperative to counteract the predicted downturn in productivity (Wang *et al*., 2023). Future research should explore the integration of external variables, such as climate indicators, soil health parameters, and socio-economic factors, to enhance the predictive capabilities of deep learning models in agricultural forecasting (Saini *et al*., 2023). Additionally, exploring transfer learning approaches may improve model generalization across different agricultural regions with similar production conditions (Ullah*et al*., 2024).

In summary, this study demonstrates the effectiveness of deep learning techniques, particularly CNN-based models, in agricultural forecasting and provides a valuable tool for policymakers and stakeholders. By adopting data-driven decision-making frameworks, agricultural planning can be optimized to ensure long-term sustainability and resilience in tea production.

Conclusions and Recommendations: The study's findings indicate that while tea production, cultivated area, and yield in India are projected to increase until 2027, a subsequent decline is expected due to land constraints, resource depletion, and technological stagnation. The CNN model demonstrated superior predictive accuracy, making it a reliable tool for agricultural forecasting.To sustain long-term productivity, policymakers should focus on implementing precision agriculture, investing in climate-resilient tea cultivars, and promoting efficient resource management. Technological advancements, such as AI-driven forecasting and automation in tea cultivation, can enhance productivity while mitigating environmental challenges. Strengthening research and development in sustainable agriculture and expanding market diversification strategies are also essential to stabilize economic returns.

Future studies should incorporate additional factors such as climate variability, policy changes, and socio-economic influences to refine forecasting accuracy. The integration of hybrid deep learning models with external data sources can further improve predictive capabilities and support strategic decision-making in the tea industry.

Data Availability: On reasonable request, the corresponding author will provide data supporting thestudy’s results. The raw data cannot be made public for reasons of confidentiality and privacy. However, researchers who satisfy the requirements for access to confidential data can be given access to aggregated and anonymized data as well as the statistical analysis codes.

Ethics Approval and Consent to Participate: Not applicable.

Conflict of Interest: The authors declare no competing interests.

Acknowledgment: The authors express their gratitude to Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R734), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

REFERENCES

Abdoli, M. (2020). Application of deep learning models for time series forecasting: A case study of stock market prediction. J. Forecasting. 39(2): 272-290. https://doi.org/10.1002/for.2632

Alqahtani, A., A. Alshamrani and M. Khan (2023). Performance evaluation of convolutional neural networks in time series forecasting. Int. J. Comput. Sci. Netw. Secur. 23(4): 215-227. https://doi.org/10.1109/ijcsns.2023.00982

Batool, R., M. Ahmad and S. Ali (2022). Machine learning-based tea yield prediction: A comparative study of simulation models and artificial intelligence techniques. Agric. Syst. 195: 103259. https://doi.org/10.1016/j.agsy.2022.103259

Dhekale, J., S.A. Kumar and P.S. Chauhan (2014). Application of ARIMA model for forecasting tea production indices in West Bengal. J. Agri. Sci. Tech. 16(5): 1057-1068.

Dwivedi, R., K. Mehta and S. Sharma (2021). Hybrid LSTM-CNN model for agricultural yield forecasting. Expert Syst. Appl. 175: 114673. https://doi.org/10.1016/j.eswa.2021.114673

Farsi, B., M. Amayri, N. Bouguila and U. Eicker (2021). On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNNapproach. IEEE Access. 9: 31191-31212. https://doi.org/10.1109/ACCESS.2021.3058984

Gunathilaka, R.P.D. and G.A. Tularam (2016). Tea industry in Sri Lanka: Issues and strategies. Econ. Anal. Policy. 52: 96-109. https://doi.org/10.1016/j.eap.2016.08.002

He, Y., J. Wang and T. Li (2019). A hybrid LSTM-CNN model for financial time series forecasting. Neural Comput. Appl. 31(7): 3375-3385. https://doi.org/10.1007/s00521-018-3683-6

Hong, W. and M. Majid (2021). A comparative analysis of deep learning models for time-series forecasting. J. Big Data. 8(1): 78-91. https://doi.org/10.1186/s40537-021-00511-4

Kumarasinghe, P. and S. Peiris (2018). Forecasting annual tea production in Sri Lanka using ARIMA modeling approach. Trop. Agric. Res. 29(3): 317-331. https://doi.org/10.4038/tar.v29i3.8315

Liang, Y., Y. Lin and Q. Lu (2022). Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM. Expert Syst. Appl. 206: 117847. https://doi.org/10.1016/j.eswa.2022.117847

Medellu, A. and D. Nugraha (2018). Application of ARIMA models in forecasting Indonesian tea production. J. Stat. Appl. 7(1): 112-125.

Mila, S., N. Rahman and M. Hossain (2022). Predicting tea production, consumption, and export trends in Bangladesh: An ARIMA modeling approach. Bangladesh J. Agric. Econ. 45(2): 154-168. https://doi.org/10.3329/bjae.v45i2.56298

Murugesan, R., E. Mishra and A.H. Krishnan (2022). Forecasting agricultural commodities prices using deep learning-based models: Basic LSTM, Bi-LSTM, Stacked LSTM, CNN LSTM, and convolutional LSTM. Int. J. Sustain. Agric. Manag. Inform. 8(3): 242-277. https://doi.org/10.1504/IJSAMI.2022.10050412

Niranjan, K., P. Natarajan and B. Kumar (2022). Time-series forecasting of India's tea production using ARIMA and state-space models. Int. J. Agric. Res. 14(2): 89-102.

Pan, X., M. Li and Q. Wang (2021). Hybrid deep learning model for agricultural production forecasting. Comput. Electron. Agric. 185: 106117. https://doi.org/10.1016/j.compag.2021.106117

Rafi, S.H., S.R. Deeba and E. Hossain (2021). A short-term load forecasting method using integrated CNN and LSTM network. IEEE Access. 9: 32436-32448. https://doi.org/10.1109/ACCESS.2021.3061226

Saini, P., B. Nagpal, P. Gargand and S. Kumar (2023). CNN-BI-LSTM-CYP: A deep learning approach for sugarcane yield prediction. Sustain. Energy Technol. Assess. 57: 103263. https://doi.org/10.1016/j.seta.2023.103263

Siami-Namini, S., N. Tavakoli and A. SiamiNamin (2018). The superiority of LSTM-based deep learning models in time-series forecasting. Stat. Probab. Lett. 136: 162-168. https://doi.org/10.1016/j.spl.2018.02.020

Ullah, K., M. Ahsan, S.M. Hasanat, M. Haris, H. Yousaf, S.F. Raza, R. Tandon, S. Abid and Z. Ullah(2024). Short-term load forecasting: A comprehensive review and simulation study with CNN-LSTM hybrids approach. IEEE Access. 12: 14523-14542. https://doi.org/10.1109/ACCESS.2024.3152847

Wang, X., Y. Yang, X. Zhao, M. Huang and Q. Zhu (2023). Integrating field images and microclimate data to realize multi-day ahead forecasting of maize crop coverage using CNN-LSTM. Int. J. Agric. Biol. Eng. 16(2): 199-206. https://doi.org/10.25165/j.ijabe.20231602.7598

Wu, Q., F. Guan, C. Lv and Y. Huang (2021). Ultra-short-term multi-step wind power forecasting based on CNN-LSTM. IET Renew. Power Gener. 15(5): 1019-1029. https://doi.org/10.1049/rpg2.12147

Yan, R., J. Liao, J. Yang, W. Sun, M. Nong and F. Li (2021). Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. Expert Syst. Appl. 169: 114513. https://doi.org/10.1016/j.eswa.2021.114513.

Wierzejska, R. (2014). Tea and health--a review of the current state of knowledge. PrzeglEpidemiol. 2014;68(3):501-6, 595-9. English, Polish. PMID: 25391016.

Wachira, F.N., Kamunya, S., Karori, S., Chalo, R. andT. Maritim (2013). The tea plants: botanical aspects. Tea in health and disease prevention, 3-17.

Yang, C.S. and J.M.Landau (2000). Effects of tea consumption on nutrition and health. The J. Nutrition, 130(10), 2409–2412

Hajiboland, R. (2017). Environmental and nutritional requirements for tea cultivation. Folia Horticulturae, 29(2), 199–220.

Kumareswaran, T., H.P. Singh, V. Kamalvanshi, K. Debashish andA. Sharma (2018). Dynamics of tea export in India. International J. Current Microbiology and Applied Sciences, 7(10), 3753–3759.

Martin, L. C. (2011). Tea: the drink that changed the world. Tuttle Publishing.

Hazarika, K. and Borah, K. (2013). Small Tea Cultivation in the Process of Self-Employment: A Study on the Indigenous People of Assam. International J. Latest Trends in Finance and Economic Sciences, 3(4), 655–660.

Shah, S.K. (2013). Tea Production in India: challenges and opportunities. J. Tea Science Research, 3(2), 45–50.

Mitra, A. (1991). Tea industry in India: problems and prospects. Economic and Political Weekly, 26(52), 2953–2955.