A Hybrid RF-Attentive BiLSTM Framework for the Agroclimatic Mapping of Arecanut Yield in Central Karnataka

S. Sushitha
BMS Institute of Technology and Management, Bangalore, Visvesvaraya Technological University, Belagavi, India
bhatsushitha@gmail.com (corresponding author)

K. Aparna
BMS Institute of Technology and Management, Bangalore, Visvesvaraya Technological University, Belagavi, India
aparnak@bmsit.ac.in

Received: 14 February 2026 | Revised: 18 March 2026 | Accepted: 26 March 2026

 DOI: https://doi.org/10.48084/etasr.18172

ABSTRACT

Climate variability, geographical disparity, and the lack of agricultural datasets make it difficult to estimate Arecanut crop production in Central Karnataka. Existing machine learning and deep learning models often fail to capture nonlinear climatic interactions as well as temporal climate dynamics. To address this, this research proposes a novel hybrid Random Forest (RF)–attentive Bidirectional Long Short-Term Memory (BiLSTM) framework for the agroclimatic mapping of the Arecanut crop. To handle the class imbalance problem and to enhance feature representation, the Synthetic Minority Over-sampling Technique combined with Edited Nearest Neighbors (SMOTE-ENN) resampling technique is employed in the present study. Here, the RF module extracts nonlinear climatic interactions, whereas the attentive BiLSTM module captures temporal climate patterns that strongly influence the Arecanut growth cycle. The use of a self-attention model helps improve the dynamic weighting of historical observations, whereas the BiLSTM model learns relevant features, improving overall performance. The effectiveness of the developed model is evaluated through comparative analysis with other state-of-the-art solutions. The hybrid model outperforms all baseline models, achieving the highest coefficient of determination (R^2 = 0.914), the lowest Root Mean Square Error (RMSE = 0.39), the lowest Mean Absolute Error (MAE = 0.30), and the lowest Mean Absolute Percentage Error (MAPE = 8.1%). The high correlation coefficient (r = 0.956) further demonstrates the strong predictive capability of the model in capturing crop–climate interactions and confirms the superiority of the proposed hybrid model in agroclimatic mapping.

Keywords-Arecanut yield prediction; Random Forest (RF); BiLSTM feature learning; climate analysis; Arecanut yield; precision agriculture

I. INTRODUCTION

The high-paced growth of the global population has intensified the demand for food resources, putting pressure on agricultural systems to sustain productivity. Climate variability and erratic weather patterns have significantly disrupted traditional farming practices, resulting in declining and unstable crop yield [1]. These climatic challenges have broadened the horizon of precision agriculture, where data-driven and intelligent decision-support systems are applied to improve agricultural yield while guaranteeing long-term sustainability. Under rapidly evolving agroclimatic conditions, effective monitoring and modeling of crop yield responses to environmental variables have become indispensable [2, 3]. This requires understanding the complex, nonlinear interactions among climatic factors such as temperature, humidity, rainfall, and soil moisture, alongside spatial attributes including land cover and geographic location. Capturing the nonlinear relationships between yield and these climatic factors remains a challenging task due to strong spatial heterogeneity, temporal variability, and inherent nonlinearity present in large-scale agricultural systems.

In the past, a few efforts have been made towards agroclimatic mapping; however, most of the state-of-the-art methods primarily relied on statistical regression techniques and classical time-series models, including multiple linear regression and autoregressive frameworks [4-6]. These linear models often fail to generalize across heterogeneous geographic regions. Subsequently, machine learning methods like Support Vector Regression (SVR), decision trees, and ensemble learners have been introduced to capture nonlinear dependencies between climatic variables and crop yield [7-9]. Although these models improve predictive accuracy, many of them inadequately address spatial non-linearity, temporal dependencies, and class imbalance issues inherent in agricultural yield datasets, particularly when applied over broad agroecological zones.

In recent years, numerous studies have proposed hybrid deep learning architectures that amalgamate convolutional, recurrent, and attention modules to jointly learn spatial patterns and temporal dynamics [10, 11]. Such hybrid models outperform standalone Convolutional Neural Network (CNN) or Long Short-Term Memory (LSTM) approaches by capturing both local feature representations and long-term temporal correlations. However, many of these methods focus predominantly on satellite imagery or weather indices and often overlook yield imbalance and regional heterogeneity issues. Despite these efforts, most of the state-of-the-art methods fail to address the data imbalance problem, which is undeniably frequent in agricultural yield prediction, particularly when modeling crops across diverse agroclimatic zones. Recent methods have emphasized the adverse impact of skewed yield distributions on model generalization [12]. Resampling methods like Synthetic Minority Over-sampling Technique (SMOTE) and Edited Nearest Neighbors (ENN) (often called SMOTE-ENN) have been shown to improve regression stability and robustness by mitigating bias toward dominant yield ranges [13, 14]. Nevertheless, their integration with deep hybrid architectures for agroclimatic mapping remains underexplored.

II. RELATED WORK

Karnataka, along with Kerala and Assam, contributes nearly 80% of India's total Arecanut production, whereas Karnataka alone accounts for a substantial share of both cultivation area and output volume [15]. Arecanut farming provides livelihood security for a large rural population and contributes significantly to regional economic development [16]. However, Arecanut yield is highly sensitive to agroclimatic variability, including fluctuations in temperature, humidity, rainfall distribution, and soil moisture [17, 18]. This sensitivity is particularly evident across Karnataka's central districts—Dakshina Kannada, Shivamogga, Davangere, Tumkur, Chikmagalur, Uttara Kannada, Chitradurga, and Udupi—which exhibit pronounced spatial diversity in climatic conditions and yield behavior.

Though, in the past, a few efforts have been made toward Arecanut modeling; yet most of these state-of-the-art approaches are focused on traditional methods such as correlation analysis [19, 20] or path analysis followed by regression models [21]. Several machine learning predictive models have been also implemented to predict yield targeting climatic factors [22-25]. Due to the lack of comprehensive spatial-temporal analysis and limited exploration to capture complex nonlinear climatic interactions, the study highlights the need of advanced spatial predictive models that integrate climate and crop dynamics.

In the context of crop-specific modeling, state-of-the-art studies on Arecanut cultivation have extensively emphasized disease incidence prediction [26, 27] and pest risk analysis using climatic variables [28]. A few recent studies have focused on assessing environmental factors [29, 30] and their impact on crops in Malnad [20] and coastal zones of Karnataka [31-33]. However, these methods largely overlook direct yield prediction and fail to generalize across heterogeneous agroclimatic regions. Furthermore, climatic variables vary significantly across Arecanut-growing districts, including central districts such as Davangere and eastern regions such as Tumkur and Chitradurga, which limits the generalizability of existing approaches over broader geographical regions. Limited efforts have examined the direct relationship between agroclimatic factors and Arecanut yield, particularly in large and heterogeneous regions. Existing studies are often restricted to coastal or high-rainfall zones and lack generalizable spatial yield mapping frameworks [34]. In addition, comprehensive agroclimatic mapping models that jointly model spatial, temporal, and nonlinear dependencies in Arecanut yield across central Karnataka remain absent in the literature.

These inferences clearly indicate that although significant progress has been made in crop yield prediction using machine learning and deep learning, a unified framework that integrates ensemble learning, temporal deep models, attention mechanisms, and imbalance-aware resampling for large-scale agroclimatic mapping is still lacking. This gap motivates the proposed hybrid Random Forest (RF)–attention-based Bidirectional Long Short-Term Memory (BiLSTM) model augmented with SMOTE-ENN resampling, designed to achieve robust, scalable, and interpretable agroclimatic yield mapping for Arecanut cultivation. This serves as the key driving force behind this study. Centered on Arecanut yield prediction, the proposed framework first extracts temporal time-series sequences comprising geographic location, year, temperature, humidity, soil moisture, area, yield, longitude, and latitude features, followed by addressing missing values and outlier analysis.

III. MATERIALS AND METHODS

This section discusses the overall proposed method and its sequential implementation.

A. Study Area

This study considers regions of central Karnataka, including Dakshina Kannada, Shivamogga, Davangere, Tumkur, Chikmagalur, Uttara Kannada, Chitradurga, and Udupi. For these regions, climatic variables such as soil moisture, humidity, and temperature were obtained from the respective geographic locations. Figure 1 illustrates the selected study area. These regions were carefully selected based on their Arecanut yield performance over the past two decades.

The dataset covers the years from 2003 to 2022, depicting annual data collected for each selected district and emphasizing the aggregate Arecanut yield in tonnes. The average Arecanut yield over the two-decade period is shown in Table I. Based on the twenty-year average yield data, the highest Arecanut production (482,774 t) and cultivation area (57,786 ha) are observed in the Shivamogga district. In contrast, the minimum yield (105,895 t) and cultivation area (10,590 ha) are observed in the Udupi district. The yield is defined in tonnes (t), whereas the cultivation area is measured in hectares (ha). The source of the Arecanut production data (Table I) is the Directorate of Economics and Statistics, Department of Agriculture and Farmer Welfare, Government of India [35].

TABLE I. AVERAGE ANNUAL ARECANUT PRODUCTION AND CULTIVATION AREA ACROSS DISTRICTS

Region	Production (t)	Cultivation area (ha)
Dakshina Kannada	474,683	46,608
Shivamogga	482,774	57,786
Davangere	343,030	39,615
Tumkur	297,439	36,358
Chikmagalur	226,377	34,942
Uttara Kannada	227,338	20,141
Chitradurga	189,495	22,340
Udupi	105,895	10,590

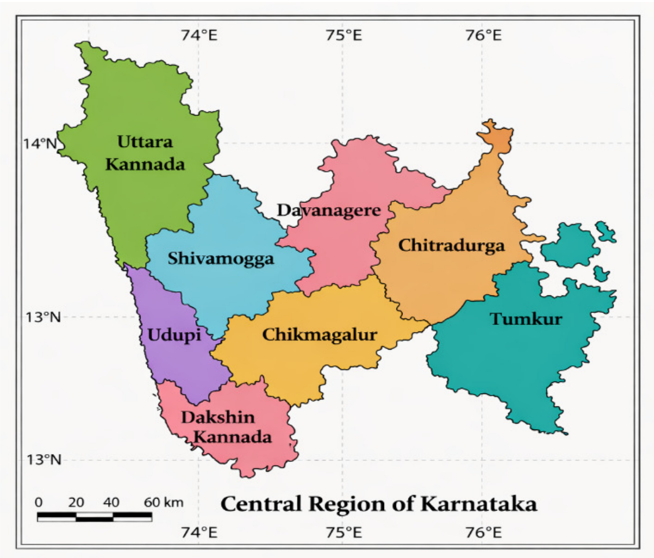

Fig. 1. Study area.

B. Data Acquisition

A long-term dataset (20 years) on the cultivation area and Arecanut production in the selected regions was collected from the Directorate of Economics and Statistics, Department of Agriculture and Farmer Welfare, Government of India [35]. The dataset provides district-level agricultural statistics on area, production, and yield, compiled from state agricultural statistics.

The data for the primary external factors of the study, namely earth skin temperature, relative humidity, and soil moisture levels, were obtained using the NASA POWER Project's Data Access Viewer (DAV) [36] over a period of twenty years for the selected regions. Data were collected under the agroclimatology category for each district. The geospatial dataset, including the longitude and latitude of the selected regions, was collected using the Q-field app through QGIS software [37]. The pertinent variables considered in this study are schematically presented in Table II.

TABLE II. AGROCLIMATIC VARIABLES OF THE STUDY

Variable	Unit	Description
Temperature	°C	Average annual temperature across the selected regions
Humidity	%	Average humidity measured in the selected study areas
Soil moisture	%	Average soil moisture levels recorded in the selected regions
Cultivation area	ha	Area under cultivation in the selected region over a specified period
Crop yield	t	Average annual Arecanut production in the selected region
Latitude	—	Represents the geographical location
Longitude	—	Indicates the longitudinal position of the selected regions

C. Data Preprocessing

Before training the machine learning and deep learning models, the data underwent the following preprocessing steps.

1) Weighted Neighborhood-Based Estimator

The missing values in the temporal (sequential) datasets were imputed using a neighborhood-based estimator. The proposed method estimates missing entries using information from the most similar samples, thereby preserving local patterns and nonlinear relationships. Outlier analysis was performed using Interquartile Range (IQR)-based outlier detection and removal.

2) Z-Score Normalization

Considering nonlinear input patterns and possible overfitting, the proposed method applies z-score normalization. The normalized temporal climatic variables were segmented into fixed-length sequences to preserve seasonal patterns.

3) SMOTE-ENN Sampling

The class imbalance problem associated with the multi-geographical agroclimatic data is addressed using an improved SMOTE resampling technique [38], referred to as SMOTE-ENN. The proposed method generates synthetic samples representing highly correlated data instances without affecting the original data distribution. It applies minority samples as input to generate the synthetic data instances, which are subsequently processed using the k-Nearest Neighbors (k-NN) algorithm. In this work, a Euclidean distance-based k-NN algorithm is applied to generate the most probable samples to be combined with the original samples.

D. Model Design

This research proposes a hybrid RF–attention-based BiLSTM model for feature extraction and learning. The RF ensemble learning model is applied to learn temporal features, whereas the attention-based BiLSTM model helps learn parametric interdependencies over large sequential (temporal) inputs. A brief description of these feature extraction methods is provided as follows.

1) Random Forest Module

In this work, an RF model is trained on the resampled data to capture nonlinear relationships amongst the agroclimatic variables. Each tree in the forest learns a different subspace of the data through bootstrapped sampling, thereby collectively capturing spatial heterogeneity. The ensemble then combines the output of all trees using a Majority Voting-based Ensemble (MVE) for classification. With a fixed number of trees operating on the input features, the RF classifier computes the maximum voting score for each data instance and uses it for prediction. For the targeted agroclimatic mapping tasks, 80% of the samples are used for training, whereas 20% of the samples are used as Out-Of-Bag (OOB) samples for further inner cross-validation to enhance classification performance. Given T decision trees, the RF prediction is defined as:

$$f_{RF}(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (1)$$

2) Self-Attentive BiLSTM

a) BiLSTM Layer

To capture temporal dependencies in climatic sequences, the proposed method applies an attention-based BiLSTM network as one of the feature models. To effectively model temporal dependencies across climatic seasons, the agroclimatic data are structured into sequences and fed into the BiLSTM network. To extract long-term dependencies across different agroclimatic parameters, this research applies a robust weighted attention-based BiLSTM that captures temporal dependencies for improved learning and prediction. The architecture of the BiLSTM layer is depicted in Figure 2.

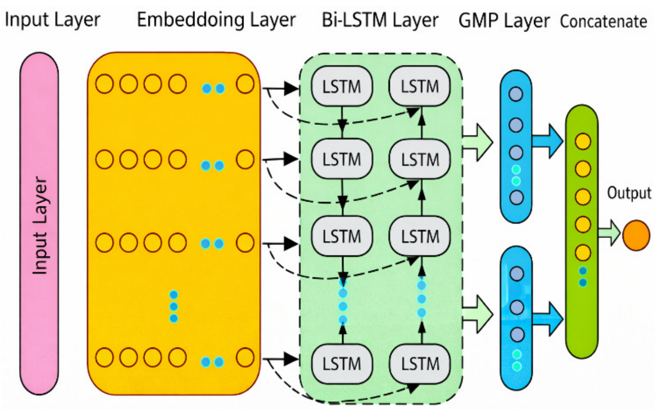

Fig. 2. BiLSTM layer.

In this work, the BiLSTM is designed with two LSTM networks, each capable of processing input vectors (i.e., extracting features) in both forward and backward directions. As depicted in Figure 2, the forward LSTM model extracts features from left to right, with the hidden layer output defined by (2):

$$\vec{h}_t = \text{LSTM}(x_t, \vec{h}_{t-1}) \quad (2)$$

Unlike the forward LSTM, the backward LSTM performs feature extraction in the right-to-left direction (Figure 2). Thus, the hidden layer output during the backward pass is obtained as:

$$\overleftarrow{h}_t = \text{LSTM}(x_t, \overleftarrow{h}_{t+1}) \quad (3)$$

The features extracted from both forward and backward directions are combined to form contextual (global) features, as represented in (4):

$$h_{t,BiLSTM} = [\vec{h}_t, \overleftarrow{h}_t] \quad (4)$$

b) Attention Layer

To emphasize the more important features within large sequential agroclimatic data, the output from the BiLSTM layer is passed to an attention layer, where weights are assigned to each data element. The attention-weighted BiLSTM embeddings are then fused with the RF-extracted features to focus on the most relevant agroclimatic inputs. Figure 3 illustrates this process.

The BiLSTM output, or feature vector h_t, is fed into a single-layer Multilayer Perceptron (MLP) to learn a hidden representation u_t. An attention score is subsequently computed using u_t and a context vector u_w. The final attention-weighted feature is obtained as follows:

$$u_t = \tanh(W_u h_t + b_w) \quad (5)$$

$$a_t = \frac{\exp(u_t^T u_w)}{\sum_t \exp(u_t^T u_w)} \quad (6)$$

$$s = \sum_t a_t h_t \quad (7)$$

The output from the hidden state of the attention-based BiLSTM cell is then fed into a fully connected layer, which utilizes a neural network with a single hidden layer.

c) Hybrid Prediction Fusion

After obtaining the outputs from the RF and self-attention-based BiLSTM models, a weighted fusion method is applied to perform prediction, as given in (8):

$$\hat{y} = w_1 f_{RF}(x) + w_2 f_{BiLSTM-Attention}(x) \quad (8)$$

subject to the constraint that $w_1 + w_2 = 1$, where $w_1, w_2 \geq 0$. In this work, the optimal weights are determined using validation-based optimization to minimize the prediction error.

E. Model Evaluation

The overall proposed model was implemented using MATLAB 2022b software. The statistical performance parameters employed are as follows:

- Coefficient of determination (R^2): It measures the explanatory power of the developed model and is represented as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (9)$$

- Root Mean Square Error (RMSE): It evaluates the standard deviation of the prediction errors and is given by:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (10)$$

- Mean Absolute Error (MAE): It measures the average magnitude of prediction errors. For actual values O_i and predicted values P_i over N observations, MAE is defined as:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|O_i - P_i| \qquad (11)$$

- Mean Absolute Percentage Error (MAPE): It evaluates the prediction error as a percentage of the actual values, making it scale-independent:

$$\text{MAPE}(\%) = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{O_i - P_i}{O_i}\right| \qquad (12)$$

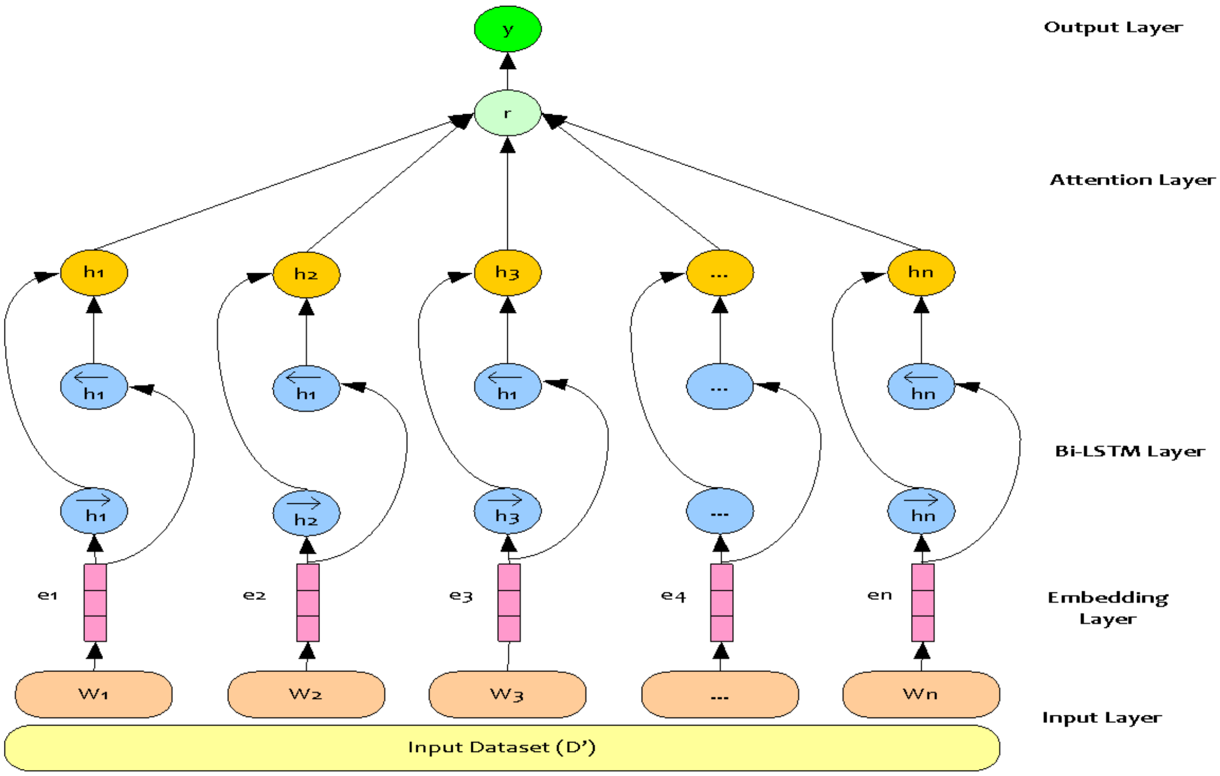

Fig. 3. Attention layer.

IV. RESULTS AND DISCUSSION

This section presents the experimental results of the proposed hybrid model. In the context of Arecanut yield prediction, the proposed model first extracts key temporal time-series sequences containing different geographic locations and corresponding agroclimatic measurements.

A. Region-Wise Hybrid Model Analysis

The developed hybrid model successfully models temporal dependencies, indicating a strong association between the Arecanut crop and climatic variables. The findings from Table III indicates that coastal and Malnad districts, such as Udupi, Dakshina Kannada, and Chikmagalur, exhibit the highest prediction accuracy, with R^2 values exceeding 0.92 and comparatively lower RMSE and MAE. This indicates that the model effectively learns stable yield–climate relationships with respect to climatic factors.

In contrast, districts such as Davangere, Tumkur, and Chitradurga show marginally lower R^2 values and slightly higher error metrics. These regions are known for greater climatic variability, which introduces nonlinearity into yield dynamics. Despite these challenges, the model still achieves R^2 values close to or above 0.89, exhibiting robust generalization capability. Notably, Shivamogga presents a balanced performance profile. The relatively moderate RMSE and MAE values suggest that the hybrid learning framework successfully captures both short-term temporal dependencies and long-range spatial patterns.

TABLE III. REGION-WISE MODEL PERFORMANCE

District	R^2	RMSE	MAE
Dakshina Kannada	0.928	0.372	0.289
Shivamogga	0.919	0.401	0.311
Chikmagalur	0.925	0.389	0.298
Davangere	0.901	0.426	0.329
Tumkur	0.893	0.441	0.341
Chitradurga	0.887	0.458	0.356
Udupi	0.931	0.365	0.281

B. Comparative Performance Analysis of the Proposed Model

To examine comparative performance, different baseline models, including SVR, MLP, standalone BiLSTM, standalone RF, and the attention-based Transformer, were designed and simulated. The comparative performance results are presented in Table IV.

TABLE IV. COMPARATIVE RESULTS WITH STATE-OF-THE-ART MODELS

Model	R^2	RMSE	MAE	MAPE (%)	r
SVR	0.781	0.612	0.481	14.92	0.884
MLP	0.804	0.574	0.452	13.68	0.897
RF	0.832	0.531	0.418	12.21	0.912
BiLSTM	0.861	0.487	0.382	10.96	0.928
Attention Transformer	0.883	0.451	0.349	9.84	0.940
Hybrid model (proposed)	0.914	0.398	0.307	8.11	0.956

1) *Performance of the Base Models*

The SVR model exhibits the lowest predictive accuracy, with an R^2 value of 0.781 and comparatively higher error magnitudes, indicating its limited ability to capture nonlinear and spatiotemporal dependencies inherent in agroclimatic datasets. The MLP shows moderate improvement ($R^2 = 0.804$) over SVR, achieving reduced error values and a higher correlation coefficient ($r = 0.89$), suggesting that shallow neural architectures can partially learn nonlinear patterns. The ensemble-based RF model further enhances predictive reliability, with an increased R^2 of 0.832. However, it remains inherently static and does not explicitly model temporal dependencies. The BiLSTM model achieves a substantial gain in performance, attaining an R^2 of 0.861 and lower percentage errors, underscoring the importance of bidirectional temporal learning in capturing yield variations driven by sequential agroclimatic factors. The attention-based Transformer model further improves performance by reducing prediction errors and strengthening correlation.

2) *Performance of the Proposed Hybrid Model*

The proposed hybrid RF–attention-based BiLSTM model consistently outperforms all individual baselines, achieving the highest R^2 value of 0.914 and the lowest RMSE, MAE, and MAPE values. The strong correlation coefficient ($r = 0.956$) confirms the robustness and reliability of the hybrid framework. The results demonstrated in Figure 4 indicate the superior performance of the proposed hybrid model.

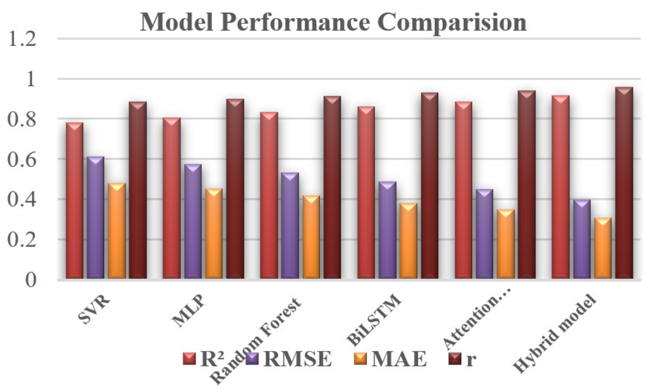

Fig. 4. Comparison of base models with the proposed model.

C. *SMOTE-ENN Resampling Results*

In the present work, SMOTE-ENN was applied to address the class imbalance problem. In this context, the simulations were performed both with and without SMOTE-ENN resampling, and the corresponding performance was examined. After integrating the SMOTE–ENN strategy, a consistent and notable improvement was observed across all evaluation measures (Table V).

TABLE V. EFFECT OF RESAMPLING TECHNIQUE ON PERFORMANCE

Dataset	R^2	RMSE	MAE	MAPE (%)
Without SMOTE-ENN	0.865	0.482	0.374	11.27
With SMOTE-ENN	0.914	0.398	0.307	8.11

D. *Error Reduction Analysis*

The error reduction analysis (Table VI) is conducted to evaluate the effectiveness of the proposed hybrid model. The proposed hybrid model outperforms all the baseline models, validating its effectiveness in improving prediction accuracy.

TABLE VI. ERROR REDUCTION ANALYSIS

Baseline model	RMSE reduction (%)	MAE reduction (%)
SVR	35.0	36.2
MLP	30.7	32.1
RF	25.0	26.6
BiLSTM	18.3	19.6
Attention Transformer	11.7	12.0

E. *Feature Importance Analysis*

To evaluate the relative importance of the selected climatic variables, feature importance analysis was conducted using the RF model along with permutation importance based on OOB samples. As presented in Figure 5, the climatic factors were assessed across five different permutation runs to ensure stability. The results clearly indicate that the temperature variable consistently exhibited the highest OOB permutation importance score (0.85–0.9), highlighting its critical influence on the Arecanut crop. In contrast, the variables humidity and soil moisture demonstrated a moderate impact on crop productivity.

F. *Optimized Blending Weights of the Proposed Hybrid Model*

The relative contribution of each individual model to the final ensemble framework is depicted in Figure 6. The optimized blending weights indicate that the BiLSTM contributed more prominently to the final outcome by effectively capturing the temporal climate patterns, whereas the RF contributed comparatively less by capturing nonlinear climatic interactions, thereby improving the overall performance of the hybrid model. Thus, the overall findings from the present study confirm that the proposed RF–attention-based BiLSTM hybrid model successfully captures temporal climatic patterns, followed by nonlinear features, which significantly influence the Arecanut crop in central regions of

Karnataka. The model estimates crop outcomes more accurately, improves prediction accuracy, and reduces model error, thereby ensuring robust performance across the selected study region.

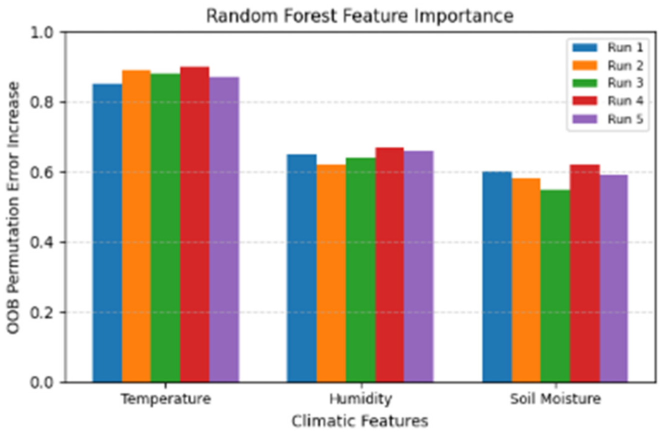

Fig. 5. Feature importance using the RF model.

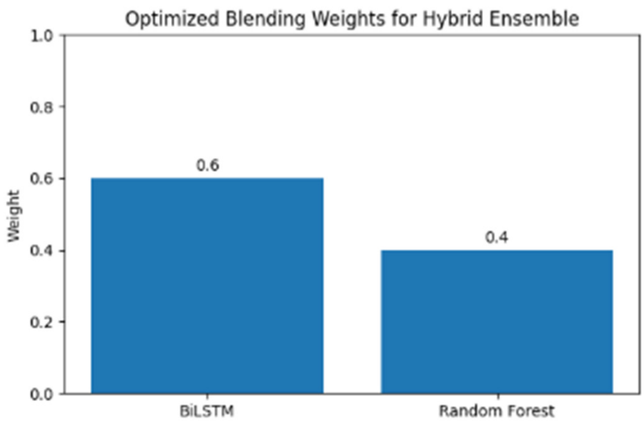

Fig. 6. Optimized blending weights (BiLSTM and RF).

V. CONCLUSION

The present study proposes a novel hybrid Random Forest (RF)–attention-based Bidirectional Long Short-Term Memory (BiLSTM) framework for agroclimatic mapping of the Arecanut crop in central Karnataka. The proposed model effectively captures the association between nonlinear climatic factors and Arecanut yield. The Synthetic Minority Over-sampling Technique combined with Edited Nearest Neighbors (SMOTE-ENN) resampling method addresses the class imbalance problem present in the agroclimatic data, ensuring data consistency and improving the predictive accuracy of the proposed model.

In this framework, the RF component captures underlying complex climatic interactions, whereas the BiLSTM component captures temporal patterns. The quantitative evaluation demonstrates the superiority of the hybrid framework over baseline models, including Support Vector Regression (SVR), Multilayer Perceptron (MLP), RF, BiLSTM, and the attention-based Transformer. The highest coefficient of determination (R^2) of 0.914, along with the lowest Root Mean Square Error (RMSE) (0.39), Mean Absolute Error (MAE) (0.30), and Mean Absolute Percentage Error (MAPE) (8.1%), highlights the strong predictive performance of the model. The high correlation coefficient ($r = 0.956$) further confirms the robustness of the hybrid framework.

The region-wise analysis confirms the spatial generalizability of the model ($R^2 > 0.90$), with the highest performance observed in Udupi ($R^2 = 0.931$) and Dakshina Kannada ($R^2 = 0.928$) districts. Furthermore, the feature importance analysis using RF highlights the critical influence of the temperature variable on Arecanut yield. The consistent reduction in error metrics (RMSE and MAE) strongly confirms that the integration of the ensemble-based RF encoder with the attention-based BiLSTM model enables high prediction accuracy. Hence, these findings establish the proposed hybrid framework as a reliable model for crop–climate mapping across heterogeneous climatic zones.

DECLARATION OF COMPETING INTERESTS

Not applicable to this work.

ACKNOWLEDGMENT

Not applicable to this work.

FUNDING SOURCES

The authors received no funding for this work.

DATA AVAILABILITY

The utilized data can be found in [35-37].

REFERENCES

[1] T. F. Fotabong *et al.*, "Soil Analysis Using Deep Learning for Precision Agriculture," in *2024 First International Conference on Pioneering Developments in Computer Science & Digital Technologies*, Delhi, India, 2024, pp. 303–307, https://doi.org/10.1109/IC2SDT62152.2024.10696104.

[2] S. K. Phang, T. H. A. Chiang, A. Happonen, and M. M. L. Chang, "From Satellite to UAV-Based Remote Sensing: A Review on Precision Agriculture," *IEEE Access*, vol. 11, pp. 127057–127076, 2023, https://doi.org/10.1109/ACCESS.2023.3330886.

[3] FAO, "The future of food and agriculture: trends and challenges," Food and Agriculture Organization of the United Nations, Rome, 2017.

[4] M. R. Evans, D. Oliver, X. Zhou, and S. Shekhar, "Spatial Big Data: Case Studies on Volume, Velocity, and Variety," in *Big Data*, 2nd ed., H. A. Karimi, Ed. Boca Raton, FL, USA: CRC Press, 2024, ch. 8, https://doi.org/10.1201/9781003406969-8.

[5] P. Helber, B. Bischke, C. Packbier, P. Habelitz, and F. Seefeldt, "An Operational Approach to Large-Scale Crop Yield Prediction with Spatio-Temporal Machine Learning Models," in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, Athens, Greece, 2024, pp. 4299–4302, https://doi.org/10.1109/IGARSS53475.2024.10641218.

[6] M. J. Wellington, R. Lawes, and P. Kuhnert, "A framework for modelling spatio-temporal trends in crop production using generalised additive models," *Computers and Electronics in Agriculture*, vol. 212, Sept. 2023, Art. no. 108111, https://doi.org/10.1016/j.compag.2023.108111.

[7] B. S. Sidhu, Z. Mehrabi, N. Ramankutty, and M. Kandlikar, "How can machine learning help in understanding the impact of climate change on

crop yields?," *Environmental Research Letters*, vol. 18, no. 2, Jan. 2023, Art. no. 024008, https://doi.org/10.1088/1748-9326/acb164.

[8] L. Wickramasinghe, R. Weliwatta, P. Ekanayake, and J. Jayasinghe, "Modeling the Relationship between Rice Yield and Climate Variables Using Statistical and Machine Learning Techniques," *Journal of Mathematics*, vol. 2021, no. 1, Feb. 2021, Art. no. 6646126, https://doi.org/10.1155/2021/6646126.

[9] B. N. S. GUPTA and U. Lekhana, "Crop Recommender System Using Machine Learning Approach," *International Journal of Engineering Research and Science & Technology*, vol. 20, no. 2, pp. 1167–1177, June 2024.

[10] V. H. Kalmani, N. V. Dharwadkar, and V. Thapa, "Crop Yield Prediction using Deep Learning Algorithm based on CNN-LSTM with Attention Layer and Skip Connection," *Indian Journal Of Agricultural Research*, vol. 59, no. 8, pp. 1303–1311, Nov. 2025, https://doi.org/10.18805/IJARe.A-6300.

[11] S. M. V, J. P, G. George, P. S, and K. N, "Hybrid Deep Learning Models for Precision and Sustainable Crop Management," in *2025 IEEE 6th Global Conference for Advancement in Technology*, Bangalore, India, 2025, pp. 1–7, https://doi.org/10.1109/GCAT66372.2025.11368572.

[12] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sept. 2009, https://doi.org/10.1109/TKDE.2008.239.

[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, June 2002, https://doi.org/10.1613/jair.953.

[14] D. L. Wilson, "Asymptotic Properties of Nearest Neighbor Rules Using Edited Data," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-2, no. 3, pp. 408–421, July 1972, https://doi.org/10.1109/TSMC.1972.4309137.

[15] H. B P, A. K. K M, S. S, and A. K, "Karnataka's Arecanut Boom: Understanding the Current Scenario and Factors Driving the Shift," vol. 6, no. 8, Krishi Science – eMagazine for Agricultural Sciences, pp. 1–6, Aug. 06, 2025.

[16] P. K, S. H.L., K. Sharma, and M. S. Nayak, "Cultivating arecanut in India: challenges, opportunities and sustainable practices," *Madras Agricultural Journal*, vol. 111, no. 4, pp. 58–70, Dec. 2024, https://doi.org/10.29321/MAJ.10.601147.

[17] R. Bhat, S. Sujatha, Bhavishya, U. K. Priya, A. Gupta, and A. Uchoi, "Arecanut (Areca catechu L.)," in *Soil Health Management for Plantation Crops: Recent Advances and New Paradigms*, G. V. Thomas and V. Krishnakumar, Eds. Singapore: Springer Nature, 2024, pp. 177–206, https://doi.org/10.1007/978-981-97-0092-9_4.

[18] K. M. Sunil, V. S. Devadas, and S. P. George, "Influence of Weather Parameters on Yield and Yield Attributes of Areca Nut (Areca catechu L.)," *Journal of Agricultural Physics*, vol. 11, pp. 88–90, 2011.

[19] S. Sujatha, R. Bhat, and S. E. Apshara, "Climate change, weather variability and associated impact on arecanut and cocoa in humid tropics of India," *International Journal of Innovative Horticulture*, vol. 7, no. 1, pp. 27–37, 2018.

[20] K. B. Hebbar, A. S. P, S. Jose V, R. S v, and R. Bhat, "Predicting current and future climate suitability for arecanut (*Areca catechu* L.) in India using ensemble model," *Heliyon*, vol. 10, no. 4, Feb. 2024, Art. no. e26382, https://doi.org/10.1016/j.heliyon.2024.e26382.

[21] C. T. Jose, K. P. Chandran, K. Muralidharan, and S. Jayasekhar, "Crop weather relationship in arecanut," *Journal of Plantation Crops*, vol. 47, no. 3, pp. 145–151, 2019, https://doi.org/10.25081/jpc.2019.v47.i3.6049.

[22] J. H. Jeong *et al.*, "Random Forests for Global and Regional Crop Yield Predictions," *Plos One*, vol. 11, no. 6, June 2016, Art. no. e0156571, https://doi.org/10.1371/journal.pone.0156571.

[23] Z. K. Jabal, T. S. Khayyun, and I. A. Alwan, "Impact of Climate Change on Crops Productivity Using MODIS-NDVI Time Series," *Civil Engineering Journal*, vol. 8, no. 6, pp. 1136–1156, June 2022, https://doi.org/10.28991/CEJ-2022-08-06-04.

[24] D. B. Lobell, D. Thau, C. Seifert, E. Engle, and B. Little, "A scalable satellite-based crop yield mapper," *Remote Sensing of Environment*, vol. 164, pp. 324–333, July 2015, https://doi.org/10.1016/j.rse.2015.04.021.

[25] R. P. P. Guthu and S. R. Bellipady, "A Formal Statistical Data Modeling for Knowledge Discovery and Prognostic Reasoning of Arecanut Crop using Data Analytics," *International Journal of Software Science and Computational Intelligence*, vol. 14, no. 1, pp. 1–27, 2022, https://doi.org/10.4018/IJSSCI.311447.

[26] R. Krishna, P. K. V, and R. Gaonkar, "Areca Nut Disease Dataset Creation and Validation using Machine Learning Techniques based on Weather Parameters," *Engineered Science*, vol. 19, no. 4, pp. 205–214, June 2022.

[27] P. B. R and N. P. S, "A Literature Review of the Detection and Categorization of various Arecanut Diseases using Image Processing and Machine Learning Approaches," *International Journal of Applied Engineering and Management Letters*, vol. 5, no. 2, pp. 183–204, Dec. 2021, https://doi.org/10.5281/zenodo.5773853.

[28] K. A. Garrett *et al.*, "Complexity in climate-change impacts: an analytical framework for effects mediated by plant disease," *Plant Pathology*, vol. 60, no. 1, pp. 15–30, 2011, https://doi.org/10.1111/j.1365-3059.2010.02409.x.

[29] B. Patil *et al.*, "Exploring the Impact of Climatic Variables on Arecanut Fruit Rot Epidemic by Understanding the Disease Dynamics in Relation to Space and Time," *Journal of Fungi*, vol. 8, no. 7, July 2022, Art. no. 745, https://doi.org/10.3390/jof8070745.

[30] K. Rajashree, K. V. Prema, G. Rajath, and S. Angad, "Prediction of fruit rot disease incidence in Arecanut based on weather parameters," vol. 22, no. S1, pp. 1117–1133, Nov. 2022, https://doi.org/10.15159/AR.22.076.

[31] A. B. Tejaswini, H. S. Padmashri, D. V. Soumya, T. N. Prakash Kammardi, and H. G. Hemalatha, "Association studies of climate factors on arecanut yield in Western Ghats region Karnataka.," *Environment & Ecology*, vol. 32, no. 3A, pp. 1085–1087, July 2014.

[32] S. Jayasekhar, C. T. Jose, C. Thamban, and K. Muralidharan, "Economic impact of arecanut based cropping systems: A study of Dakshina Kannada district-Karnataka," *Journal of Plantation Crops*, vol. 40, no. 1, pp. 50–55, June 2012.

[33] P. K, G. Naik, M. K. Naik, V. Hegde, B. C. Dhananjaya, and S. M. K., "Exploring Yellow Leaf Disease Patterns in Areca Plantations in Chikkamagaluru District of Karnataka," *Saudi Journal of Pathology and Microbiology*, vol. 9, no. 3, pp. 63–70, Mar. 2024, https://doi.org/10.36348/sjpm.2024.v09i03.002.

[34] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, pp. 4559–4565, Feb. 2017, https://doi.org/10.1609/aaai.v31i1.11172.

[35] "Area, Production and Yield (APY) Statistics Portal." Directorate of Economics and Statistics, Department of Agriculture and Farmer Welfare, Government of India. https://data.desagri.gov.in/website/apy-query-report-web.

[36] "Data Access Viewer (DAV)." Nasa Power. https://power.larc.nasa.gov/data-access-viewer/.

[37] "QGIS Geographic Information System." Open-Source Geospatial Foundation Project. https://qgis.org/.

[38] N. G. Renukaradya and K. G. Rao, "A Long Short-Term Memory Guided Conditional Dynamic Variational Auto-Encoder for Crop Yield Prediction and Crop Type Classification," *Engineering, Technology & Applied Science Research*, vol. 15, no. 6, pp. 29770–29778, Dec. 2025, https://doi.org/10.48084/etasr.12586.

AUTHORS PROFILE

Sushitha S is a Research Scholar in the Department of Master of Computer Applications (MCA) at BMSIT, under Visvesvaraya Technological University, Belagavi, India. Her research interests include Machine Learning, Data Analytics, and RNN networking models.

Aparna K is a Professor at the Department of Master of Computer Applications (MCA) at BMSIT, under Visvesvaraya Technological University, Belagavi, India. She is a professional member of IEEE and ISTE.