*AgriEngineering*

MDPI

*Article*

# A Hybrid Spatiotemporal Learning for the Regional Agroclimatic Mapping and Genotypic Resilience Analysis: A Case Study of Areca Nut Yield in Central Karnataka

**Sushitha Seetharam * and Aparna Kalyanasundaram**

BMS Institute of Technology & Management Bangalore, Visvesvaraya Technological University, Belagavi 590018, Karnataka, India; aparnak@bmsit.in

* Correspondence: bhatsushitha@gmail.com

**Abstract**

The last few years have witnessed an exponential rise in the demand of precision agriculture that needs a robust agroclimatic mapping. The nonlinearity, spatial, temporal and region-specific climate interactions which governs the production makes the crop–climate mapping a challenging task. Centered on areca nut yield, in the present study, a novel spatiotemporal integrated hybrid RF attention-BiLSTM framework is developed to analyze the crop and climate associations in Central Karnataka. By integrating a comprehensive data preprocessing strategy, including missing-value estimation, outlier treatment, scale normalization, data stratification, and SMOTE–ENN resampling, the developed model captured the region-specific climate sensitive production pattern. The hybrid model achieved an $R^2$ value of 0.914, RMSE of 0.39, and correlation coefficient of 0.956, outperforming conventional models. The repeated-run, statistical, and ablation analyses further demonstrated the model consistency and the effectiveness of integrated hybrid framework. Further, a genotype analysis is performed to analyze resilience and stability of different areca nut varieties. The overall statistical results confirm robustness of the proposed framework, where the ability to amalgamate data-level optimization, attention-driven temporal learning, and ensemble modeling enables it robust and hence well-suited for complex, real-world time-series agroclimatic mapping and prediction tasks for areca nut precision agriculture decisions.

**Keywords:** machine learning; areca nut yield; crop modeling; BiLSTM learning; prediction analysis; climate analysis

Check for updates

Academic Editor: Zhengjun Qiu

Received: 12 February 2026
Revised: 11 September 2026
Accepted: 14 September 2026
Published: 20 September 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

## 1. Introduction

In recent years climate variability, unplanned urbanization, soil degradation, and irregular weather conditions have decisively disrupted traditional farming practices, resulting in declining and unstable crop yields [1]. Such challenges have expanded the scope of precision agriculture (PA), where data-driven and intelligent decision-support systems are applied to improve agricultural yield productivity while guaranteeing long-term sustainability. In the recent years, the world has witnessed substantial alterations in vegetation patterns, soil characteristics, land-use dynamics, and microclimatic conditions that cumulatively have exerted a profound impact on crop yield behavior [2]. The recent reports from the United Nations (UN) highlight a widening gap between agricultural production and consumption demands, emphasizing the urgency of deploying advanced technologies,

including the machine learning, artificial intelligence and derived analytical frameworks capable of supporting climate-resilient agriculture decisions [3,4]. Despite efficacy towards agriculture prediction systems, the diversity of climatic variables and crop-specific yields heterogeneity remains as a significant challenge.

Under the changing agroclimatic conditions, effective monitoring and modeling of crop yield responses to environmental variables have become essential for improving agricultural sustainability [4,5]. An effective sustainable agricultural planning decisively relies on 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 [5]. The precision agriculture (PA) leverages these multidimensional datasets to enable adaptive crop selection, optimized resource allocation, and improved land management strategies [6]. A critical component of PA is agroclimatic mapping, which seeks to establish spatially explicit relationships between agroclimatic variables and crop yield to support informed, location-specific farming decisions [7]. Yet, accurately capturing these relationships remains a challenging task due to the spatiotemporal and nonlinear behavior of the agricultural systems. In the past, a few efforts have been made towards agroclimatic mapping; however, most of the state-of-art methods primarily focused on statistical regression techniques and classical time-series models, embodying multiple linear regression and autoregressive frameworks [8–10]. While these methods provide interpretability, they operate largely as local or linear models and often fail to generalize across heterogeneous geographic regions. Subsequently, ML methods like artificial neural networks, support vector regression, decision trees, and ensemble learners have been introduced to capture nonlinear dependencies between climatic variables and crop yield [11–14]. Although these models improve predictive accuracy, many of them inadequately address spatial and temporal nonlinearity and data imbalance issues inherent in agricultural yield datasets, particularly when applied over broad agroecological zones. In recent years, the efforts were made to design more advanced spatial regression techniques, such as the geographically weighted regression (GWR) model, which have demonstrated improved performance by applying spatial dependencies into the modeling process [15,16]. Nevertheless, these techniques remain limited in their ability to simultaneously model nonlinear relationships, temporal dependencies, and spatial heterogeneity at scale. Recent studies suggest that hybrid learning paradigms, by amalgamating ensemble learning with deep neural architectures can provide a promising pathway to overcome aforesaid limitations [17]. These key inferences are considered as the motivation behind this study, where we intend to develop a robust agroclimatic mapping and genotypic resilience analysis tool for areca nut cultivation in Central Karnataka regions.

Areca nut (*Areca catechu* L.), also known as betel nut, is an economically important plantation crop in India, where the state of Karnataka contributes substantially to the nation's areca nut production by supporting the livelihoods of large farming communities [18]. The long production cycle of the areca nut crop makes the crop performance highly sensitive and responsive to various changing environmental conditions, such as variations in temperature, acute rainfall conditions, increases in humidity levels and soil moisture availability [19]. Under such drastic climatic conditions, a reliable crop production assessment system necessitates a systematic model more than a conventional trend analysis, which is capable of handling the increasing climatic variability. It demands advanced crop models which are capable of identifying the impact of climatic influences that evolve over time and differ across different production environments. Hence, this requirement is central to agroclimatic mapping, where agroclimatic and the geographical information is combined and interpreted in order to characterize the spatially and temporally varying crop and climate associations and to support location-specific agricultural impacts.

One of the major challenges in modeling areca nut production, particularly in Karnataka, an Indian state, is the uneven distribution of climatic conditions among its production regions. The eight study regions considered in the present study—Dakshina Kannada, Udupi, Uttara Kannada, Shivamogga, Chikkamagaluru, Davangere, Tumkur, and Chitradurga—span coastal, Malnad, and central regions, and relatively drier interior environments may be represented [20]. As a result, the external environmental conditions that govern areca nut production are neither spatially uniform nor temporally at a steady state. Under such conditions, the crop and climate associates learned primarily from humid or from the higher temperature variation zones may not remain the same in other interior regions that are characterized by different temperature regimes. Hence, spatial variability is one of the major characteristics of this agroclimatic analysis and it is a fundamental modeling constraint. Apart from this, these crop–climate associations also vary over time, where interannual environmental variations can alter the areca nut production performance across its growing periods. Moreover, these climatic factors do not impact crop performance in a linear manner; instead, the response is uneven, curved and sometimes even sudden. This nonlinear behavior along with the spatiotemporal variability that exists between crop and climate interactions is therefore a fundamental modeling constraint with respect to the agroclimatic areca nut dataset. Hence, an effective agroclimatic mapping must integrate this spatiotemporal and nonlinear behavior and provide robust representations across the production environments rather than being optimized for a particular climatic zone.

Therefore, the multiregional agroclimatic areca nut dataset poses interconnected modeling challenges: nonlinear crop and climate interactions, temporal dependency followed by spatial or geographical heterogeneity and the uneven representation of the production conditions. The climatic patterns do not affect the production system independently; instead, their combined influence varies with location and time, producing a complex yield response pattern. To learn the long-term crop behavior pattern, one should observe the monthly and year-to-year climatic variations that produce a dynamic change in crop growth. Apart from this, the sparsely represented areca nut production conditions and the existing regional differences create a substantial variation with respect to data distributions, where the less frequently observed production patterns may be presented inadequately in model learning. Hence, the integrated spatiotemporal and nonlinear crop and climate interactions that exist in an arecanut system challenges the traditional single-model frameworks and motivate an integrated approach which is capable of capturing the combined learning patterns under diverse agroclimatic conditions.

The existing investigations provide a partial solution to these crop–climate challenges. Although the statistical regression [21,22] and the conventional time-series frameworks [23] provide interpretable crop and climate associations, they are constrained when the areca nut response becomes strongly nonlinear in nature and varies spatiotemporally. Many of the existing machine learning approaches, including support vector regression, decision trees, artificial neural networks, and ensemble methods, have demonstrated improved capability in capturing nonlinear predictions, yet provide a limited representation of long-term sequential climatic distributions [21,24–27]. On the other hand, the deep learning frameworks, including LSTM and BiLSTM, have addressed these limitations by effectively learning the temporal behavior [28]. However, these approaches predominantly target temporal dynamics in the agroclimatic dataset rather than the regional differences. Hence, their limited focus towards the spatial heterogeneity limits their ability to represent the crop–climate mapping system under geographically varying agroclimatic environments.

In spite of the growing research scope on other major crops like rice and wheat, the areca nut crop was selected in this study because of its regional importance in Karnataka and also its sensitivity towards various external environmental and soil factors. Although,

in the past, many studies have tried to address this climatic sensitivity, a critical methodological gap exists in agroclimatic mapping under climatically heterogeneous regions. First, the previous investigations have largely focused on *Areca catechu* L. crop disease incidence datasets [25,26,28,29], pest risk (Tirathaba mundella, Carvalhoia) [30] and production behavior with respect to climatic sensitivity within geographically restricted environments. That is, the majority of these investigations are concentrated on coastal, Western Ghats, and Malnad regions [17,28,31–34] and notably very few studies have explored the interior areca nut regions like Davangere and Tumkur, which represent a drier zone with greater climatic variability with respect to temperature and soil moisture conditions. Hence, such region-specific studies limit the spatial generalizability of the learned crop–climate relationships. Second, many of the crop production models rely on annual aggregated climatic observations rather than seasonal (monthly) climatic fluctuations. The annual aggregated climatic data may help in long-term trend analysis, but it limits actual learning of the temporal crop–climate dynamics of areca nut production system. Hence, the key methodological challenge is not just to predict areca nut production but to learn the spatiotemporal and nonlinear areca nut crop and climate relationship under heterogeneous agroclimatic conditions. Thus, the objective of the present investigation extends beyond areca nut forecasting. Hence, the present study aims at the advanced predictive understanding and mapping of areca nut crop–climate relationships across the heterogeneous agroclimatic regions of Central Karnataka.

Therefore, the existing single model or modeling methods for specific regions are difficult in dealing with the complex spatiotemporal variability and data imbalance of agricultural climate in the central Karnataka region, which provides a clear motivation for developing a hybrid integrated learning framework that can simultaneously capture spatial heterogeneity, time dependence and nonlinear characteristics. The developed hybrid novel approach addresses the combined challenges of spatiotemporal nonlinear crop–climate interactions. Since areca nut production formulates a regression problem with a continuous target variable, initially, the data distribution pattern is toughly examined in order to identify the uneven production pattern. Further, data stratification is performed to address this uneven production distribution, where the long-term production values were stratified as high, medium and low production categories. The SMOTE-ENN algorithm is applied to the stratified dataset to handle the data imbalance, which is a key challenge in the agroclimatic dataset. This algorithm was extensively applied as a preprocessing strategy and the final production observations after the pre-processing stage remained as a continuous value.

After the data pre-processing, a hybrid RF-BiLSTM model is developed to analyze the multi-regional crop and climate behavioral patterns. The ensemble random forest model is employed to capture the complex nonlinear and spatial interactions and the BiLSTM model analyzes the temporal variations that exist in the sequential dataset. Further, a weighted hybrid prediction-fusion mechanism is employed to analyze the relative contribution of both model outputs. Thus, the proposed hybrid framework represents an integrated approach which effectively captures the spatiotemporal nonlinear crop–climate associations in the climatically contrasted regions of Central Karnataka. The key contribution of this investigation is learning from two decades of multi-regional crop–climatic observations across the Central Karnataka regions and delivering a robust unified agroclimatic mapping system for areca nut production.

Another key contribution of the present study is areca nut genotype analysis, which is conducted along with the hybrid model development. Different type of areca nut genotypes may exhibit varying production rates under changing climatic conditions termed as genotypic resilience of the crop. Therefore, the present investigation evaluates the

genotypic resilience and stability by using yield variability, mean areca nut yield and regression deviation through a mathematical resilience score. As limited studies have analyzed the genotypic characteristics of the areca nut crop [35–37], the current analysis provides an additional insight into the areca nut production behavior.

## 2. Materials and Methods

### 2.1. Related Analysis

Precision agriculture has increasingly relied on data-driven modeling methods to understand and predict crop yield under complex agroclimatic variability. Early studies primarily applied statistical regression and time-series models to relate agroclimatic variables like rainfall, temperature, and soil moisture with crop productivity [38,39]. Despite these state-of-arts offering interpretability, they were confined in capturing nonlinear associations and spatiotemporal dependencies inherent in large-scale agricultural systems. To address such limitations, ML methods like support vector regression (SVR), random forests (RF), gradient boosting, and artificial neural networks (ANN) have been extensively applied for crop yield prediction [40–43]. Amongst the state-of-art methods, RF models have gained attention because of their robustness to noise and capability to model nonlinear interactions between agroclimatic variables [44]. Yet, traditional ML models usually treat observations as independent samples and fail to exploit temporal continuity in crop growth cycles. Unlike traditional ML methods, recurrent neural networks (RNNs), especially LSTM networks, have been primarily explored to capture temporal dependencies in weather and vegetation time-series data [45,46]. Bi-LSTM architectures, due to their ability to perform forward and backward learning of the temporal dependencies, further enhance prediction accuracy for seasonal crop yield modeling [47,48]. Despite robustness, LSTM-based models struggle to identify the most influential temporal features when long sequences and multivariate inputs are involved. This, as a result, confines their scalability over large geographics, which can have the different agroclimatic conditions impacting on crop yield and genotypic resilience.

Recent investigations have developed a hybrid model using deep learning frameworks which integrates convolutional, recurrent, and attention mechanisms to handle the spatial and temporal interactions present in agricultural prediction [49,50]. However, many of these approaches primarily rely on satellite imagery or weather index datasets. With respect to the major challenges like data imbalance and spatial heterogeneity, skewed production distributions can also impact model generalization [51,52]. The resampling methods like SMOTE have shown their capability in reducing bias toward dominant production ranges [53]. Nevertheless, the integration of such resampling techniques with the hybrid deep learning framework for the regional crop–climate analysis is still limited. Most of the areca nut studies have focused on disease incidence datasets, while less attention is given to crop–climatic mapping. Hence, to address this gap, the present study develops a hybrid RF attention-BiLSTM framework, which integrates ensemble learning along with temporal attention. With SMOTE-ENN preprocessing, it mitigates data imbalance and improves prediction across different agroclimatic regions.

### 2.2. Geographical Study Region and Data Acquisition

As stated in the previous sections, this research primarily focuses on emphasizing agroclimatic mapping for areca nut crop in central Karnataka, India. More specifically this research considers the districts including Dakshina Kannada, Shivamogga, Davangere, Tumkur, Chikkamagaluru, Uttara Kannada, Chitradurga, and Udupi. In reference to the aforesaid study regions, the agroclimatic variables, including the soil moisture, humidity, temperature, cultivation area and the geographical coordinates such as longitude and

latitude, were obtained from the different geographics. Figure 1 represents the targeted study region of the present analysis.

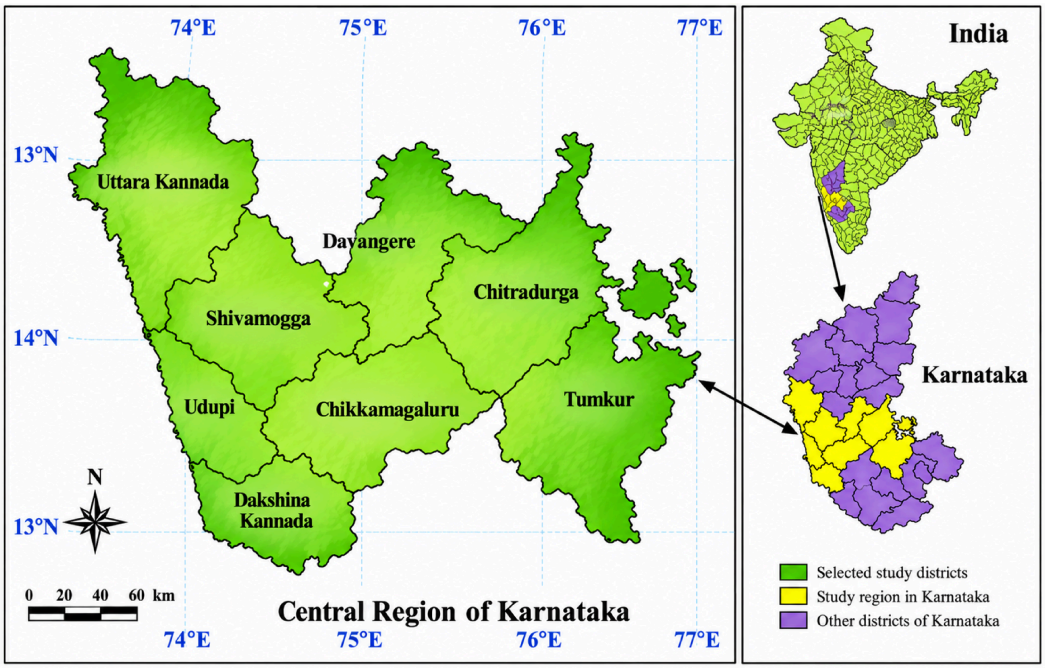


**Figure 1.** Geographical location of the selected study districts. The map was prepared using QGIS (version 3.4) and graphically formatted using Inkscape (version 1.4).

These regions were carefully selected on the basis of their areca nut yield status over the past 20 years. In addition, these geographies were selected due to their consistent and decisive contribution over the period of time. It provides a reliable basis for assessing long-term areca nut yield patterns. The dataset embodies the years from 2003 to 2022, depicting annual data collected for each selected district, emphasizing on the aggregate areca nut yield in tonnes. The average areca nut yield over the two-decade period is shown in Table 1. On the basis of the 20-year average yield data (Table 1), the highest areca nut production (482,774 tonnes) and cultivation area (57,786 hectares) are observed in the Shivamogga district. In this manner, the minimum yield rate (105,895 tonnes) and cultivation area (10,590 hectares) were observed in the Udupi district. The yield rate is defined in tonnes, while the cultivation area is defined in hectares.

**Table 1.** Average annual areca nut production and cultivation area across study regions.

| Region | Production | Cultivation Area |
|---|---|---|
| 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 |

### *2.3. Data Source*

The study used a 20-year dataset spanning 2003–2022 for the selected districts, which included agroclimatic and geospatial data. The Directorate of Economics and Statistics, Department of Agriculture and Farmers Welfare, Government of India, provided district-level

annual statistics on areca nut cultivation area and production [54]. These data include detailed information on area productivity derived from the relevant regions. Annual/monthly data on major external agroclimatic parameters, such as temperature, relative humidity, and soil moisture, were collected from the NASA POWER Project Data Access Viewer (DAV) in the Agroclimatology category [55]. The geographical coordinates (latitude and longitude) of the selected districts were collected using the QField application connected with QGIS [56], which allowed for spatial representation and analysis of the research locations. Pertinent variables of the investigation are schematically represented in Table 2.

**Table 2.** Agroclimatic and geographical variables of the study.

| Variable Name | Temporal Resolution | Spatial Resolution |
|---|---|---|
| Temperature (°C) | Monthly and Annual | District-representative location |
| Humidity (%) | Monthly and Annual | District-representative location |
| Soil Moisture (%) | Monthly and Annual | District-representative location |
| Area under cultivation (Hectares) | Annual | District-specific |
| Crop yield (Tonnes) | Annual | District-specific |
| Latitude | Static | District-specific |
| Longitude | Static | District-specific |

### 2.4. Crop Genotype Data

Different areca nut crop genotype information for the selected study regions is collected from published studies [55,56]. Both the genotype and their associated scientific names are illustrated in Tables 3 and 4. These details are further employed for genotype resilience analysis.

**Table 3.** Areca nut genotype details.

| Selected Study Region | Areca nut Genotype |
|---|---|
| Dakshina Kannada | Mangala, South Kanara Local, Ratnagiri, Saigon (swarnamangala), Sumangala, VTLAH, VTLAH–1, Sreemangala, vital-local. |
| Shivamogga | Thirthahalli local, Mangala, Sumangala, Sreemangala and Mohit Nagar, Ratnagiri |
| Davangere | Chengiri local, Sagar local, |
| Tumkur | Hirehalli Local |
| Chikkamagluru | Thirthahalli local, Mangala, Sumangala, Sree Mangala and Mohit Nagar, Ratnagiri |
| Chitradurga | Sreemangala |
| Udupi | Mangala, South Kanara Local, Sumangala, Sreemangala |
| Uttara Kannada | Sagar, Sumangala, Sirisi local, Swarna Mangala |

**Table 4.** Areca nut genotype's scientific nomenclature.

| Common Variant Name | Scientific Representation |
|---|---|
| Mangala | *Areca catechu* L. cv. Mangala |
| Sumangala | *Areca catechu* L. cv. Sumangala |
| Sreemangala | *Areca catechu* L. cv. Sreemangala |
| Mohitnagar | *Areca catechu* L. cv. Mohitnagar |
| Kahikuchi | *Areca catechu* L. cv. Kahikuchi |
| Thirthahalli Local | *Areca catechu* L. (local cultivar) |

**Table 4.** *Cont.*

| Common Variant Name | Scientific Representation |
|---|---|
| Sirsi Local | *Areca catechu* L. (local cultivar) |
| Vittal Areca | *Areca catechu* L. cv. Vittal |
| Hirehalli Dwarf | *Areca catechu* L. cv. Hirehalli Dwarf |
| South Kanara Local | *Areca catechu* L. (South Kanara ecotype) |

### *2.5. Data Pre-Processing*

1. Missing values, normalization, and data preparation

Let $D = \{(x_i)\}_{i=1}^{N}$ be the original agroclimatic areca nut dataset with $x_i \in R^d$ as climatic, soil and geographical features. Once collecting the above discussed data, the proposed method executes pre-processing tasks. Particularly, the proposed method performs the following pre-processing methods:

- Weighted neighborhood-based estimator for data imputation (missing values imputation);
- Outlier analysis;
- z-score normalization.

The agroclimatic dataset represents sequential data collected on the different time instances; however, being a multi-parametric environment, there can be inconsistency amongst the data elements. Training any ML or deep learning models over aforesaid missing elements or outliers can impact learning efficiency and accuracy. It can impact reliability of the proposed areca nut agroclimatic mapping. To address this, the proposed model performs pre-processing tasks to address missing values and outliers. Additionally, the proposed method executes z-score normalization over the large sequential data to alleviate overfitting problems. A brief of the pre-processing tasks is given in the subsequent sections.

To address outliers in the considered temporal (sequential) datasets, at first missing values were imputed by means of neighborhood-based estimator. This method is extensively applied as a data-driven method to alleviate missing values in sequential inputs. Functionally, it exploits the local similarity structure of the dataset to perform data imputation. Unlike traditional mean or median-based global feature-driven methods, the developed method inputs the missing values based on the similar samples, thereby retaining local patterns and nonlinear relationships. Consider $x_i \in R^d$ as the agroclimatic and genotypic dataset and allied data instance with missing attributes. Thus, the proposed neighborhood-based estimator searches a set of k-NN N_k($x_i$) from the complete data by means of Euclidian distance method. In other words, it estimates k-NN$N_k(x_i)$ elements by performing Euclidian distance estimation on $x_i$. For a missing feature $\hat{x}_{ij}$, it measures the neighborhood-based estimated value $x_{ij}$, as per Equation (1)

$$\hat{x}_{ij} = \frac{1}{k}\sum_{x_m \in N_k(x_i)} x_{mj} \tag{1}$$

A distance-weighted formulation as defined in (2) can also be applied to assign higher influence to the closer neighbor.

$$\hat{x}_{ij} = \frac{\sum_{x_m \in N_k(x_i)} w_m x_{mj}}{\sum_{x_m \in N_k(x_i)} w_m} \tag{2}$$

where

$$w_m = \frac{1}{dist(x_i, x_m)} \tag{3}$$

Once data imputation is completed, outlier analysis is performed. More specifically, the proposed method applies inter-quartile-range (IQR)-based outlier detection and removal. The proposed method applied Equation (4) to perform outlier removal.

$$P(x) = \begin{cases} x, & if\ Q_1 - 1.5\ IQR \leq x \leq Q_3 + 1.5 \times IQR \\ Reject & Otherwise \end{cases} \quad (4)$$

Here, $x$ states the data instance within the feature vector present in $n-$dimensional data space, while $x \in R^n$. The quartile $Q_1$, $Q_3$ states the IQR measurements for the first quartile and third quartile, respectively. In this work, the values of (1) $Q_1$, $Q_3$ and $IQR$are the subset of $R^n$. Thus, based on the above formulation those data instances which do not fall into the stated criteria (i.e., $Q_1 - 1.5\ IQR \leq x \leq Q_3 + 1.5 \times IQR$) are dropped. Considering non-linear input patterns and possible over-fitting, the proposed method executes z-score normalization, as defined in (5).

$$x_{ij}^{Norm} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (5)$$

In (5), $\mu_j$ and $\sigma_j$ refer to the mean and standard deviation of the jth feature, correspondingly. The normalized temporal climatic variables were segmented into fixed-length sequences to preserve seasonal and phenological patterns. The normalized dataset was later processed for data stratification and SMOTE-ENN-driven resampling technique used to alleviate the data-imbalance problem. A brief of the same method is given in the subsequent section.

2. Data Stratification

In the present study, the areca nut production is a continuous regression target variable. This production data markedly exhibited uneven and right positive skewed data distributions (observed) as shown in Figure 2. Very few samples were observed at the lower and higher ranges. This type of data imbalance will bias the entire learning process towards the dominant areca nut production pattern and increase the risk of poor generalization and overfitting to certain observed ranges. Hence, to address this data imbalance challenge, the continuous areca nut production values were stratified into low, high and medium categories using predefined thresholds. These categories were further used to identify and balance the production regimes. This method provided a strong basis to implement SMOTE-ENN resampling algorithm.

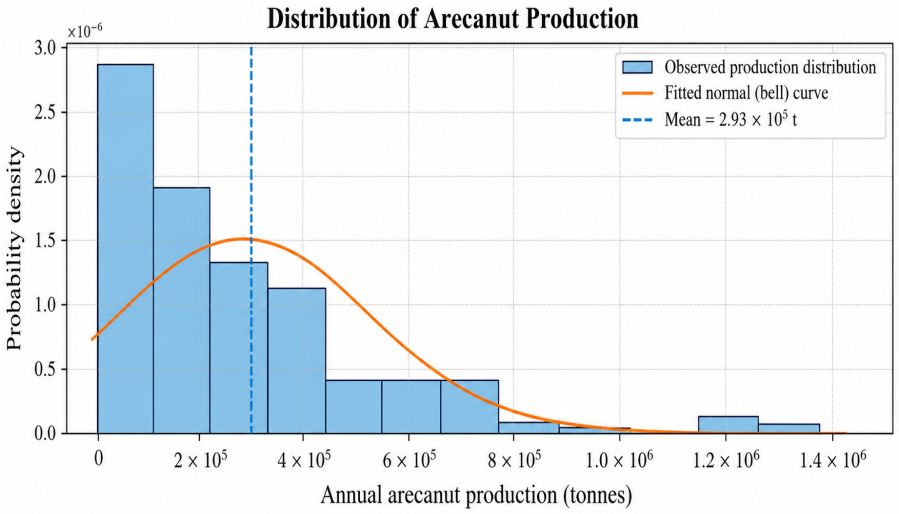


**Figure 2.** Distribution of areca nut production.

3. SMOTE-ENN Resampling

Data imbalance is one of the fundamental problems in multi-regional agroclimatic datasets that can impact overall learning and prediction results. Training any ML or deep learning model over imbalanced datasets can skew overall learning outcomes such as the false positive or false negative, and, therefore, there is an inevitable need for data resampling technique. Though, in the past, different traditional resampling techniques such as the random sampling and up-sampling are applied, their inherent probability of iterative hotspot creation cannot be ruled out. Despite superiority of SMOTE over traditional resampling methods, the ambiguous data distribution over large distributed data space can impact their efficacy or reliability.

Hence, this research applies an improved SMOTE -ENN method on the stratified dataset to alleviate the data-imbalance problem. Unlike traditional resampling methods, the proposed method generates synthetic samples depicting highly correlated data instances, without affecting the original data distribution. It applied minority samples as input to generate the synthetic data instances or samples, which was later processed with the k-NN classifier. In this work, Euclidean distance-based k-NN algorithm was applied to generate the most probable samples to be combined with the original (pre-processed sample). It has to be noted that the SMOTE-ENN method is subsequently applied within data stratified space and the synthetic observation is not assigned the categorical stratum as its final target. The low-, medium- and high-production categories were used only as temporary strata for data imbalance handling. In this method, edited nearest neighbors (ENN) method removes ambiguous and noisy samples by eliminating instances which differ from the majority of their k-NN neighbors. Mathematically it is represented as,

$$x_{new} = x_i + \lambda(x_{nn} - x_i), \quad \text{here, } \lambda \in (0,1) \tag{6}$$

Once the resampling process is completed, the high, low and medium labels are completely discarded and synthetic samples were generated within underrepresented strata, and their continuous production targets were obtained through regression-oriented interpolation, as shown in Equation (7)

$$Y_{syn} = Y_i + \lambda(Y_{nn} - Y_i) \tag{7}$$

*2.6. Hybrid Model Development*

Unlike traditional standalone feature-driven agroclimatic mapping models or crop-yield prediction techniques, this research proposes a hybrid approach for crop–climate mapping. Here, RF ensemble learning model is applied to learn temporal features, while attention-based Bi-LSTM helps to learn parametric inter-dependencies over large sequential (temporal) inputs. A brief of these feature extraction methods is given as follows.

1. Random Forest Model for Spatial Nonlinearity

In the present study, the ensemble random forest model is employed to capture the spatial nonlinear interactions that exist between crop and climatic factors. Unlike a single decision tree, the current approach generates an ensemble of regression trees using bootstrap with replacement sampling method. The incorporation of the geographical coordinates, latitude and longitude helps the ensemble to generate location-specific decision trees using randomly selected feature subsets. Hence, this will reduce the variance as well as overfitting tendency. The samples from the original dataset which were not selected for tree bootstrap are further moved to OOB samples and used for further internal cross-validation. Once all trees are generated, the regression ensemble average method is used to

find the average of all tree predictions. For the targeted agroclimatic mapping tasks, 80% of samples were used for training, while 20% of samples were applied as the out-of-bag samples for further inner cross-validation to enhance results. The overall working of RF is presented in Figure 3.

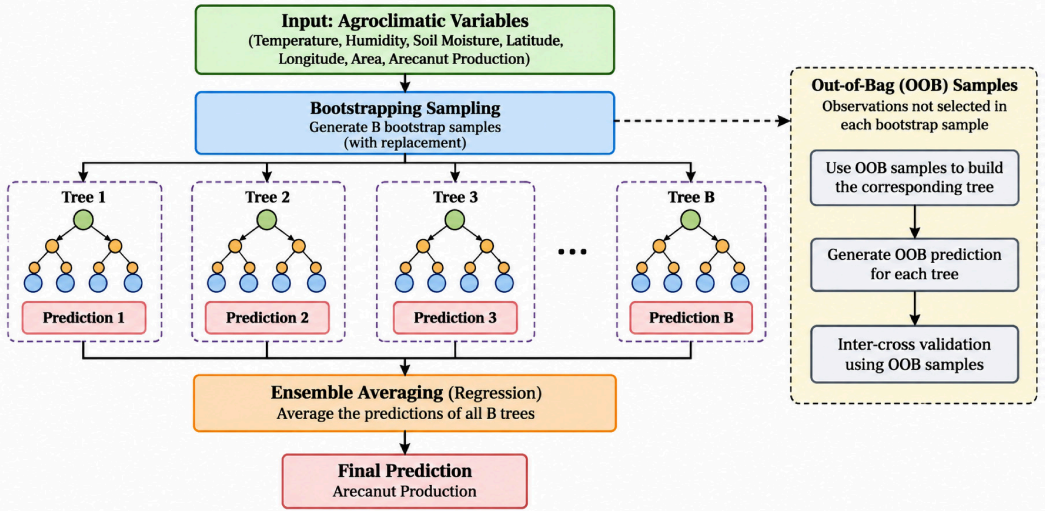


**Figure 3.** Working of Random Forest model. The figure was graphically prepared using Inkscape (version 1.4).

Given $T$ decision trees, the RF prediction is defined as per Equation (8).

$$f_{RF}(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \tag{8}$$

In (9), $h_t$ represents the output of the $t$-th decision tree. Thus, the use of RF ensemble method estimated the implicit feature importance, which helped for the strong generalization for static feature representations. In this manner, it served high-level feature representations $f_{RF}(x)$ for ensemble learning. Consider $X \in R^{N\times d}$ as the feature matrix and $y$ as the yield vector. Then, the RF outputs the following (9).

$$\hat{y}_{RF} = f_{RF}(X) \tag{9}$$

These predictions serve both as baseline estimates as well as a high-level spatial descriptor, which is subsequently applied for deep learning stages.

2. Attention-Based BiLSTM Model for Temporal Variations

To retrieve temporal dependencies in the climatic sequences, the proposed method applies attention-based BiLSTM model. To capture temporal dependencies across climatic seasons, the monthly agroclimatic data (January to December of 20 years) are structured into sequences and fed into a BiLSTM network. The BiLSTM processes sequences in both forward and backward directions, enabling it to model long-term temporal correlations. In the present study, the attention operates across different time stamps (rather than feature dimensions).

Noticeably, the traditional deep learning models, such as CNNs and DNNs, are often criticized due to their inability to address inherent adversaries like gradient vanishing, gradient explosion, lack of contextual information, long-term reliance, etc., which keeps mounting over increasing convolution layers or feature dimension sizes. Moreover, exceedingly high reliance onto the local hierarchical features confines their suitability towards our targeted temporal sequential inputs, where they fail to extract long-term dependencies

over the different agroclimatic parameters. To address this, this research applied a robust attention-based BiLSTM model that retrieves temporal dependencies for better learning and prediction. Figure 4 represents the architecture of proposed attention-BiLSTM model.

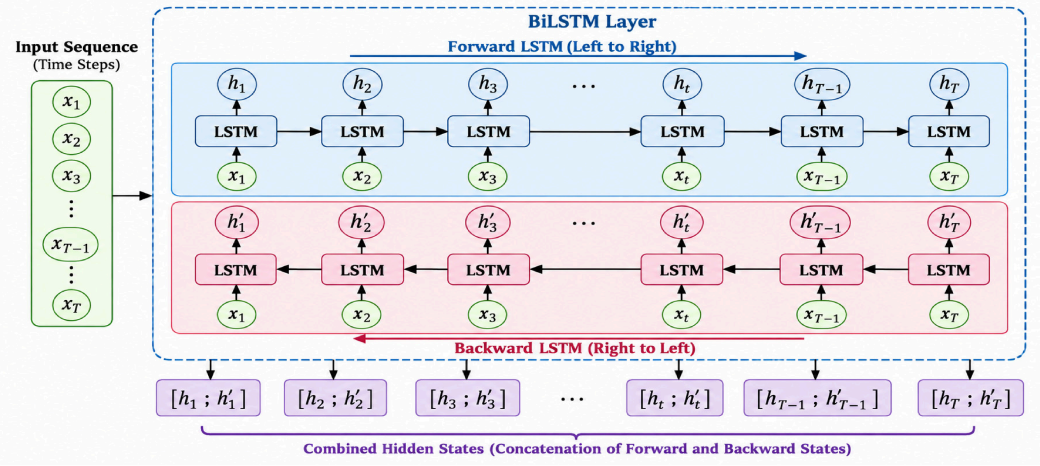


**Figure 4.** Working of attention-BiLSTM model. The figure was graphically prepared using Inkscape (version 1.4).

In the present study, for each year, the monthly climatic observations are represented as,

$$X = [x_1, x_{2.....}x_{12}] \tag{10}$$

The BiLSTM combines forward and backward hidden states as,

$$h_{t,Bi-LSTM} = \left[\overrightarrow{h}_t, \overleftarrow{h}_t\right] \tag{11}$$

Further, the attention weight for each month is calculated and normalized using Equations (12) and (13). The final temporal representation is given by Equation (14).

$$u_t = tanh(W_u h_t + b_w) \tag{12}$$

$$a_t = \frac{exp(u_t^T u_w)}{\sum_t exp(u_t^T u_w)} \tag{13}$$

$$s = \sum_t a_t h_t \tag{14}$$

where $a_t$ represents the importance assigned to each monthly time step. Thus, this developed model captures the bidirectional temporal dependencies and analyzes the most affected climatic periods for areca nut production.

3. Hybrid Prediction Fusion

Once obtaining outputs from the RF and self-attention-based Bi-LSTM, weighted fusion method was applied to perform prediction. More specifically, we applied the formulation, given in (15), to perform weighted prediction fusion, which is later processed for backpropagation-based training for prediction.

$$\hat{y} = w_1 f_{RF}(x) + w_2 f_{BiLSTM-Attention}(x) \tag{15}$$

conditioned at w_1 + w_2 = 1, where w_1, w_2 $\geq$ 0. In this work, the optimal weights are measured by applying validation-based optimization to minimize prediction error.

Due to the availability of the limited agroclimatic dataset of areca nut, the Leave-One-Region-Out Cross-Validation (LORO-CV) validation is employed to assess the model's generalization.

### 2.7. Genotypic Resilience and Stability Analysis

In the present analysis, the genotypic resilience and stability analysis is performed by using a composite score obtained from mean yield, coefficient of variation (CVi), yield stability index (YSi), and deviation from regression. Mathematically it is represented in Equation (16)

$$\text{Genotype Resilience} = \frac{\left(\frac{y_{si} \times \text{Mean Yield}}{50}\right)}{(CV_i + 1) - \text{abs}(s^2_{di})} \quad (16)$$

The resilience score, calculated as (YSi × Mean Yield/50)/(CVi + 1) − abs($s^2$_di), serves as a comprehensive, single-value metric to evaluate the overall resilience and reliability of areca nut genotypes under variable environmental conditions, particularly relevant for semi-arid central Karnataka regions, where drought spells and irrigation dependency pose significant challenges. By multiplying the stability index YSi (which rewards consistent performance across environments) with the average mean yield and normalizing it against a typical yield threshold of 50 g/fruit, the formula emphasizes genotypes that deliver both high productivity and dependable stability.

Dividing this product by (CVi + 1) introduces a smooth penalty for yield variability; lower CVi values (indicating greater consistency) result in higher scores, while high variability substantially reduces the contribution. Finally, subtracting the absolute value of the deviation from regression abs($s^2$_di)—derived from the Eberhart–Russell model—further penalizes genotypes exhibiting large, unpredictable fluctuations that cannot be explained by environmental trends. This balanced approach ensures top-ranked genotypes.

### 2.8. Model Evaluation

Further, the developed hybrid approach is evaluated using the following statistic metrics as shown below:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_I - P_I)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \quad (17)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_I - O_i)^2} \quad (18)$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|O_I - P_I| \quad (19)$$

$$\text{MAPE}(\%) = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{O_I - P_I}{O_I}\right| \quad (20)$$

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n[\sum x^2 - (\sum x)^2] - n\sum y^2 - (\sum y)^2]}} \quad (21)$$

### 2.9. Data Splitting and Execution Environment

The proposed framework was implemented in MATLAB 2022b software tool using the required ML and deep learning toolboxes. The complete pre-processed dataset was portioned into 80% training and 20% testing subsets. The SMOTE-ENN was applied to the training dataset and, further, the hybrid model was simulated for the pre-processed agroclimatic data. For the RF model, the tree-specific hyperparameters, including the number of trees, maximum tree depth, and minimum leaf size, were automatically tuned

by the MATLAB implementation according to the training dataset characteristics while maintaining the robustness and generalization capability of the ensemble model. The clear and informative diagrams for the obtained outputs (from MATLAB) are further generated using Python Matplotlib (version 3.8.2) library.

For BiLSTM, network hyperparameters such as the hidden layer setup, learning rate, batch size, dropout values and the number of training epochs were chosen through iterative trials for stable convergence, efficient learning of temporal features and better generalization. The final model was trained with a starting learning rate of 0.0001, a batch size of 32 and 200 training epochs, which provided a reasonable trade-off between training stability and prediction accuracy for the sequential monthly climate data.

During the preparation of this manuscript, the authors used QGIS (version 3.4) for the preparation of the geographical study-area map and Inkscape (version 1.4) for graphical formatting of the map and preparation of schematic figures. These software tools were operated manually for figure preparation and graphical editing. Adobe Firefly (Firefly Image 4) was used only to improve the visual presentation of the Graphical Abstract. The prompt used was: "Enhance the visual quality, clarity, and layout of this scientific graphical abstract without altering the scientific content, data, text, or results". The authors have reviewed and edited the output and take full responsibility for the content of this publication.

## 3. Results and Discussion

### 3.1. Results of Resampling Approach

Realizing a probable data imbalance problem in at-hand agroclimatic mapping, in this work, SMOTE-ENN was applied to address the data imbalance problem. In this reference, the simulations were done with and without SMOTE-ENN resampling, and corresponding performances were examined. Table 5 presents relative performance outputs.

**Table 5.** Results of resampling method.

| Dataset | $R^2$ | RMSE | MAE | MAPE (%) | r |
|---|---|---|---|---|---|
| Hybrid Model without SMOTE-ENN | 0.86 | 0.48 | 0.37 | 11.27 | 0.92 |
| Hybrid Model with SMOTE-ENN | 0.91 | 0.39 | 0.30 | 8.11 | 0.95 |

The outcome from Table 5 demonstrates the effectiveness of SMOTE-ENN preprocessing in improving the prediction accuracy. Without resampling, the model achieved an $R^2$ of 0.86 and an r of 0.92, with relatively higher error rates in RMSE, MAE, and MAPE. With SMOTE-ENN preprocessing, the value of $R^2$ increased to 0.91 by decreasing the error metrics. Hence, these results indicate the capability of SMOTE-ENN in addressing the data imbalance and noise within unevenly distributed production ranges. Consistent improvements reported in the previous agricultural investigations [57,58] further highlight the value of the SMOTE-ENN method in addressing the data imbalance issues through balanced sample distribution, thereby enhancing the predictive performance in agroclimatic datasets.

### 3.2. Predictive Performance Analysis of the Hybrid Model

To evaluate the predictive behavior of the developed hybrid model, the model predicted values were compared with the actual areca nut production data using scatter plot and frequency distribution. The close agreement between the actual and the predicted production values, which is illustrated in Figure 5, demonstrated the strong predictive capability of the developed model. Through the distribution plot presented in Figure 6, it is observed that most of the observations are concentrated in a lower region, with fewer val-

ues at a higher production level. In both the scatter and the distribution plot, the predicted production values were in close agreement with the observed areca nut production values. Therefore, these findings support the quantitative performance result by demonstrating the proposed hybrid model's capability.

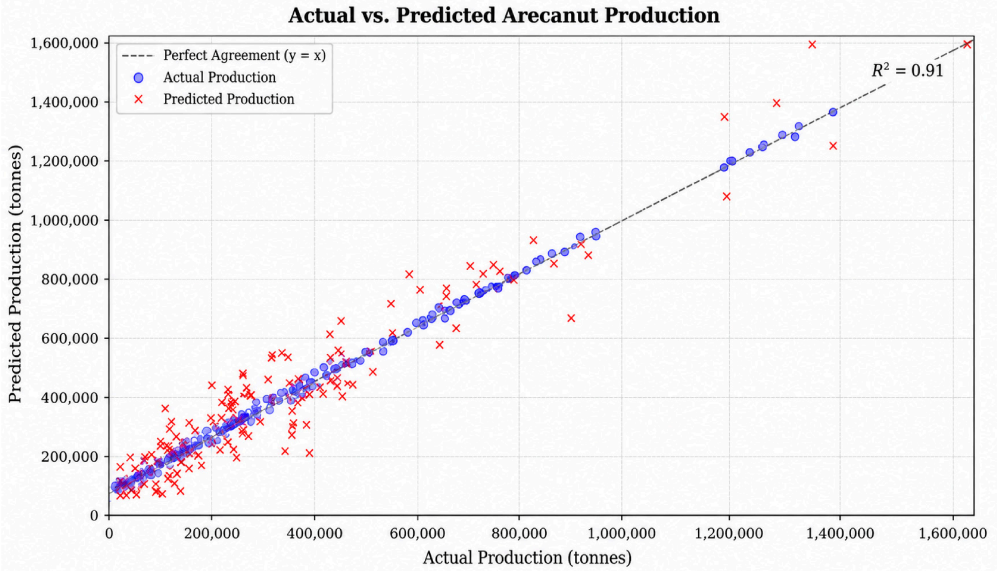


**Figure 5.** Predictive performance analysis result.

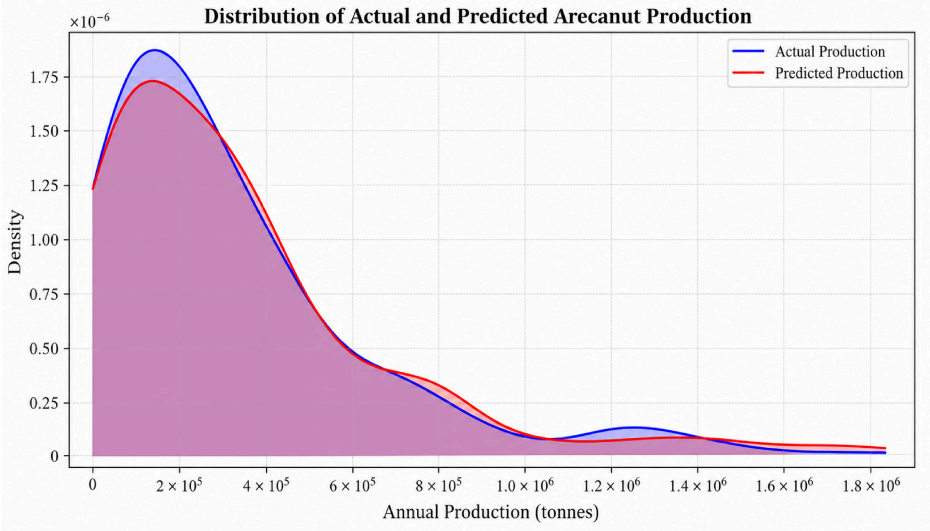


**Figure 6.** Distribution of actual vs. predicted production.

Further, to analyze the proposed model's performance across the selected study regions of Central Karnataka, the observed and the predicted areca nut production values were compared across the study regions. The result of this analysis is presented in Figure 7. A close alignment between actual and predicted values was observed in Shivamogga, Dakshina Kannada and Tumkur regions, where Udupi and Uttara Kannada indicated greater differences. Hence, these results demonstrate the spatial variability and highlight the importance of region-wise analysis.

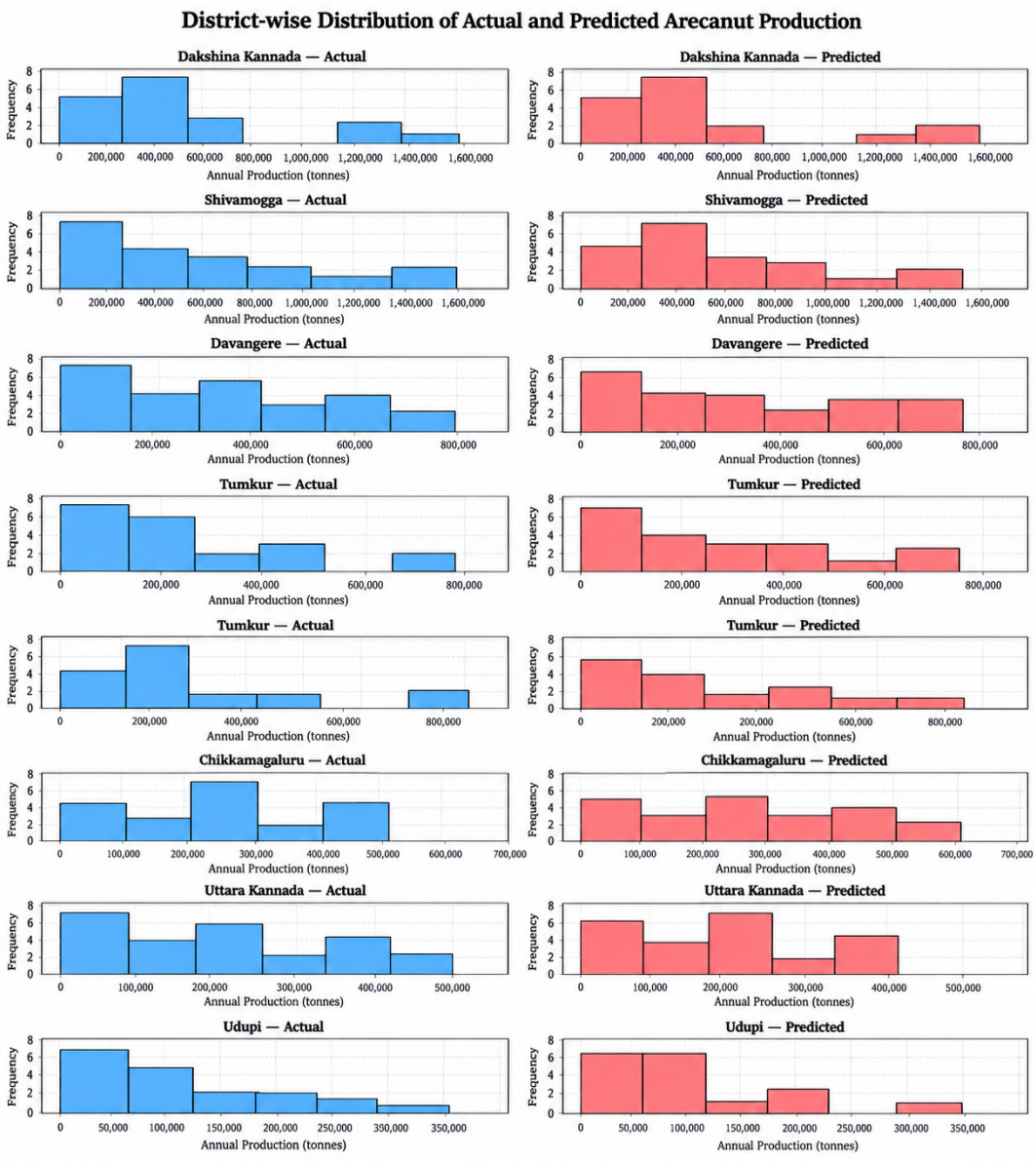


**Figure 7.** Region-wise distribution of actual vs. predicted production.

### *3.3. Relative Performance Analysis of the Hybrid Model Using Baseline Approaches*

To examine the relative performance of the developed hybrid model, different base-line models, including support vector regression (SVR), multi-layer perceptron (MLP), RF, standalone BiLSTM, attention transformer and the gradient boosting model XGBoost and CatBoost, were developed. The relative performance of these models is illustrated in Table 6.

**Table 6.** Comparisoin with base-line models.

| Model | $R^2$ | RMSE | MAE | MAPE (%) | r |
|---|---|---|---|---|---|
| SVR | 0.78 | 0.61 | 0.48 | 14.92 | 0.88 |
| MLP | 0.80 | 0.57 | 0.45 | 13.68 | 0.89 |
| Random Forest (RF) | 0.83 | 0.53 | 0.41 | 12.21 | 0.91 |
| BiLSTM | 0.86 | 0.48 | 0.38 | 10.96 | 0.92 |
| Attention Transformer | 0.88 | 0.45 | 0.34 | 9.84 | 0.94 |
| XGBoost | 0.78 | 0.51 | 0.39 | 10.8 | 0.88 |
| CatBoost | 0.70 | 0.57 | 0.44 | 14.2 | 0.83 |
| Proposed Hybrid model | 0.91 | 0.39 | 0.30 | 8.11 | 0.95 |

The quantitative comparison presented in the table highlights a clear and consistent performance progression across the evaluated models, reflecting their increasing capability to capture complex agroclimatic–yield relationships. Among the conventional learning approaches, the SVR, XGBoost and CatBoost models exhibit the lowest predictive accuracy, with an $R^2$ value below 0.80 and with higher error metric. This indicates the behavior of gradient boosting and the kernel-based models to handle the spatiotemporal dependencies. Further, MLP showed a moderate improvement over SVR, achieving reduced error values and a higher correlation coefficient, suggesting that shallow neural architectures can partially learn nonlinear patterns but remain constrained by their feedforward structure.

The ensemble-based random forest (RF) model further enhances predictive reliability, as evidenced by the increase in $R^2$ to 0.832 and noticeable reductions in RMSE and MAE. However, RF remains inherently static and does not explicitly model temporal dependencies. In contrast, the Bi-LSTM model achieves a substantial gain in performance, attaining an $R^2$ of 0.861 and lower percentage errors, which underscores the importance of bidirectional temporal learning in capturing yield variations. By dynamically emphasizing relevant temporal segments, the transformer improves representation learning under heterogeneous climatic conditions. It was clearly observed that the developed hybrid approach performed a superior performance with $R^2$ of 0.91 and with lower error rates (RMSE of 0.90, MAE of 0.30 and MAPE of 8.11%) and with correlation coefficient t of 0.95 by successfully capturing the spatiotemporal nonlinear relationship between areca nut crop and climatic factors.

The proposed hybrid model of the study demonstrated superior performance compared with the baseline approaches by indicating the advantage of combined learning mechanisms that enable more effective representations of complex agroclimatic production patterns. This improvement can be attributed to the combined ability of RF to model complex nonlinear agroclimatic interactions and the attention-guided BiLSTM to effectively capture the temporal dependencies in climatic patterns. The similar improvements of hybrid frameworks reported in the previous agricultural prediction investigations [59,60], where integrated models outperformed standalone approaches, further support the potential of the hybrid framework to capture complex crop and climate associations, thereby improving accuracy.

### *3.4. Ablation Analysis of the Proposed RF–Attention BiLSTM Model*

To quantify the contributions of each core component of the developed hybrid RF attention-based BiLSTM model, an ablation analysis was systematically performed. This analysis basically removes one part of the developed approach and carefully examines the outcome. The existing studies through ablation experiments have demonstrated that excluding temporal or attention mechanisms leads to measurable performance degradation and highlights their critical roles in improving crop-yield prediction [61,62]. In the present analysis, initially, the RF model was removed and the expected outcomes were noted. Similarly, in the next iteration, the attention model is removed and, finally, the BiLSTM model was removed. The combinations of the executed models and their final outcomes are illustrated in Table 7.

The results of the analysis indicate that the RF, attention and BiLSTM make a meaningful contribution. The observed performance degradation after removing an individual component confirms their contributions in capturing nonlinear, spatiotemporal associations. The superior performance of the proposed hybrid model once again confirms the advantage of integrating ensemble learning with attention-based bidirectional temporal modeling.

**Table 7.** Outcome of ablation analysis.

| Model Configuration (Without) | Model Configuration (With) | $R^2$ | RMSE | MAE | MAPE (%) |
|---|---|---|---|---|---|
| Full RF, Attention, BiLSTM | Full RF, Attention, BiLSTM | 0.91 | 0.39 | 0.30 | 8.11 |
| Without RF | With Attention, BiLSTM | 0.87 | 0.46 | 0.35 | 9.8 |
| Without Attention | With RF, BiLSTM | 0.88 | 0.44 | 0.34 | 9.3 |
| Without BiLSTM | With RF, Attention | 0.88 | 0.49 | 0.30 | 10.7 |

### *3.5. Performance Stability Analysis of the Comparative Models*

In order to assess the consistency of the evaluated models, the individual developed model was independently executed more than 10 times. The obtained outcome across the repeated runs was summarized using the mean and standard deviation (Mean ± SD) for $R^2$, RMSE, MAE, MAPE, and correlation coefficient ($r$). This repeated run evaluation provides an effective assessment of the model performance by carefully analyzing the variations across multiple executions. This type of stability/repeated-run analysis have demonstrated a consistent and reliable performance in previous wheat and soybean yield prediction [63,64] studies by supporting its suitability for evaluating model stability. Hence, the present analysis employes this method to evaluate the model robustness and consistency across repeated training runs.

The result of the performance stability analysis is presented in Table 8. It was clear evidence from the stability analysis that the proposed hybrid model achieved the best overall mean performance by maintaining relatively low variability. It obtained the highest $R^2$ and r along with the lowest RMSE, MAE, and MAPE. The observed small standard deviation across the error metrics further highlights the consistent predictive behavior of the model when compared to the baseline approaches. Thus, this analysis clearly demonstrates the benefit of integrating nonlinear feature learning along with the bidirectional temporal learning.

**Table 8.** Result of the stability analysis.

| Model | $R^2$ (Mean ± SD) | RMSE (Mean ± SD) | MAE (Mean ± SD) | MAPE (%) (Mean ± SD) | r (Mean ± SD) |
|---|---|---|---|---|---|
| SVR | 0.774 ± 0.021 | 0.618 ± 0.029 | 0.486 ± 0.024 | 15.08 ± 0.72 | 0.876 ± 0.015 |
| MLP | 0.796 ± 0.019 | 0.578 ± 0.025 | 0.457 ± 0.021 | 13.82 ± 0.64 | 0.887 ± 0.014 |
| RF | 0.826 ± 0.017 | 0.537 ± 0.022 | 0.416 ± 0.018 | 12.34 ± 0.56 | 0.907 ± 0.012 |
| BiLSTM | 0.856 ± 0.015 | 0.486 ± 0.019 | 0.385 ± 0.016 | 11.08 ± 0.49 | 0.917 ± 0.011 |
| Attention Transformer | 0.876 ± 0.013 | 0.456 ± 0.017 | 0.346 ± 0.014 | 9.96 ± 0.43 | 0.936 ± 0.009 |
| XGBoost | 0.783 ± 0.020 | 0.516 ± 0.024 | 0.396 ± 0.019 | 10.92 ± 0.52 | 0.884 ± 0.016 |
| CatBoost | 0.706 ± 0.026 | 0.579 ± 0.029 | 0.447 ± 0.023 | 14.37 ± 0.68 | 0.834 ± 0.019 |
| Proposed model | 0.906 ± 0.010 | 0.394 ± 0.012 | 0.304 ± 0.011 | 8.18 ± 0.32 | 0.947 ± 0.008 |

### *3.6. District-Wise Performance Analysis of the Hybrid Model*

The district-wise performance analysis reveals that the proposed hybrid model illustrated in Table 9 maintains strong and consistent predictive capability across diverse agroclimatic regions of central Karnataka. Coastal and Malnad districts such as Udupi, Dakshina Kannada, and Chikkamagaluru exhibit the highest prediction accuracy, with $R^2$ values exceeding 0.92 and comparatively lower RMSE and MAE. This indicates that the model is able to effectively learn stable yield–climate relationships in regions character-

ized by relatively uniform rainfall patterns, higher humidity, and favorable soil moisture conditions, which reduce uncertainty in yield behavior.

**Table 9.** District-wise performance analysis.

| 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 |
| Uttara Kannada | 0.902 | 0.380 | 0.294 |

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, irregular rainfall distribution, and higher susceptibility to moisture stress, which introduce nonlinearity and uncertainty into yield dynamics. Despite these challenges, the model still achieves $R^2$ values close to or above 0.89, exhibiting robust generalization even under heterogeneous and stress-prone agroclimatic conditions. Notably, Shivamogga presents a balanced performance profile, reflecting its transitional agroclimatic nature between coastal and interior regions. 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. Overall, this district-level evaluation confirms the scalability and adaptability of the proposed model, highlighting its suitability for region-specific yield prediction.

### *3.7. Spatial Generalization Using Leave-One-Region-Out Cross-Validation Approach*

In order to evaluate the ability of the developed approach on unseen regions, the leave-one-region-out (LORO-CV) cross-validation was employed. In each analysis (run), one district or region was completely considered as a 'test' region and the data from the other remaining seven regions were used to 'train' the model. Further, the model trained on the other seven regions is used to predict the production of the excluded region for each year from 2003 to 2022. The entire process was repeated by changing each region each time. Thus, the region which was excluded during the training provides a reliable spatial generalization. The LORO-CV approach has been employed in previous agricultural prediction studies (almond and oil palm yield prediction) [65,66] in order to analyze the model generalization across different production environments. In the present investigation the LORO-CV approach is employed to analyze the hybrid model's ability for areca nut production prediction in unseen regions. The result of this validation is presented in Table 10.

The region-wise test outcomes clearly demonstrated the model's predictive capability by achieving $R^2$ values from 0.802 to 0.884. Districts like Davangere, Tumkur and Chitradurga achieved lower error rates, whereas the Uttara Kannada and Chikkamagaluru regions obtained comparatively higher error rates. Moderate error rates were seen in Dakshina Kannada and Shivamogga regions. Hence, these variations clearly demonstrate the strong explanatory capability of the model even when predictions were made for districts completely excluded from training. These findings highlight the model's spatial robustness and its suitability for multi-regional agroclimatic prediction.

**Table 10.** Performance of the proposed model under leave-one-region-out cross-validation.

| Run | District Excluded from Training | Training Data | Testing Data | $R^2$ | RMSE | MAE |
|---|---|---|---|---|---|---|
| 1 | Dakshina Kannada | Remaining 7 regions | Dakshina Kannada | 0.802 | 0.36 | 0.27 |
| 2 | Shivamogga | Remaining 7 regions | Shivamogga | 0.809 | 0.31 | 0.26 |
| 3 | Davangere | Remaining 7 regions | Davangere | 0.841 | 0.29 | 0.21 |
| 4 | Tumkur | Remaining 7 regions | Tumkur | 0.884 | 0.26 | 0.17 |
| 5 | Chikmagalur | Remaining 7 regions | Chikmagalur | 0.784 | 0.49 | 0.49 |
| 6 | Uttara Kannada | Remaining 7 regions | Uttara Kannada | 0.668 | 0.48 | 0.32 |
| 7 | Chitradurga | Remaining 7 regions | Chitradurga | 0.859 | 0.28 | 0.19 |
| 8 | Uttara Kannada | Remaining 7 regions | Uttara Kannada | 0.839 | 0.28 | 0.20 |

### *3.8. Regional Performance Variability Analysis*

To examine the weaker performance of some districts, that is, the weaker prediction performance in inland regions, the regional prediction errors were carefully analyzed.

1. Sample Representation

Initially, the impact of the sample representation was analyzed to determine whether resulted weaker prediction performance in inland regions was due to the limited sample. All the regions had comparable observations over the study periods, except for the weaker performance in the inland regions. Hence, the observed weaker performance of the inland regions cannot be attributed to differences in sample size. This indicates that the regional climatic variability has a crucial impact on the prediction differences.

2. Climatic Variability

In the next step, the climatic fluctuations were analyzed to determine the cause of the weaker inland performance. The interannual variability in temperature, humidity and soil moisture variables was calculated from the monthly climatic data. For an individual climatic factor, first the standard deviation (SD) is computed for each month across 20 years. Further, the resulting SD values were averaged to obtain the regional variability. The result of the same is presented in Table 11.

**Table 11.** Climatic variability and regional prediction performance.

| Study Regions | Temperature Variability | Humidity Variability | Soil-Moisture Variability | RMSE |
|---|---|---|---|---|
| Udupi | 0.588 | 2.945 | 3.278 | 0.365 |
| Dakshina Kannada | 0.522 | 2.746 | 2.039 | 0.372 |
| Uttara Kannada | 0.782 | 4.535 | 2.601 | 0.380 |
| Chikmagalur | 0.907 | 5.008 | 6.060 | 0.389 |
| Shivamogga | 1.159 | 5.731 | 6.168 | 0.401 |
| Davangere | 1.128 | 6.069 | 3.517 | 0.426 |
| Tumkur | 1.143 | 6.074 | 3.684 | 0.441 |
| Chitradurga | 1.178 | 6.176 | 4.039 | 0.458 |

The result from Table 9 shows that the regions (Udupi, Dakshina Kannada and Uttara Kannada) with lower temperature and humidity variations generally achieved lower prediction errors. On the other hand, the regions such as Davangere, Tumkur and Chitradurga recorded the highest variability in these climatic factors and they exhibited the highest prediction error too. It was also observed that the soil variable and soil moisture variable did not show the same consistent pattern. The region Shivamogga and Chikkamagaluru exhibited low soil moisture variability but maintained lower RMSE compared to

some inland regions. Hence, the observed regional differences in the model performance are mainly related to the temperature and humidity variations rather than soil moisture. This suggests that greater variations in inland regions may increase the complexity of the temporal production pattern and make accurate model prediction more challenging.

3. The Temporal Attention Behavior

To investigate the further effect, the monthly attention weights of the proposed hybrid (RF attention-BiLSTM) model were examined across the 12-month climatic sequence over the study period (20 years). The attention stability was analyzed by comparing the monthly attention patterns across the year. This analysis helped to determine the association between model consistency and temporal periods. The outcome is illustrated in Table 12.

**Table 12.** Regional-temporal attention behavior.

| District | Attention Variability (SD) | Temporal Attention Behavior | Regional RMSE |
|---|---|---|---|
| Udupi | 0.012 | Stable | 0.365 |
| Dakshina Kannada | 0.014 | Stable | 0.372 |
| Uttara Kannada | 0.017 | Relatively stable | 0.380 |
| Chikmagalur | 0.019 | Relatively stable | 0.389 |
| Shivamogga | 0.022 | Moderate variation | 0.401 |
| Davangere | 0.028 | Increased variation | 0.426 |
| Tumkur | 0.034 | High variation | 0.441 |
| Chitradurga | 0.039 | High variation | 0.458 |

This analysis represented a stable temporal behavior in coastal regions and a higher variation in attention behavior is recorded in the inland regions. The Tumkur and Chitradurga districts showed higher variability along with greater climatic variability too. This suggests the stronger climatic variations in the inland regions may make the key temporal patterns more challenging. Overall, the outcome of this analysis clearly indicates that the observed weaker performance in the inland regions is due to greater climatic variability (variations in temperature and humidity) than to the sample size. The observed attention behavior further highlights the potential of region-adaptive temporal modeling for handling greater climatic variability.

Overall, the outcome of the regional performance variability analysis indicated that the variations in predictive performance are influenced not only by sample distribution but also by climatic and temporal variability across the areca nut production regions. Similar regional variations reported from the previous investigations (corn and mango) [67,68] further demonstrate the role of spatial heterogeneity in crop prediction. Collectively, these results highlight the need for region-specific agroclimatic dynamics, which are essential to achieve reliable and generalizable crop prediction across heterogeneous production environments.

### *3.9. Genotypic Resilience and Stability Analysis Outcome*

To examine genotypic resilience of the different areca nut variants in the study region, genotypic resilience score was obtained by exploiting the genotype mapping over the different locations (i.e., Dakshina Kannada, Shivamogga, Davangere, Tumkur, Chikmagalur, Uttara Kannada, Chitradurga, and Udupi). The genotypic resilience performance obtained is given in Table 13. The genotypic resilience performance obtained is given in Table 14.

**Table 13.** Genotype mapping over the different locations.

| District | Genotype/ Variant | Mean Yield (kg/Fruit or Equiv.) | CVi (%) (est./ Proxy) | b (Regression) (est.) | $s^2$_di (Deviation) (est.) | YSi/Stability (Proxy) | Resilience Notes/Suitability |
|---|---|---|---|---|---|---|---|
| Chikkamagaluru | Thirthahalli local | ~55–60 (high quality) | Low-Moderate | ~1 (stable) | Low | High (+) | Traditional; high stability in humid Malnad; good under consistent rain. |
| Chikkamagaluru | Mangala | ~50–55 | Low | ~1 | Low | High (+) | Early stabilization; consistent yield. |
| Chikkamagaluru | Sumangala | ~55–60 | Low-Moderate | ~1 | Low | High (+) | High fruit set; stable in hilly zones. |
| Chitradurga | Hirehalli Local | ~45–55 | Moderate | <1 (less responsive) | Moderate | Medium-High (+) | Maidan/dry preferred; stable under irrigation; drought-tolerant proxy. |
| Chitradurga | SK1 (trial) | 57.2 | 1.69 | −0.48 | −1.91 | 16 (+) | Top resilience; climate-adapted; suitable for semi-arid with irrigation. |
| Chitradurga | Betara | 54.52 | 1.62 | 0.59 | −2.3 | 14 (+) | High stability; drought-tolerant; ideal for plateau stress. |
| Dakshina Kannada | Mangala/South Kanara Local | ~50–55 | Low | ~1 | Low | High (+) | Coastal; high stability; nematode-tolerant. |
| Dakshina Kannada | Sumangala/Sreemangala | ~55–60 | Low | ~1 | Low | High (+) | High yield; consistent in humid areas. |
| Davangere | Chengiri local/Sagar local | ~50–55 | Moderate | ~1 | Moderate | Medium (+) | Transitional; stable with irrigation. |
| Davangere | SK2 | 51.21 | 4.67 | 1.73 | 0.16 | 10 (+) | Stable high-performer; good for expanding irrigated areas. |
| Shivamogga | Thirthahalli local | ~55–60 | Low | ~1 | Low | High (+) | Top quality; stable in Malnad. |
| Shivamogga | Mohit Nagar | ~55–58 | Low-Moderate | ~1 | Low | High (+) | Uniform bunches; consistent bearing. |
| Tumkur | Hirehalli Local | ~45–55 | Moderate | <1 | Moderate | Medium-High (+) | Dry plateau; irrigation-essential; stable performer. |
| Tumkur | Galangsuka | 51.47 | 4.21 | 1.75 | −1.37 | 11 (+) | Adaptable to temp swings; suitable for variable conditions. |
| Udupi | Mangala/South Kanara Local | ~50–55 | Low | ~1 | Low | High (+) | Coastal; high stability. |
| Udupi | Sumangala | ~55–60 | Low | ~1 | Low | High (+) | Consistent in humid coastal. |
| Uttara Kannada | Sagar/Sirsi local | ~50–55 | Moderate | ~1 | Moderate | Medium-High (+) | Malnad/coastal mix; stable locals. |
| Uttara Kannada | Swarna Mangala | ~55–60 | Low | ~1 | Low | High (+) | High yield; good processing quality. |

**Table 14.** Genotypic resilience performance.

| Genotype | Avg Mean Yield (g/Fruit) | Avg CVi (%) | Avg b | Avg $s^2$_di | Avg YSi | Resilience Score | Notes (Central Karnataka Focus) |
|---|---|---|---|---|---|---|---|
| Mohit Nagar | 55.0 | 3.8 | 0.9 | 4.0 | 14 | 4.45 | Top resilience; early stabilization; good for variable conditions. |
| Thirthahalli local | 52.0 | 5.5 | 1.0 | 10.0 | 12 | 1.98 | High quality/stable in humid; moderate in central with irrigation. |
| Sumangala | 50.5 | 4.5 | 1.1 | 6.0 | 11 | 1.52 | Consistent high yielder; adaptable to semi-arid. |
| Sreemangala | 49.0 | 4.2 | 1.0 | 5.5 | 10 | 1.35 | Bold nuts; stable bearing. |
| Mangala | 48.0 | 4.0 | 1.2 | 5.0 | 10 | 1.28 | Early; high in initial years; good for central expansion. |
| Hirehalli Local | 48.0 | 7.0 | 0.8 | 15.0 | 8 | −13.85 | Drought proxy; stable in dry maidan (central focus). |
| South Kanara Local | 45.0 | 6.5 | 1.0 | 10.0 | 7 | −9.32 | Traditional; uniform but lower resilience in dry. |

The genotypic resilience score, as depicted in Figure 8, reveals a clear performance hierarchy, stating that the Mohit Nagar stands out as the top-ranked genotype with the least negative score (−0.79), reflecting its superior combination of high yield, strong stability (high YSi), low variability (low CVi), and minimal unpredictable deviation. In contrast, most genotypes exhibit strongly negative scores ranging from approximately −2.6 to −14.0, primarily due to substantial penalties from high $s^2$_di values (unpredictable fluctuations) and moderate to high CVi (yield inconsistency). This pattern highlights that improved CPCRI varieties (e.g., VTLAH, Sumangala, and Shatamangala) generally perform better than traditional local types (e.g., Hirehalli Local, Chengiri local, and South Kanara Local), which suffer from greater environmental sensitivity and deviation. These findings suggest that, for reliable cultivation, especially in the variable, irrigation-dependent conditions of central Karnataka, prioritizing genotypes with the least negative resilience scores (such as Mohit Nagar and VTLAH series) would be most advantageous, while traditional locals may require enhanced management to mitigate their higher risk of inconsistent performance.

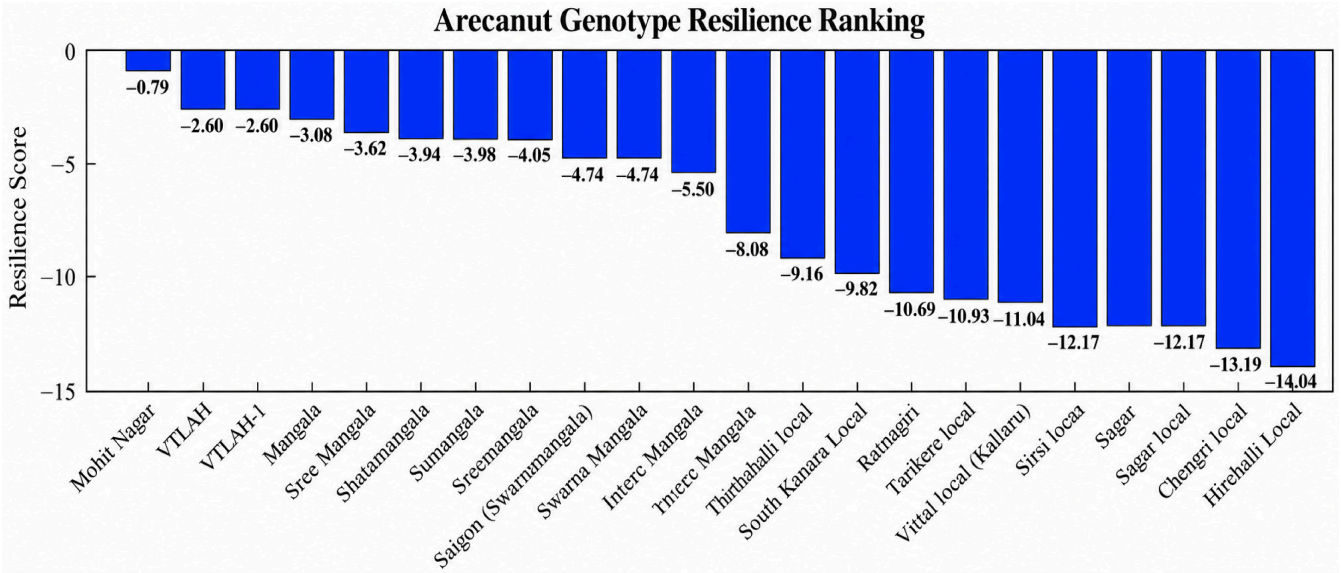


**Figure 8.** Genotypic resilience scores.

Overall, the genotype analysis demonstrated that resilience and stability differ across areca nut genotypes under changing climatic conditions. Further, a comparable stability pattern was reported in cashew [69], where different genotypes (18 genotypes across

four contrasting environments) showed varying performance and yield stability under varying climatic conditions. Thus, these results clearly provide evidence that integration of genotype stability along with the agroclimatic assessment can effectively help to identify the genotypes with greater resilience and consistent performance and also supports a climate-resilient crop production system.

## 4. Conclusions

This study developed a novel hybrid RF attention-BiLSTM framework to analyze the nonlinear spatiotemporal impact of external climatic factors on areca nut production in Central Karnataka. The study presented a robust framework for agro-climatic mapping and for areca nut yield prediction under imbalanced and noise-prone data conditions. A structured preprocessing pipeline comprising missing-value estimation, outlier handling, normalization, and SMOTE-ENN resampling ensured statistical consistency and data balance prior to learning. From a quantitative perspective, the superiority of the proposed framework is substantiated through consistent error reduction and correlation improvement across all baseline models. The hybrid model achieved an $R^2$ value of 0.914, outperforming conventional models. Furthermore, the inclusion of SMOTE-ENN resampling contributed to measurable performance gains, increasing $R^2$ from 0.865 to 0.914, while reducing RMSE and MAE by 17.4% and 17.9%. Statistically, the proposed hybrid model exhibited strong linear agreement between predicted and observed yields, achieving a Pearson correlation coefficient (r = 0.956), which is substantially higher than that of individual baselines. The converging training and validation MSE curves demonstrated the robustness of the model with limited overfitting. Region-wise evaluation further confirmed the robustness and spatial generalizability of the model. The repeated-run assessment, ablation analysis, statistical comparison and regional evaluation further confirmed the stability of the developed model beyond accuracy.

The genotypic resilience analysis revealed that the genotypes with the highest resilience scores, such as Mohit Nagar (4.45) and Thirthahalli local (1.98), demonstrate superior balance of high yield, strong stability (high YSi), and low variability—making them the most promising choices for reliable performance. Lower-scoring genotypes like Hirehalli Local (−13.85) and South Kanara Local (−9.32), while regionally adapted, exhibit greater unpredictability and are better suited to more stable humid environments rather than semi-arid plateau expansion. Thus, these statistical outcomes validate that the integration of attention-guided temporal learning, ensemble-based feature encoding, and data-level imbalance correction followed by genotype analysis enables the proposed framework to deliver significantly superior predictive accuracy, stability, and generalization. Hence, beyond crop yield prediction, the developed approach provides a strong foundation for effective crop–climate mapping towards region-wise analysis and supports the precision agricultural planning under increasing climatic variability.

## 5. Limitations

The present study is limited to region-level historical data and restricted to the spatial coverage of eight selected regions of Central Karnataka. Future analysis may incorporate wider geographical coverage to enhance the model's generalization.

**Author Contributions:** The major efforts including the model (algorithm) development, paper drafting, etc., were contributed by the first author S.S., while the review, guidance and rectification measures were suggested by the second author A.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Some of the data elements and climatic variables are publicly available. Though, the research specific data representing the agroclimatic measurements related to the study regions are private and can be availed with written permission. Since, these datasets are the part of ongoing PhD research, it is not made publicly accessible; though, would be made open-access post PhD completion.

**Acknowledgments:** During the preparation of this manuscript, the author(s) used QGIS (version 3.4) for the preparation of the geographical study-area map and Inkscape (version 1.4) for graphical formatting of the map and preparation of schematic figures. These software tools were operated manually for figure preparation and graphical editing. Adobe Firefly (Firefly Image 4) was used only to improve the visual presentation of the Graphical Abstract. The prompt used was: "Enhance the visual quality, clarity, and layout of this scientific graphical abstract without altering the scientific content, data, text, or results." The authors have reviewed and edited the outputs and take full responsibility for the content of this publication.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| RF | Random Forest |
| CV | Coefficient of Variation |
| PA | Precision Agriculture |
| MLP | Multi-layer Perceptron |
| LSTM | Long Short-Term Memory |
| RMSE | Root Mean Square Error |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| SVR | Support Vector Regression |
| ENN-SMOTE | Extended Nearest Neighbor Synthetic Minority Oversampling Technique |
| MVE | Maximum Voting Based Ensemble |
| ML | Machine Learning |
| k-NN | k-Nearest Neighbour |
| IQR | Inter Quartile Range |
| ANN | Artificial Neural Network |
| RNN | Recurrent Neural Network |
| CNN | Convolutional Neural Networks |
| Bi-LSTM | Bidirectional long and short-term memory |

## References

1. Fotabong, T.F.; Muhammad, A.; Keisham, M.; Mehrotra, T.; Singh, R.K.; Singh, S.P.; Agrawal, A.P. Soil Analysis Using Deep Learning for Precision Agriculture. In Proceedings of the 2024 First International Conference Pioneering Developments in Computer Science & Digital Technologies (IC2SDT), Delhi, India, 2–4 August 2024; pp. 303–307. [CrossRef]
2. Nolasco, M.; Sgró, M.A.; Zader, P.; Buss, E.F.; Rubio, J.; Scavuzzo, M. Performance Evaluation of Satellite Temporal Aggregation Methods and Sample Size Training Data for Crop Mapping. In Proceedings of the 2024 IEEE Biennial Congress of Argentina (ARGENCON), San Nicolás de los Arroyos, Argentina, 18–20 September 2024; pp. 1–7. [CrossRef]
3. Chen, W.; Liu, G. A Novel Method for Identifying Crops in Parcels Constrained by Environmental Factors Through the Integration of a Gaofen-2 High-Resolution Remote Sensing Image and Sentinel-2 Time Series. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 450–463. [CrossRef]
4. Phang, S.K.; Chiang, T.H.A.; Happonen, A.; Chang, M.M.L. From Satellite to UAV-Based Remote Sensing: A Review on Precision Agriculture. *IEEE Access* **2023**, *11*, 127057–127076. [CrossRef]
5. Food and Agriculture Organization (FAO). *The Future of Food and Agriculture: Trends and Challenges*; FAO: Rome, Italy, 2017.

6. Tamasiga, P.; Ouassou, E.H.; Onyeaka, H.; Bakwena, M.; Happonen, A.; Molala, M. Forecasting Disruptions in Global Food Value Chains to Tackle Food Insecurity: The Role of AI and Big Data Analytics—A Bibliometric and Scientometric Analysis. *J. Agric. Food Res.* **2023**, *14*, 100819. [CrossRef]
7. Hu, T.; Zhang, X.; Khanal, S.; Wilson, R.; Leng, G.; Toman, E.M.; Wang, X.; Li, Y.; Zhao, K. Climate Change Impacts on Crop Yields: A Review of Empirical Findings, Statistical Crop Models, and Machine Learning Methods. *Environ. Model. Softw.* **2024**, *179*, 106119. [CrossRef]
8. Evans, M.R.; Oliver, D.; Zhou, X.; Shekhar, S. *Spatial Big Data*; Informa: London, UK, 2024; pp. 115–138. [CrossRef]
9. Helber, P.E.; Bischke, B.; Packbier, C.; Habelitz, P.; Seefeldt, F. An Operational Approach to Large-Scale Crop Yield Prediction with Spatio-Temporal Machine Learning Models. In Proceedings of the IGARSS 2024—IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 5 September 2024; pp. 4299–4302. [CrossRef]
10. Wellington, M.J.; Lawes, R.; Kuhnert, P. A Framework for Modelling Spatio-Temporal Trends in Crop Production Using Generalized Additive Models. *Comput. Electron. Agric.* **2023**, *212*, 108111. [CrossRef]
11. Cheung, C. Spatial and Temporal Synergy: Advanced Autoregression Models for Global Agricultural Development Insights. *Am. J. Stud. Res.* **2024**, *2*, 52–63. [CrossRef]
12. Kamatchi, C.B.; Muthukumaravel, A. Enhancing Crop Productivity Through Land Dataset Analysis on Selecting the Right Crops for the Right Land. In *Optimizing Intelligent Systems for Cross-Industry Application*; IGI Global: Hershey, PA, USA, 2024; pp. 267–288. [CrossRef]
13. Rowe, F.; Lovelace, R.; Dennett, A. Spatial Interaction Modelling: A Manifesto. In *A Research Agenda for Spatial Analysis*; Edward Elgar Publishing: Cheltenham, UK, 2024; pp. 177–196. [CrossRef]
14. Hebbar, K.B.; Ramesh, S.V.; Bhat, R. Predicting Current and Future Climate Suitability for Arecanut (*Areca catechu* L.) in India Using Ensemble Model. *Heliyon* **2024**, *10*, e26382. [CrossRef] [PubMed]
15. Mishra, P.; Arti; Mondal, B.; Sharma, R.; Kumari, B.; Biswas, T.; Ray, S. Forecasting of Arecanut in India Using Time Series Model. *Agric. Sci. Dig.* **2025**, *45*, 907–913. [CrossRef]
16. Borse, K.; Agnihotri, P.G. Adaptive Neuro-Fuzzy Inference System-Based Yield Forecast Using Climatic Variables. In *Climate Change Impacts on Water Resources*; Jha, R., Singh, V.P., Singh, V., Roy, L.B., Thendiyath, R., Eds.; Water Science and Technology Library; Springer: Cham, Switzerland, 2021; Volume 98, pp. 775–790. [CrossRef]
17. Pakkala, P.G.R.; Rai, B.S.; Rai, P. Statistical Driven Feature Selection for Prognostic Reasoning and Insight Exploration of Areca Nut Crop Using Data Analytics Approach. *Indian J. Sci. Technol.* **2023**, *16*, 1495–1504. [CrossRef]
18. Harish, B.P.; Arvind; Karuna, K.M.; Sushma, S.; Anil, K. Karnataka's Arecanut Boom: Understanding the Current Scenario and Factors Driving the Shift. *Krishi Sci.* **2025**, *6*, 1–6.
19. Vasundhara, R.; Prakash, N.; Kumar, A.; Hegde, R. Characterization and Classification of Arecanut-Growing Soils of Karnataka. *J. Plant. Crops* **2020**, *48*, 91–102. [CrossRef]
20. Simpson, R.; Abhilash, K.P.; Kavya, R.; Hiremath, S. Examining two decades of climate variability in Karnataka: Insights from temperature and rainfall trends. *Int. J. Environ. Clim. Change* **2024**, *14*, 97–109. [CrossRef]
21. Jose, C.T.; Chandran, K.P.; Muralidharan, K.; Jayasekhar, S. Crop weather relationship in arecanut. *J. Plant. Crops* **2019**, *47*, 145–151.
22. Sujatha, S.; Bhat, R.; Elain Apshara, S. Climate change, weather variability and associated impact on arecanut and cocoa in humid tropics of India. *Int. J. Innov. Hortic.* **2018**, *7*, 27–37.
23. Patil, B.; Hegde, V.; Sridhara, S.; Narayanaswamy, H.; Naik, M.K.; Patil, K.K.R.; Rajashekara, H.; Mishra, A.K. Exploring the Impact of Climatic Variables on Arecanut Fruit Rot Epidemic by Understanding the Disease Dynamics in Relation to Space and Time. *J. Fungi* **2022**, *8*, 745. [CrossRef] [PubMed]
24. Guthu, R.P.P.; Bellipady, S.R. A formal statistical data modeling for knowledge discovery and prognostic reasoning of arecanut crop using data analytics. *Int. J. Softw. Sci. Comput. Intell.* **2022**, *14*, 1–27. [CrossRef]
25. Bharathi, K.; Ashwini, B.; Rao, S.S.; Prathibha, V.H. Lag Aware Ensemble Learning Model for Weather Based Forecasting of Arecanut Fruit Rot Disease. *Appl. Eng. Agric.* **2026**, *42*, 271–284. [CrossRef]
26. Krishna, R.; Prema, K.V.; Gaonkar, R. Areca nut disease dataset creation and validation using machine learning techniques based on weather parameters. *Eng. Sci.* **2022**, *19*, 205–214. [CrossRef]
27. Lingananda, N.K.; Naik, G.B.; Chittaragi, A.; Patil, B.; Sridhara, S.; Vasanad, N.C.; Patil, S.; Naik, M.K. Hybrid Machine Learning Framework for Predicting Bacterial Leaf Stripe Disease in Arecanut under Complex Agroclimatic Conditions. *Microb. Risk Anal.* **2026**, *31*, 100371. [CrossRef]
28. Krishna, R.; Prema, K.V. Constructing and optimizing RNN models to predict fruit rot disease incidence in areca nut crop based on weather parameters. *IEEE Access* **2023**, *11*, 110582–110595. [CrossRef]
29. Hegde, A.; Sadanand, V.S.; Hegde, C.G.; Naik, K.M.; Shastri, K.D. Identification and Categorization of Diseases in Areca Nut: A Machine Learning Approach. *Indones. J. Electr. Eng. Comput. Sci.* **2023**, *31*, 1803–1810. [CrossRef]

30. Mohan, C.; Josephrajkumar, A.; Thube, S.H.; Saneera, E.K.; Rajkumar. Pests and Their Management in Arecanut. In *Trends in Horticultural Entomology*; Mani, M., Ed.; Springer: Singapore, 2022. [CrossRef]
31. Rajashree, K.; Prema, K.V.; Rajath, G.; Angad, S. Prediction of fruit rot disease incidence in Arecanut based on weather parameters. *Agron. Res.* **2022**, *20*, 1117–1133. [CrossRef]
32. Kanan, L.V.; Kumar, M.S. Arecanut Yield Disease Forecast using IoT and MachineLearning. *Int. J. Sci. Res. Eng. Technol.* **2021**, *1*, 6–8.
33. Sharath Kumar, K.R.; Mohan, K.; Nirisha. Arecanut crop disease prediction using IoT and machine learning. *J. Sci. Technol.* **2020**, *5*, 160–165. [CrossRef]
34. Manjunatha, M.; Parkavi, A. Estimation of arecanut yield in various climatic zones of Karnataka using data mining technique: A survey. In *2018 International Conference on Current Trends Towards Converging Technologies (ICCTCT)*; IEEE: Piscataway, NJ, USA, 2018; pp. 1–4. [CrossRef]
35. Hegde, M.R.; Bhoomika, H.R.; Maheswarappa, H.P.; Ravikumar, M.; Dhananjaya, B.C. Performance of different arecanut (*Areca catechu* L.) cultivars under Malnad region of Karnataka. *J. Pharmacogn. Phytochem.* **2021**, *10*, 373–375.
36. Sane, A.; Ananda, K.S.; Kumar, S.N.S.; Sannamarappa, S.; Ramanujam, B. Yield performance of arecanut (*Areca catechu* L.) varieties in maidan region of Karnataka. *J. Plant. Crops* **2002**, *30*, 22–26.
37. Nagaraja, N.R.; Ananda, K.S.; Rajesh, M.K. Genetic variation among varieties, wild species and related genera of arecanut (*Areca catechu* L.) as revealed by microsatellite markers. *Int. J. Plant Soil Sci.* **2022**, *34*, 846–855. [CrossRef]
38. Lobell, D.B.; Asner, G.P. Climate and management contributions to recent trends in US agricultural yields. *Science* **2003**, *299*, 1032. [CrossRef] [PubMed]
39. Foley, J.A.; Ramankutty, N.; Brauman, K.A.; Cassidy, E.S.; Gerber, J.S.; Johnston, M.; Mueller, N.D.; O'Connell, C.; Ray, D.K.; West, P.C.; et al. Solutions for a cultivated planet. *Nature* **2011**, *478*, 337–342. [CrossRef] [PubMed]
40. Jeong, J.H.; Resop, J.P.; Mueller, N.D.; Fleisher, D.H.; Yun, K.; Butler, E.E.; Timlin, D.J.; Shim, K.-M.; Gerber, J.S.; Reddy, V.R.; et al. Random Forests for Global and Regional Crop Yield Predictions. *PLoS ONE* **2016**, *11*, e0156571. [CrossRef] [PubMed]
41. Joshua, S.V.; Priyadharson, A.S.M.; Kannadasan, R.; Khan, A.A.; Lawanont, W.; Khan, F.A.; Rehman, A.U.; Ali, M.J. Crop Yield Prediction Using Machine Learning Approaches on a Wide Spectrum. *Comput. Mater. Contin.* **2022**, *72*, 5663–5679. [CrossRef]
42. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
43. Khaki, S.; Wang, L. Crop yield prediction using deep neural networks. *Front. Plant Sci.* **2019**, *10*, 621. [CrossRef] [PubMed]
44. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
45. You, J.; Li, X.; Low, M.; Lobell, D.; Ermon, S. Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. *Proc. AAAI Conf. Artif. Intell.* **2017**, *31*, 4559–4566. [CrossRef]
46. Tian, H.; Wang, P.; Tansey, K.; Zhang, J.; Zhang, S.; Li, H. An LSTM Neural Network for Improving Wheat Yield Estimates by Integrating Remote Sensing Data and Meteorological Data in the Guanzhong Plain, PR China. *Agric. For. Meteorol.* **2021**, *310*, 108629. [CrossRef]
47. Chaudhary, Y.; Pathak, H. CYPBL: Crop Yield Prediction Using Bi-Directional LSTM under PySpark Interface. *Multimed. Tools Appl.* **2024**, *83*, 75781–75800. [CrossRef]
48. Kolipaka, V.R.R.; Namburu, A. Integrating Temporal Fluctuations in Crop Growth with Stacked Bidirectional LSTM and 3D CNN Fusion for Enhanced Crop Yield Prediction. *Int. J. Recent Innov. Trends Comput. Commun.* **2023**, *11*, 376–383. [CrossRef]
49. Gavahi, K.; Abbaszadeh, P.; Moradkhani, H. DeepYield: A Combined Convolutional Neural Network with Long Short-Term Memory for Crop Yield Forecasting. *Expert Syst. Appl.* **2021**, *184*, 115511. [CrossRef]
50. Nejad, S.M.M.; Abbasi-Moghadam, D.; Sharifi, A.; Farmonov, N.; Amankulova, K.; László, M. Multispectral Crop Yield Prediction Using 3D-Convolutional Neural Networks and Attention Convolutional LSTM Approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 254–266. [CrossRef]
51. He, H.; Garcia, E.A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [CrossRef]
52. Stiller, S.; Grahmann, K.; Ghazaryan, G.; Ryo, M. Improving Spatial Transferability of Deep Learning Models for Small-Field Crop Yield Prediction. *ISPRS Open J. Photogramm. Remote Sens.* **2024**, *12*, 100064. [CrossRef]
53. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
54. Directorate of Economics and Statistics, Ministry of Agriculture and Farmers Welfare, Government of India. Crops APY Report Web. Available online: https://data.desagri.gov.in (accessed on 18 May 2023).
55. NASA POWER a Project's Data Access Viewer (DAV). Available online: https://power.larc.nasa.gov/data-access-viewer/ (accessed on 20 September 2023).
56. QGIS.Org. QGIS Geographic Information System, Open-Source Geospatial Foundation Project, 2023. Available online: https://qgis.org/ (accessed on 24 October 2023).

57. Han, L.; Yang, G.; Yang, X.; Song, X.; Xu, B.; Li, Z.; Wu, J.; Yang, H.; Wu, J. An explainable XGBoost model improved by SMOTE-ENN technique for maize lodging detection based on multi-source unmanned aerial vehicle images. *Comput. Electron. Agric.* **2022**, *194*, 106804. [CrossRef]
58. Sun, J.; Sun, C.; Li, Z.; Qian, Y.; Li, T. Prediction method of sugarcane important phenotype data based on multi-model and multi-task. *PLoS ONE* **2024**, *19*, e0312444. [CrossRef] [PubMed]
59. Deepika, N.M.; Sreeama Murthy, K. AgroAttenNet: A CNN–BiLSTM–Attention–Random Forest hybrid model for temporal groundnut yield forecasting. *J. Adv. Inf. Technol.* **2026**, *17*, 723–736. [CrossRef]
60. Pandit, P.; Sagar, A.; Ghose, B.; Dey, P.; Paul, M.; Alqadhi, S.; Mallick, J.; Almohamad, H.; Abdo, H.G. Hybrid time series models with exogenous variable for improved yield forecasting of major Rabi crops in India. *Sci. Rep.* **2023**, *13*, 22240. [CrossRef] [PubMed]
61. Xiang, W.; Long, L.; Liu, Z.; Dai, F.; Zhang, Y.; Li, H.; Cheng, L. A crop model based on dual attention mechanism for large area adaptive yield prediction. *Smart Agric. Technol.* **2025**, *11*, 100957. [CrossRef]
62. Cheema, I.A.; Hanif, M.K.; Khokhar, A.M.A.; Rehman, M.A.U.; Zaheer, R. Enhancing crop yield prediction accuracy with a novel interpretable deep learning model: MHCNN-LSTM-MHA. *Sci. Rep.* **2026**, *16*, 23515. [CrossRef] [PubMed]
63. Cao, J.; Zhang, Z.; Luo, Y.; Zhang, L.; Zhang, J.; Li, Z.; Tao, F. Wheat yield predictions at a county and field scale with deep learning, machine learning, and Google Earth Engine. *Eur. J. Agron.* **2021**, *123*, 126204. [CrossRef]
64. Shook, J.; Gangopadhyay, T.; Wu, L.; Ganapathysubramanian, B.; Sarkar, S.; Singh, A.K. Crop yield prediction integrating genotype and weather variables using deep learning. *PLoS ONE* **2021**, *16*, e0252402. [CrossRef]
65. Quintanilla-Albornoz, M.; Miarnau, X.; Pamies-Sans, M.; Bellvert, J. Almond yield prediction at orchard scale using satellite-derived biophysical traits and crop evapotranspiration combined with machine learning. *Front. Agron.* **2025**, *7*, 1667674. [CrossRef]
66. Jantaraprasit, N.; Promchote, P.; Wang, S.-Y.S.; Daengnui, S.; Motlagh, S.K.; de Lima Moraes, A.G.; França, L.; Yoon, J.-H.; Phumichai, C.; Kittipadakul, P. A machine learning–coupled APSIM model pipeline for projected oil palm yield in Surat Thani, Thailand. *PLoS ONE* **2026**, *21*, e0349782. [CrossRef] [PubMed]
67. Li, Y.; Guan, K.; Yu, A.; Peng, B.; Zhao, L.; Li, B.; Peng, J. Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the U.S. *Field Crops Res.* **2019**, *234*, 55–65. [CrossRef]
68. Kouadio, L.; Kulanthaivel, B.; Mai, T.; Nguyen-Huy, T.; Wang, Q.J.; Byrareddy, V.M.; Mushtaq, S.; Senthil, A.; Newlands, N.K.; Geethalakshmi, V. Quantifying error in fine-scale crop yield forecasts to guide data and algorithm improvements: Case study of mango in Tamil Nadu, India. *Comput. Electron. Agric.* **2025**, *236*, 110450. [CrossRef]
69. Eradasappa, E.; Mohana, G.S.; Poduval, M.; Sethi, K.; Aneesa Rani, M.S.; Lourdusamy, I.K.; Velmurugan, S.; Manjusha, M.; Raviprasad, T.N.; Anilkumar, C. Analysis of stability for nut yield and ancillary traits in cashew (*Anacardium occidentale* L.). *Sci. Rep.* **2024**, *14*, 2127. [CrossRef] [PubMed]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.