Journal of Information Systems Engineering and Management
2025, 10(10s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Exploring RNN Model Implementations for Predicting Sunflower Disease Outbreaks

Rupali Sarode[1], Dr. Arti Deshpande[2]
[1]*Assistant Professor, Thadomal Shahani Engineering College, Bandra West, Mumbai, Computer Engineering Department, rupali.patil@thadomal.org*
[2]*Associate Professor, Thadomal Shahani Engineering College, Bandra West, Mumbai, Computer Engineering Department arti.deshpande@thadomal.org*

ARTICLE INFO

Received: 05 Nov 2024

Revised: 25 Dec 2024

Accepted: 11 Jan 2025

ABSTRACT

The development of diseases in oilseed fields is influenced by meteorological conditions and the pathogen's predilection for susceptible hosts. This research examines forecasting models for two sunflower diseases: Alternaria leaf blight and powdery mildew. The disease percentage for the Alternaria leaf blight and powdery mildew is predicted for the Kharif and Rabi seasons, respectively. For prediction, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Bi-directional Long Short-Term Memory (BILSTM), and Simple Recurrent Neural Network (SimpleRNN) models are employed. These models incorporated six parameters: precipitation, maximum temperature, minimum temperature, maximum and minimum relative humidity, and disease percentage. Meteorological data and disease percentages were obtained from India (Marathwada region). After the experimental study, the results indicate that the SimpleRNN model demonstrated superior performance for Alternaria Leaf Blight with a Mean Squared Error (MSE) value of 0.11%, while for Powdery Mildew, LSTM exhibited the best performance with an MSE value of 0.32%. Each model exhibits unique performance characteristics, and all models are evaluated using the same dataset. A non-parametric Friedman test is employed to statistically validate the differences in performance, followed by a Nemenyi Test as a post-hoc analysis. This approach enables a side-by-side comparison of the average performance across all models.

Keywords: Alternaria Leaf Blight, Powdery Mildew, Long Short-Term Memory, (LSTM) Gated Recurrent Unit (GRU), Bidirectional LSTM(BILSTM)

1. Introduction

Early-stage plant disease forecasting is crucial for agricultural management, enabling farmers to make informed decisions on sowing and prevention, enhancing their capacity to manage outbreaks. Disease forecasting helps growers to determine season-wise sowing, disease spread, and control measures and is beneficial for anticipating disease occurrence in a location. Sunflower, introduced to India as an oilseed crop in 1969, is significant in the nation. Sunflower seeds contain 48–53 percent high-quality vegetable oil (Agrovista Farming 2024) and have 40–52 percent edible oil content with low cholesterol (Josipovic J et.al 2006). Sunflower sowing can occur in Kharif (the First Fortnight of July), Rabi (the Second Fortnight of October), and Zaid (the First Fortnight of March). Diseases like Powdery Mildew, Alternaria leaf blight, and Downy Mildew have affected sunflower plants in various parts of India Shivani R(2023), including Kurnool, Hyderabad, Akola, Latur, Raichur, and Coimbatore. These diseases, caused by fungi, viruses, and bacteria, have reduced sunflower yield. Sunflowers are mainly cultivated in southern India, including Andhra Pradesh, Karnataka, Maharashtra, and Tamil Nadu. Key fungal diseases impacting sunflowers include Alternaria leaf spot (Alternaria helianthi), rust (Puccinia helianthi), and downy mildew (Plasmopara halstedii Farl.) These diseases reduce yield and affect oil quality.

<u>Review of Literature:</u>

The author (A. Bheemaraya et.al 2028) provided valuable insights into the relationship between powdery mildew development in sunflowers and various meteorological conditions. The findings highlight the complex interplay of factors such as temperature, relative humidity, vapor pressure, and cloud cover influencing disease progression.

Morning relative humidity correlates negatively with powdery mildew disease, but maximum temperature, vapor pressure, minimum temperatures, and evening relative humidity are critical. Also, the presence of clouds is causing the disease to develop more. Rainfall does not influence the powdery mildew. The experimental results contribute to a better understanding of the environmental conditions that favor or inhibit powdery mildew in sunflower crops, potentially informing future disease management strategies and cultivation practices in agricultural settings.

The researcher (Rakesh Kaundal et al. 2006) investigated the use of a Support Vector Machine (SVM) for creating weather-based prediction models for rice blast disease. The study used data gathered over a five-year period (2000-2004) from five sites in Himachal Pradesh, India. Temperature, relative humidity, and rainfall were among the meteorological characteristics that were noted. Cross-location and cross-year models were created by the authors and compared SVM with other approaches like Backpropagation Neural Network, Multiple Regression, and Generalized Regression Neural Network. Performance evaluation metrics included Mean Absolute Error, correlation coefficient, and coefficient of determination. Results showed that SVM outperformed existing techniques, demonstrating lower error rates and higher correlation coefficients. A forewarning model proposed by the researcher (V. B. AKASHE et al. 2016) for predicting sunflower thrips populations based on a 10-year field experiment conducted in Solapur, India. The research involved monitoring thrips incidence on sunflower crops from the 30th to 39th meteorological weeks. The thrips population was correlated with weekly mean weather parameters over eight years (2004-2011), and a regression equation was developed for forecasting. The model was evaluated using data from year 2012 and 2013. Results showed that thrips appeared in the 29th-30th weeks and persisted until the 38th-39th weeks. Peak thrips activity occurred during the 32nd-34th weeks, coinciding with specific temperature, humidity, and rainfall conditions. The population showed a positive correlation with maximum temperature and negative correlations with humidity and rainfall. Another study (Katti P et al. 2011) reported higher thrips populations in the first week of the Kharif season. The study (N.S. PANKAJA et al. 2011) examined the relationship between sunflower necrosis disease, thrip populations, and weather conditions for two sunflower varieties (i.e., KBSH-44 and Morden). Key findings include that maximum temperature was positively correlated with disease incidence in both varieties. 2. Thrips palmi populations increased with higher maximum temperatures and longer sunshine durations. 3. Relative humidity, minimum temperatures, and rainy days negatively correlated with thrip populations. 4. Precipitation had minimal impact on thrip occurrence. 5. Regression analysis showed disease incidence was 75% for KBSH-44 and 67% for Morden. 6. Thrip population was 72% for KBSH-44 and 62% for Morden. These results highlight the influence of weather parameters on sunflower necrosis disease and thrip populations. The analyst (Vijaykumar N Ghante et.al 2020) examined the relationship between whitefly population dynamics and various weather parameters on sunflower hybrid DRSH-1 from 2011 to 2019. Key findings include: Non-significant positive correlations with rainfall and relative humidity. Modest negative correlations with sunshine hours and maximum temperature. Bright Sunshine Hours, evening relative humidity, and temperature extremes significantly influence leafhopper population growth. Weather factors collectively account for 92% of whitefly population variations. These results highlight the importance of climate conditions in predicting and managing whitefly populations in sunflower crops. The investigator (Kittakorn Sriwanna 2022) proposed a method for forecasting rice blast disease using weather data. The top 10 features outperformed the five classification models—MLP, SVM, NB, DT, and KNN in classification. Ensemble methods ranked 15 weather features—K5 cross-validation addressed for an uneven class sample. Evaluation measures include geometric mean, balanced accuracy, ROC-AUC score, and F1 score. The main meteorological factors included in the study are rainfall amount, sunshine hours, maximum wind speed, average visibility, and number of rainy days. This approach aims to improve rice blast disease prediction using weather data and machine learning techniques. In literature, variants of the Recurrent Neural Network (RNN) were used for rice crops to forecast the presence of disease. Also, many authors have worked to show how the occurrence of disease correlates with different weather parameters for sunflower oilseed. For Rainfall prediction (Manoj Chhetri et al. 2020) and stock market predictions (Ya Gao et al. 2021), LSTM, BILSTM, and GRU were employed by many authors. According to this literature, RNN variants are not yet used to predict disease by incorporating weather parameters for any crop.

2. Materials and Methods:

Disease Outbreak Methodology:

The methodological framework for forecasting disease is depicted in Figure 1. The study utilizes Latur Oilseed Research Center data, focusing on two sunflower diseases: Alternaria Leaf Blight and Powdery Mildew. This approach encompasses five essential stages: 1) Data Gathering and Initial Processing, 2) Data Normalization and K-Fold Validation, 3) Model Construction, 4) Disease Prediction, and 5) Comparative Analysis of Models using performance metrics, including Mean Square Error (MSE), Root Mean Square Error (RMSE), and Correlation Coefficient (CC).

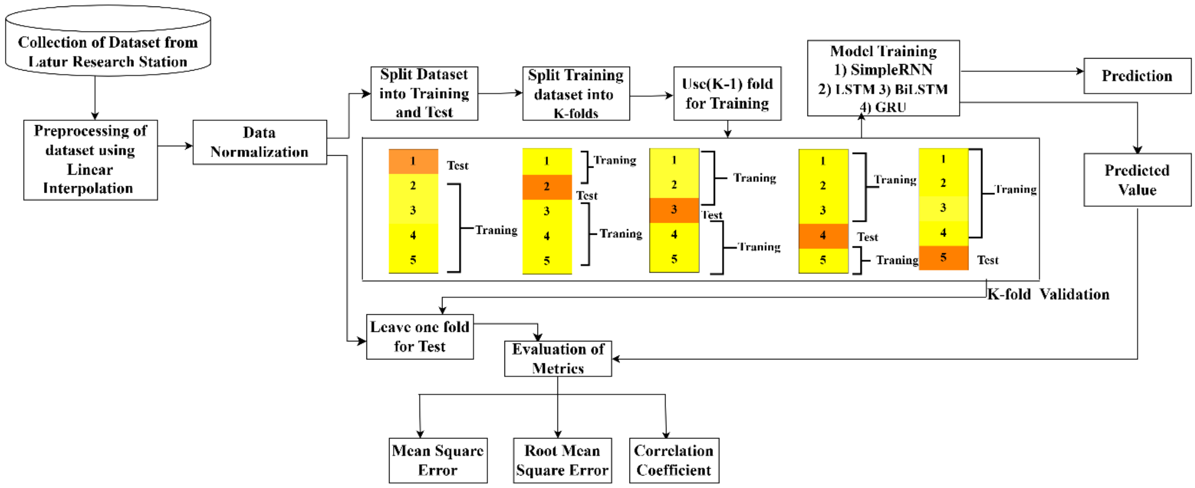


Figure 1: Methodological block diagram of Disease Forecasting

2.1 Dataset Collection and Preprocessing:

The research encompasses two categories of data collection: disease percentage and meteorological information. The Latur Oilseed Research Centre in the Marathwada region of India serves as the source for the dataset. This investigation aims to predict two sunflower diseases: Alternaria Leaf Blight and Powdery Mildew. In Latur, the Kharif season runs from July to October, while the Rabi season extends from December to January. Alternaria leaf blight data was gathered during the Kharif season, whereas powdery mildew information was collected in Rabi. To explore the correlation between weather conditions and disease incidence, five weather-related variables were recorded alongside the date and disease percentage. The compiled dataset spans a decade, covering the years 2014 to 2023.

Table 1: Sample of collected weather data with 3 days gap for the Rabi period(Powdery Mildew).

Date (2022-23)	Powdery%	Max temp	Min temp	Rainfall	Max RH	Min RH
24 Dec - 27 Dec	6.3	32.3	19.3	0.0	90.3	54.0
28 Dec - 31 Dec	7.1	36.2	23.2	0.0	97.0	63.0
1 Jan- 4 Jan	9.5	35.1	22.4	0.0	94.3	58.5

Observations of disease and weather parameters were conducted at the field scale with a three-day interval, as presented in Table 1. Subsequently, the data were pre-processed to impute missing weather parameters and disease percentage values. A linear interpolation imputation technique was employed to fill in the missing percentage values. Linear interpolation estimates values between two known points in one-dimensional data. This method estimates the data value using the two adjacent data points in a one-dimensional sequence to interpolate the desired value. The basic first-order linear interpolation (Guilin Huang 2021) is as follows:

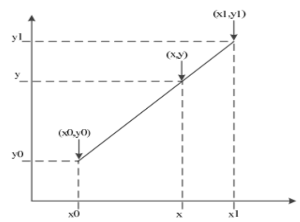


Figure 2: Schematic diagram of linear interpolation

Following the completion of the imputation process, the datasets about both diseases are exhibited in Tables 2 and 3.

Table 2 a) Sample of data received for Powdery Mildew (Rabi) from Latur

Date(2022-23)	Powdery %	Max temp	Min temp	Rainfall	Max RH	Min RH
24 Dec - 27 Dec	6.3	32.3	19.3	0.0	90.3	54.0
28 Dec - 31 Dec	7.1	36.2	23.2	0.0	97.0	63.0
1 Jan- 4 Jan	9.5	35.1	22.4	0.0	94.3	58.5
5 Jan- 9 Jan	9.8	31.8	20.0	0.0	86.0	45.0

Table 2 b) Dataset for Powdery Mildew dataset after preprocessing for date field and with Linear Interpolation.

Table 3 a) Sample of data received for Alternaria Leaf Blight (Kharif) from Latur

Date	Powdery%	Max temp	Min temp	Rainfall	Max RH	Min RH
24-12-2022	6.3	32.3	19.3	0	90.3	54
25-12-2022	6.5	33.28	20.27	0	91.98	56.25
26-12-2022	6.7	34.25	21.25	0	93.65	58.5
27-12-2022	6.9	35.22	22.22	0	95.32	60.75

Date(2021)	ALB%	Max temp	Min temp	Rainfall	Max RH	Min RH
10 Sept- 13 Sept	28.88	24.17	25.01	0.95	92.74	57.31
14 Sept- 17 Sept	33.33	24.04	25.14	0.96	92.9	57.89
18 Sept- 21 Sept	42.22	23.9	25.27	0.98	93.06	58.47

Table 3 b) Dataset for Alternaria Leaf Blight dataset after preprocessing for date field and with Linear Interpolation.

Date(2021)	ALB%	Max temp	Min temp	Rainfall	Max RH	Min RH
10-09-2021	28.88	24.17	25.01	0.95	92.74	57.31
11-09-2021	29.99	24.14	25.04	0.95	92.78	57.46
12-09-2021	31.1	24.1	25.08	0.96	92.82	57.6
13-09-2021	32.22	24.07	25.11	0.96	92.86	57.74

2.2 Data Normalization and K-fold Validation:

The disease forecasting dataset is in CSV format. After data preprocessing, values are normalized using a standard scalar operation to establish a balanced scale ranging from 0 to 1. Subsequently, dataset train and test data frames are generated using Python panda's library. After preprocessing, both disease datasets are further organized for k-fold validation sets. The experimentation performance was evaluated using a k-fold cross-validation method, specifically by employing k values of 4 and 5. Figure 1 illustrates the division of the dataset into k-folds for model training and testing. K-1 folds are utilized for training, while one-fold is reserved for testing. All models are trained and tested for experimental analysis with k-fold cross-validation, incorporating varying numbers of neurons, dropouts, and epochs.

Table 4: Data preprocessing per fold for Powdery Mildew Alternaria Leaf Blight

Disease name	Instances	Training	Forecasting (in day)	K4		K5	
				Train	Test	Train	Test
Powdery Mildew	3683	3639	30	2729	910	2911	728
Alternaria Leaf Blight	3591	3547	30	2660	887	2838	709

2.3 Model Building:

Disease forecasting is accomplished by utilizing four deep learning models: Simple Recurrent Neural Network (Simple RNN) (Tsantekidis, A et.al 2022) (Ibomoiye Domor Mienye et.al 2024) (Mienye, I.D et.al 2024) (Mienye, I.D. et.al 2023), Long Short-Term Memory (LSTM), Bidirectional LSTM (BILSTM) (Benjamin Lindemann et.al 2021), and Gated Recurrent Unit (GRU) (Poluru Eswara et.al 2023) All models are trained and evaluated using K4 and K5 cross-validation. Each fold comprised training and testing instances, as illustrated in Table 4. Before model input, the dataset is structured in the appropriate format of total instances X timesteps X number of features (e.g., 3693 X 14 X 6). This study employs the preceding 14 values to predict the subsequent value, and the disease percentage for the ensuing 30 days is forecasted. The Alternaria Leaf Blight and Powdery Mildew datasets contain 3591 and 3683 instances, respectively. Various hyperparameters, including batch size, neurons, epochs, learning rate, and dropout, are employed in this research to obtain precise results. Table 5 presents the details of the hyperparameters utilized.

Table 5: Hyperparameters Used

Hyperparameter	Values
Epochs	100,200,300
Batch size	32
Learning rate	0.0001
Dropout	0.5 and 0.4
Activation function	ReLu
Neurons used	64,128,256

Upon completion of model training, the models predict disease outbreaks for the subsequent 30-day period. Before calculating model performance, dataset values undergo denormalization using the Inverse Scalar operation to restore the original values. Subsequently, utilizing the predicted values and actual values, model performance is evaluated using various performance metrics: Mean Square Error (MSE), Root Mean Square Error (RMSE), and Correlation Coefficient (CC).

2.4 Model Performance Hypothesis:

The statistical significance of unique mean performance across all four models is demonstrated through the Friedman test, followed by a post-hoc analysis using the Nemenyi Test (Lijie Chen et.al 2025)(Hamza M 2024). This study incorporates multiple forecasting models utilizing the same dataset; consequently, the Friedman test is employed to test the hypothesis. The test posits that empirical evidence from Mean Squared Error (MSE) calculations suggests at least one neural network architecture among LSTM, BILSTM, SimpleRNN, and GRU exhibits statistically significant performance compared to the other models. The hypothesis is mathematically expressed as: H0: The median ranks of the MSE for all models are equal, and Ha: The median ranks of the MSE for at least one model are different. The proposed hypothesis utilizes median ranks because the Friedman test is a non-parametric test that assigns ranks to MSE values and calculates the average of ranks. The test then computes a statistic based on these rankings to determine significant model differences. To establish the significance of the Friedman test, a post-hoc analysis (Nemenyi Test) is conducted to obtain pairwise comparison results, visualized using a heatmap. The post-hoc test evaluates the critical distance factor as follows:

$$CD = q_{\alpha}\sqrt{\frac{k(k+1)}{6n}} \qquad \text{.......} (3)$$

Where:q_{α}: Critical value from the Studentized range distribution (depends on the confidence level and number of models), k: Number of model, n: Number of trials.

For each pair of models, calculate the absolute difference in their average ranks, which is referred to as the p-value for post-hoc analysis (Nemenyi Test):

$$|R_i - R_j| \qquad \text{.......} (4)$$

The aforementioned difference value is compared as $|R_i - R_j| > CD$ which shows the difference is significant; otherwise, it is insignificant. The difference value is compared as significant or not significant. Statistical significance for each pairwise comparison is determined by analyzing rank difference magnitudes relative to the established critical distance. Figure 3 shows the flowchart for the Friedman test, followed by the post-hoc test.

3 Results:

The research aims to identify the optimal model for disease forecasting. All four models underwent training with varying epochs: 100, 200, and 300, with neuron counts of 64, 128, and 256, utilizing k-fold validation (k=4 and k=5). Table 6(a-d) presents all trials for forecasting Alternaria Leaf Blight using four models and their respective performance metrics.

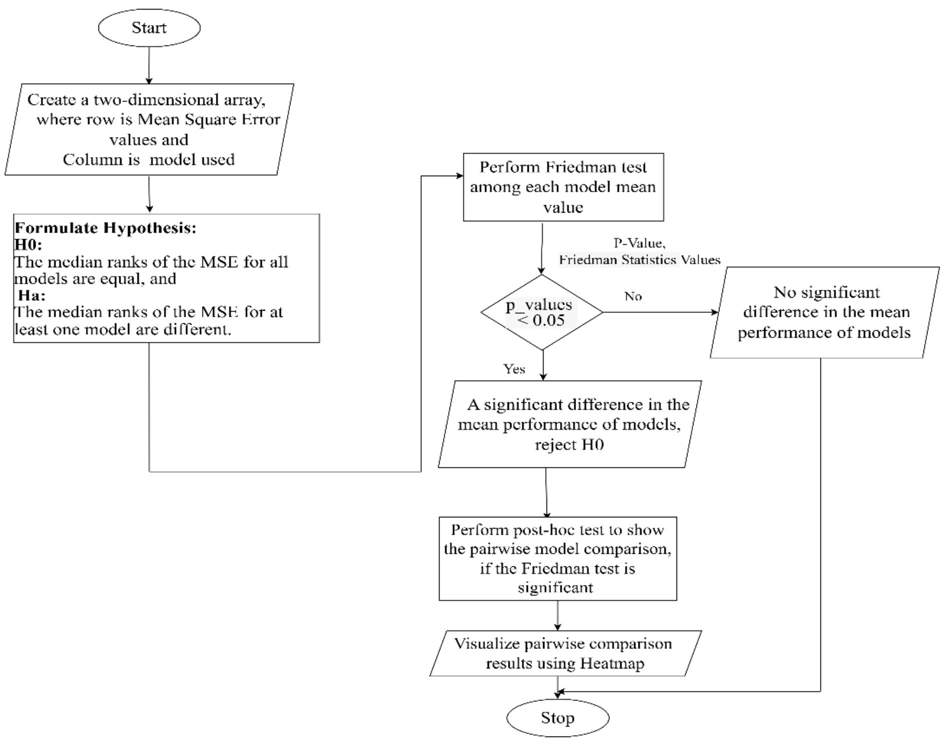


Figure 3: Flowchart for Friedman Test followed by post-hoc test

Table 6 a): Alternaria Leaf Blight forecasting using SimpleRNN

Model	Kfolds	Neurons	Epocs	MSE	RMSE	CC
SimpleRNN Batch size 32 Dropout 0.5	4	64	100	7.93	2.82	1.00
			200	14.08	3.75	1.00
			300	13.39	3.66	1.00
		128	100	12.48	3.53	1.00
			200	0.11	0.34	1.00
			300	3.24	1.80	1.00
		256	100	0.29	0.54	1.00
			200	0.62	0.79	1.00
			300	2.79	1.67	1.00
	5	64	100	15.89	3.99	1.00
			200	17.71	4.21	1.00
			300	17.12	4.14	1.00
		128	100	8.53	2.92	1.00
			200	26.36	5.13	1.00
			300	8.03	2.83	1.00
		256	100	0.58	0.76	1.00
			200	0.98	0.99	1.00
			300	11.42	3.38	1.00

Table 6 b): Alternaria Leaf Blight forecasting using LSTM

Model	Kfolds	Neurons	Epocs	MSE	RMSE	CC
LSTM Batch size 32 Dropout 0.5	4	64	100	10.28	3.21	0.99
			200	14.91	3.86	0.99
			300	9.16	3.03	0.99
		128	100	22.27	4.72	0.99
			200	17.74	4.21	0.99
			300	26.34	5.13	0.99
		256	100	5.21	2.28	0.99
			200	1.86	1.36	0.99
			300	1.91	1.38	0.97
	5	64	100	16.82	4.10	1.00
			200	26.18	5.12	1.00
			300	24.34	4.93	1.00

		128	100	18.98	4.36	1.00
			200	7.93	2.82	0.99
			300	8.39	2.90	0.98
		256	100	4.39	2.10	0.99
			200	8.52	2.92	0.96
			300	18.44	4.29	0.99

Table 6 c): Alternaria Leaf Blight forecasting using BILSTM

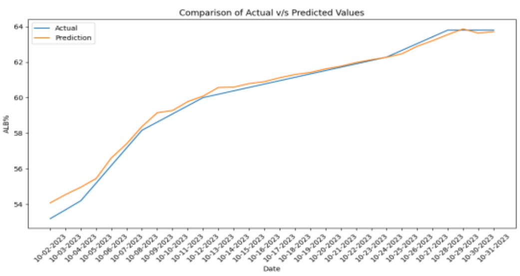
Model	Kfolds	Neurons	Epocs	MSE	RMSE	CC
BILSTM Batch size 32 Dropout 0.4	4	64	100	5.02	2.24	0.99
			200	7.00	2.65	0.99
			300	0.35	0.59	1.00
		128	100	2.12	1.46	0.99
			200	3.61	1.90	1.00
			300	0.89	0.94	0.99
		256	100	2.05	1.43	0.99
			200	0.52	0.72	0.99
			300	0.56	0.75	0.98
	5	64	100	6.06	2.46	0.99
			200	2.46	1.57	0.98
			300	0.33	0.57	0.99
		128	100	2.11	1.45	0.99
			200	4.26	2.06	0.98
			300	1.31	1.14	0.98
		256	100	1.32	1.15	0.98
			200	10.52	3.24	0.99
			300	5.02	2.24	0.99

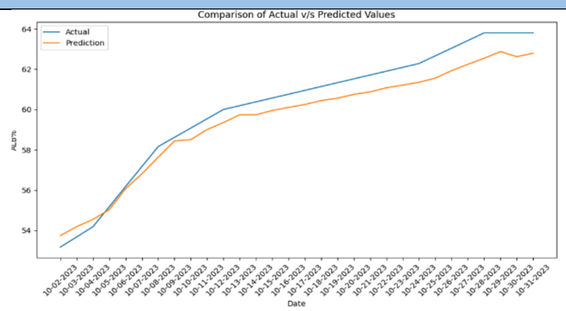
Table 6 d): Alternaria Leaf Blight forecasting using GRU

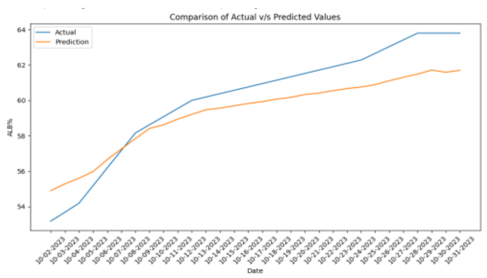
Model	Kfolds	Neurons	Epocs	MSE	RMSE	CC
GRU Batch size 32 Dropout 0.5	4	64	100	12.72	3.57	1.00
			200	21.57	4.64	1.00
			300	4.33	2.08	1.00
		128	100	8.16	2.86	1.00
			200	8.23	2.87	1.00

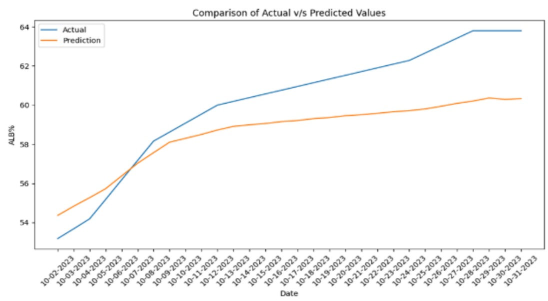
			300	10.80	3.29	1.00
		256	100	0.21	0.46	0.99
			200	2.09	1.45	0.99
			300	1.19	1.09	0.99
	5	64	100	22.18	4.71	1.00
			200	13.11	3.62	1.00
			300	16.39	4.05	1.00
		128	100	4.21	2.05	1.00
			200	2.45	1.57	1.00
			300	0.96	0.98	0.99
		256	100	2.27	1.51	1.00
			200	1.06	1.03	1.00
			300	2.28	1.51	1.00

Figure 8(a-h) shows the plots of the best models, with a smaller mean square error and a 30-day prediction of Alternaria Leaf Blight (Kharif Season).

<table>
<tr><th>Model</th><th>Forecasting plots at K4</th><th>Forecasting plots at K4</th></tr>
<tr><td>Simple RNN</td><td>


(a)128 neurons, 100 epochs</td><td>


(b 256 neurons,100 epochs</td></tr>
<tr><td>LSTM</td><td>


(c) 256 neurons,200 epochs</td><td>


(d) 256 neurons,100 epochs</td></tr>
</table>

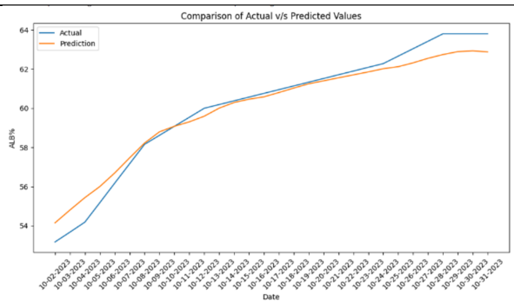
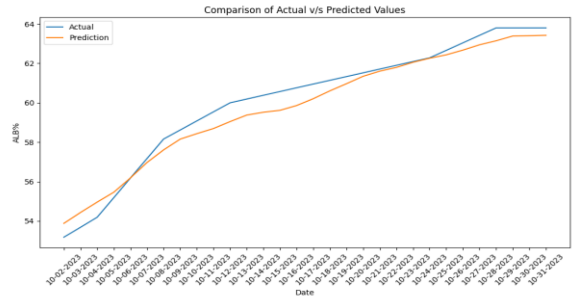
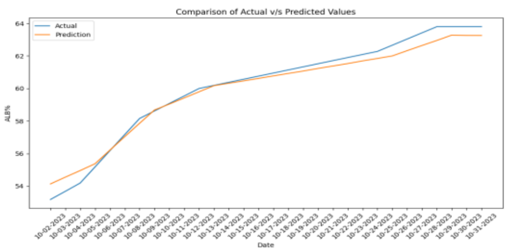
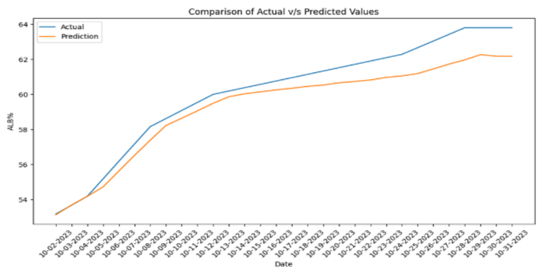
BILSTM	 (e) 64 neurons,300 epochs	 (f) 64 neurons,300 epochs
GRU	 (g) 256 neurons,100 epochs	 (h) 128 neurons,300 epochs

Figure 8(a-h): Forecasting for the subsequent 30-day period of October 2023 is presented on the x-axis, with the Alternaria Leaf Blight Disease value (in %) on the y-axis. The observed and predicted values are represented by blue and orange lines, respectively. Similarly, for Powdery Mildew, all four models were trained with varying epochs: 100, 200, and 300, with neuron counts of 64, 128, and 256, and with k-fold cross-validation (k=4 and k=5). Table 7(a-d) presents all trials for forecasting Powdery Mildew (Rabi Season) utilizing four models and their corresponding performance metrics.

Table 7 a): Powdery Mildew forecasting using SimpleRNN

Model	Kfolds	Neurons	Epocs	MSE	RMSE	CC
SimpleRNN Batch size 32 Dropout 0.5	4	64	100	1.54	1.24	0.98
			200	3.61	1.90	0.94
			300	4.15	2.04	0.93
		128	100	0.77	0.88	0.99
			200	3.04	1.74	0.95
			300	5.34	2.31	0.91
		256	100	1.19	1.09	0.98
			200	4.26	2.06	0.92
			300	7.45	2.73	0.86
	5	64	100	1.15	1.07	0.98
			200	2.89	1.70	0.95
			300	4.12	2.03	0.95

		128	100	1.40	1.18	0.97
			200	4.63	2.15	0.92
			300	6.07	2.46	0.90
		256	100	3.08	1.75	0.94
			200	5.97	2.44	0.89
			300	6.91	2.63	0.87

Table 7 b) Powdery Mildew forecasting using LSTM

Model	Kfolds	Neurons	Epocs	MSE	RMSE	CC
LSTM Batch size 32 Dropout 0.5	4	64	100	2.97	1.72	0.97
			200	1.31	1.14	0.99
			300	0.32	0.56	0.99
		128	100	2.18	1.48	0.98
			200	0.64	0.80	0.99
			300	0.66	0.81	0.99
		256	100	2.36	1.54	0.97
			200	1.08	1.04	0.98
			300	1.09	1.04	0.98
	5	64	100	1.61	1.27	0.99
			200	0.79	0.89	0.99
			300	0.70	0.83	0.99
		128	100	2.24	1.50	0.98
			200	0.51	0.72	0.99
			300	0.91	0.95	0.98
		256	100	1.29	1.14	0.98
			200	1.12	1.06	0.99
			300	0.63	0.79	0.99

Table 7 c) Powdery Mildew forecasting using BILSTM

Model	Kfolds	Neurons	Epocs	MSE	RMSE	CC
BILSTM Batch size 32 Dropout 0.4	4	64	100	3.44	1.86	0.96
			200	1.42	1.19	0.98
			300	2.15	1.46	0.97
		128	100	4.53	2.13	0.94
			200	2.64	1.63	0.96

			300	3.73	1.93	0.95
		256	100	9.47	3.08	0.83
			200	4.66	2.16	0.93
			300	5.32	2.31	0.92
	5	64	100	4.03	2.01	0.96
			200	3.42	1.85	0.97
			300	1.92	1.39	0.98
		128	100	4.00	2.00	0.93
			200	2.32	1.52	0.96
			300	8.09	2.84	0.89
		256	100	6.28	2.51	0.93
			200	4.19	2.05	0.96
			300	3.99	1.99	0.94

Table 7 d) Powdery Mildew forecasting using GRU

Model	Kfolds	Neurons	Epocs	MSE	RMSE	CC
GRU Batch size 32	4	64	100	2.34	1.53	0.98
			200	0.99	0.99	0.99
			300	1.62	1.27	0.99
		128	100	0.67	0.82	0.99
			200	0.71	0.85	0.99
			300	1.68	1.30	0.98
		256	100	1.33	1.15	0.98
			200	1.42	1.19	0.97
			300	1.53	1.24	0.98
	5	64	100	0.44	0.66	0.99
			200	0.46	0.68	0.99
			300	1.02	1.01	0.99
		128	100	1.24	1.11	0.99
			200	0.58	0.76	0.99
			300	0.48	0.69	0.99
		256	100	0.63	0.80	0.99
			200	1.17	1.08	0.99
			300	0.40	0.64	0.99

Figure 9(a-h) shows the plots of best models having Means Square Error smaller and with 30 days prediction of Powdery Mildew (Rabi Season)

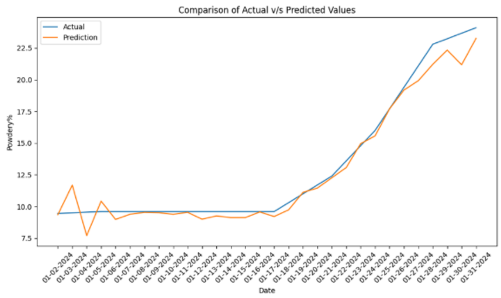
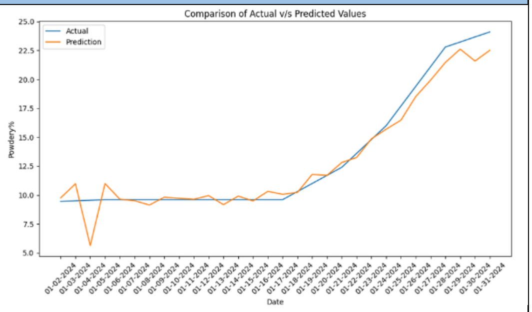
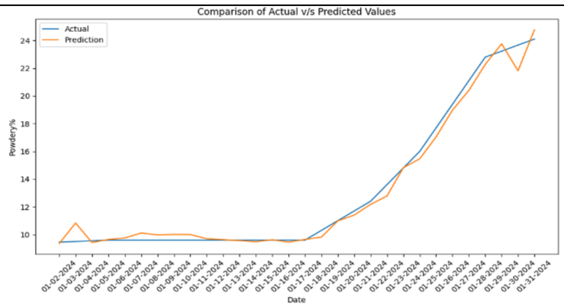
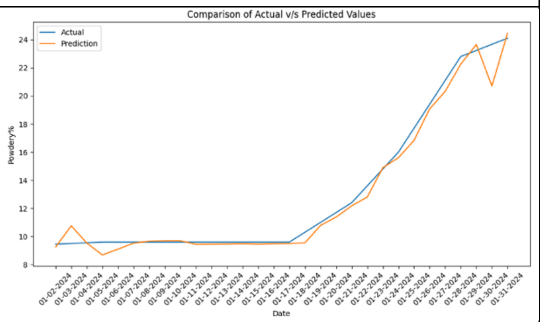
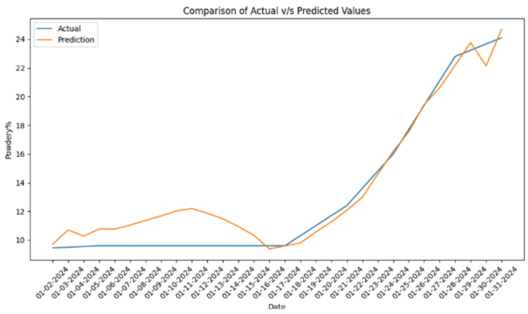
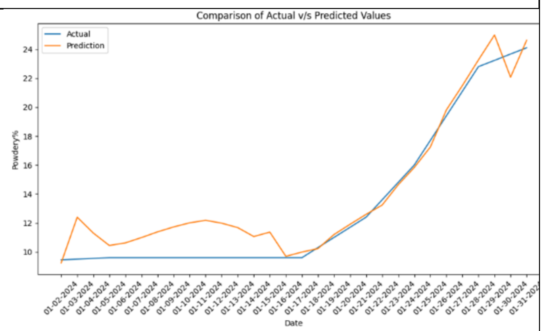
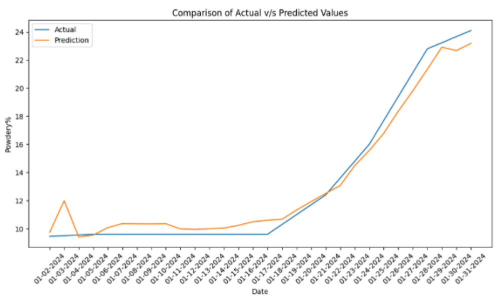
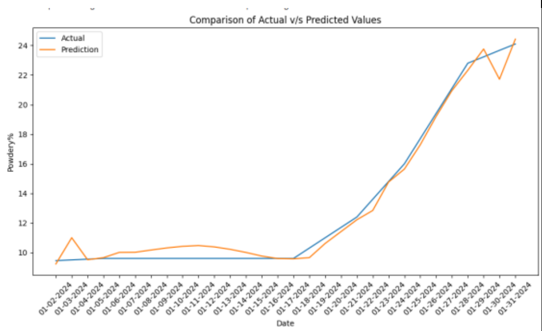
Model name	Forecasting plots at K4	Forecasting plots at K4
Simple RNN	 (a) 128 neurons,100 epochs	 (b) 64 neurons,100 epochs
LSTM	 (c) 64 neurons, 300 epochs	 (d)128 neurons,200 epochs
BILSTM	 (e) 64 neurons,200 epochs	 (f) 64 neurons ,300 epochs
GRU	 (g) 128 neurons,100 epochs	 (h) 256 neurons,300 epochs

Figure 9(a-h): Forecasting of the next 30-day period of January 23 on the x-axis and the Powdery Mildew Disease value (in %) on the y-axis, respectively. The actual values and predicted values are depicted using a blue line and an orange line, respectively.

4. Discussions:

After applying various models, the performance results for Alternaria Leaf Blight disease are presented in Table 6 (a-d). The optimal model prediction results are highlighted in yellow, while the suboptimal performance of models is highlighted in green. Table 8 provides a summary of the most effective models. For Alternaria Leaf Blight, SimpleRNN at k=4 yielded a lower mean square error, whereas LSTM at k=4 produced a higher mean square error. Figure 10 illustrates a comprehensive plot of performance metrics in relation to the models.

Table 8: Summary of performance metrics for Alternaria Leaf Blight

Model	MSE	RMSE	CC
SimpkeRNN K=4	0.11	0.34	1.00
SimpleRNN K=5	0.58	0.76	1.00
LSTM K=4	1.86	1.36	0.99
LSTM K=5	4.39	2.10	0.99
BiLSTM K=4	0.35	0.59	1.00
BiLSTM K=5	0.33	0.57	0.99
GRU K=4	0.21	0.46	0.99
GRU K=5	0.96	0.98	0.99

For Powdery Mildew, the model performance is presented in Table 7 (a-d). The optimal prediction results are highlighted in yellow, while suboptimal results are highlighted in green. Table 9 provides a summary of the most effective models. In the case of Powdery Mildew, LSTM at k=4 yielded a lower mean square error, whereas BILSTM at k=5 produced a higher mean square error. Figure 11 illustrates a comprehensive plot of performance metrics in relation to the various models.

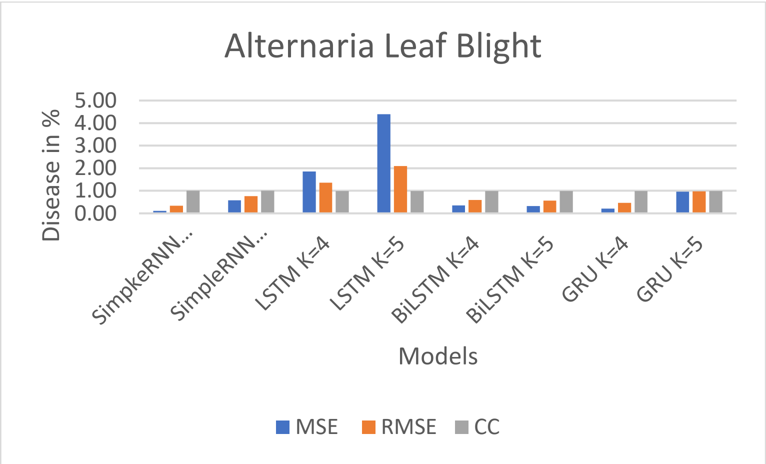


Figure 10: MSE, RMSE, and CC values of models for Alternaria Leaf Blight

Table 9: Summary of performance metrics for Powdery Mildew

Model	MSE	RMSE	CC
SimpkeRNN K=4	0.77	0.88	0.99
SimpleRNN K=5	1.15	1.07	0.98
LSTM K=4	0.32	0.56	0.99
LSTM K=5	0.51	0.72	0.99

BiLSTM K=4	1.42	1.19	0.98
BiLSTM K=5	1.92	1.39	0.98
GRU K=4	0.67	0.82	0.99
GRU K=5	0.40	0.64	0.99

The performance of the four models shows statistically significant differences, demonstrated by the Friedman test followed by the Nemenyi post-hoc test. Results are visualized using a

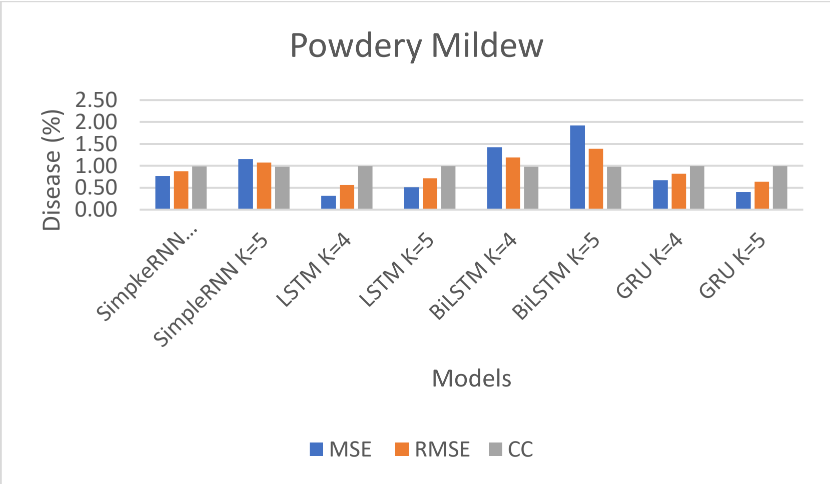


Figure 11: MSE, RMSE, and CC values of models for Powdery Mildew

heatmap illustrating pairwise comparison of models' mean performance. This study uses a critical value of α=0.05 with the two-tailed test. If p-value < 0.05, the proposed alternate hypothesis (in Figure 3) is accepted, and the Nemenyi Post-Hoc Test elucidates significant differences between each model's performance. For Alternaria Leaf Blight, regarding mean square error values at K=4, a significant difference is observed between the mean performance of SimpleRNN, LSTM, BILSTM, and GRU models, with a Friedman test statistic value of 12.5999 and p-value of 0.00558 (p-value < 0.05). For Powdery Mildew, the mean performance of these models yields a Friedman test statistic value of 11.8000 and p-value of 0.0081 (p-value < 0.05). The heatmap of p-values illustrates pairwise comparison of models with p-value < 0.05, indicated by green color, demonstrating acceptance of the proposed hypothesis shown in Table 10,11,12,13.

Table 10: Alternaria Leaf Blight heatmap for K=4 MSE values

Model	LSTM	BILSTM	GRU	SimpleRNN
LSTM	1	0.0029	0.0069	0.03543
BILSTM	0.0029	1	0.0082	0.02611
GRU	0.0069	0.0082	1	0.0095
SimpleRNN	0.0035	0.02611	0.0095	1

Table 11: Alternaria Leaf Blight heatmap for K=5 MSE values

Model	LSTM	BILSTM	GRU	SimpleRNN
LSTM	1	0.0103	0.0313	0.0262
BILSTM	0.0103	1	0.0098	0.0057

GRU	0.0313	0.0098	1	0.0197
SimpleRNN	0.0262	0.0057	0.0197	1

Table 12: Powdery Mildew heatmap for K=5 MSE values

Model	LSTM	BILSTM	GRU	SimpleRNN
LSTM	1	0.0314	1	0.0182
BILSTM	0.0314	1	0.0313	0.0884
GRU	1.00	0.0313	1	0.0182
SimpleRNN	0.0184	0.0884	0.0184	1

Table 13: Powdery Mildew heatmap for K=5 MSE values

Model	LSTM	BILSTM	GRU	SimpleRNN
LSTM	1	0.0314	0.0079	0.0125
BILSTM	0.0314	1	0.0015	0.0094
GRU	0.0079	0.0015	1	0.0103
SimpleRNN	0.0125	0.0094	0.0103	1

5. Conclusion:

This study proposes sunflower disease forecasting using Recurrent Neural Network variants. To forecast disease prevalence, five features were used to elucidate the relationship between disease percentage and weather parameters. The dataset for Alternaria Leaf Blight and Powdery Mildew was obtained from the Latur Oilseed Research Center. Alternaria Leaf Blight values were collected for the Kharif period (July to October), and Powdery Mildew values for the Rabi Period (October to January). The dataset underwent preprocessing and normalization. Four models (Simple RNN, LSTM, BILSTM, and GRU) were trained and evaluated using K-fold cross-validation. Results demonstrated disease forecasting for the subsequent 30 days in both periods. For Alternaria Leaf Blight, SimpleRNN performed optimally with a lower MSE value of 0.11, while LSTM showed superior performance for Powdery Mildew with a lower MSE value of 0.32. The significant mean performance of paired models differed for both diseases, with p-values illustrated in the heatmap. This outcome supports the proposed hypothesis using the Friedman Test, followed by a post-hoc test. Future research may enable disease forecasting for extended periods by acquiring datasets from additional historical years. Forecasting model accuracy can be enhanced by incorporating factors such as soil temperature, soil moisture, unexpected weather fluctuations, and the diverse effects of global warming on climate.

References:

[1] Agrovista Farming (2024), Origin and History of the sunflower crop | Sunflower Cultivation in India; A complete information Guide. Wordpress. https://agrovistafarming.com/2021/01/sunflower-cultivation-in-india-a-complete information-guide/field-crops/ Accessed on 29th August 2024.

[2] Joksimovic J, Atlagic Jovanka, Marinkovic, R. and Jovanovi, D, (2006), Genetic control of oleic and linoleic acid contents in sunflower. Helia, 29(44): 33-40. https://doi.org/10.2298/hel0644033j

[3] Shivani R(2023), Diseases of Sunflower: Necrosis, Leaf Blight, Mildew and Other Diseases, Biology Discussions, https://www.biologydiscussion.com/plants/plant-diseases/diseases-of-sunflower-necrosis-leaf-blight-mildew-and-other-diseases/43079 Accessed on : 2023

[4] A. Bheemaraya, M.M. Jamadar and Shalini Huilgol (2018) Correlation between Weather Parameters and Sunflower Powdery Mildew Caused by Erysiphe cichoracearum DC, International Journal of Current

Microbiology and Applied Sciences ISSN: 2319-7706 Volume 7 Number 09. https://doi.org/10.20546/ijcmas.2018.709.434.

[5] Rakesh Kaundal, Amar S Kapoor & Gajendra PS Raghava (3rd Nov 2006) Machine Learning techniques in disease forecasting: case study on Rice blast prediction BMC Bioinformatics, Methodology article, Article number: 485. https://doi.org/10.1186/1471-2105-7-485.

[6] V. B. AKASHE, J. D. JADHAV, V. R. BAVADEKAR, P. B. PAWAR and V. M. AMRUTSAGAR, (June 2016), Forewarning model for sunflower thrips (Thrips palmi Karny) in western Maharashtra scarcity zone" Journal of Agrometeorology 18(1):68-70 June 2016 DOI:10.54386/jam.v18i1.902.

[7] Katti, P., Naganagoud, A., Reddy, M. and Chandranath, H. T. (2011). Seasonal incidence of sunflower thrips and their relationship with weather factors. Plant Archives., 11(1) 339-341.DOI:10.54386/jam.v18i1.902

[8] N.S. PANKAJA, G.V. HARISH BABU1 AND NAGARAJU (January 2011) Epidemiology of sunflower necrosis disease International Journal of Agricultural Sciences, Vol. 7 Issue 1147-151.

[9] Vijaykumar N Ghante, Shantappa Duttarganvi, Umesh MR and Vikas Kulkarni (October 2020) Influence of weather parameters on population of whitefly, Bemisia tabaci Bemisia tabaci in Sunflower ,Journal of Entomology and Zoology Studies, JEZS 2020; 8(6):1729-1734,E-ISSN: 2320-7078 P-ISSN: 2349-6800.

[10] Kittakorn Sriwanna, Weather-based rice blast disease forecasting (11th January 2022) School of Computer and Information Technology, Chiang Rai Rajabhat University, Chiang Rai 57100, Thailand, 0168-1699, https://doi.org/10.1016/j.compag.2022.106685.

[11] Manoj Chhetri, Sudhanshu Kumar, Partha Pratim Roy and Byung-Gyu Kim,Deep (2020) BLSTM-GRU Model for Monthly Rainfall Prediction: A Case Study of Simtokha, Bhutan , *Remote Sens*, *12*(19), 3174; https://doi.org/10.3390/rs12193174.

[12] Ya Gao, Rong Wang, and Enmin Zhou (2021) Stock Prediction Based on Optimized LSTM and GRU Model, Hindawi ,Scientific Programming , Article ID 4055281, 8 pages, https://doi.org/10.1155/2021/4055281

[13] Guilin Huang Missing data filling method based on linear interpolation and lightgbm (2021), Journal of Physics: Conference Series, ISPECE 2020, IOP Publishing, 1754 012187, doi:10.1088/1742-6596/1754/1/012187.

[14] Tsantekidis, A.; Passalis, N.; Tefas, A (2022) A. Recurrent Neural Networks. In Deep Learning for Robot Perception and Cognition, Elsevier: Amsterdam, The Netherlands; pp. 101–11

[15] Mienye, I.D.; Jere, N.(2024) Deep Learning for Credit Card Fraud Detection: A Review of Algorithms, Challenges, and Solutions. IEEE Access doi: 10.1109/ACCESS.2024.3426955

[16] Mienye, I.D.; Sun, Y. (2023) A machine learning method with hybrid feature selection for improved credit card fraud detection. Appl. Sci. https://doi.org/10.3390/app13127254.

[17] Ibomoiye Domor Mienye, Theo G. Swart and George Obaido, (2024) Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications, Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications. Information , https://doi.org/10.3390/info15090517.

[18] Benjamin Lindemann, Timo Müller, Hannes Vietz, Nasser Jazdi, Michael Weyrich (2021) A survey on long short-term memory networks for time series prediction Procedia CIRP Volume 99, Pages 650-655 https://doi.org/10.1016/j.procir.2021.03.088

[19] Poluru Eswaraiah and Hussain Syed (2023) A Hybrid Deep Learning GRU based Approach for Text Classification using Word Embedding EAI Endorsed Transactions on Internet of Things DOI:10.4108/eetiot.459

[20] Lijie Chen , Daofei Li , TaoWang , Jun Chen and Quan Yuan (2025) Driver Takeover Performance Prediction Based on LSTM-BiLSTM-ATTENTION Model, MDPI Article, System,https://doi.org/10.3390/systems13010046.

[21] Hamza Mubarak Abdallah Abdellatif , Shameem Ahmad,Mohammad Zohurul Islam, S.M. Muyeen, Mohammad Abdul Mannan,Innocent Kamwa (2024) Day-Ahead electricity price forecasting using a CNN-BiLSTM model in conjunction with autoregressive modeling and hyperparameter optimization International Journal of Electrical Power & Energy SystemsVolume 161, 110206 https://doi.org/10.1016/j.ijepes.2024.110206

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The assistance of Dr. R.D. Prasad, Principal Scientist of Plant Pathology at ICAR-Indian Institute of Oilseeds Research, was indispensable to completing this study. I am grateful to Dr. Maharudra Ghodke, former head of the Latur Oilseed Research Centre, Santosh Waghmare, senior plant pathologist, and Sangita Aradwad, junior plant pathologist.