Agronomy Research **20**(S1), 1117–1133, 2022
https://doi.org/10.15159/AR.22.076

Prediction of fruit rot disease incidence in Arecanut based on weather parameters

K. Rajashree[1,*], K.V. Prema[2], G. Rajath[3] and S. Angad[1]

[1]Manipal Academy of Higher Education, Manipal Institute of Technology, Department of Computer Science and Engineering, Madhav Nagar, Manipal, IN576104 Karnataka, India
[2]Manipal Academy of Higher Education, Manipal Institute of Technology, Department of Computer Science and Engineering, BSF Campus, Yelahanka, Bangalore, IN560064 Karnataka, India
[3]PES University, PES Institute of Technology, Department of Electronics and Communication, Dwaraka Nagar, Banashankari, Bangalore, IN560085 Karnataka, India
[*]Correspondence: raji.krish@manipal.edu

Received: July 19th, 2022; Accepted: October 20th, 2022; Published: November 22nd, 2022

Abstract. The occurrence of pests and diseases in arecanut crops has always been an important factor affecting the total production of arecanut. Arecanut is always dependent on environmental factors during its growth. Thus monitoring and early prediction of the occurrence of the disease would be very helpful for prevention and therefore more crop production. Here, we propose artificial intelligence-based deep learning models for fruit rot disease prediction. Historical data on fruit rot incidence in representative areas of arecanut production in Udupi along with historical weather data are the parameters used to develop region-specific models for the Udupi district. The fruit rot disease incidence score value is predicted using recurrent neural network variants (i.e., Vanilla LSTM, Vanilla GRU, stacked LSTM, and Bidirectional LSTM) for the first time. The predictive performance of the proposed models is evaluated by mean square error (MSE) along with the 5-fold cross-validation technique. Further, compared to other deep learning and machine learning models, the Vanilla LSTM model gives 1.5 MSE, while the Vanilla GRU model gives 1.3 MSE making it the best prediction model for arecanut fruit rot disease.

Key words: crop disease, deep learning, prediction, cross-validation, fruit rot disease.

INTRODUCTION

The different technologies like robotics platforms, the Internet of things, and remote sensing belonging to the industry are now applied to solve agriculture challenges, leading to intelligent farming. Smart farming is essential to meet consumer demands, as well as for sustainability and productivity. According to the annual report of 2018–19 from the Department of Cooperation, Agriculture, and Farmers welfare, 54.6% of the entire workforce is affianced in allied and agricultural sector activities. Farming the arecanut is a financially rewarding endeavor for farmers. Over 90% of all arecanuts produced in India are produced in Karnataka, Assam, Kerala, Mizoram, Meghalaya, West Bengal, and Nagaland. Arecanut is exported worldwide and 700 million people depend on it for many reasons (Balanagouda et al., 2021).

Crop disease always depends on three factors: environmental fluctuation, host, and pathogen (Fenu & Malloci, 2021). Arecanut is exceptionally vulnerable to several pathogens like bacteria, fungi, and viruses. Heavy rainfall, humidity, and high-intensity rains reduce pollination and increase the disease rate reducing yield (Sujatha et al., 2018). Arecanut in India is especially threatened by fruit rot disease which causes a 10–70% loss in crop yield and sometimes leads to the tree collapsing in itself. The fruit rot disease (FRD) problem is escalating yearly because of the lack of human resources available to spray fungicides. This increases the disease severity and the ailment spreads to nearby regions. Hence, the scattering of diseases was assessed across many locations and the risk correlated with FRD (Balanagouda et al., 2021), which will help the extension officers and farmers to take control actions and prevent the further expropriation of FRD into adjacent places. Farmers rely more on forewarning systems at an initial stage as they can avoid the roll-out of disease and also minimize chemical usage. Literature shows that arecanut is comprehensively used in medicinal practice in India, China, and other countries in Asia (Peng et al., 2015). It is used as traditional medicine for abdominal health issues and for treating parasitic diseases, gastrointestinal diseases, and edematous diseases.

Deep learning is a modern tool that was later integrated with the agriculture domain, which has already succeeded in other disciplines. The applications of deep learning in agriculture are surveyed by a few authors (Kamilaris & Prenafeta-Boldú, 2018; Santos et al., 2019; Ren et al., 2020). They summarized that deep learning models give superior performance in terms of accuracy when compared with machine learning models. Image-based disease detection, weed detection, land cover classification, and yield prediction are some of the more common uses of DL. Even then, it is rarely used for early disease detection in crops based on weather. Recent research states that long short-term memory (LSTM) performs better in sequential data processing when compared with deep feedforward neural networks (Kim et al., 2017).

Researchers have developed a variety of weather-based crop disease forecasting models using deep learning techniques. Along with that, many crop disease detection models are developed using images. Analysis of agriculture big data through machine learning and deep learning has emerged lately. The analysis and classification of different algorithms used recently for plant and crop disease forecasting are explored (Fenu & Malloci, 2021). The study categorizes the forecasting models into image-based data, weather-based information, and data from heterogeneous sources, in which weather-based forecasting is more examined. Support vector machine, artificial neural network (Malicdem & Fernandez, 2015; Fenu et al., 2019), LSTM (Kim et al., 2017), extreme learning machine(ELM) (Anshul Bhatia et al., 2020), random forest (Jawade et al., 2020), and multi-layer perceptron (de Oliveira Aparecido et al., 2020) techniques are used to predict crop and plant disease based on weather parameters. Similarly, CNN, MLP, RF, ELM, and support vector regression techniques are frequently used to forecast crop and plant disease based on images (Zhu et al., 2017; Duarte-Carvajalinoet al., 2018). Along with images and weather data, data from different sources - i.e. remotely sensed variables like vegetation index, water stress index, land surface temperature, leaf area index, and soil temperature - are used to predict crop and plant diseases in their early stages.

The enhanced LSTM model is used in the rice crops' pest attack forecasting with the help of weather data (Wahyono et al., 2020). Daily temperature, humidity, and rainfall data are the inputs given for the model over 16 years. Vanilla LSTM, stacked LSTM, and bidirectional LSTM models were compared with the sliding window concept. They found that bidirectional LSTM performed better due to the forward and backward processing of the data.

Rice blast disease (RBD) is forecasted before occurrence using LSTM RNN (Recurrent Neural Network) (Kim et al., 2017). To test the model, they considered 17 different rice varieties from 12 regions in South Korea. The proposed LSTM model is evaluated by varying the input parameters relative humidity, air temperature, sunshine hours, and rice blast disease score. Among several LSTM models, the BlastTHS LSTM model gives more accuracy because it contains all the input parameters to predict the disease in the Cheolwon region.

The LSTM model is used to predict the pest and disease occurrence in the cotton crop based on weather factors (Xiao et al., 2018; Xiao et al., 2019). The Apriori algorithm is used to find the association rules between the weather parameters and disease data. The advantages of the LSTM model are shown by comparing it with the support vector method (SVM) and random forest method. Relative humidity, rainfall, minimum and maximum temperature, wind speed, and sunshine hours are the parameters used for the model development along with 63 datasets of cotton pests and diseases.

A new deep LSTM method is proposed to predict the rice crop pest with the help of anomaly climate data (Wahyono et al., 2021). In this study, two prediction models are developed and prediction results are compared. Among the two models, the first model uses climate data with anomaly as input, and the second model uses data without anomaly. Good accuracy was seen in the first model when experimented with 100 epochs. An attempt was made to predict the rice crop disease using an artificial intelligence-based hybrid CNN (Convolution Neural Network)-LSTM model (Jain & Ramesh, 2021). The study presents both region-specific prediction and classification models, where SVM is used to classify yellow stem border disease in rice crops. The model takes relative humidity, temperature, rainfall, evaporation, sunshine hours, and pest value captured from the light trap as validation parameters.

Recently, ANN and SVM learning models were developed to forecast the incidence and severity of RBD to prevent or mitigate the escalating of the disease (Malicdem & Fernandez, 2015). The PCA algorithm is used to select the features contributing more to disease occurrence. SVM gives a more accurate prediction when the result is compared with the ANN model. A system was proposed to predict the RBD based on weather parameters like rainfall, temperature, wind speed, wind gust, sun hours, visibility, humidity, and UV index (Sriwanna, 2022). The agriculture office has documented the RBD occurrence between 2013 and 2019. The class 1 label is given to the recorded disease data elements but the class 0 label is given to ‘no disease’ data elements. After combining the blast data and weather data, they used ensemble feature rankings to select the weather features that have a meaningful effect on disease occurrence. The classification performance was measured using the F1 score, ROC AUC score, balanced accuracy, and geometric mean over DT, KNN, SVM, NB, and MLP classification algorithms using the top 10 features.

LSTM and multilayer perceptron algorithms are used to develop a predictive model for RBD in the Davangere region (Varsha et al., 2021). The data for blast disease is considered based on the literature study and climate data collected from the Karnataka state natural disaster monitoring station. Hyperparameters are adjusted and based on the readings shown by the authors, the final results have outstanding accuracy. Image-based deep convolution encoder network model is proposed to predict and classify the diseases in different seasonal crops (Khamparia et al., 2020). The crop leaf images from the Plant Village dataset are considered for the experiment for five various diseases. Due to the combination of autoencoders and CNNs, the model is in hybrid mode. This network performed better than the conventional techniques.

A CNN model was developed to identify the tomato crop disease using leaf images. In comparison to different machine learning models like Naive Bayes, kNN, and decision trees, the proposed model has 98.4% accuracy, which is relatively high (Agarwal et al., 2020). The proposed model is tested using a different dataset from other domains, and additional model variants are used in the experiments. Accuracy, F1 score, and AUC-ROC metrics are used to measure the model's achievement. Rice diseases are identified through leaf images using the deep CNN model (Lu et al., 2017). The proposed model achieves 95.48% accuracy under a 10-fold cross-validation strategy, which is relatively high compared to conventional machine learning models such as SVM and particle swarm optimization. 10 different rice diseases are identified from the model. Apple scab disease was detected at its early stage using CNN and transfer learning methods (Kodors et al., 2021). The imbalanced dataset containing images of apple fruit and apple leaves is considered for the experiment.

A summary was made on detecting and categorizing various arecanut diseases using image processing and machine learning methods (Puneeth et al., 2021). The results of the study indicate that not much work is being done on the early identification of arecanut disease. The authors found no weather-based prediction model for arecanut disease as part of their review process. Authors have also detected and classified wheat disease using improved deep convolution architecture (Goyal et al., 2021). The spike and leaf parts are the most affected in the wheat plant. The proposed method detects nine diseases and one healthy class from the Large wheat disease classification dataset 2020. The proposed CNN model is compared with VGG16 and RESNET50 CNN models and found high testing accuracy of 97.88%.

Current state-of-the-art ML models are applied to predict the fruit rot disease in arecanut crops after data collection (Krishna et al., 2022). The dataset is created by integrating disease data and weather data. The decision tree regression (DTR), multilayer perceptron regression (MLPR), random forest regression (RFR), and support vector regression (SVR) models are used to predict the disease. RFR model gives the best performance with 0.9 as Mean Absolute Error and 1.9 as Mean Square Error.

The development of image-based arecanut crop disease identification models has recently gained much traction. For instance, there was an identification of different arecanut diseases using a multi-gradient direction-based deep learning model (Mallikarjuna et al., 2021; Mallikarjuna et al., 2022). In the proposed method, a multi-Sobel directional kernel is applied to each input image to generate multi-gradient directional images. These images are given as input to the ResNet CNN architecture for disease identification. The author has generated the dataset with a total of 281 images including healthy, rot, split, and rot-split images. Precision, F-measure, and recall performance metrics are used to evaluate the generated results.

The literature shows that disease identification and prediction models are developed for different plants and crops. To list citrus, cucumber (Liu et al., 2022), potato, olive, apple, tomato, tobacco, mango, barley, coffee, cotton, grape, rice, wheat, orange, strawberry, oil palm, ginger, and sugar beet can be considered. The main contribution of this research work is as follows:

- Even though many prediction models exist in the crop disease management domain for various crops, there is no deep learning-based prediction model for the arecanut crop. Hence the present study is the first effort to develop weather-based arecanut crop disease forecasting using deep learning models.
- A novel aspect of this study is integrating historical weather data with arecanut crop disease data to create a unique dataset.
- The performance of four deep learning methods is compared and analyzed.

Consequently, sincere effort and research are required to develop novel solutions to prevent or mitigate the effect of crop disease on yield loss at an early stage.

MATERIALS AND METHODS

The flow diagram for the proposed work is shown in Fig. 1. First, the disease and weather data are collected from different sources and integrated. Second, the data is pre-processed to make it fit for further processing. After this, different deep-learning models are applied to predict the disease score value. Finally, based on the validation loss and training loss, all the models are compared and the results are analyzed.

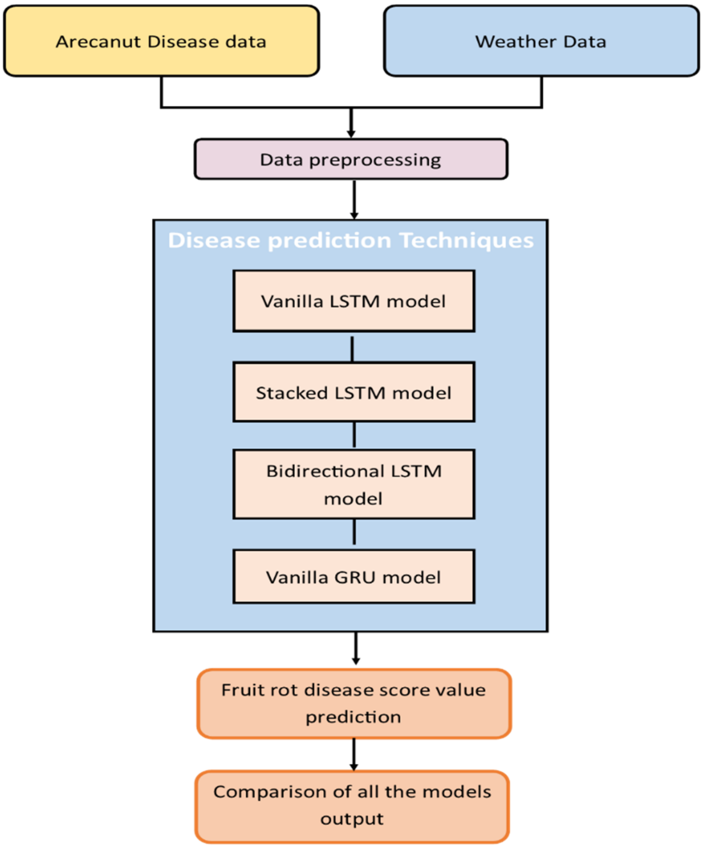

Figure 1. Flow diagram of the proposed work.

Sources of Data

In this research work, arecanut fruit rot disease data and historical climate data are used to generate an FRD forewarning model. The arecanut disease data is collected from literature resources, arecanut disease recommendations, and field surveys with nearly 60 farmers (Krishna et al., 2022) from the Udupi region in Karnataka state, India. Zone Agricultural and Horticultural Research Station, Brahmavar, Udupi has provided 21 years of weather data from 2000 to 2020. Only temperature, rainfall, humidity, sunshine hours, and cloud direction parameters are considered in the experiment. The data before pre-processing/integration is shown in Table 1 for September 2000.

Table 1. The historical weather data sample for September 2000 before preprocessing. The parameters are: Rainfall (RF), Maximum temperature (Max Temp), Minimum Temperature (Min. Temp), Relative Humidity in the morning (RH - I), Relative Humidity in the evening (RH - II), Cloud measure (I & II in oktas)

Date	RF	Max. Temp	Min. Temp	RH-I	RH-II	Cloud I	Cloud II	Sunshine Hours
01.09.2000	0	30	22	97	97	6	6	0
02.09.2000	0	30	22	98	95	8	7	0
03.09.2000	0	30	23	93	91	8	6	0
04.09.2000	0	30	22	100	97	8	6	0
05.09.2000	0	30	23	95	91	8	8	0
06.09.2000	0	31	22	92	90	8	8	8
07.09.2000	6	31	22	96	91	8	8	9
08.09.2000	12	31	23	94	91	8	6	9
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
29.09.2000	0	31	23	94	94	8	6	4
30.09.2000	0	30	25	92	95	8	8	9

Data Pre-processing

The FRD occurs due to heavy rainfall, high humidity, and low temperature. It is around May through October that the climate will be conducive to disease activity. Hence from 21 years of collected data, only these months are considered. Arecanut diseases usually directly depend on the weather pattern. Therefore, the disease and weather data are integrated as shown in Algorithm 1 to create the final dataset, which is used to develop the prediction model. This study represents the relationship between weather parameters and FRD in arecanut crops by predicting the score value. The score value varies from 0 to 35 based on the weather pattern. As the score value increases, there is a greater chance of disease outbreaks and spreading. The following rules are used to calculate the disease score value.

Algorithm 1: Rules used to calculate disease score value

```
If Rainfall > 15 mm and Temperature < 24 °C and Humidity > 90,
then Score value increments
if Rainfall > 5 mm and Sunshine > 5 hrs,
then Score value increments
if Rainfall < 10 mm and Temperature > 24 °C,
then Score value decreases.
In all other cases score value remains same.
```

After integrating/preprocessing the fruit rot disease data and weather data, the final data can be used to predict the fruit rot disease score value. It is possible to forecast disease severity for farmers based on the range of score values. The range from 0 to 15 indicates no disease, 15–25 indicates a medium risk of getting a fungal infection, 25–34 shows an occurrence of disease and a score value above 35 indicates a high severity of the infection. Once the infection starts it will spread to the entire farm with the help of wind. Table 2 shows the final integrated data after preprocessing for September 2000.

Table 2. The historical weather data sample for September 2000 and score value after integration/ preprocessing through algorithm 1

Date	RF	Max. Temp	Min. Temp	RH-I	RH-II	Cloud I	Cloud II	Sunshine Hours	Score value
01.09.2000	0	30	22	97	97	6	6	0	16
02.09.2000	0	30	22	98	95	8	7	0	15
03.09.2000	0	30	23	93	91	8	6	0	14
04.09.2000	0	30	22	100	97	8	6	0	13
05.09.2000	0	30	23	95	91	8	8	0	12
06.09.2000	0	31	22	92	90	8	8	8	11
07.09.2000	6	31	22	96	91	8	8	9	12
08.09.2000	12	31	23	94	91	8	6	9	13
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
29.09.2000	0	31	23	94	94	8	6	4	3
30.09.2000	0	30	25	92	95	8	8	9	2

Disease prediction Techniques

A recurrent neural network is the best technology to work with sequential time series data (Samarawickrama & Fernando, 2017; Zhang & Dong, 2020). Hence the present study has experimented with different variants of RNN, that include LSTM and GRU. A detailed explanation of the models is given in the next section.

LSTM model

Recently LSTM has become a state-of-the-art model for different time series prediction problems. It is a part of an RNN capable of learning long-term data dependencies, hence it is more useful in time series sequential analysis. The LSTM cell is shown in Fig. 2. It is also called a memory cell because it stores some information that is essential for the next decision. This LSTM cell contains three gates; each has its functionality; which provides read, write, and reset operations for the cell.

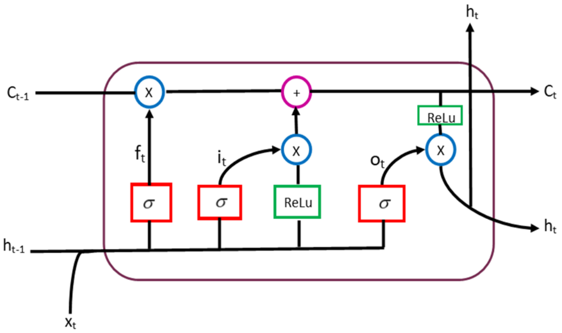

Figure 2. LSTM cell with three gates: input, output and forget.

Forget Gate: Decides what information has to be discarded from the cell.

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

Input Gate: Decides the values from the inputs to update the cell.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i \qquad (2)$$

Output Gate: Decides what to output based on the input and content of the cell.

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o \qquad (3)$$

In the equation, *W* is the weight given to the neuron, *h* is the output from the previous cell, *b* is the bias, C_t is the cell state and *x* is the input at time step t.

Vanilla LSTM model: As shown in Fig. 3(a), the model with a single hidden layer is called the vanilla LSTM model (Wu et al. 2018) used the vanilla LSTM model to estimate the remaining useful life of engineered systems.

Stacked LSTM model: In this model, multiple LSTM layers are stacked one after the other as shown in Fig. 3(b). According to the literature, stacked LSTM performs better than vanilla LSTM.

Bidirectional LSTM model: As shown in Fig. 3(c), in this model, instead of training a single model, two models are trained with forwarding LSTM using input sequence and backward LSTM using reversed input sequence. The output from the forward and backward LSTM is concatenated to feed to the dense layer.

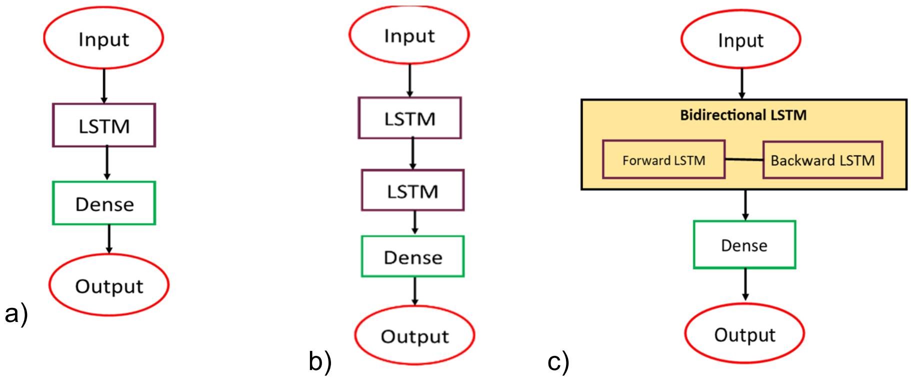

Figure 3. Variants of RNN (a) Vanilla LSTM (b) Stacked LSTM (c) Bidirectional LSTM.

GRU Model

It is similar to LSTM but a new variant of RNN. The GRU has not maintained the cell state like LSTM. It controls the flow of information through the reset gate and the update gate. The GRU unit is shown in Fig. 4.

Update Gate: It combines the input and the forget gates into a single update gate.

$$z_t = \sigma(w_z.[h_{t-1}, x_t]) \qquad (4)$$

Reset Gate: It stores the short-term memory of the network that is the hidden state.

$$r_t = \sigma(w_r.[h_{t-1}, x_t]) \qquad (5)$$

In the equation, *W* is the weight given to the neuron, *h* is the output from the previous cell, *b* is the bias, and *x* is the input at time step t.

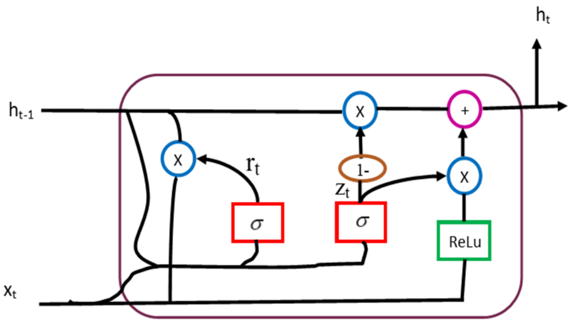

Figure 4. GRU cell with two gates.

Both LSTM and GRU performance varies from case to case, but GRU has only two gates, so it processes faster than LSTM. K-fold cross-validation is used to validate the model. It is a statistical method that estimates the performance of the learning models on new data samples. Here the parameter K refers to the number of splits in the dataset. K value 5 is taken in the present experiment. In this method, the dataset is divided into 5 subsets; from the subset, one set is used for testing, and the remaining four sets are used for training. By using this technique, we ensure that the model's score does not depend on the selected subset for training and testing.

A diagram of the neural network layers for the different deep learning models used in the study is presented in Fig. 3. A detailed parameter list for the models used in the experiment are shown in Table 3. ReLu activation function is used in all four models, along with a 0.001 learning rate. A dataset with step size 15 is used to train the model. A Python environment and the TensorFlow library are used for the experiment. TensorFlow supports deep learning and machine learning applications.

Table 3. Parameter list for the different DL models

Units	First layer		Dense layer	Dropout layer
Vanilla LSTM	64		32	0.2
Stacked LSTM	64 (Lstm 1)	32 (Lstm 2)	32	0.2
Vanilla GRU	64		32	0.2
Bidirectional LSTM	64 (Lstm 1)	64 (Lstm 2)	32	0.2

RESULTS AND DISCUSSION

The present study compares the DL models to predict the arecanut crop disease score value. According to the current literature, deep learning models were not applied to predict the arecanut FRD. Hence for the first time, vanilla LSTM, vanilla GRU, stacked LSTM, and bidirectional LSTM are used in this present experiment. Models are measured through validation loss and training loss with the help of mean square error (MSE) as shown in Eq. 6.

$$MSE = \frac{\sum(actual\ value - predicted\ value)^2}{total\ number\ of\ test\ samples} \qquad (6)$$

Table 4 shows the validation loss of different DL algorithms with the help of the 5-fold cross-validation technique. Similarly, Table 5 shows the training loss of different DL models. In training, the loss represents how well the model fits the training samples, while in the validation, the loss represents how well the model fits newly arrived samples.

Similarly, Table 4 shows the training loss of different DL models. Observation from Tables 4 and 5 shows that the loss can vary depending on the subset of the sample.

Table 4. Comparison of validation loss with different learning models and training sets

	Validation Loss				
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Vanilla LSTM	2.9	1.5	2.2	2.2	2.9
Stacked LSTM	3.7	1.8	3.0	2.2	2.4
Vanilla GRU	2.0	1.5	2.1	1.3	1.6
Bidirectional LSTM	2.2	5.5	2.5	1.9	1.8

Vanilla LSTM and stacked LSTM give better accuracy (lower validation loss) in their second fold, whereas vanilla GRU performed best in its fourth fold and bidirectional LSTM gives the best performance in the fifth fold. The measured loss

MSE is the mean data of the squared differences between actual and predicted values. MSE performance metric is always used in regression problems. The training loss and validation loss of Vanilla LSTM for 5-fold cross- validation are shown in Fig. 5.

Table 5. Comparison of training loss with different learning models and training sets

	Training Loss				
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Vanilla LSTM	9.6	7.2	6.3	5.8	4.7
Stacked LSTM	7.1	6.6	5.4	5.1	4.3
Vanilla GRU	8.3	7.1	6.3	5.5	4.3
Bidirectional LSTM	6.7	1.7	4.6	3.5	3.0

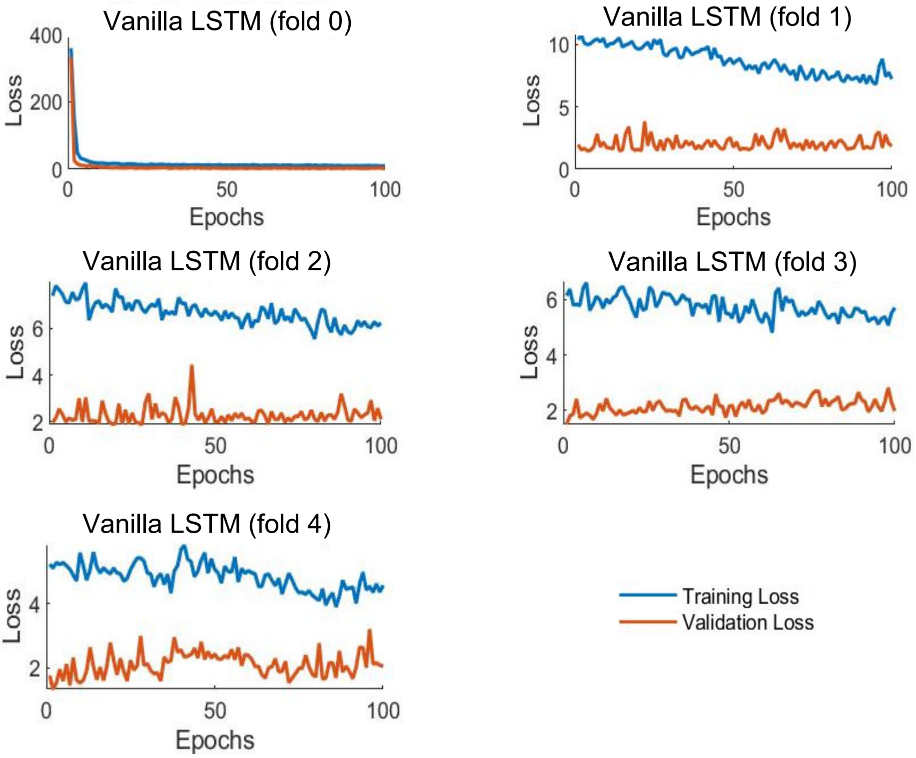

Figure 5. Training loss and validation loss for Vanilla LSTM at different fold cross-validation stage.

The training loss is slightly higher than the validation loss in all the validation folds when 100 epochs are executed. As shown in the figure, it is observed that as the epochs increase, the training loss is decreasing. When 1,000 epochs are executed, the training loss is almost equal to the validation loss, which is the best fit. The training loss and validation loss of Vanilla GRU for 5-fold cross-validation are shown in Fig. 6. In the K-fold cross-validation method each data sample is used at least once in testing, and (K-1) times in training. Hence the result generated from cross-validation reduces bias and variance. The training loss and validation loss of stacked LSTM for 5-fold cross-validation are shown in Fig. 7. The difference between training loss and validation loss for a bidirectional model is lower than in other models. The bidirectional model performed well when the K-fold cross-validation technique was applied. Fig. 8 shows the loss obtained during the experimentation with a bidirectional model.

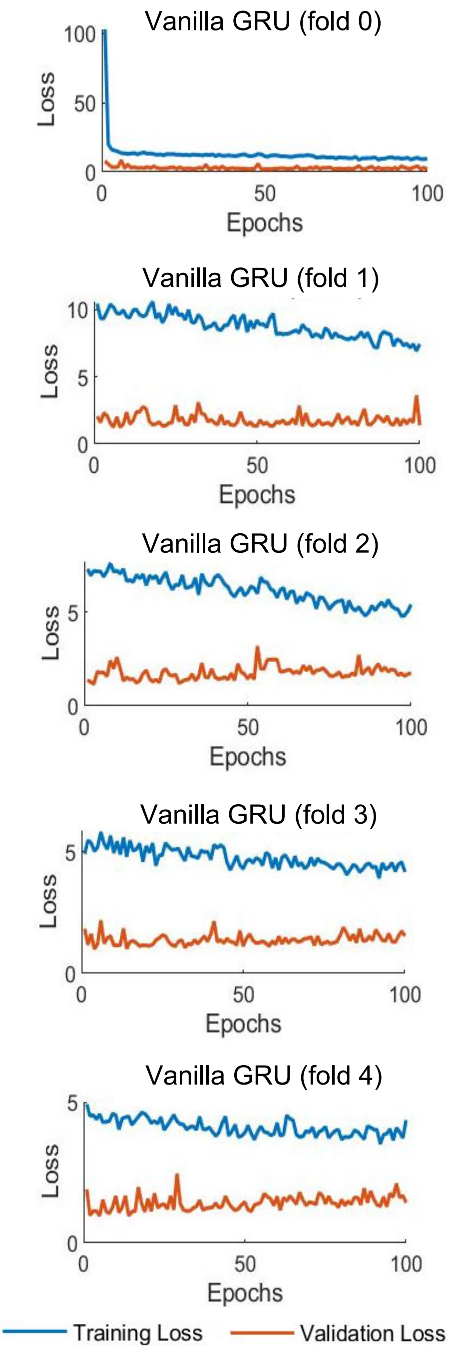

Figure 6. Training loss and validation loss for Vanilla GRU at different fold cross-validation stage.

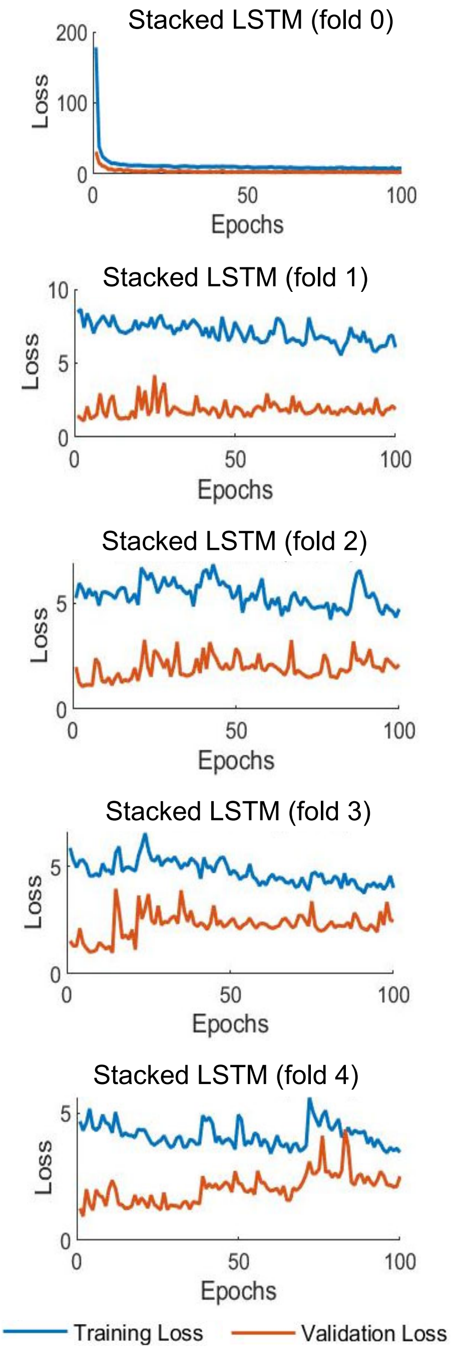

Figure 7. Training loss and validation loss for stacked LSTM at different fold cross-validation stage.

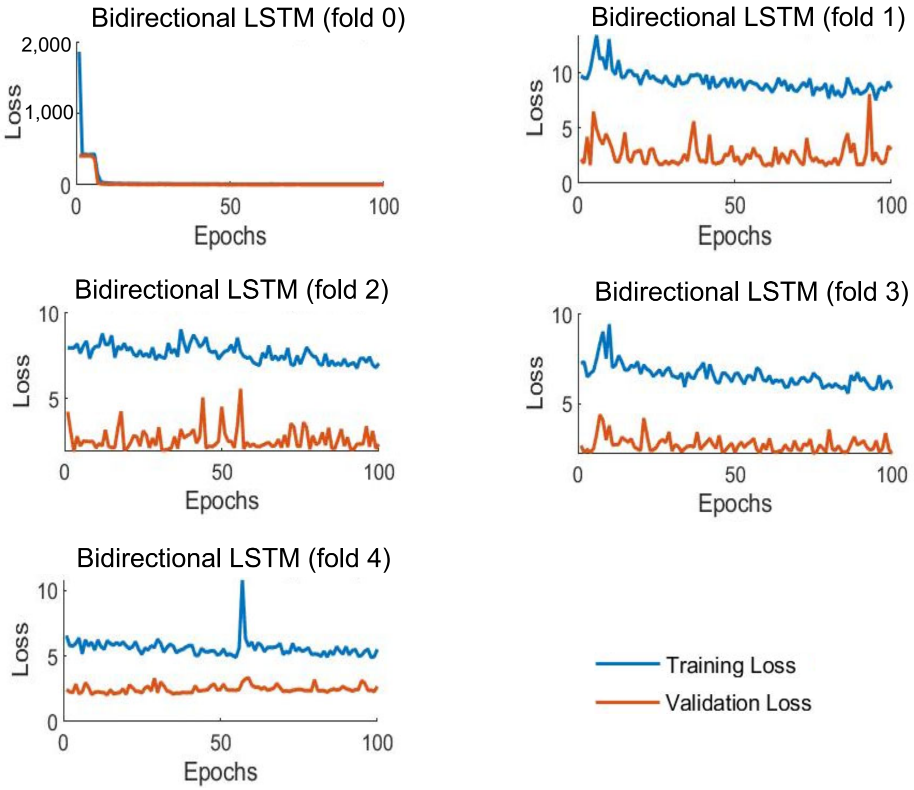

Figure 8. Training loss and validation loss for bidirectional LSTM at different fold cross-validation stage.

From Figs 5, 6, 7, and 8, it can be observed that as the number of epochs increases the training loss decreases. The training is done with 1000 epochs and it is observed that the training (TL) and validation loss (VL) is almost equal as shown in Fig. 9. In the conducted experiment, the predicted value is slightly different from the actual value. The graph for prediction values versus actual values is shown in Fig. 10.

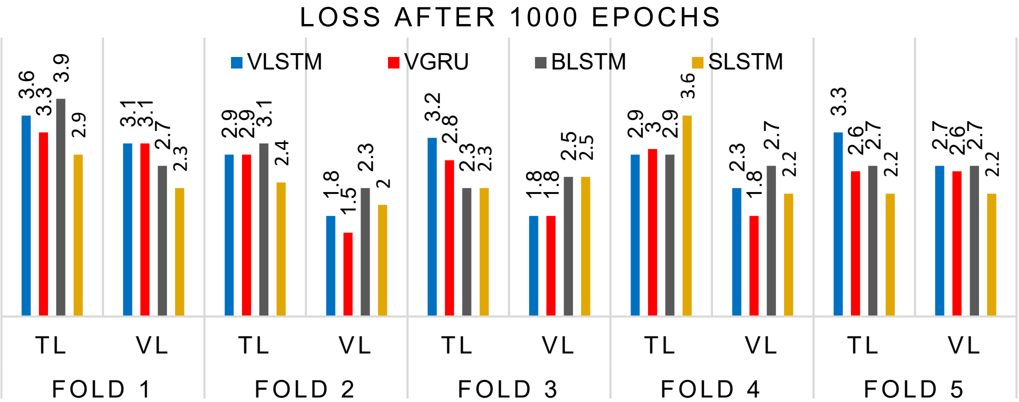

Figure 9. Loss rate after 1000 epochs training with 5-fold cross-validation.

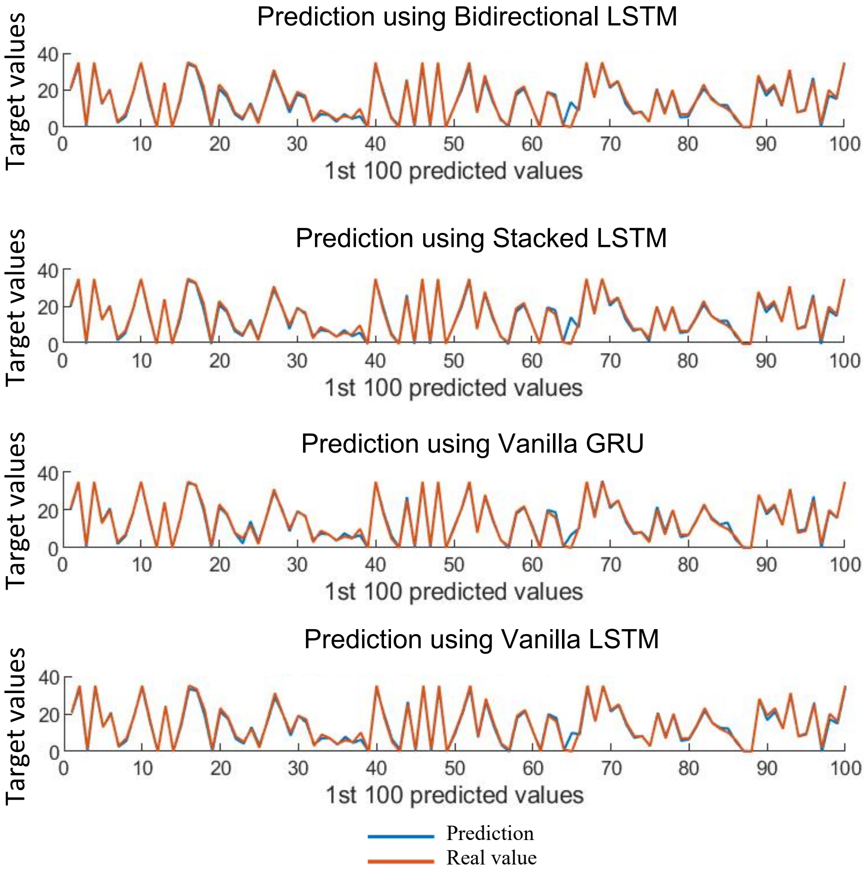

Figure 10. Plotting of first 100 actual values vs prediction value: Bidirectional LSTM, Stacked LSTM, Vanilla GRU, and Vanilla LSTM.

There are 3,152 samples in the dataset, 80% of samples are used for training, and 20% are used for testing purposes. From the 630 testing samples, only 100 predicted values are shown in the graph. The disease score value is predicted from the model and using this score value, disease incidence severity classification can be done. The proposed study and the current study were compared in Table 6 based on different crops and methods used to make crop disease predictions. It is difficult to compare the results with other related works since the use of deep learning models in arecanut crop disease forecasting is a novel approach. In contrast, LSTM models are used to predict disease in crops such as rice and cotton.

Rice blast disease is forecasted at an early stage based on RH, sunshine, temperature, and disease score data, with 67.4% prediction accuracy (Kim et al., 2017). The model is developed using only the LSTM technique. Similarly, the present study experimented with a different crop, and the rainfall parameter plays an important role here. Along with LSTM, the GRU model is also considered for the study, which gives a 1.3 MSE value. Arecanut disease classification is done with the help of CNN architecture (Mallikarjuna et al., 2022). The model takes only images as input and produces an 86.8% accuracy rate on average. However, in the present study, weather data is incorporated as input, allowing for an earlier disease diagnosis. Arecanut disease data and weather data are integrated as novel data sets, which are used as input to the different machine learning models (Krishna et al., 2022). This article considers the same dataset but experiments on different deep-learning models.

Table 6. The proposed study and the existing study are compared concerning crops and methods

Reference	Crop type	Input parameters	Techniques used	Prediction
(Kim et al., 2017)	Rice	Relative humidity, sunshine, temperature, disease score data	LSTM	Rice blast disease
(Xiao et al., 2019; Xiao et al., 2018)	Cotton	Rainfall, wind speed, humidity, sunshine, temperature	SVM, RF, LSTM	Pest and disease
(Wahyono et al., 2021)	Rice	Temperature, rainfall, humidity, climate anomaly	Deep LSTM	Pest attack
(Mallikarjuna et al., 2022; Mallikarjuna et al., 2021)	Arecanut	Images	ResNet CNN	Rot, split, rot split, and healthy
(Krishna et al., 2022)	Arecanut	Rainfall, Relative humidity, sunshine, temperature, disease score data	DTR, SVR, RFR, MLPR	FRD score value
Proposed study	Arecanut	Rainfall, Relative humidity, sunshine, temperature, disease score data	LSTM, GRU, Bidirectional LSTM	FRD score value

The DL algorithm results are compared with the ML algorithm results for the arecanut disease prediction. Table 7 shows the MSE values of ML and DL algorithms applied to the arecanut dataset to predict fruit rot disease score values.

Table 7. Compares MSE loss with machine learning and deep learning models

Model	Algorithm	MSE value
ML (Krishna et al., 2022)	Support Vector Regression	6.1
	Random Forest Regression	1.9
	Decision Tree Regression	3.4
	Multi-layer Perceptron Regression	3.3
DL	Vanilla LSTM	1.5
	Stacked LSTM	1.8
	Vanilla GRU	1.3
	Bidirectional LSTM	1.8

DL models give better accuracy in prediction when compared to ML models. Vanilla GRU is the best model compared with all other models, with a low error rate of 1.3 and less processing time due to its fewer gates.

CONCLUSIONS

In the present study, deep learning approaches are first used to predict the fruit rot disease in arecanut based on weather parameters. This novel approach primarily focused on the relationship between weather parameters and fruit rot disease in arecanut crops.

Historical weather data is taken from the agriculture research station in Brahmavar, India, and disease data is generated by referring to the arecanut disease recommendations. Vanilla GRU gives a lower error value of 1.3 MSE compared to different LSTM models. Bidirectional LSTM does not show promising results with the current dataset since its MSE value is 1.8. The stacked LSTM model also shows the same MSE value. K-fold cross-validation gives a less biased model than a single training and testing data set.

Since the validation loss is not very low, it can be surmised that the accuracy of the model is not very high. Nevertheless, it is a meaningful starting point, being the first attempt to predict arecanut crop disease using deep learning techniques based on weather parameters. This effort will help farmers take precautionary measures and prevent the spread of crop disease. In future work, disease and weather data from different regions can be considered to develop the model. Different optimization techniques can be used to increase the accuracy of the model.

REFERENCES

Agarwal, M., Gupta, S.K. & Biswas, K.K. 2020. Development of Efficient CNN model for Tomato crop disease identification. *Sustainable Computing: Informatics and Systems* **28**, 100407.

Anshul, B., Anuradha, C. & Amit, P.S. 2020 Application of extreme learning machine in plant disease prediction for highly imbalanced dataset, *Journal of Statistics and Management Systems* **23**(6), 1059–1068. doi: 10.1080/09720510.2020.1799504

Aparecido, de O., Lucas, E., Glauco, de S.R., Jose, S.C.de M., Cicero, T.S.C. & Paulo, S.de S. 2020 Machine learning algorithms for forecasting the incidence of Coffea arabica pests and diseases. *International Journal of Biometeorology* **64**(4), 671–688. doi: 10.1007/s00484-019-01856-1

Balanagouda, P., Sridhara, S., Shil, S., Hegde, V., Naik, M.K. & Narayanaswamy, H. 2021. Assessment of the Spatial Distribution and Risk Associated with Fruit Rot Disease in Areca Catechu L. *Journal of Fungi* **7**(10). 797.

Duarte-Carvajalino, J.M., Diego, F.A., Andrés, A.R., Santa-Sepulveda, J.D. Fajardo-Rojas, A.E. & Soto-Suárez, M. 2018. Evaluating late blight severity in potato crops using unmanned aerial vehicles and machine learning algorithms. *Remote Sensing* **10**(10), 1513. doi: 10.3390/rs10101513

Fenu, G. & Malloci, F.M. 2021. Forecasting plant and crop disease: an explorative study on current algorithms. *Big Data and Cognitive Computing* **5**(1), 2.

Fenu, G. & Malloci, F.M. 2019 An application of machine learning technique in forecasting crop disease. In *Proceedings of the 2019 3rd International Conference on Big Data Research*, 76–82. doi: 10.1145/3372454.3372474

Goyal, L., Sharma, C.M., Singh, A. & Singh, P.K. 2021. Leaf and spike wheat disease detection & classification using an improved deep convolutional architecture. *Informatics in Medicine Unlocked* **25**, 100642.

Jain, S. & Ramesh, D. 2021, July. AI based hybrid CNN-LSTM model for crop disease prediction: An ML advent for rice crop. *In 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* 1–7. IEEE.

Jawade, P.B., Dattatray Chaugule, Devashri Patil & Hemendra Shinde. 2020. Disease prediction of mango crop using machine learning and IoT. In *Advances in Decision Sciences, Image Processing, Security and Computer Vision*, 254–260. doi: 10.1007/978-3-030-24322-7_33

Kamilaris, A. & Prenafeta-Boldú, F.X. 2018. Deep learning in agriculture: A survey. *Computers and electronics in agriculture* **147**, 70–90.

Khamparia, A., Saini, G., Gupta, D., Khanna, A., Tiwari, S. & de Albuquerque, V. H. C. 2020. Seasonal crops disease prediction and classification using deep convolutional encoder network. *Circuits, Systems, and Signal Processing* **39**(2), 818–836.

Kim, Y., Roh, J.H. & Kim, H.Y. 2017. Early forecasting of rice blast disease using long short-term memory recurrent neural networks. *Sustainability* **10**(1), 34.

Kodors, S., Lacis, G., Sokolova, O., Zhukovs, V., Apeinans, I. & Bartulsons, T. 2021. Apple scab detection using CNN and Transfer Learning. *Agronomy Research* **19**(2), 507–519.

Krishna, R., Prema, K.V. & Gaonkar, R. 2022. Arecanut Disease Dataset Creation and Validation using Machine Learning Techniques based on Weather Parameters. *Engineered Science* **19**, 205–214.

Liu, K., Zhang, C., Yang, X., Diao, M., Liu, H. & Li, M. 2022. Development of an Occurrence Prediction Model for Cucumber Downy Mildew in Solar Greenhouses Based on Long Short-Term Memory Neural Network. *Agronomy* **12**(2), 442.

Lu, Y., Yi, S., Zeng, N., Liu, Y. & Zhang, Y. 2017. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* **267**, 378–384.

Malicdem, A.R. & Fernandez, P.L. 2015. Rice blast disease forecasting for northern Philippines. *WSEAS Trans. Inf. Sci. Appl.* **12**, 120–129.

Mallikarjuna, S.B., Shivakumara, P., Khare, V., Basavanna, M., Pal, U. & Poornima, B. 2022. Multi-gradient-direction based deep learning model for arecanut disease identification. *CAAI Transactions on Intelligence Technology*. doi: 10.1049/cit2.12088

Mallikarjuna, S.B., Shivakumara, P., Khare, V., Kumar, V., Basavanna, M., Pal, U. & Poornima, B. 2021. CNN based method for multi-type diseased Arecanut Image Classification. *Malaysian Journal of Computer Science* **34**(3), 255–265.

Peng, W., Liu, Y.J., Wu, N., Sun, T., He, X.Y., Gao, Y.X. & Wu, C.J. 2015. Areca catechu L.(Arecaceae): A review of its traditional uses, botany, phytochemistry, pharmacology and toxicology. *Journal of ethnopharmacology* **164**, 340–356.

Puneeth, B.R. & Nethravathi, P.S. 2021. A literature review of the detection and categorization of various arecanut diseases using image processing and machine learning approaches. *International Journal of Applied Engineering and Management Letters (IJAEML)* **5**(2), 183–204.

Ren, C., Kim, D.K. & Jeong, D. 2020. A survey of deep learning in agriculture: techniques and their applications. *Journal of Information Processing Systems* **16**(5), 1015–1033.

Samarawickrama, A.J.P. & Fernando, T.G.I. 2017. A recurrent neural network approach in predicting daily stock prices an application to the Sri Lankan stock market, *IEEE International Conference on Industrial and Information Systems (ICIIS)*, 1–6, doi: 10.1109/ICIINFS.2017.8300345

Santos, L., Santos, F.N., Oliveira, P.M. & Shinde, P. 2019, November. Deep learning applications in agriculture: A short review. *In Iberian Robotics conference*, Springer, Cham, pp. 139–151.

Sriwanna, K. 2022. Weather-based rice blast disease forecasting. *Computers and Electronics in Agriculture* **193**, 106685.

Sujatha, S., Bhat, R. & Apshara, S.E. 2018. Climate change, weather variability and associated impact on arecanut and cocoa in humid tropics of India. *International Journal of Innovative Horticulture* **7**(1), 27–37.

Varsha, M., Poornima, B., Vinutha, H.P. & Pavan Kumar, M.P. 2021. Predictive Model for Rice Blast Disease on Climate Data Using Long Short-Term Memory and Multi-Layer Perceptron: An Empirical Study on Davanagere District. *Annals of the Romanian Society for Cell Biology* **25**(6), 4703–4722.

Wahyono, T., Heryadi, Y., Soeparno, H. & Abbas, B.S. 2020. Enhanced lstm multivariate time series forecasting for crop pest attack prediction. *ICIC Express Lett* **10**, 943–949.

Wahyono, T., Heryadi, Y.A.Y.A., Soeparno, H.A.R.Y.O.N.O. & Abbas, B.S. 2021. Crop pest prediction using climate anomaly model based on deep-LSTM method. *ICIC express letters. Part B, Applications: an international journal of research and surveys* **12**(4), 395–401.

Wu, Y., Yuan, M., Dong, S., Lin, L. & Liu, Y. 2018. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing* **275**, 167–179.

Xiao, Q., Li, W., Chen, P. & Wang, B. 2018, August. Prediction of crop pests and diseases in cotton by long short term memory network. *In International Conference on Intelligent Computing*. Springer, Cham, pp. 11–16,

Xiao, Q., Li, W., Kai, Y., Chen, P., Zhang, J. & Wang, B. 2019. Occurrence prediction of pests and diseases in cotton on the basis of weather factors by long short term memory network. *BMC bioinformatics* **20**(25), 1–15.

Zhu, H., Bingquan, C., Zhang, C., Fei, L., Linjun, J. & Yong, H. 2017 Hyperspectral imaging for presymptomatic detection of tobacco disease with successive projections algorithm and machine-learning classifiers. *Scientific reports* **7**(1), 1–12. doi: 10.1038/s41598-017-04501-2s

Zhang, Z. & Yuan Dong, Y. 2020. Temperature forecasting via convolutional recurrent neural networks based on time-series data. *Complexity*. doi: 10.1155/2020/3536572