IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 August 2023, accepted 29 August 2023, date of publication 4 September 2023, date of current version 12 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3311477

RESEARCH ARTICLE

Constructing and Optimizing RNN Models to Predict Fruit Rot Disease Incidence in Areca Nut Crop Based on Weather Parameters

RAJASHREE KRISHNA[1] **AND K. V. PREMA**[2]

[1]Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
[2]Department of Computer Science and Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal 576104, India

Corresponding author: K. V. Prema (prema.kv@manipal.edu)

This work was supported in part by the Zonal Agricultural and Horticultural Research Station, Brahmavara, Karnataka; and in part by the Central Plantation Crops Research Institute, Kasaragod, Kerala, India.

ABSTRACT A farmer faces several challenges associated with fruit rot disease in the areca nut crop. Weather factors, including rainfall and temperature, largely influence the disease severity and spread of infection in crops. Significant growth has been achieved using RNN models for time series forecasting in the past few decades, but these models are not applied to fruit rot disease prediction. This study introduces Vanilla GRU, Stacked GRU, Bidirectional GRU, and Bidirectional LSTM models for the inaugural prediction of fruit rot disease scores using past weather data. The current investigation also involves the mitigation and comparison of forecast inaccuracies by utilising weight optimisation algorithms like Adam, Adagrad, RMSprop, and Genetic algorithms. Meteorological and disease score data are gathered from the Agricultural Research Station in Brahmavar, Karnataka, and CPCRI Kasargod, Kerala. These datasets are combined using a rule-based algorithm to train and evaluate the proposed model. Empirical outcomes reveal that the vanilla GRU model, when fine-tuned with the Adam algorithm, exhibits a diminished Mean Squared Error (MSE) value of 0.0009, an exceptionally minimal Mean Absolute Error (MAE) value of 0.02, and an elevated R-squared (R2) score of 0.99. Similarly, the Bidirectional LSTM model, optimised through RMSprop, yields an impressively low Root Mean Squared Error (RMSE) value of 0.033. Generally, the optimised Deep Learning (DL) models consistently demonstrate enhanced predictive precision compared to alternative models. In conclusion, the anticipation of areca nut disease, as facilitated by this study, stands to aid farmers in curtailing unnecessary fungicide application and achieving more favourable yields.

INDEX TERMS Areca nut, deep learning, disease prediction, fruit rot disease, genetic algorithm, gated recurrent unit, optimisation, weather.

I. INTRODUCTION

The areca nut is the predominant industrial crop giving economic security to many people in India. Areca nut cultivation offers many job opportunities for small-scale industries. India produces 57% of total production, while China, Bangladesh, Myanmar, and other countries produce the remaining 43% of areca nut [32]. There are only a few locations in the country where areca nuts are grown, but the consumption of areca nuts is widespread. The country's areca nuts are grown in the south and northeast, comprising Karnataka, West Bengal, Kerala, Assam and Tamil Nadu. Every part of the areca catechu is beneficial for humankind. Areca flowers are used for worship; areca leaf is used for plate, bowl and spoon making; areca palm is used for furniture making and many more.

Many pharmacological benefits of areca nut have been documented, including anti-bacterial, anti-parasitic, anti-inflammatory, anti-fungal, and analgesic properties [8], [35]. There has been growing evidence that areca catechu is effective in controlling fascioliasis due to its molluscicidal properties against snails [22]. The by-product of areca nut

The associate editor coordinating the review of this manuscript and approving it for publication was Kuo-Ching Ying.

is areca tannin which is used as a colouring agent in many products, for dying clothes and tanning leather. However, the areca nut is highly prone to attack at all stages due to biotic and abiotic stress. The areca nut is highly susceptible to fungal diseases since climatic conditions promote fungal growth and activity there. Fruit rot disease is one among them, and it results in the complete death of the palm individually or attacks the entire plantation area. Nagappa et al. [27] discussed the problems experienced by farmers in areca nut cultivation. A significant production constraint in areca nut growing is high costs, limited labour availability, and a lack of knowledge about managing pests and diseases. Hence, predicting fruit rot disease is essential to take precautionary measures.

Deep learning algorithms generalise the data and make predictions on data that has not yet been seen. In these algorithms, the inputs are mapped to the outputs. An optimisation algorithm aims to find the weights that minimise the error when mapping inputs to outputs. Hence the optimisation algorithm is commonly used to select the best hyperparameter combination for neural network training. Different types of optimisation algorithms can be used to train the network. Gradient-based algorithms (Adam, RMSprop, Adagrad, Adadelta, stochastic gradient, etc.) and swarm intelligence-based algorithms (ant colony optimisation, genetic algorithms, Particle swarm optimisation, etc.) are commonly used to optimise the neural network (NN).

A. MOTIVATION OF THE RESEARCH

Being predominantly agrarian, India relies heavily on cash crops as a primary source of sustenance for its populace. The application of pesticides becomes imperative due to the susceptibility of areca nut yields to fruit rot disease, particularly during the monsoon season. Given the farmers' strong reliance on these yields, they resort to recurrent pesticide spraying to boost production and counteract disease incidence. However, this practice leads to a cascade of issues:

1) FINANCIAL INEFFICIENCY

Repetitive pesticide application results in monetary wastage.

2) ENVIRONMENTAL CONTAMINATION

The environment suffers from increased pollution due to excessive pesticide use.

3) SOIL DEGRADATION

Soil quality deteriorates due to the accumulation of pesticides.

4) HEALTH CONCERNS

Health problems arise for both farmers and consumers due to pesticide exposure.

The sole solution to mitigate the cycle of excessive spraying rests in predicting fruit rot disease occurrences through forecasted weather data. Consequently, the current research introduces a pioneering methodology: predicting the areca nut disease score by leveraging weather data via deep learning models. Notably, the accuracy of these predictions is enhanced by implementing diverse weight optimisation techniques for the first time.

Several aspects of this present study that contribute original knowledge to the literature are:

- A GRU network-based approach was used for the first time in this study to forecast areca nut disease. These networks have successfully predicted many other problems due to their robustness, efficiency and reliability.
- Different optimisation techniques are used to compare the model's performance.
- For the first time, a genetic algorithm has been proposed as an optimisation tool for forewarning fruit rot incidence.
- Implementation of GRU networks involved incrementing neurons in the hidden layer and several epochs. Thus it determines the impact of the number of neurons and epoch number on forecast accuracy and computation time in the hidden layer.
- This study uses data from the (Brahmavar) Karnataka and (Kasargod) Kerala regions between May and October for 50 years to analyse the weather and disease patterns for the first time.
- Using the proposed approach, researchers should also be able to predict fruit rot disease incidence in other regions.
- This study can solve the problem of disease uncertainty in farming by making the precautionary decision to spray fungicides based on requirements.

The main aim of this research is to use the GRU-based deep learning models in areca nuts' fruit rot disease incidence score prediction and optimise these models to get more accuracy in prediction. Our main aim is to forecast the areca nut disease rate as it reduces crop yield worldwide. Section I describes the importance of areca nut and the research requirement in this domain. Section II states the existing models in crop disease prediction weight optimisation models used in neural network and RNN deep learning models applications. Section III depicts the methodology used in areca nut disease prediction. Section IV represents the proposed work's experimental analysis and is compared with the existing deep learning models. The final conclusion is given in the V section.

II. LITERATURE REVIEW

Much research is being done on areca nut-related issues, such as yield prediction, disease management, chemical and analytical aspects, image-based disease prediction, detection, and classification. Still, very little research is being done on weather-based areca nut disease prediction. Recently authors [20] and [34] detected areca nut disease using a convolution neural network model and classified it as a healthy and diseased crop based on images. In recent years many journals and conference articles have been published on the different optimisation algorithms and deep learning techniques used to solve real-world problems.

A. OPTIMIZATION ALGORITHMS

Haji and Abdulazeez [16] compared the different types of gradient descent optimisation techniques and published a review article. Stochastic gradient, Adagrad, Adam, and RmsProp are compared in terms of their training speed, convergency rate, performance and pros and cons. The efficiency of the algorithm varies based on the dataset used. Jais et al. [21] evaluated the adam optimisation algorithm effect on broad and deep neural networks. They considered the breast cancer dataset from the UCI ML repository to find the best accuracy.

Wibowo et al. [39] used different stochastic gradient optimisation algorithms to train the neural network for classifying cancer microRNA biomarkers. They concluded with Adam and RMSProp as the best optimisation techniques for the given dataset, with an accuracy of 98.5%. Huk [19] used the RMSprop stochastic algorithm to optimise contextual neural networks and multilayer perceptron to solve real-life classification problems. The generalised backpropagation (GBP) algorithm is merged with RMSprop optimisation. The Armstrong, Golub, SRBCT, Sonar, Heart C, and Crx datasets from the UCI ML repository are considered for experimentation. The author concludes that GBP RMSprop performs better than GBP stochastic gradient descent optimisation technique. Shao et al. [36] forecasted wind speed using the LSTM network, optimised by the fireworks algorithm. Initially, they predicted the wind speed using the LSTM model with Adam optimiser. The two hidden LSTM layers are used to experiment with loopback value 15. Then the fireworks algorithm is used as an optimiser to fine-tune the network attributes like weights and learning rate to reduce the losses. Compared to other optimisers, fireworks have a fast convergence rate. They found promising results with 0.64 as the RMSE value and 0.46 as the MAE value. Krishna et al. [24], [33] used machine learning and deep learning techniques to predict areca nut fruit rot disease based on weather parameters. The authors have created and validated the dataset by integrating disease data and historical weather data. MLR, SVR, RFR, DTR, and LSTM models are used and achieve good disease prediction accuracy.

Suksri and Kimpan [37] designed a temperature forecasting model using historical weather data. The model is developed on neural network technology. The network is optimised with the help of the fireworks algorithm. They conducted multiple experiments by varying the neural network parameters and firework algorithm. The prediction accuracy with the training set is 81.48%, and with testing, it is 73.79%.

The best optimiser is selected for bushfire occurrence prediction using deep learning by Halgamuge et al. [17]. Real-time and historical weather data like temperature, pressure, wind speed, direction, daily rain and humidity are collected from Weather Underground API from 2012 to 2017. Different optimisers like Adam, Adagrad, Nadam, RMSprop and SGD are used to train the neural network, and finally selected Adagrad as the best optimiser because of its highest processing speed and accuracy.

Duchi et al. [14] introduced the Adagrad algorithm, Adaptive Subgradient Method for Online Learning and Stochastic Optimization. The author has given many abstract examples in the paper to show that the adaptive method is better than the non-adaptive method for sparse data. They considered adaptive proximal functions, diagonal matrix proximal functions and full matrix proximal functions to update the parameters. Different regularisation is incorporated naturally with the AdaGrad family of algorithms, resulting in very sparse solutions that perform similarly to dense solutions.

Kingma and Ba [23] introduced the Adam algorithm, which uses adaptive estimates of lower-order moments to optimise first-order stochastic objective functions. This algorithm works well for hyperparameter tuning and is appropriate for noisy and sparse gradients. The algorithm combines the advantages of two algorithms, namely AdaGrad to handle sparse gradients and RMSProp to handle non-stationary objectives.

Using the fireworks algorithm, Pang et al. [31] developed a prediction model for electric vehicle relay lifetime. They have shown how the mapping and mutation function of the FWA can be modified to increase the neural network model's convergence ability and running speed. The prediction results from the grey model, grey neural network (GNN) model, FWA GNN model and improved FWA GNN model are compared and declared the best model. The fireworks optimisation did not give good prediction results with the proposed areca nut dataset; hence, we reviewed the genetic algorithm and its applications for optimisation.

Chung and Shin [12] used a genetic algorithm to optimise the LSTM network for stock market prediction. They optimised the window size and the number of LSTM units because these two parameters play a vital role in prediction accuracy. They used daily Korean Stock Price Index data for the experiment, showing that the GA-integrated LSTM model outperforms other benchmark models. Abdolrasol et al. [1] published a review article on artificial neural network-based optimisation techniques. Several parameters have been optimised in this review, including weight optimisations, initial weight, bias and learning rate optimisations, hidden layers, hidden nodes, and activation functions. The techniques like particle swarm optimisation, Artificial bee colony, Backtracking search algorithm, and genetic algorithm are used to optimise the ANN and are discussed by taking various engineering applications.

In buildings, energy consumption is predicted by Luo et al. [25] using deep neural networks optimised by genetic algorithm. To extract the features of weather data daily, clustering techniques were used. A genetic algorithm selects the best architecture for each set of datasets in a DNN submodel. The proposed prediction model is unique because

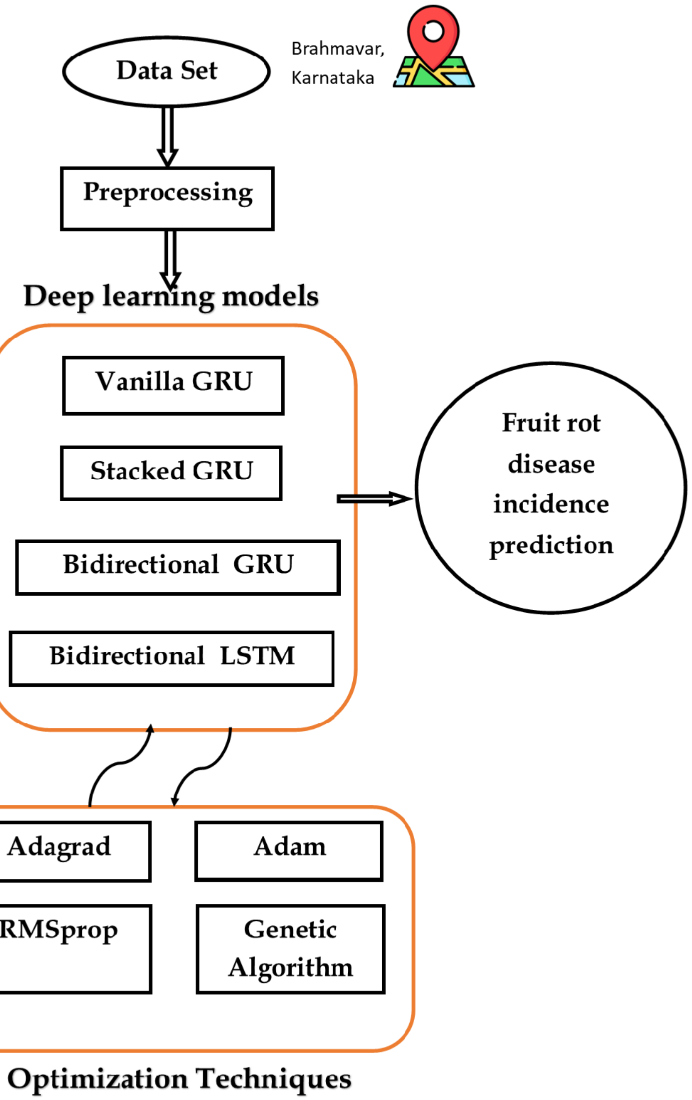

FIGURE 1. Overall system architecture of the present study.

it combines three artificial intelligence approaches: clustering for feature extraction, genetic algorithms for architecture optimisation, and deep neural networks for energy prediction.

Genetic algorithm-based weighted ensembles of deep convolutional neural networks have been developed by Ayan et al. [9] to classify crop pests. The top three best-performing CNN models, Inception-V3, Xception, and MobileNet, were ensembled to increase the performance using weighted voting. The genetic algorithm is utilised to determine the weights, and the model achieved a classification accuracy of 98.81%.

Noh et al. [28] forecasted the product demand in a supply chain management system using the gated recurrent unit with a genetic algorithm. In their approach, GA finds five GRU hyperparameters: window size, number of neurons in a hidden state, batch size, epoch size, and initial learning rate. The proposed GA GRU is compared with k-fold cross-validation, RNN, GA LSTM, ARIMA, and ANN. Using the multi-layer perceptron technique Ecer et al. [15] predicted the stock price index. Genetic algorithm and particle swarm optimisation techniques are used to optimise the weights and biases of the MLP network. RMSE and MAPE measure the performance of both models.

B. RNN MODELS

Cho et al. [11] proposed a gated recurrent unit first time in 2014 on the task of translation from French to English. The idea behind the study is to scale up the computation and memory requirements of neural networks. Chung et al. [13] performed an empirical evaluation of LSTM and gated recurrent neural networks on sequence modelling. The authors evaluated these models on speech signal modelling and polyphonic music modelling tasks. Despite this, the authors could not draw a definitive conclusion about which gating unit was better.

Hochreiter et al. [18] developed the LSTM model to overcome the problem of the traditional RNN model of storing information over extended time intervals. In addition, the LSTM algorithm solves tasks with artificial long-time lags that were previously impossible to solve using previous recurrent network algorithms.

LSTM and GRU techniques were used by Ozdemir et al [29] to forecast the medium- to long-term nickel price. From the study, it is found that the computational time taken for GRU is less than the LSTM network for the given dataset. Based on the author's calculations, LSTM networks averaged 7.060% mean absolute percentage error (MAPE), while GRU networks averaged 6.986% MAPE. Based on the model, nickel prices will be predicted for 10 years between 2022 and 2031, with 2026 providing the best forecast performance.

Yamak et al used ARIMA, GRU and LSTM models to compare the time series forecasting results [40]. Bitcoin's price dataset is considered to make the comparisons. ARIMA is a statistical model that performed well with the dataset, but GRU performed better than LSTM among deep learning models. Alguliyev et al. [4] used infected leaves to detect the disease of 14 species. The aim behind this idea is early detection of plant disease will prevent disease from spreading in large areas. The PlanVillage dataset of images is considered for experimentation, along with CNN and GRU deep learning models. The proposed CNN+GRU model performs better compared to ResNet and VGGnet models.

Stock market values are forecasted by. Althelaya et al. [7] using stacked and bidirectional LSTMs and GRUs. The historical data S&P500 downloaded from Yahoo Finance is used as input to the model, and the Tensorflow package is used for the implementation. Across short-term and long-term forecasts, Stacked LSTM demonstrated the highest performance. Wazrah and Alhumoud [3] used a stacked GRU model for sentiment analysis of Arabic tweets. The model is compared with LSTM, SVM model and an ensemble of these models, in which the ensemble method outperforms all other methods with an accuracy above 90%.

The stacked bidirectional GRU is used to predict the COVID-19 cases in India by Ahuja et al [2]. The proposed model predicts recovery rate, death cases, health index, and positive cases for the next 30 days. In comparison with other models, the model is found to be the most accurate. Alsyaibani et al [6] used bidirectional LSTM and GRU to increase the accuracy of the intrusion detection system model. The dataset used in the study is CIC IDS 2017. The Adam optimisation method is used because of its performance. Among all implemented algorithms, bidirectional GRU showed the highest accuracy 97.8% with ReLu as an activation function.

The GRU model is tremendously used in the agriculture domain for different purposes, namely crop yield estimation [5], crop variety recommendation [26], data from satellite imagery can be used to detect soybean sudden death syndrome early [10] etc.

The research gaps observed from the literature are listed below:

- The stacked and bidirectional GRU and LSTM can be used for time series data forecasting but have not experimented with the areca nut crop disease dataset.
- Optimization algorithms will help improve the model's accuracy that needs to be compared.
- Genetic algorithm helps to optimise the weights required for neural network training but is not used for fruit rot disease forewarning.
- Generalizing the model requires evaluating its performance with different region-specific datasets.

This study is the first attempt to compare the different suitable optimisation algorithms to predict the areca nut disease incidence. Firstly the dataset is preprocessed and made suitable input for the DL models. Secondly, the disease prediction model using a gated recurrent unit (GRU), LSTM and its variants are established. Thirdly, the optimisation algorithms are used to find the best suitable hyperparameters to optimise the RNN model. Finally, the model predicts the fruit rot disease incidence score value.

III. METHODOLOGY

The overall methodology of the present study is shown in figure 1.

As shown in the figure, the historical weather and fruit rot disease data are taken from the agriculture department and preprocessed. The data is given as input to the different DL models with different optimisation algorithms. DL models are vanilla, stacked and bidirectional GRU, and bidirectional LSTM has solid literature in sequential time series forecasting. Optimisation techniques are Adam, Adagrad, RMSprop and genetic algorithm, in which genetic algorithm plays an influential role in neural network weight optimisation. At last, the results in terms of MSE value, MAE value and R2 score value are compared and published.

A. DATA SET

In the study, data from two different regions are considered for experimentation. Historical weather data and areca nut disease data are collected from Agricultural College Brahmavar, Udupi, from 2000 to 2020. Also, historical weather and areca nut disease data are collected from KPCRI Kasargod from 1960 to 2014. There are no differences in the weather patterns of the two regions. So the dataset contains 13555 records after combining data from both regions. The features used in the proposed study are listed in table 1 with their range of values.

The weather and disease data are integrated using the rule-based algorithm shown in Algorithm 1. The rule-based classifier uses various if…else rules to generate the score value in the areca nut crop. The disease score value is considered the target value in the proposed models.

Table 2 shows the sample records with dependent (disease score value) and independent variables (weather data) from the dataset.

TABLE 1. List of features used in the dataset and its range.

Feature Name	Proper Name	Range
RF	Rainfall	0 to 260 (mm)
Max T	Maximum temperature	18 to 39 (Celcius)
Min T	Minimum temperature	4 to 31 (Celcius)
RH 1	Relative humidity (Morning)	36 to 140 (percentage)
RH 2	Relative humidity (Evening)	36 to 100 (percentage)
Cl 2	Wind speed	0 to 8 (km/h)
SS	Sunshine	0 to 11 (hrs)
DS	Disease score value	0 to 35 (Unit)

Algorithm 1 Dataset Integration

if RF $>$ 15 and Max T $<$ 24°C and RH 1 $>$ 90 **then**
    DS++;
else
    if RF $>$ 5 and SS $>$ 5 **then**
        DS++;
    else
        if RF $<$ 10 and Max T$>$ 24°C **then**
            DS$-$;
        else
            DS = DS
        end if
    end if
end if

TABLE 2. Weather parameters and disease score of the dataset.

Date	RF	Max T	Min T	RH 1	RH 2	Cl 2	SS	DS
01-05-1960	0.0	34.3	26.7	83.0	78	23.2	10.5	0
02-05-1960	17.8	34.8	29.0	76.0	78	23.2	9.3	1
03-05-1960	9.1	35.2	24.1	88.0	78	23.2	8.0	2
04-05-1960	42.4	32.1	24.1	80.0	78	23.2	4.2	2
05-05-1960	5.3	33.2	23.4	91.0	78	23.2	4.5	2

Data preprocessing is essential in learning as the model's outcome heavily depends on the proper data. So, the weather data in our dataset contained the missing values; all the missing values are filled with column mean. As listed in the table, the features have different ranges of values. So to make learning easy, the min-max scalar is used to shrink the data from 0 to 1.

B. TECHNIQUES USED

The GRU deep learning model and its variants are used to predict the disease incidence score in the areca nut crop. Stochastic algorithms like RmsProp, Adagrad, genetic algorithm and Adam are used to optimise the model. The optimisation and learning algorithms are explained below.

1) ADAGRAD

This algorithm uses a different learning rate for each iteration. It is constructive because the dataset contains sparse and dense features. Adagrad reaches convergence at a very high speed. The formula used to calculate the new weights is given in equation 1.

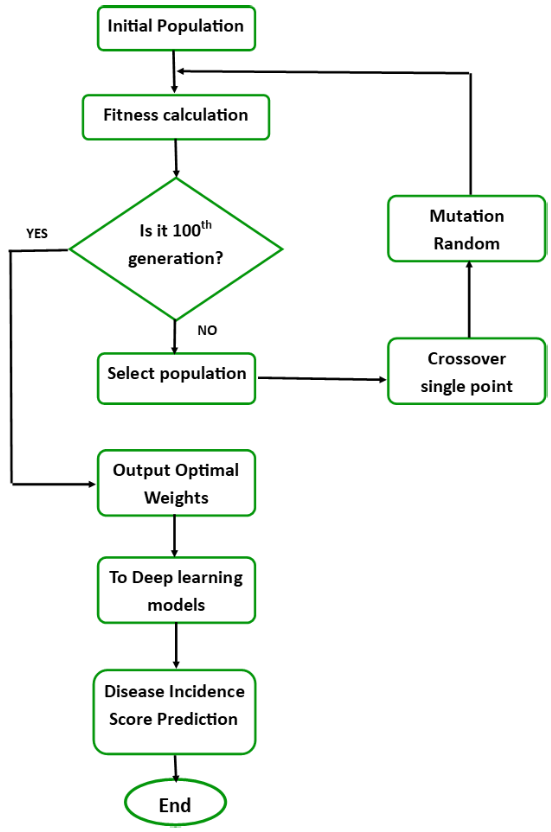

FIGURE 2. Flow diagram of genetic algorithm.

$$w_i^{(t+1)} = w_i^{(t)} - \frac{\eta}{\sqrt{\sum_{\tau=1}^{t} g_{\tau,i}^2}} g_{t,u} \tag{1}$$

The η is a constant and g_t is the learning rate for each iteration. The learning algorithm receives a single example at each iteration in the Adagrad optimisation setting [14]. At the end of a single pass through the training data, the learning algorithm outputs a predictor measured by online loss and test set performance.

2) ADAM

Adam derives its name from adaptive moment estimation. The update of network weights is realised using this optimisation algorithm, a variant of stochastic gradient descent. Each network weight is updated individually by the Adam optimiser. This algorithm inherits both the features of Adagrad and RMSprop algorithms. This algorithm works best with extensive data and/or parameter problems [23]. Adagrad's performance should theoretically be similar to Adam's with $1/\sqrt{t}$ decay on its stepsize.

3) RMSPROP

The RMSprop (Root Mean Square propagation) optimiser also contributes to the advancement of AdaGrad since it reduces the monotonically decreasing learning rate. This algorithm accelerates the optimisation process.

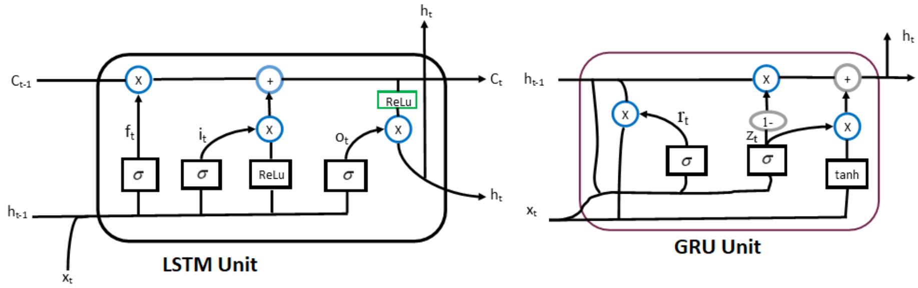

FIGURE 3. Gated recurrent unit and long short-term memory cell with its gate.

TABLE 3. The MSE value of four deep learning techniques when multiple optimisation techniques are used.

DL Technique	Adam	RMSprop	Adagrad	GA
Vanilla GRU	0.0009	0.001	0.003	0.0014
Stacked GRU	0.001	0.001	0.002	0.0015
Bidirectional GRU	0.0013	0.0012	0.008	0.0014
Bidirectional LSTM	0.0012	0.0009	0.0023	0.0025

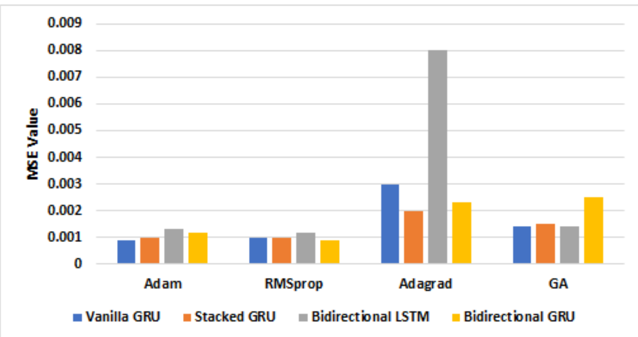

FIGURE 4. MSE value for different DL models.

It was proposed by Tieleman and Hinton [38] to decouple coordinate-adaptive learning rates from rate scheduling. In RMSprop, the learning rate has to be applied manually. The formula for calculating the weight using the RMSprop technique is given in equation 2.

$$w_{t+1}^{j} = w_{t}^{j} - \frac{\alpha}{\sqrt{v_{t}^{j} + \varepsilon}} \left(\frac{\partial E_j}{\partial w_j} \right)_t \tag{2}$$

Gradient descent is adaptively adjusted by taking the partial derivative of error E over connection weight w.

4) GENETIC ALGORITHM

A genetic algorithm was first introduced by Holland in 1975. This technique is based on the mechanisms of natural selection and uses stochastic global adaptive search optimisation. Using GA, operators can mimic the crossover and mutation processes found in nature. Many researchers use the GA for a variety of purposes. Genetic algorithms calculate the fitness value F as shown in equation 3 based on the neural network topology structure and predicted results [12].

$$F = k\left(\sum_{i=1}^{n} (y_i - o_i)^2\right) \tag{3}$$

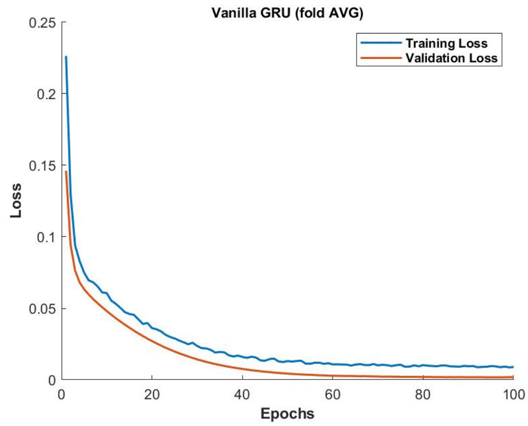

FIGURE 5. Adagrad: Loss in vanilla GRU model.

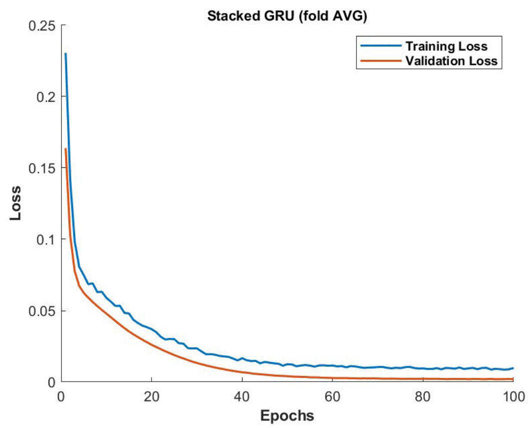

FIGURE 6. Adagrad: Loss in stacked GRU model.

Figure 2 illustrates six stages in processing the GA: initialisation, fitness calculation, termination condition check, selection, crossover, and mutation.

Here, the model weights are considered as the initial population. The GA model gives the best weights as output solutions after the selection, crossover and mutation process. It stops as soon as the algorithm reaches the maximum number of iterations.

5) RNN MODELS

Gradient explosion or gradient disappearance can be solved using a GRU, a type of recurrent neural network. It is

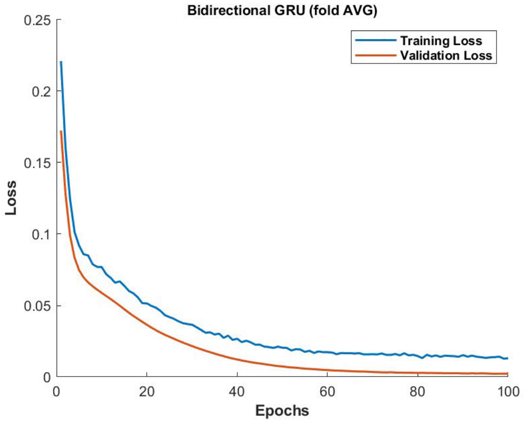

FIGURE 7. Adagrad: Loss in bidirectional GRU model.

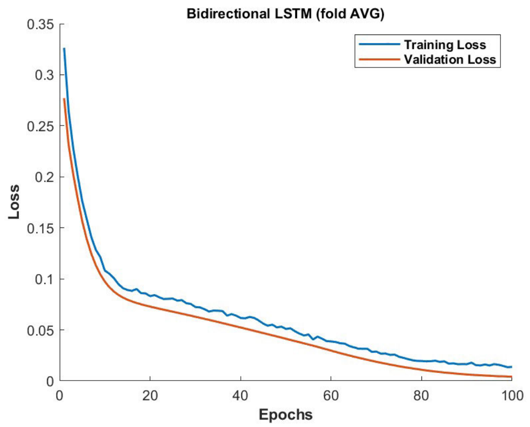

FIGURE 8. Adagrad: Loss in bidirectional LSTM model.

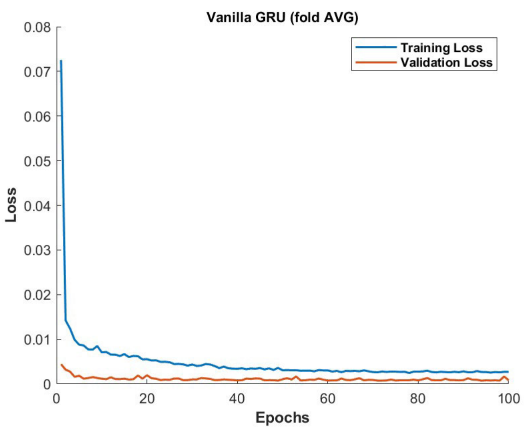

FIGURE 9. Adam: Loss in vanilla GRU model.

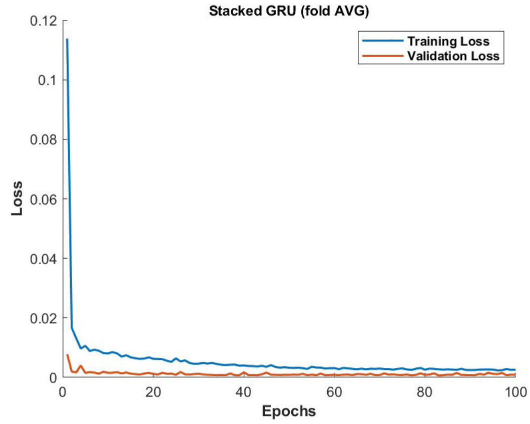

FIGURE 10. Adam: Loss in stacked GRU model.

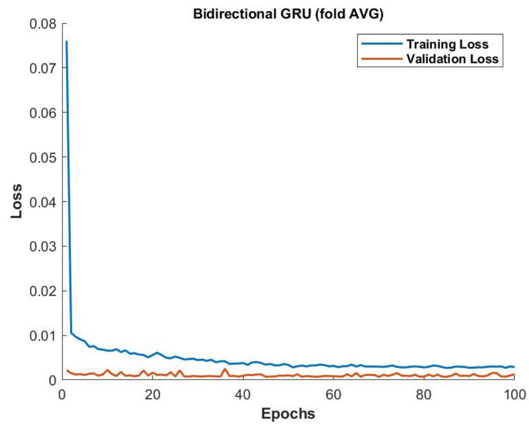

FIGURE 11. Adam: Loss in bidirectional GRU model.

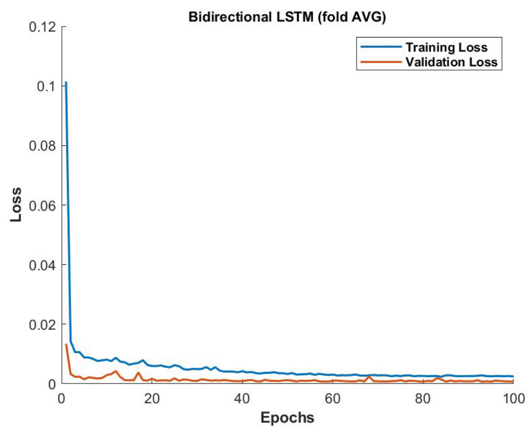

FIGURE 12. Adam: Loss in bidirectional LSTM model.

difficult for traditional neural networks to process time series information. Still, GRU and LSTM can combine historical data with current data to predict future events based on historical data. The GRU unit is shown in figure 4.

The LSTM has a feedback connection and three gates to maintain the state over time. Among the three gates, the input gate updates the cell status, the output gate gives the value of the next hidden state and the forget gate decides which information has to be ignored. GRU contains an update and reset gate to link the hidden information with the future prediction. It is possible that the GRU unit will discard some historical information if the reset gate is close to 0, which means that it will not preserve the current output of the hidden layer. In the update gate, the amount of information to store in the output of the hidden layer at time t is determined [30]. The layers in the DL model use the L1 regularisation to simplify

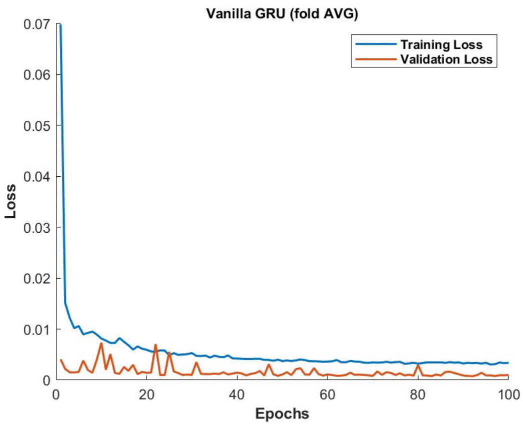

FIGURE 13. RMSprop: Loss in vanilla GRU model.

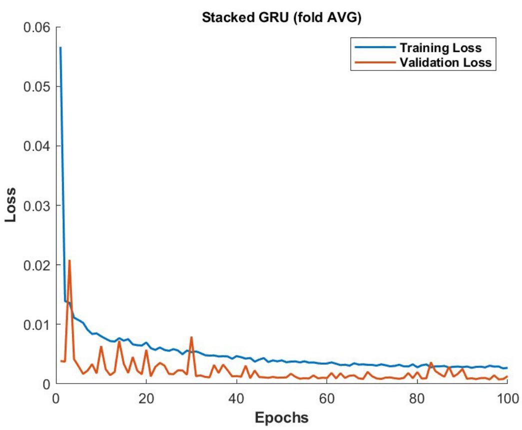

FIGURE 14. RMSprop: Loss in stacked GRU model.

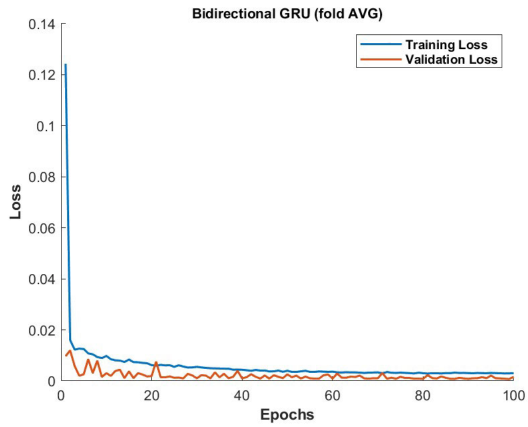

FIGURE 15. RMSprop: Loss in bidirectional GRU model.

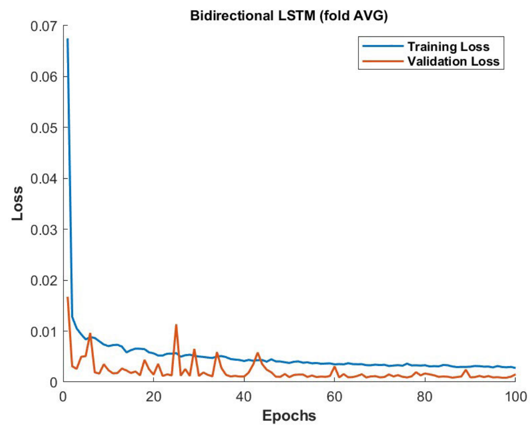

FIGURE 16. RMSprop: Loss in bidirectional LSTM model.

the network model and to prevent the overfitting problem. The dropout regularisation technique is used with a value of 0.2. In this technique, it randomly selects and removes some nodes at every iteration. So, a model with dropout performs better than a normal neural network model. 50 epochs are considered to train and test the model.

Stacked GRU: Several GRU units create stacked GRU in multiple layers. If the hidden layers are more, data overfitting is possible. Therefore, fewer GRU layers are placed one after another to create a stacked GRU.

Bidirectional GRU: This contains two GRU layers, one taking the input in the forward direction and the other in the backward direction. Here, the data sequence is processed in both forward and backward directions. Hence, it is more suitable for large-scale data.

Bidirectional LSTM: This contains two LSTM layers, one taking the input in the forward direction and the other taking the input in the backward direction. Here, the data sequence is processed in both forward and backward directions. Hence it is more suitable for large-scale data. LSTM units contain one more gate than GRU units; hence, the time taken for processing is longer in the bidirectional LSTM model.

TABLE 4. The R2 score value of four deep learning techniques when multiple optimisation techniques are used.

DL Technique	Adam	RMSprop	Adagrad	GA
Vanilla GRU	0.99	0.98	0.96	0.98
Stacked GRU	0.98	0.98	0.97	0.98
Bidirectional GRU	0.98	0.99	0.89	0.98
Bidirectional LSTM	0.98	0.99	0.97	0.97

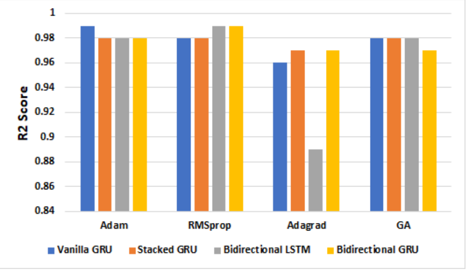

FIGURE 17. R2 score value for different DL algorithms.

C. PERFORMANCE INDEX

Measurement is necessary to evaluate our proposed model's prediction performance against other state-of-the-art methods. The mean absolute error (MAE), mean square error

TABLE 5. The MAE values of four deep learning techniques when multiple optimisation techniques are used.

DL Technique	Adam	RMSprop	Adagrad	GA
Vanilla GRU	0.02	0.027	0.04	0.026
Stacked GRU	0.023	0.022	0.03	0.029
Bidirectional GRU	0.022	0.021	0.03	0.039
Bidirectional LSTM	0.023	0.023	0.07	0.027

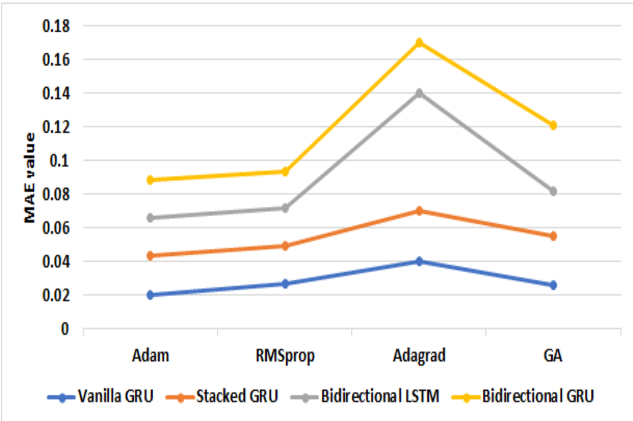

FIGURE 18. MAE value for different DL algorithms.

TABLE 6. The RMSE values of four deep learning techniques when multiple optimisation techniques are used.

DL Technique	Adam	RMSprop	Adagrad
Vanilla GRU	0.034	0.034	0.047
Stacked GRU	0.033	0.034	0.051
Bidirectional GRU	0.034	0.039	0.06
Bidirectional LSTM	0.034	0.033	0.045

(MSE), root mean square error (RMSE) and R2 score are used to evaluate the benefits and drawbacks of the DL models. R2 score is closely related to MSE.

The definition of MSE is given in equation 4.

$$MSE = 1/n \sum_{i=1}^{n} (X_i - Y_i)^2 \tag{4}$$

where X is the actual value, Y is the predicted value, and n is the number of observations. In addition, MSE is measured in units equal to the square of the target variable, while RMSE is measured in units equivalent to the target variable's value. The formula used to calculate RMSE is given in the equation 5.

$$RMSE = \sqrt{1/n \sum_{i=1}^{n} (X_i - Y_i)^2} \tag{5}$$

The R2 score is:

$totalVarianceByModel/totalVariance$

So, the higher the value, the higher the correlation between actual output and predicted output.

TABLE 7. MSE value of proposed optimised DL model is compared with other benchmark models developed using ML and DL technique.

ML [24]	Support vector regression	6.1
	Random forest regression	1.9
	Decision tree regression	3.4
	MLP regression	3.3
DL [33]	Vanilla LSTM	1.5
	Stacked LSTM	1.8
	Vanilla GRU	1.3
	Bidirectional LSTM	1.8
Proposed Models	Vanilla GRU	**0.0009**
	Stacked GRU	**0.001**
	Bidirectional GRU	**0.0012**
	Bidirectional LSTM	**0.0009**

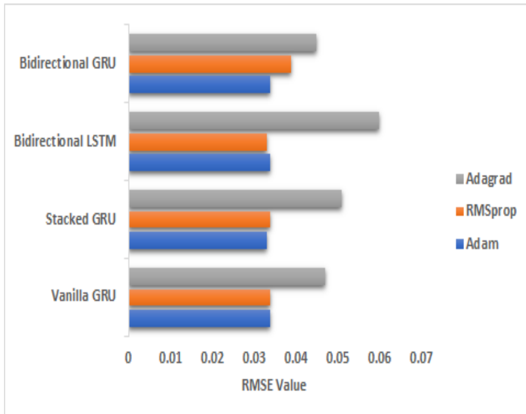

FIGURE 19. RMSE value for different DL algorithms.

MAE is the main component to measure the regression model, and the formula used is shown in equation 6.

$$MAE = (X_i - Y_i)/n \tag{6}$$

where X is the actual value, Y is the predicted value, and n is the number of observations.

The following are the parameters and tools used for the experiment:

- Learning rate - 0.001 and 0.005
- Step size - 4 and 5
- Activation function - ReLu
- Training and testing ratio - 70:30
- No of generation in GA - 100
- Platform - Tensorflow
- IDE - Jupyter Notebook

IV. RESULTS AND ANALYSIS

DL models like vanilla GRU, stacked GRU, bidirectional GRU and bidirectional LSTM have experimented on two region-specific fruit rot disease datasets from Karnataka and Kerala. Different optimisation algorithms, like Adam, Adagrad, RMSprop and Genetic algorithms, are applied to these models to compare the performance. The training loss and validation loss for different models is shown graphically. The dataset contains feature values in different ranges and units. Feature scaling is necessary for deep learning

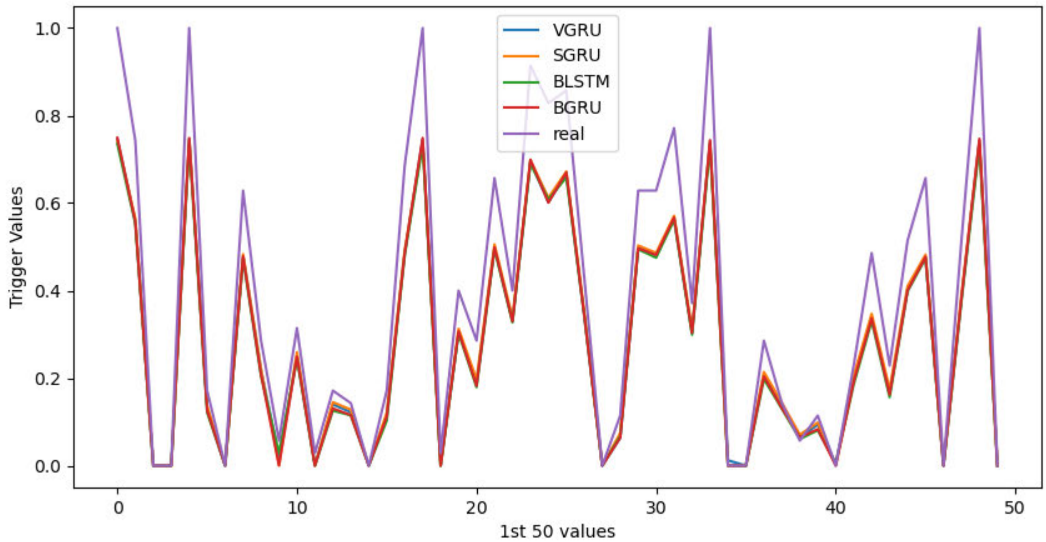

FIGURE 20. Actual and predicted value of areca nut fruit rot disease incidence when Adam optimiser used.

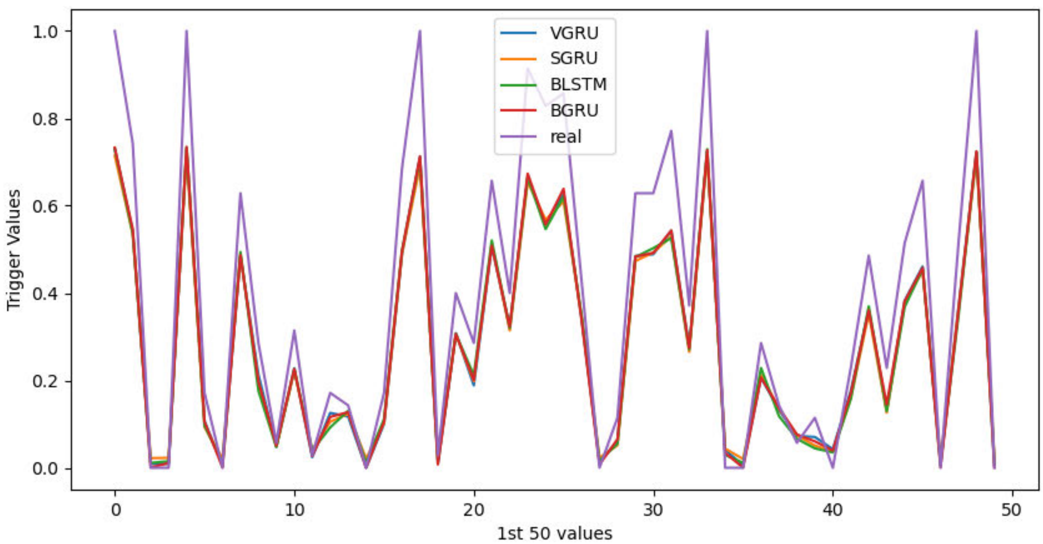

FIGURE 21. Actual and predicted value of areca nut fruit rot disease incidence when Adagrad optimiser used.

models to interpret these features similarly. Therefore, it is preprocessed using a min-max scalar method. Five-fold cross-validation is used to estimate the performance of the DL models. It helps to avoid the problem of overfitting. Table 3 contains the MSE values of four deep learning techniques when multiple optimisation techniques are used. As five-fold cross-validation is used, the average error values are considered. The table shows that the Adam algorithm gives less MSE value of 0.0009 with the vanilla GRU model, hence more accurate fruit rot disease incidence prediction results. Adagrad algorithm gives the highest MSE value of 0.008 with the bidirectional GRU model. Hence it can not be suggested further. Bidirectional LSTM with RMSprop algorithm is the best combination to get less error value of 0.0009 for the present dataset. According to our study, the genetic algorithm is comparable to all of the DL models used in the experiment in terms of accuracy.

The graphical representation of the MSE value is shown in figure 4.

Figs 5,6,7 and 8 show the graph for validation and training losses with respect to multiple epochs in vanilla, stacked and bidirectional GRU and LSTM when Adagrad optimisation is used. The graph shows that training and validation loss decreases as the epochs increase, but the validation loss is slightly less than the training loss.

Figs 9,10,11, and 12 show the graph for validation and training losses for multiple epochs in vanilla, stacked and bidirectional GRU and LSTM when Adam optimisation is

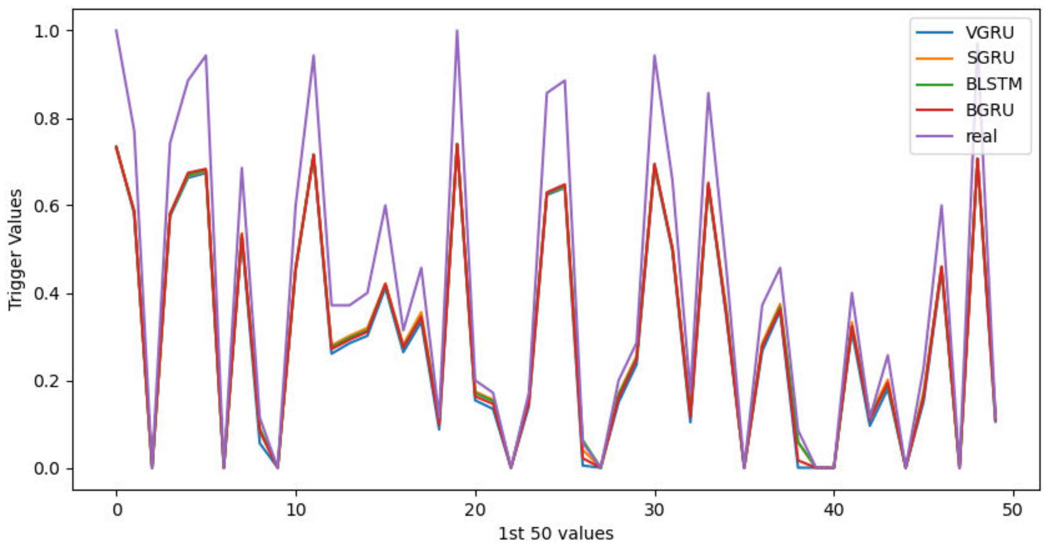

FIGURE 22. Actual and predicted value of areca nut fruit rot disease incidence when RMSprop optimiser used.

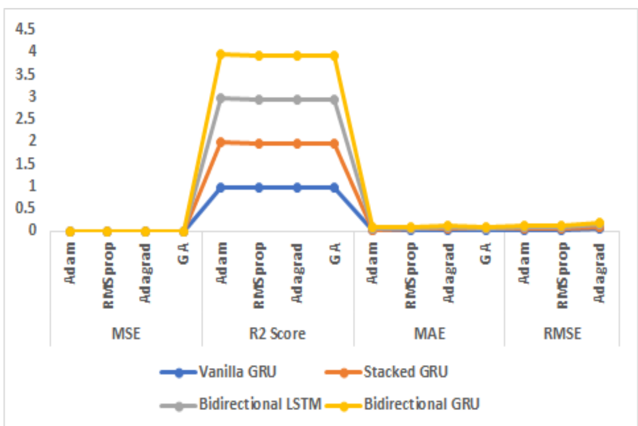

FIGURE 23. MSE, RMSE, R2 score and MAE value for different DL algorithms.

used. The graph shows that training and validation loss decreases as the epochs increase; the validation loss is almost the same as the training loss.

Figs 13,14,15, and 16 show the graph for validation and training losses concerning multiple epochs in vanilla, stacked and bidirectional GRU and LSTM when RMSprop optimisation is used. The graph shows that training and validation loss decreases as the epochs increase; the validation loss is slightly less than the training loss. But it has been experimented that loss is stable after 100 epochs.

Table 4 contains the R2 score values of four deep learning techniques when multiple optimisation techniques are used. R squared is a statistical method of measuring the regression model. If the R2 score value is 100%, then the actual and predicted values are appropriately correlated. The table 4 shows that bidirectional GRU and bidirectional LSTM models give a good R2 value of 0.99 with RMSprop optimisation and the Vanilla GRU model with Adam algorithm. But bidirectional GRU optimised by Adagrad gives less r2 score value of 0.89. All the DL models used in the experiment perform well with the genetic algorithm. Figure 17 depicts the graphical visualisation of the R2 scores across various DL models, each paired with different optimisation techniques.

Table 5 contains the MAE values of four deep learning techniques when multiple optimisation techniques are used. The table shows that Vanilla GRU gives less MAE error value of 0.02 with the Adam optimisation algorithm, and the Adagrad optimisation algorithm gives more MAE value of 0.07 with the bidirectional LSTM model. A study of MAE readings indicates that Adagrad does not perform well with the present dataset when DL models are experimented with. Despite this, the genetic algorithm delivered competitive results in weight optimisation comparison.

The graphical representation of the MAE value is shown in figure 18 below.

When multiple optimisation techniques are used, the RMSE values of four deep learning techniques can be found in Table 6. From table 6, RMSE results indicate that Adam is the best-optimised algorithm for a given set of DL models, and Adagrad is unsuitable for the same models. Figure 19 shows the graphical representation of the RMSE value.

The graph which shows the actual value and the predicted value of fruit rot disease incidence is shown in the figure. Figure 20 shows the output when Adam optimisation is used; similarly, figure 21 for Adagrad optimisation and figure 22 for RMSprop optimisation. 30% of the total dataset is used for testing the proposed model; among them, only the first 50 values are displayed in the graph.

The proposed model can be compared with disease prediction using ML and DL models [24], [33]. The table 7 will list the MSE value of all the experimented models. It is evident from Table 7 that the variants of GRU models outperformed the ML and LSTM-based DL models in terms of MSE. Also, there is a huge change in the results generated by the proposed and existing methods.

Figure 23 shows the graphical representation of all four performance indices, MSE, RMSE, MAE, and R2.

In summary, adopting the proposed model brings about a transformative shift in agricultural practices by leveraging technology to benefit both farmers and the environment. It aligns with sustainability, efficiency, and responsible resource management principles in modern agriculture.

V. CONCLUSION AND FUTURE WORK

The emergence of crop disease prediction models has made notable strides in the agricultural sector. Nonetheless, a noticeable research gap exists in forecasting fruit rot disease within areca nut crops through applying deep learning methodologies. The envisaged predictive model empowers farmers with proactive measures, enabled by anticipatory insights gleaned from forecasted weather data. This, in turn, augments the potential for elevated areca nut crop yields. The crux of this study revolves around utilising Recurrent Neural Network (RNN)–based deep learning models to forecast the incurrence score of areca nut diseases. What sets this research apart is the utilisation of multiple optimisation techniques employed and compared to curtail predictive inaccuracies effectively. The experimental and testing facets of the study are enriched by datasets drawn from diverse geographic regions. When juxtaposed with conventional machine learning approaches, the outcomes underscore the potency of RNN-based deep learning methodologies in heightening accuracy by mitigating errors. Performance evaluation metrics, including MAE, MSE, RMSE, and the R2 value, gauge the efficacy of a regression model in fitting a given dataset.

Remarkably, the vanilla GRU and bidirectional LSTM models exhibit the most impressive outcomes, yielding a substantially reduced MSE value of 0.0009 when coupled with the Adam and RMSprop optimization techniques, respectively. In parallel, bidirectional GRU and LSTM models, when optimized using RMSprop, as well as the Vanilla GRU optimized with Adam, produce the highest R2 score of 0.99. Notably, models fine-tuned via the genetic algorithm consistently yield an R2 score of 0.98. However, Adagrad's performance seems to falter when employed in conjunction with DL models and the existing dataset. The analysis reveals a noteworthy minimum RMSE value of 0.033 in the stacked GRU model with Adam optimization, as well as in the bidirectional LSTM model utilizing RMSprop optimization. Furthermore, the bidirectional GRU model, when optimized through the RMSprop algorithm, and the vanilla GRU model with Adam optimization, showcase commendable performance, each yielding a minimal MAE value of 0.02.

Certainly, here are the potential avenues for future work that can be explored based on the developed prediction model:

- Regional Generalization: Extend the applicability of the developed prediction model to regions characterised by diverse climate patterns compared to those examined in the current study. This would offer insights into the model's adaptability across different environmental conditions.
- Enhanced Optimization: Investigate the possibility of enhancing and validating the accuracy of network models by experimenting with a broader range of optimisation algorithms. This exploration could lead to discovering even more effective ways to fine-tune the models.
- Severity Classification: Extend the predictive capabilities by delving into the classification of disease severity based on the score values. This expansion could give farmers a more nuanced understanding of disease risk and potential impact on crop yield.
- Real-time Validation: Incorporate the real-time weather data collection to validate and refine the existing fruit rot disease prediction models. This would bridge the gap between model predictions and real-world observations, enhancing the model's practical utility.

These proposed directions for future research can contribute to the continued advancement and applicability of the developed prediction model in addressing agricultural challenges and supporting farmers' decision-making processes.

REFERENCES

[1] M. G. M. Abdolrasol, S. M. S. Hussain, T. S. Ustun, M. R. Sarker, M. A. Hannan, R. Mohamed, J. A. Ali, S. Mekhilef, and A. Milad, "Artificial neural networks based optimization techniques: A review," *Electronics*, vol. 10, no. 21, p. 2689, 2021.

[2] S. Ahuja, N. A. Shelke, and P. K. Singh, "A deep learning framework using CNN and stacked Bi-GRU for COVID-19 predictions in India," *Signal, Image Video Process.*, vol. 16, pp. 1–8, Jan. 2022.

[3] A. A. Wazrah and S. Alhumoud, "Sentiment analysis using stacked gated recurrent unit for Arabic tweets," *IEEE Access*, vol. 9, pp. 137176–137187, 2021.

[4] R. Alguliyev, Y. Imamverdiyev, L. Sukhostat, and R. Bayramov, "Plant disease detection based on a deep model," *Soft Comput.*, vol. 25, no. 21, pp. 13229–13242, Nov. 2021.

[5] K. Alibabaei, P. D. Gaspar, and T. M. Lima, "Crop yield estimation using deep learning based on climate big data and irrigation scheduling," *Energies*, vol. 14, no. 11, p. 3004, May 2021.

[6] O. M. A. Alsyaibani, E. Utami, S. Raharjo, and A. D. Hartanto, "Stacked LSTM-GRU model for traffic anomalies detection," *Telematika*, vol. 15, no. 2, pp. 81–91, 2022.

[7] K. A. Althelaya, E. M. El-Alfy, and S. Mohammed, "Stock market forecast using multivariate analysis with bidirectional and stacked (LSTM, GRU)," in *Proc. 21st Saudi Comput. Soc. Nat. Comput. Conf. (NCC)*, Apr. 2018, pp. 1–7.

[8] A. Ansari, T. Mahmood, P. Bagga, F. Ahsan, A. Shamim, S. Ahmad, M. Shariq, and S. Parveen, "Areca catechu: A phytopharmacological legwork," *Food Frontiers*, vol. 2, no. 2, pp. 163–183, Jun. 2021.

[9] E. Ayan, H. Erbay, and F. Varçın, "Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks," *Comput. Electron. Agricult.*, vol. 179, Dec. 2020, Art. no. 105809.

[10] L. Bi, G. Hu, M. M. Raza, Y. Kandel, L. Leandro, and D. Mueller, "A gated recurrent units (GRU)-based model for early detection of soybean sudden death syndrome through time-series satellite imagery," *Remote Sens.*, vol. 12, no. 21, p. 3621, Nov. 2020.

[11] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," 2014, *arXiv:1409.1259*.

[12] H. Chung and K.-S. Shin, "Genetic algorithm-optimized long short-term memory network for stock market prediction," *Sustainability*, vol. 10, no. 10, p. 3765, Oct. 2018.

[13] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[14] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, no. 7, pp. 1–39, 2011.

[15] F. Ecer, S. Ardabili, S. S. Band, and A. Mosavi, "Training multilayer perceptron with genetic algorithms and particle swarm optimization for modeling stock price index prediction," *Entropy*, vol. 22, no. 11, p. 1239, Oct. 2020.

[16] S. H. Haji and A. M. Abdulazeez, "Comparison of optimization techniques based on gradient descent algorithm: A review," *PalArch's J. Archaeol. Egypt/Egyptol.*, vol. 18, no. 4, pp. 2715–2743, 2021.

[17] M. N. Halgamuge, E. Daminda, and A. Nirmalathas, "Best optimizer selection for predicting bushfire occurrences using deep learning," *Natural Hazards*, vol. 103, no. 1, pp. 845–860, Aug. 2020.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[19] M. Huk, "Stochastic optimization of contextual neural networks with RMSprop," in *Proc. Asian Conf. Intell. Inf. Database Syst.* Cham, Switzerland: Springer, 2020, pp. 343–352.

[20] B. Jagadeesha and R. Hegde, "Detection and classification of areca nut disease using convolutional neural network," in *Proc. 4th Int. Conf. Commun. Cyber Phys. Eng.* Cham, Switzerland: Springer, 2022, pp. 441–459.

[21] I. K. M. Jais, A. R. Ismail, and S. Q. Nisa, "Adam optimization algorithm for wide and deep neural network," *Knowl. Eng. Data Sci.*, vol. 2, no. 1, pp. 41–46, 2019.

[22] P. Jaiswal, P. Kumar, V. K. Singh, and D. K. Singh, "*Areca catechu* L.: A valuable herbal medicine against different health problems," *Res. J. Medicinal Plant*, vol. 5, no. 2, pp. 145–152, Feb. 2011.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[24] R. Krishna and R. Gaonkar, "Areca nut disease dataset creation and validation using machine learning techniques based on weather parameters," *Engineered Sci.*, vol. 19, pp. 205–214, Jan. 2022.

[25] X. J. Luo, L. O. Oyedele, A. O. Ajayi, O. O. Akinade, H. A. Owolabi, and A. Ahmed, "Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings," *Renew. Sustain. Energy Rev.*, vol. 131, Oct. 2020, Art. no. 109980.

[26] E. Manjula and S. Djodiltachoumy, "Efficient prediction of recommended crop variety through soil nutrients using deep learning algorithm," *J. Postharvest Technol.*, vol. 10, no. 2, pp. 66–80, 2022.

[27] N. Desai, T. S. Sukanya, and B. Mamatha, "Problems experienced by farmers in arecanut cultivation," *Asian J. Horticulture*, vol. 11, no. 2, pp. 301–305, Dec. 2016.

[28] J. Noh, H.-J. Park, J. S. Kim, and S.-J. Hwang, "Gated recurrent unit with genetic algorithm for product demand forecasting in supply chain management," *Mathematics*, vol. 8, no. 4, p. 565, Apr. 2020.

[29] A. C. Ozdemir, K. Buluş, and K. Zor, "Medium-to long-term nickel price forecasting using LSTM and GRU networks," *Resour. Policy*, vol. 78, Sep. 2022, Art. no. 102906.

[30] F. Pan, J. Li, B. Tan, C. Zeng, X. Jiang, L. Liu, and J. Yang, "Stacked-GRU based power system transient stability assessment method," *Algorithms*, vol. 11, no. 8, p. 121, Aug. 2018.

[31] X. Pang, Z. Li, M.-L. Tseng, K. Liu, K. Tan, and H. Li, "Electric vehicle relay lifetime prediction model using the improving fireworks algorithm–grey neural network model," *Appl. Sci.*, vol. 10, no. 6, p. 1940, Mar. 2020.

[32] B. Patil, V. Hegde, S. Sridhara, H. Narayanaswamy, M. K. Naik, K. K. R. Patil, H. Rajashekara, and A. K. Mishra, "Exploring the impact of climatic variables on arecanut fruit rot epidemic by understanding the disease dynamics in relation to space and time," *J. Fungi*, vol. 8, no. 7, p. 745, Jul. 2022.

[33] K. Rajashree, K. Prema, G. Rajath, and S. Angad, "Prediction of fruit rot disease incidence in arecanut based on weather parameters," *Agronomy Res.*, vol. 20, no. 1, pp. 1–17, 2022.

[34] A. Rajendra, N. Rajkumar, and P. Shetty, "Areca nut disease detection using image processing," in *Soft Computing: Theories and Applications*. Cham, Switzerland: Springer, 2020, pp. 925–931.

[35] M. Rashid, S. Shamsi, R. Zaman, and A. Ilahi, "Areca catechu: Enfolding of historical and therapeutical traditional knowledge with modern update," *Int. J. Pharmacognosy*, vol. 2, no. 5, pp. 221–228, 2015.

[36] B. Shao, D. Song, G. Bian, and Y. Zhao, "Wind speed forecast based on the LSTM neural network optimized by the firework algorithm," *Adv. Mater. Sci. Eng.*, vol. 2021, pp. 1–13, Sep. 2021.

[37] S. Suksri and W. Kimpan, "Neural network training model for weather forecasting using fireworks algorithm," in *Proc. Int. Comput. Sci. Eng. Conf. (ICSEC)*, Dec. 2016, pp. 1–7.

[38] T. Tieleman and G. Hinton, "RMSProp: Divide the gradient by a running average of its recent magnitude," *COURSERA Neural Netw. Mach. Learn*, vol. 17, pp. 26–31, Jun. 2012.

[39] A. Wibowo, P. W. Wiryawan, and N. I. Nuqoyati, "Optimization of neural network for cancer microRNA biomarkers classification," *J. Phys., Conf. Ser.*, vol. 1217, no. 1, May 2019, Art. no. 012124.

[40] P. T. Yamak, L. Yujian, and P. K. Gadosey, "A comparison between ARIMA, LSTM, and GRU for time series forecasting," in *Proc. 2nd Int. Conf. Algorithms, Comput. Artif. Intell.*, Dec. 2019, pp. 49–55.

RAJASHREE KRISHNA is currently an Assistant Professor-Senior Scale with the Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal. She has 15 years of teaching experience. She has published many papers in reputed journals and conferences. Her research interests include machine learning and deep learning in data analysis.

K. V. PREMA is currently the Associate Director, a Professor, and the Head of the Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Bengaluru. She has 30 years of teaching experience and 22 years of research experience. She has published more than 140 papers in reputed journals and conferences. Her research interests include soft computing, computer networks, and security.