International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Analysis of AI-ML Models for Prices Forecasting of Agriculture and Horticultural Commodities

Laveti Krishna Babu*

Department of Computer Science and Engineering (Artificial Intelligence & Machine Learning), GMR Institute of Technology, Rajam-532127, Andhra Pradesh, India

DOI : https://doi.org/10.55248/gengpi.5.1124.3405

A B S T R A C T

Agriculture plays an essential role in stabilizing markets and providing food security to the population. Price forecasting of agriculture commodities helps traders, farmers, and policymakers make informed decisions. This study motivates using an AI-ML-based approach. By analyzing historical data, AI-ML models can predict future prices and identify the key factors that influence them. The model may include some external factors like economic conditions and environmental factors. These predictions help farmers judge what kind of crop to grow to reduce loss and also help to guide market strategies and help to mitigate risk from sudden price fluctuations. It also ensures market stability and everyone has access to safe food. The insights gained from the research will steer policymakers in the right direction. This shows how important AI and machine learning are in forecasting prices for agricultural products.

Keywords: *Agriculture, Price forecasting, AI & Machine Learning, Market stability, Food security, Policymaking.*

1. Introduction

Agricultural products are part and parcel of our daily activities. They also heavily influence the overall economy. When there is fluctuation in the prices, it normally effects the consumers and farmers. This has the effect of causing an economic imbalance. Recently, unusual trends in weather have made such predictions of prices nearly impossible. This is why such policies need to be put in place by the government that keep balance between what is supplied and what is required. Changes in Agricultural Commodity Prices Impact the Economy Changes in commodity prices directly affect the economy, mainly in rural areas. In the case of lowering prices, they might damage their families and communities seriously. The agricultural sector faces lots of drawbacks. It is climate change, coupled with short water supplies and low productivity. These factors sum up as ups and downs in commodity prices. This research paper shall study how AI and ML models can better predict agricultural commodity prices. The historical dataset of commodity prices can be used to find insights that would guide farmers, traders, and policymakers to make smart choices. For this, machine learning methods can be applied. All of these, starting from Auto-Regressive Integrated Moving Average (ARIMA) up to Random Forest (RF), Long Short-Term Memory (LSTM), Artificial Neural Networks (ANN), Back Propagation (BP), Recurrent Neural Networks (RNN), Convolution Neural Networks (CNN), Support Vector Machines (SVM), Decision Trees (DT), advanced, and hybrid approaches, have been embraced and used. Price trends and analyzed, and model performance is evaluated through means of error metrics, including Mean Absolute Error (MAE), R-squared values, Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Since predicting the price of agricultural products precisely can reduce the risks of sudden price fluctuations, it is considerable to obtain accurate price predictions. This might lead to better economic results. The paper aims to demonstrate exactly how advanced analytics is used to enhance the forecasting of agricultural commodity prices, thus appropriately assisting the agricultural sector.

2. Literature Survey

Forecasting agricultural commodity prices in [1] has improved by using a big data framework to help farmers and traders make informed decisions in Pakistan. Here advanced technique LSTM (Long Short-Term Memory) outperformed traditional algorithms such as ARIMA (Auto-Regressive Integrated Moving Average) and RF (Random Forest) with R-squared score of 0.8012 followed by RF with 0.585 and ARIMA with -0.8.

The price forecasting was improved by considering external factors like metrological data along with historical data of Cabbage and Radish. Because of these metrological and historical data as inputs, forecasting agricultural commodity prices became accurate. DIA-LSTM (Dual Input Attention Long Short-Term Memory) was proposed [2] to reduce risk associated with price fluctuations and to maintain a good balance between supply and demand in South Korea.

The data was collected automatically through a program called web crawler. Many factors have been affecting the prices of horticultural products, such as environmental conditions and economic factors. In [3], ARIMA is compared with Neural Network algorithms, such as Back Propagation (BP) and Recurrent Neural Network (RNN). When compared to ARIMA and BP, RNN performed significantly better.

A study made in [4] cited that mainly the ability of Deep Learning techniques can increase the predictability of future soybean price. In this regard, CNN, ILSTM, and SA mechanisms have been used for the better prediction and efficiency. ILSTM is an improved version of LSTM. CNN-ILSTM-SA performed well over other models with Mean Absolute Error (MAE) of 39.0441, R-squared score of 0.97487, and a training time about 157.64 seconds.

In [5], Radial Basis Function (RBF) Neural Network algorithm is used to forecast the agricultural commodity prices like Garlic and Pork in China. Sudden decreases in prices lead to reducing the cost of families and lower production cost for farmers. Here only eight influential factors were taken into account. Performance was evaluated using Absolute error. The proposed model observed high prediction accuracy, particularly after initial months.

Traditional models may fail when the data consist of both linear and non-linear patterns. In [6], Hongbing Ouyang et al. proposed an advanced model LSTNet (Long and Short-Term Time Series Network) to improve accuracy of the agricultural product price prediction considering both short-term and long-term influences. Here multivariate time series data was analyzed. LSTNet outperformed some traditional models like ARIMA and vector Auto-Regression (VAR) in achieving lower Relative Square Error (RSE) and Relative Absolute Error (RAE).

By achieving better price predictions for agricultural products, farmers can determine the optimal time to sell their produce, and the government can be assisted in post-harvest storage in India. For that, a hybrid forecasting model is proposed in [7] called additive-ARIMA-ANN. In this study, the author proposed two addictive models and five multiplicative hybrid models, out of which Addictive-ARIMA-ANN performed well, indicating a minimum Root Mean Square Error (RMSE) of 1.9925, Mean Absolute Error (MAE) of 1.5696, and a Symmetric Mean Absolute Percentage Error (SMAPE) of 6.6665.

Traditional models cannot handle the non-linear and non-stationary data very effectively. Thus, the models have to be shifted to Artificial Intelligence. In [8], an author proposed a hybrid model including Variational Mode Decomposition, Ensemble Empirical Mode Decomposition, and LSTM. The two decomposition methods that were applied are VMD and EEMD, which decreased the data complexity. The LSTM is then applied wherein it was trained using a learning rate of 0.005 along with 200 epochs using an optimal time-step search technique. The RMSE, MAE, and MAPE error measures are used to assess the model's performance.

This study [9] aims to develop a web-based automated system for forecasting agriculture product prices in Malaysia using historical data. The forecasting results were displayed in graph forms for better interpretation. Five different machine learning models were compared in this study: LSTM, eXtreme Gradient Boosting (XGBoost), Prophet, Support Vector Regressor (SVR), and ARIMA.Among all LSTM eclipsed with MSE of 0.304.

Suicide rates among farmers are increasing every year. Early knowledge of price trends can improve farmer's Rate of Investment (ROI) and overall economic stability. In [10], a decision-making support system was built that helps farmers know the average price forecast by logging into the system. This study highlights ARIMA, Kalman filters, and later integrated with data mining techniques.

In [11], conducted a study on existing research and stated that Regression, Clustering, Bayesian, Decision Trees and Neural Network are used. Mostly Daily, Weekly, Monthly, Quarterly observations used in studies. This study highlights The field of philosophy known as epistemology examines the nature, sources, and extent of knowledge. The positivist paradigm (the world can be understood by studying facts and numbers). Additionally, Neural Network models are widely used for price prediction of agricultural products, followed by statistical models and Support Vector Machines.

Quite lesser attention has been provided to interval forecasting, even though much research has been performed on predicting future prices. In [12], the VECM and MSVR were utilized to develop the proposed model. It is thus obvious that the VECM-MSVR accurately captures both linear and nonlinear trends. The proposed model works well with the lowest MAPE value when trained over a number of prediction horizons.
The two models BPNN and SMA, therefore, compared in this research [13] to find the former for forecasting chili prices in Indonesia. Here BPNN was tested with one hidden layer with a variety of numbers of neurons and different activation functions. BPNN is trained with 1000 epochs and with purelin activation function. BPNN is preferable when there are a lot of price changes and SMA for stable datasets.

According to [14], India has 2/3 population that directly or indirectly depends on agriculture. Crop prices can be predicted for forecasting-based recommendation systems as well as for crop cultivation. Chatbots can be built to answer the queries of farmers in their languages. It is important to include other external factors like government policies, festivals, and holidays while forecasting agricultural commodity price prediction along with historical data. Girish Hedge et al. stated that it is more important to give importance to developing the agriculture sector in India and help farmers to get profits.

Agricultural commodity price prediction must be done to aid decision-making and economic stability. These days, AI models outperform traditional and statistical models. A meta-learning strategy was presented by D. Zhang et al. in [15] to link algorithm performance to data features. Features are extracted, features are chosen, and features are classified in three processes. The best forecast model was identified by extracting twenty-nine features. The classification models are RF and SVM, while the candidate models were ANN, SVR, and ELM (Extreme Learning Machine). Simple Model Averaging (SMA) is used to evaluate the proposed model selection framework.

Apart from this, quality of the crop is also required to attain the best price for the crop. Anticipating the disease in the crop may also save the farmers from a loss of yield of the crop. Crop Grown in Andhra Pradesh: Above all, the most common crop grown in Andhra Pradesh is paddy. The most effective algorithms applied for the identification of rice plant disease include RWW-NN [16], Deep RNN or Recurrent Neural Network-based RSW

[17], RHGSO-based Deep Neuro-Fuzzy Network (DNFN) [18], deep learning or SSSO-based Shuffled Shepherd Social Optimization [19], hybrid deep learning or Ex-RHGSO [20], Using machine learning models and image processing techniques produced precise outcomes [21], Zipper S-Convoluted Neural Network (CNN) up to 99.89 accuracy [22], and KNN-Butterfly algorithm up to 98.9 accuracy [23].

3. Methodology

This review will discuss various models applied in crop price prediction. First, data is input to the model wherein it preprocesses and transforms into a form suitable for which it could predict well. Then predictions are run on testing data and the model's assessment that was conducted with performance metrics. Models have been categorized into statistical, machine learning, deep learning, and hybrid models making comprehension easier.

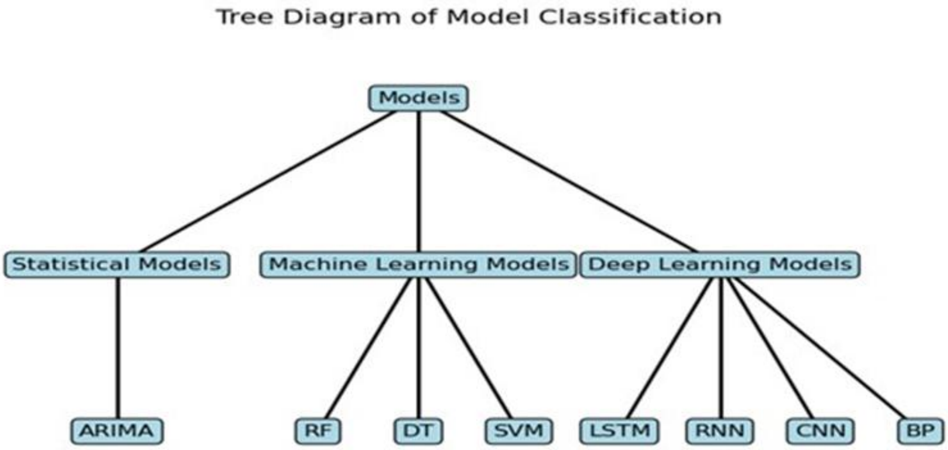

Fig 3.1: Model classification

Statistical models:

Auto-Regressive Integrated Moving Average (ARIMA): It is a time series forecasting model. They are regression analysis. They predict future values by considering past values and subtracting values with lags (previous observations) to make data stationary; they also improve predictions and reduce noise by using previous forecast errors. ARIMA makes the assumption that future trends will resemble historical tendencies. ARIMA performed well for stationary data and struggles for non-linear relationships. The parameters of ARIMA are p, d, q. where ‘p’ indicates the number of past values considered to predict future values, ‘d’ indicates the number of times the model takes the difference between the values to make time series data stationary, and ‘q’ indicates the number of lagged errors the model will use to forecast the future values. Optimal parameters are needed for better performance of ARIMA. For example, optimal parameters used in the base paper [1] are (1,1,1) (0,0,0).

Machine learning models:

Decision Tree (DT): This is a type of supervised learning model that can be applied both to regression and classification tasks. It has tree-like structure comprising a root node, inner nodes, edges also known as branches, and leaf nodes. At root node, the initial decision is made; intermediate decisions are represented by the internal nodes; branch nodes represent the output of a decision; and the leaf node contains the ultimate output made by the decision tree. By using metrics like Gini impurity and entropy, the tree finds the best feature to split. A decision tree is known for its simplicity but may lead to overfitting. Furthermore, it shows bias towards the feature that has greater levels than others.

Random Forest (RF): Random Forest is an ensemble method for machine learning and works well with both classification and regression problems. Ensemble models combine multiple model’s output to get an accurate and precise final prediction. Random Forest combines several decision trees In a random forest, each of the decision trees is trained on a random subset of the data and makes use of random subsets of the features for predictions. This randomness ensures that trees are as diverse as possible; therefore, surely not overfitting. It aggregates the results of the decision trees by performing Voting (for classification) and Averaging (for regression).

Support Vector Machines (SVM) is a supervised learning model put to classification use; it finds an optimal hyperplane, which maximizes the margin, meaning it maximizes the separation between the hyperplane and its nearest data points commonly referred to as support vectors. SVM can perform tasks on both linear and non-linear classifications. If hyperplane can separate the classes without any misclassification (linearly separable), then it is called Hard Margin SVM. For Soft Margin SVM allows to classify the data that are not linearly separable. A hyperplane is a line in 2D space and a plane in n-dimensional space. Whereas for non-linear classification, a kernel function is used in SVM to map high-dimensional data onto a lower dimension data. The most often utilized kernel functions are the RBF, sigmoid, linear, and polynomial kernels. Some of the applications of SVM are text classification, image classification, etc.

Support Vector Regressor (SVR) is an extension of SVM and used for regression type tasks like time series data forecasting. Initially, Training and testing data are the two categories into which the data is separated. The SVM model can be trained over the whole of the training dataset, and its performance can be assessed on the testing dataset. It can be measured in terms of accuracy, precision, recall, and the F1-score. Hyperparameter tuning would be the final step in building the model.

Deep Learning models:

Long and Short-Term Memory (LSTM): It is an extension of RNN where a memory unit is introduced to remember information for an extended length of time. Each memory cell has three components: three different types of gates: input, forget, and output. Information flows into and out of the memory cell through these gates. The input gate controls what comes into the memory cell. The output gate controls what goes out of the memory cell. The forget gate controls which information might be removed from the memory cell. Long sequences of data are stored, updated, and retrieved by LSTM with the help of these gates.

Recurrent Neural Networks (RNN): RNN is also another type of Artificial Neural Network that attempts to model the human brain in emulating the same and processes the sequence data like time series data to identify the patterns. The only thing different between RNN algorithm and a traditional Neural Network is that the former one has its memory state. They can store the patterns for a short period of time. They can't retain information over long sequences. Since earlier time steps' information are retained by the network's structure, it has created excellent functionality for price prediction at changes according to previous data. This algorithm builds an adaptive system that learns from mistakes to improve constantly. They usually send feedback to itself. Backpropagation Through Time (BPTT) is used by RNN to calculate error and adjust weights continuously by moving from right to left after a forward pass. The real-world applications are translations, word suggestions, and sentiment analysis. The major setback of RNN is the vanishing gradient, which means the gradient becomes very small when the number of layers increases so that the weights hardly change.

Convolutional Neural Network (CNN): CNN is also called ConvNet. Object identification, image classification, detection, and segmentation are its primary applications. CNN is different from other models and is more important because it is used for feature extraction. These layers constituting CNN include convolutional, pooling layers and fully connected layers. Convolutional layers involve applying a kernel to carry on convolutions, which helps in detection of edges; pooling layers downsample the input's spatial dimension; and ReLU (Rectified Linear Unit) is used to learn complex relationships that include non-linearity to the network. Predictions are made by fully connected layers using high-level features that the preceding layers have learned.

Backpropagation (BP): Backpropagation is an efficient method of training ANNs, especially feed-forward neural networks. The performance of the Neural Network can be improved using Backpropagation. The BP neural network consists of an input layer, hidden layer, and an output layer. Each node in each layer connects to all nodes in the next layer. This can be done by updating the bias and weights to reduce the loss function for every epoch (an iteration in the training process). Backpropagation sets the weights in the opposite direction to the gradient of the cost function in order to minimize the error. The two most widely used optimization algorithms are gradient descent and stochastic gradient descent. Two passes are used in backpropagation: forward and backward passes. In forward pass, inputs are given to train Neural Network (as normal Neural network model), but in backward pass, the mistake is back-propagated through the network to improve the performance of this neural network model by re-calibrating internal weights.

The performance metrics for evaluating such models, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared score. In [1], the study compared three models ARIMA, RF, LSTM. ARIMA is a statistical model and LSTM and RF are Deep learning models. The models are compared based on performance metrics such as MAE, RMSE showed in the following figure.

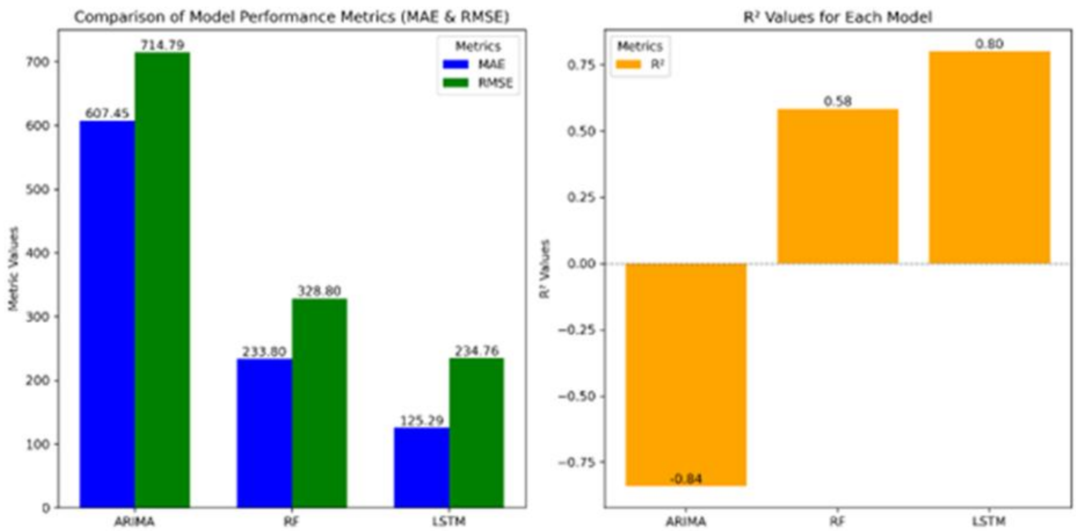

Fig 3.2 (a): Comparison of MAE and RMSE of ARIMA, LSTM, RF (b): Comparison of R2-values of ARIMA, LSTM, RF

4. Conclusion

The agriculture sector accounts for 45% of the employment rate of India. 18-19% of India's GDP is contributed by the agriculture industry. Hence, the need to predict agricultural commodities prices is essential to aid the farming fraternity in informed decision-making. From the research, Deep Learning models were better as compared to the conventional and statistical models. The integration of two or more models was effective in capturing trends and seasonality and error. The linear model can be used in combination with a non-linear model such that it captures both the linear pattern and the non-linear pattern so as to provide precise forecasting. In most studies, only historical data relating to the prices were considered. However other factors such as metrological data, market risk, economic factors and festivals and holidays have impacts on the predictions relating to the prices of agriculture products. Artificial Intelligence and Machine Learning Systems have potential to provide the necessary building blocks.

References

[1] H. Rana, M. U. Farooq, A. K. Kazi, M. A. Baig, and M. A. Akhtar, "Prediction of Agricultural Commodity Prices using Big Data Framework", Eng. Technol. Appl. Sci. Res., vol. 14, no. 1, pp. 12652–12658, Feb. 2024.

[2] Gu, Y.H.; Jin, D.; Yin, H.; Zheng, R.; Piao, X.; Yoo, S.J. "Forecasting Agricultural Commodity Prices Using Dual Input Attention LSTM". Agriculture 2022, 12, 256. https://doi.org/10.3390/ agriculture12020256

[3] Y. Weng, X. Wang, J. Hua, H. Wang, M. Kang and F. -Y. Wang, "Forecasting Horticultural Products Price Using ARIMA Model and Neural Network Based on a Large-Scale Data Set Collected by Web Crawler," in IEEE Transactions on Computational Social Systems, vol. 6, no. 3, pp. 547-553, June 2019, doi: 10.1109/TCSS.2019.2914499.

[4] H. Liu, Y. Pan, W. Liu and J. Wang, "A Prediction Model for Soybean Meal Futures Price Based on CNN-ILSTM-SA," in IEEE Access, vol. 12, pp. 111663-111672, 2024, doi: 10.1109/ACCESS.2024.3443023.

[5] Wang, Y. (2023). Agricultural products price prediction based on improved RBF neural network model. Applied Artificial Intelligence, 37(1). https://doi.org/10.1080/08839514.2023.2204600

[6] Hongbing Ouyang, Xiaolu Wei & Qiufeng Wu (2019) "Agricultural commodity futures prices prediction via long- and short-term time series network", Journal of Applied Economics, 22:1, 468-483, DOI: 10.1080/15140326.2019.1668664

[7] Purohit, S. K., Panigrahi, S., Sethy, P. K., & Behera, S. K. (2021). Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods. Applied Artificial Intelligence, 35(15), 1388–1406. https://doi.org/10.1080/08839514.2021.1981659

[8] Sun, C.; Pei, M.; Cao, B.;Chang, S.; Si, H. "A Study on Agricultural Commodity Price Prediction Model Based on Secondary Decomposition and Long Short-Term Memory Network". Agriculture 2024, 14, 60. https://doi.org/10.3390/ agriculture14010060

[9] Z. Chen, H.S. Goh, K.L. Sin, K. Lim, N.K.H. Chung, X.Y. Liew "Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques", Advances in Science, Technology and Engineering Systems Journal, vol. 6, no. 4, pp. 376-384 (2021).

[10] A. Vohra, N. Pandey and S. K. Khatri, "Decision Making Support System for Prediction of Prices in Agricultural Commodity," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 345-348, doi: 10.1109/AICAI.2019.8701273.

[11] S. Bayona-Oré, R. Cerna, and E. Tirado Hinojoza, "Machine Learning for Price Prediction for Agricultural Products," Journal of Universidad Nacional Mayor de San Marcos, 2021.DOI: 10.37394/23207.2021.18.92

[12] Xiong, Tao et al. "A combination method for interval forecasting of agricultural commodity futures prices." Knowl. Based Syst. 77 (2015): 92-102.

[13] Haviluddin et al., "A backpropagation neural network algorithm in agricultural product prices prediction," 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), Surabaya, Indonesia, 2021, pp. 399-404, doi: 10.1109/EIConCIT50028.2021.9431897.

[14] G. Hegde, V. R. Hulipalled and J. B. Simha, "A Study On Agriculture Commodities Price Prediction and Forecasting," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 2020, pp. 316-321, doi: 10.1109/ICSTCEE49637.2020.9277401.

[15] D. Zhang, S. Chen, L. Liwen and Q. Xia, "Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons," in IEEE Access, vol. 8, pp. 28197-28209, 2020, doi: 10.1109/ACCESS.2020.2971591.

[16] T. Daniya, S. Vigneshwari, "Rider Water Wave-enabled deep learning for disease detection in rice plant", Advances in Engineering Software, Volume 182,2023,103472,ISSN 0965-9978,https://doi.org/10.1016/j.advengsoft.2023.103472.

[17] T Daniya, S Vigneshwari, "Deep Neural Network for Disease Detection in Rice Plant Using the Texture and Deep Features", *The Computer Journal*, Volume 65, Issue 7, July 2022, Pages 1812–1825, https://doi.org/10.1093/comjnl/bxab022.

[18] Daniya, T., & Vigneshwari, S. (2023). "Rice Plant Leaf Disease Detection and Classification Using Optimization Enabled Deep Learning". *JOURNAL OF ENVIRONMENTAL INFORMATICS, 42*(1), 25-38.

[19] Daniya, T., & Srinivasan, V. (2022). Shuffled Shepherd Social Optimization Based Deep Learning for Rice Leaf Disease Classification and Severity Percentage Prediction. WILEY online library (2022).

[20] Daniya, T. & Vigneshwari, S.. (2022). Exponential Rider-Henry Gas Solubility optimization-based deep learning for rice plant disease detection. International Journal of Information Technology. 14. 10.1007/s41870-022-01022-8.

[21] Daniya, T. & Vigneshwari, Dr. (2019). A Review on Machine Learning Techniques for Rice Plant Disease Detection in Agricultural Research. 28. 49-62.

[22] Daniya, T. & Vigneshwari, S.. (2021). Rice leaf Disease Recognition based on deep learning using Zippier S-Method. 1-5. 10.1109/i-PACT52855.2021.9696561.

[23] Daniya, T. & Vigneshwari, S.. (2022). Amalgamation of Rice Plant Disease Classification and Detection with KNN and Butterfly Algorithm. 10.1007/978-981-16-8987-1_14.