ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

CROP YIELD PREDICTION BASED ON CLIMATIC CONSTRAINTS USING A COMBINED LSTM-RNN APPROACH

Mr. Rama Nandan Tripathi[1], Abdul Haseeb[2], Nishant[3], Mohd Atif Haider[4] and Shua Hayat[5]
[1]Assistant Professor MCA, Dr. Ram Manohar Lohia Avadh University Ayodhya U.P.
[2,3,4,5]Research Scholar MCA, Computer Application Department, Dr. Ram Manohar Lohia Avadh University Ayodhya U.P.
[1]sonu.ramanandan@gmail.com, [2]haseebsid6870@gmail.com,
[3]nishantmishra5344@gmail.com,[4]atifhaider30@gmail.com and [5]shuahayat76@gmail.com

ABSTRACT
This study presents a novel approach for predicting crop yields under climatic constraints using a combined Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN) method. The proposed model leverages the strengths of both LSTM and RNN architectures to effectively capture temporal dependencies and complex patterns in climatic data. By integrating historical climate data with crop yield information, the model learns to correlate specific weather patterns with crop productivity, enabling accurate yield predictions. The LSTM component of the model facilitates long-range memory retention, crucial for capturing seasonal and yearly variations in climate that impact crop growth. Meanwhile, the RNN component enhances the model's ability to process sequential data and extract meaningful patterns over time. Experimental results on real-world crop yield datasets demonstrate the effectiveness of the proposed LSTM-RNN combined approach compared to traditional methods. The model achieves high prediction accuracy and robustness even in scenarios with varying climatic conditions and demonstrates its potential for assisting farmers and policymakers in making informed decisions for sustainable agriculture and food security.

Index Terms - Crop Yield Prediction, Climatic Constraints, LSTM, RNN, Combined Approach, Sustainable Agriculture.

INTRODUCTION
Agriculture is the backbone of many economies worldwide, playing a pivotal role in ensuring food security and economic stability. However, the unpredictability of climate patterns poses significant challenges to agricultural productivity, making accurate crop yield prediction a crucial endeavor. Addressing this challenge requires advanced computational methods that can effectively leverage historical climatic data to forecast future crop yields. In recent years, machine learning techniques, particularly deep learning models like Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN), have shown promise in handling sequential data and capturing complex temporal dependencies. By integrating these techniques into a combined approach, we aim to enhance the accuracy and reliability of crop yield predictions while considering the impact of climatic constraints.

The proposed research focuses on developing a novel LSTM-RNN combined approach tailored specifically for crop yield prediction under varying climatic conditions. This approach harnesses the capabilities of LSTM for capturing long-range dependencies and RNN for processing sequential data, thus enabling the model to learn intricate relationships between climatic factors and crop productivity over time. By training the model on historical climate and crop yield datasets, we seek to create a robust predictive tool that can assist farmers, policymakers, and agricultural stakeholders in making informed decisions related to crop management, resource allocation, and risk mitigation strategies.

Through extensive experimentation and validation on real-world datasets, we aim to evaluate the effectiveness and practical utility of the proposed LSTM-RNN combined approach. The study aims to contribute valuable insights into leveraging advanced machine learning techniques for sustainable agriculture, enhancing food security, and adapting to the challenges posed by climate change in agricultural landscapes.

METHODOLOGY

In this section will discuss about three main deep learning model use for crop yield prediction in literature paper.

Recurrent Neural Networks:

Recurrent neural networks (RNN) [14] are a special type of neural networks for learning sequential data. RNN can remember an encoded representation of its past, thus making it suitable for modeling sequential data Given a sequential data x1, x2, . . ., xT for T time steps, the output yt at time step t, is a function of the input at time step xt and the hidden state zt−1 at time step t−1, can be defined as follows

$$z_t = f(w^T_t x_t + u^T z_{t-1}) \quad (1)$$

$$y_t = g(v^T z_t) \quad (2)$$

where, w, u and v are the weights applied on xt, zt−1 and zt respectively and f and g are the non-linear activation functions. As, output is dependent on the hidden states of the previous time steps, the back propagation through time algorithm for updating the weights can result in the problem of vanishing or exploding gradients [6].

Long Short-Term Memory (LSTM):

LSTM a special kind of RNN, were introduced to overcome this issue by integrating a gradient superhighway in the form of a cell state c, in addition to the hidden state h. The LSTM model has gates for providing the ability to add and remove information to the cell state. The forget gate decides the information to be deleted from the cell state and can be defined as follows

$$f_t = \sigma(w^T_f [h_{t-1}, x_t] + b_f \quad (3)$$

The input gate that determines the information that should be added to the cell state is defined as

$$i_t = \sigma(w^T_i [h_{t-1}, x_t] + b_i) \quad (4)$$

The cell state ct is obtained by using both ft and it in the following manner

$$\tilde{c}_t = \tanh(w^T_c [h_{t-1}, x_t] + b_c) \quad (5)$$

$$c_t = f^T_t c_{t-1} + i^T_t \tilde{c}_t \quad (6)$$

Similarly, the hidden state ht and output state ot of the LSTM are defined as

$$o_t = \sigma(w^T_o [h_{t-1}, x_t] + b_o) \quad (7)$$

$$h_t = o^T_t \tanh(c_t) \quad (8)$$

LSTM is more effective in modeling longer sequences than a simple RNN due to a more effective gradient flow during backpropagation.

Q-learning:

Q-Learning is a strategy that performs expert assessment, based on job-respecting work. It selects the limitation of status and performs certain functions in that province. It could be a major improvement in support studies by improving the calculation of non-randomized variance controls. Q-Learning assesses the work of honoring a public service with an objective plan that determines the choice of the most important occupation. Volume Q accepts the contribution as the current status and function ‘a’ and returns the normal compensation for that function to that condition. In the basic steps before clearing the weather, the Q power provides intelligently determined properties. Later on, with better research, the Q work provides a higher level of the appropriate power of the job ‘a’ in the government’. Q work continues with positively profitable renewal. The technician will play a continuation of the tasks that will ultimately create the most complete prize.

LITERATURE STUDY

Crop yield prediction is a critical area of research in agricultural science, aiming to provide insights into future crop production levels based on various environmental and agronomic factors. With advancements in machine learning techniques and the availability of vast datasets, researchers have been able to develop sophisticated models to accurately forecast crop yields. In recent years, numerous studies have explored different machine learning algorithms and methodologies to enhance the precision and reliability of crop yield predictions.

One of the key areas of focus in crop yield prediction is the integration of weather data. Reddy and Kumar [1] delved into this aspect by employing machine learning techniques for weather-based crop yield prediction. They analyzed how climatic factors such as temperature, humidity, rainfall, and sunlight impact crop growth and yield. By leveraging historical weather data and advanced machine learning models, they were able to develop predictive models that could forecast crop yields with a high degree of accuracy.

Arun et al. [2] took a different approach by developing a machine learning-based agricultural yield forecasting system. Their system focused on predicting crop/plant yield before planting, which is crucial for farmers to make informed decisions about crop selection and resource allocation. By utilizing machine learning algorithms and historical agricultural data, their system could provide early insights into potential yield outcomes, allowing farmers to optimize their farming strategies.

In the realm of neural networks, ThangaSelvi and Sathish [3] proposed an optimal bidirectional gated recurrent neural network model for crop yield prediction. Neural networks, particularly recurrent neural networks (RNNs) and variants like LSTM (Long Short-Term Memory) networks, are well-suited for analyzing sequential data such as time-series crop yield data. By incorporating bidirectional architecture and advanced gating mechanisms, their model could capture complex temporal dependencies and make accurate predictions.

Helber et al. [4] introduced an operational approach to crop yield modeling using machine learning models at both field and subfield levels. Their approach involved gathering detailed data about soil properties, crop types, irrigation practices, and weather conditions at a granular level. By applying machine learning algorithms to this comprehensive dataset, they could create predictive models tailored to specific agricultural scenarios, improving the accuracy of yield predictions.

The significance of environmental factors such as soil moisture in crop yield prediction was explored by G et al. [5]. Their study focused on predicting crop yield based on soil moisture levels, which play a crucial role in determining plant health and productivity. By integrating machine learning algorithms with soil moisture data, they could estimate future crop yields more effectively, considering the impact of varying soil moisture conditions.

In a related study, R et al. [6] developed a machine learning-based approach for crop yield prediction and fertilizer recommendation. Their model not only predicted crop yields but also provided recommendations regarding optimal fertilizer usage based on soil characteristics and historical yield data. This integrated approach aimed to improve crop productivity while minimizing resource wastage, showcasing the potential of data-driven strategies in precision agriculture.

Remote sensing technologies have also been instrumental in crop yield prediction. Rananavare and Chitnis [7] utilized Sentinel-2 satellite data for monocot crop yield prediction. Satellite imagery provides valuable insights into crop health, growth patterns, and environmental conditions. By analyzing multispectral bands and other satellite data, their model could predict crop yields with a high degree of accuracy, facilitating informed decision-making for farmers and policymakers.

A regional focus on crop yield prediction was demonstrated by Chaudhary et al. [8] in their study on Punjab state. They developed a classification-based interactive model specifically tailored to Punjab's agricultural landscape. By considering region-specific factors such as soil types, crop varieties, and climatic conditions, their model could provide localized yield predictions, aiding stakeholders in the region to make targeted interventions for agricultural improvement.

The studies by R et al. [9], Gajula et al. [10], Patki and Wazurkar [11], and Suresh et al. [12] contributed further insights into the application of various machine learning algorithms for crop yield prediction. These studies explored algorithms such as stochastic gradient descent, random forest, and neural networks, showcasing the versatility of machine learning in handling diverse agricultural datasets and predicting crop yields accurately.

Satellite imagery and remote sensing data have been extensively utilized in crop yield prediction models. Shahrin et al. [13] and Nishant et al. [14] highlighted the importance of integrating satellite imagery with machine learning algorithms for precise yield predictions. By leveraging multispectral bands and other satellite-derived information, these models could capture detailed information about crop health, nutrient levels, and growth stages, enhancing the accuracy of yield forecasts.

Deep learning techniques have also gained traction in crop yield prediction research. Terliksiz and Altýlar [15] applied deep neural networks for soybean yield prediction in Lauderdale County, Alabama, USA. Deep learning models, with their ability to learn complex patterns from large datasets, are well-suited for analyzing agricultural data and making accurate predictions. Their study demonstrated the effectiveness of deep learning in crop yield forecasting, paving the way for more advanced modeling techniques in agriculture.

Overall, the literature on crop yield prediction using machine learning techniques reflects a diverse range of methodologies, algorithms, and applications. From weather-based models to neural network architectures, from regional predictions to satellite-driven analyses, researchers have explored various avenues to improve the accuracy and effectiveness of crop yield predictions. These studies contribute valuable insights into the evolving landscape of precision agriculture and data-driven decision-making in farming practices.

PROPOSED SYSTEM

As shown in figure 1 proposed system block diagram consist of LSTM-RNN combine network with Q-learning approach for crop yield prediction.

The algorithm for preparing LSTM-RNN based Q-Learning involves several steps. First, in the pre-training phase for **LSTM-RNN:**

1. Enable the recurring memory capacity as N.
2. Launch the LSTM-RNN network with random instruments ǿi for i = 1 to I.
3. Train the hidden layer and keep the ith limits hidden layer.
4. Initialize an action value network with hidden layer parameters, excluding the output layer.
5. Initialize the function of the target action value Q' with the same parameters as Q.

Next, in the Q-learning training phase:

1. Start the watch sequence by subtracting randomly predicted harvest for each event from 1 to M.
2. Choose an unplanned action, possibly e, at each time step t.
3. Perform the chosen action and receive a reward rt.
4. Randomly produce the following country (t + 1).
5. Keep memory D as (st, at, rt, s(t + 1)).
6. Make the gradient decrease in network parameters ǿ for rt-Q(st, at, ǿ)^2.
7. Reset Q' to Q after each iteration.

These steps outline the process of utilizing LSTM-RNN and Q-Learning in a sequential manner to train and optimize the model for effective prediction and decision-making in dynamic environments.

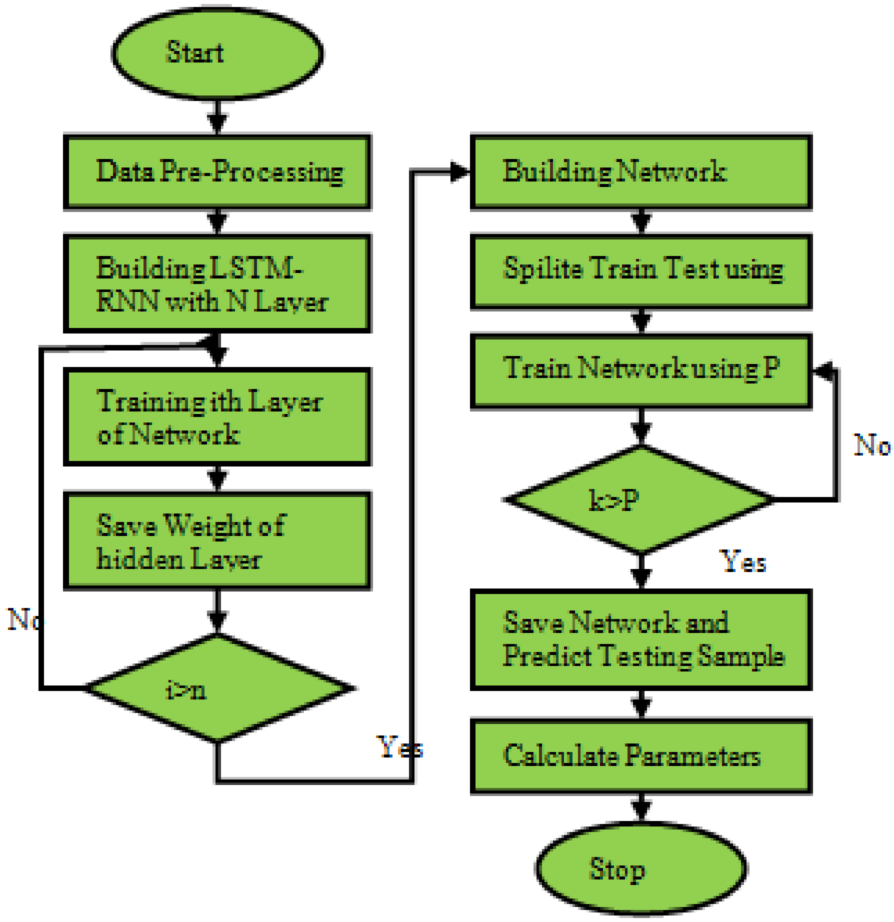

Figure 1: Proposed System Flow

The initial phase, depicted in Figure-1, begins with pre-processing accomplished through fundamental Python functions. This stage involves eliminating zero, blank, or erroneous data values from the system. Subsequently, the construction commences with the development of the LSTM sequential network followed by the creation of the RNN network. Integration of LSTM-RNN is achieved by feeding the output of the LSTM model into the RNN model. The proposed model comprises four distinct layers, culminating in a prepared and structured network ready for training. The next step involves data division into training and testing subsets. Finally, the network undergoes training iterations until either the maximum epoch is reached or the minimum gradient is attained.

RESULTS ANALYSIS

Datasets

The study area encompasses Ponai, Arcot, Sholinghur, Ammur, Timiri, and Kalavai in the Vellore region of southern India. Paddy cultivation is a significant economic activity in this region, making it a focal point of the study. The database is structured around typical climatic and soil conditions, including ecosystem characteristics, soil types, groundwater levels, and fertilizer usage specific to the crops under investigation. Parameters such as evaporation rates, recurring ground ice, groundwater fluctuations, frequency of wet days, and seasonal attributes have been excluded from the analysis discussed in this text.

The following sections present the anticipated rice crop yields using various predictive models.

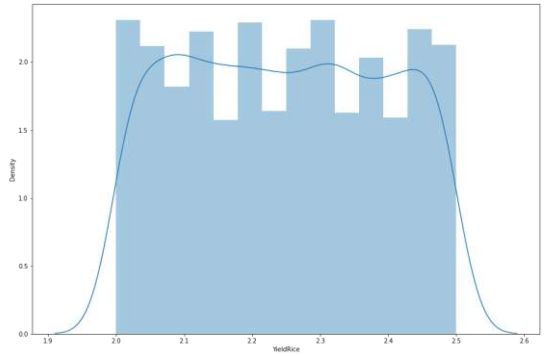

Figure 2: Original Data

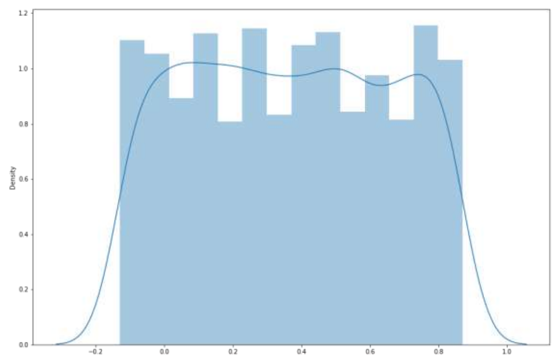

Figure 3: Predicted Data Distribution Using Rnn-Q Learning

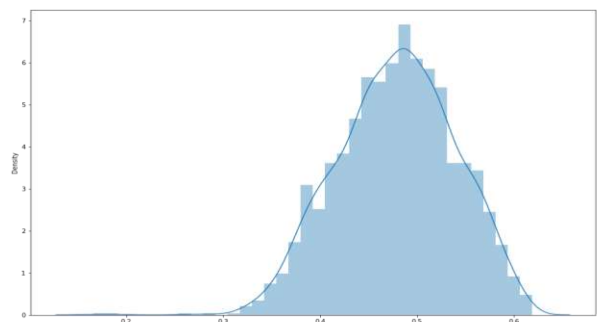

Figure 4: Predicted Data Distribution Using Rnn

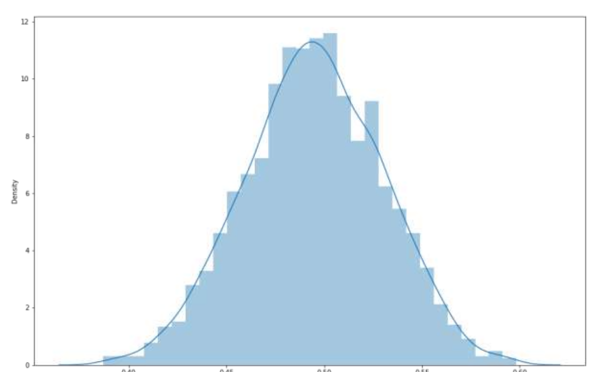

Figure 5: Predicted Data Distribution Using Lstm

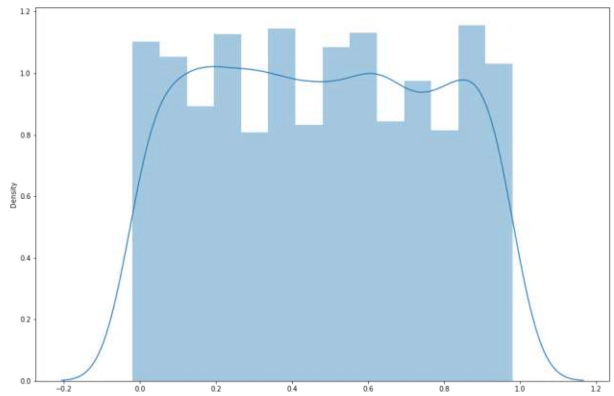

Figure 6: Predicted Data Distribution Using Proposed

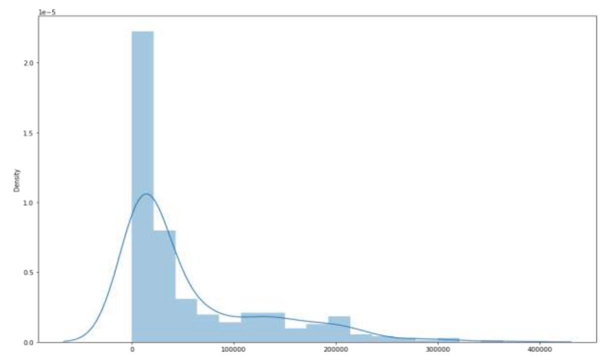

Figure 7: Original Data

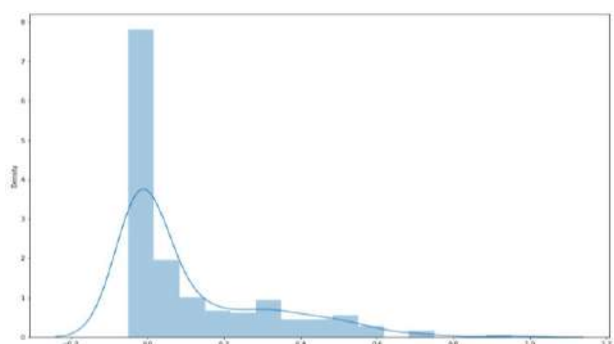

Figure 8: Predicted Data Distribution Using Rnn-Q Learning

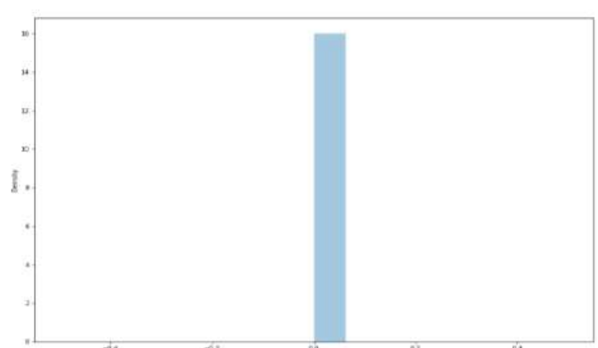

Figure 9: Predicted Data Distribution Using Rnn

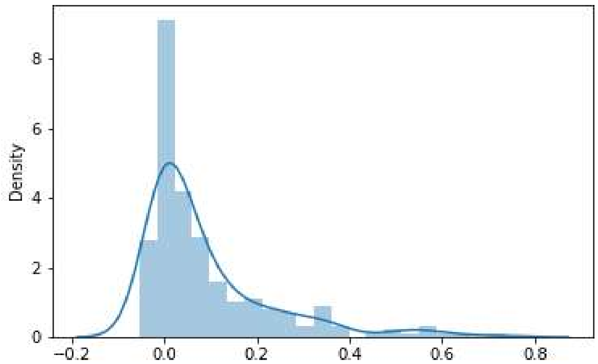

Figure 10: Predicted Data Distribution Using Lstm

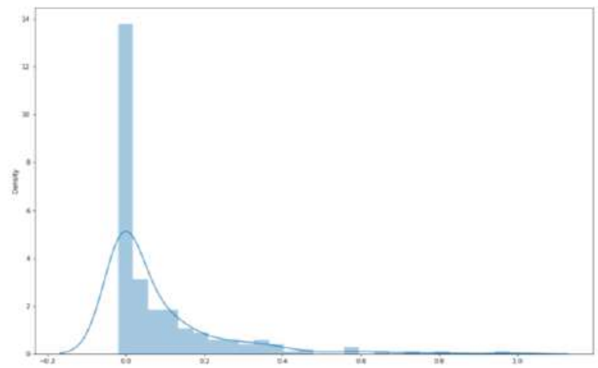

Figure 11: Predicted Data Distribution Using Proposed

Table I: Vallore Rise Crop

Model	Training				Validation			
	MSE	MAE	RMSE	R2	MSE	MAE	RMSE	R2
Linear Regression	0.02	0.125	0.144	0.57	0.022	0.131	0.15	0.57
Support Vector Machine	0.021	0.126	0.145	0.378	0.022	0.131	0.15	0.295
K-Nearest Neighbour	0.362	0.486	0.602	0.449	0.034	0.154	0.185	0.521
Random Forest	0.02	0.125	0.144	0.4	0.022	0.131	0.15	0.466
Gradient Boosting	0.015	0.107	0.125	0.104	0.023	0.133	0.153	0.249
Neural Network	0.362	0.486	0.602	0.449	0.321	0.451	0.566	0.388
Deep Reinforcement Learning	0.001	0.012	0.012	0.803	0.001	0.012	0.012	0.814
RNN	0.007	0.023	0.027	0.90	0.009	0.026	0.030	0.845
Long Short-Term Memory	0.008	0.0248	0.028	0.607	0.009	0.026	0.030	0.877
Proposed (LSTM-RNN Q-learning)	0.0003	0.019	0.019	0.9956	0.0003	0.019	0.020	0.9953

Table II: Gujarat Rise Crop

Model	Train				Validation			
	MSE	MAE	RMSE	R2	MSE	MAE	RMSE	R2
Deep Reinforcement Learning	0.2787	0.1801	0.276	0.7022	0.267	0.0201	0.028	0.7025
RNN	0.415	0.367	0.387	0.6408	0.6545	0.2378	0.342	0.6426

Long Short-Term Memory	0.0011	0.0501	0.031	0.9426	0.012	0.0524	0.037	0.9426
Proposed (LSTM-RNN Q-learning)	0.0003	0.019	0.019	0.9956	0.0003	0.019	0.020	0.9953

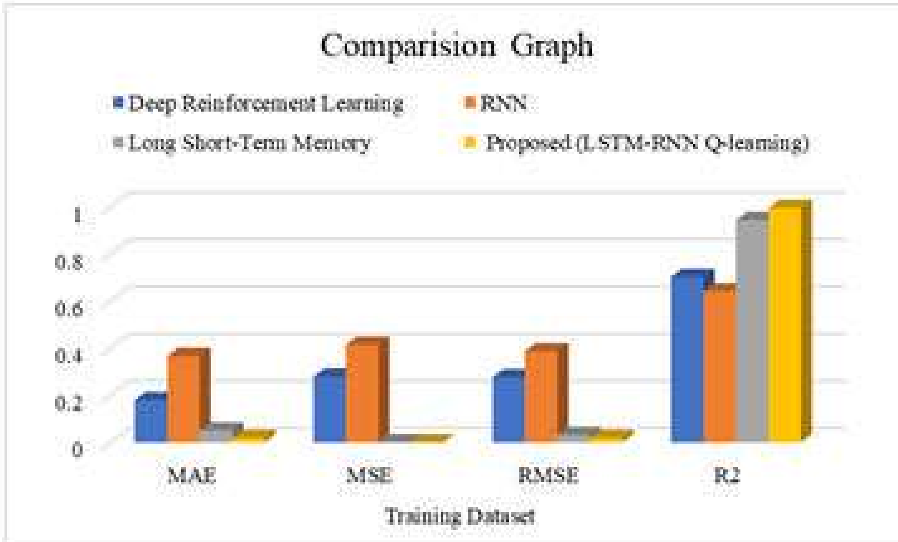

Figure 12: Comparision Graph

CONCLUSION

Agricultural analyses are instrumental in aiding rural communities to make informed and beneficial decisions for farmers. In this study, models were meticulously developed and implemented using Python, ensuring optimal perspectives and tested under similar software and hardware environments to ensure fair comparisons. The error metric is utilized to gauge the performance level during model execution. Residuals, derived from the assessments and representing the variance between actual and predicted values, serve as a measure of error. Ultimately, by examining the magnitude of residual spread, accuracy, as well as model effectiveness, can be determined.

In terms of accuracy and efficacy, the proposed support vector model evidently outperforms the LSTM-RNN with q-learning deep learning models, achieving an impressive accuracy rate of 99.59% and improved error measures specifically for datasets from Gujarat. This suggests promising prospects for utilizing the framework with datasets from other states in the future.

REFERENCES

[16] K. L. Reddy and A. P. S. Kumar, “Machine Learning Techniques for Weather based Crop Yield Prediction,” in 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), 2023, pp. 1263–1268. doi: 10.1109/ICAIS56108.2023.10073740.

[17] R. A. Arun, T. Saraswathi, D. Meeha, and G. Sugeerthi, “A Machine Learning-based Agricultural Yield Forecasting System for Predicting Crop/Plant Yield before Planting the Crops/Plants,” in 2023 12th International Conference on Advanced Computing (ICoAC), 2023, pp. 1–6. doi: 10.1109/ICoAC59537.2023.10249364.

[18] R. ThangaSelvi and M. Sathish, “An Optimal Bidirectional Gated Recurrent Neural Network Model for Crop Yield Prediction,” in 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), 2023, pp. 1302–1305. doi: 10.1109/ICSSIT55814.2023.10060890.

[19] P. Helber et al., “Crop Yield Prediction: An Operational Approach to Crop Yield Modeling on Field and Subfield Level with Machine Learning Models,” in IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, 2023, pp. 2763–2766. doi: 10.1109/IGARSS52108.2023.10283302.

[20] S. M. G, S. Paudel, R. Nakarmi, P. Giri, and S. B. Karki, "Prediction of Crop Yield Based-on Soil Moisture using Machine Learning Algorithms," in 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), 2022, pp. 491–495. doi: 10.1109/ICTACS56270.2022.9988186.

[21] R. J., D. H., and B. P., "A Machine Learning-based Approach for Crop Yield Prediction and Fertilizer Recommendation," in 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022, pp. 1330–1334. doi: 10.1109/ICOEI53556.2022.9777230.

[22] L. B. Rananavare and S. Chitnis, "Monocot Crop Yield Prediction using Sentinal-2 Satellite Data," in 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), 2022, pp. 1–6. doi: 10.1109/GCAT55367.2022.9971820.

[23] S. Chaudhary, S. Mongia, S. Sharma, and N. Singh, "Classification based Interactive Model for Crop Yield Prediction: Punjab State," in 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), 2022, pp. 678–682. doi: 10.1109/SMART55829.2022.10046782.

[24] J. R., V. K. G. Kalaiselvi, A. Sheela, D. S. D, and G. J., "Crop Yield Prediction Using Machine Learning Algorithm," in 2021 4th International Conference on Computing and Communications Technologies (ICCCT), 2021, pp. 611–616. doi: 10.1109/ICCCT53315.2021.9711853.

[25] A. k. Gajula, J. Singamsetty, V. C. Dodda, and L. Kuruguntla, "Prediction of crop and yield in agriculture using machine learning technique," in 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021, pp. 1–5. doi: 10.1109/ICCCNT51525.2021.9579843.

[26] V. Patki and P. Wazurkar, "Application of Data Sampling for Crop Yield Prediction using Stochastic Gradient Descent Neural Networks," in 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), 2021, pp. 685–689. doi: 10.1109/CSNT51715.2021.9509682.

[27] N. Suresh et al., "Crop Yield Prediction Using Random Forest Algorithm," in 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, vol. 1, pp. 279–282. doi: 10.1109/ICACCS51430.2021.9441871.

[28] F. Shahrin, L. Zahin, R. Rahman, A. J. Hossain, A. H. Kaf, and A. K. M. A. M. Azad, "Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning," in 2020 11th International Conference on Electrical and Computer Engineering (ICECE), 2020, pp. 21–24. doi: 10.1109/ICECE51571.2020.9393066.

[29] P. S. Nishant, P. S. Venkat, B. L. Avinash, and B. Jabber, "Crop Yield Prediction based on Indian Agriculture using Machine Learning," in 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1–4. doi: 10.1109/INCET49848.2020.9154036.

[30] A. S. Terliksiz and D. T. Altýlar, "Use Of Deep Neural Networks For Crop Yield Prediction: A Case Study Of Soybean Yield in Lauderdale County, Alabama, USA," in 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 2019, pp. 1–4. doi: 10.1109/Agro-Geoinformatics.2019.8820257.

AUTHOR INFORMATION

Mr. Rama Nandan Tripathi serves as an Assistant Professor in the MCA Department at Dr. Ram Manohar Lohia Avadh University, Ayodhya, U.P.

Abdul Haseeb is a Research Scholar in the MCA Department at Dr. Ram Manohar Lohia Avadh University, Ayodhya, U.P.

Nishant is a Research Scholar in the MCA Department at Dr. Ram Manohar Lohia Avadh University, Ayodhya, U.P.

Mohd Atif Haider is a Research Scholar in the MCA Department at Dr. Ram Manohar Lohia Avadh University, Ayodhya, U.P.

Shua Hayat is a Research Scholar in the MCA Department at Dr. Ram Manohar Lohia Avadh University, Ayodhya, U.P.