An Adaptive Deep Learning Model for Crop Yield Prediction

Ashwin Gadupudi
Computer Science
California State University, Fresno,
United States of America (USA).
agadupudi@mail.fresnostate.edu

Rudra Yamini Rani
Information Technology,
VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad,
India. yaminirani_r@vnrvjiet.in

B Jayaram
Computer Science and Business System
RMK Engineering College,
Kavaraipettai, Tamil Nadu, India
jayaramb05@gmail.com

Nipun Sharma
Electronics and Communication Engineering, Presidency University,
Bangalore, Karnataka, India.
nipun.sharma@presidencyuniversity

Jotiram Krishna Deshmukh
Instrumentation Engineering, Bharati Vidyapeeth College of Engineering, Navi Mumbai, India.
jotiram.deshmukh@bharatividyapeeth.edu

*Pundru Chandra Shaker Reddy
School of Computer Science and Artificial Intelligence, SR University,
Warangal-506371, Telangana, India.
chandu.pundru@gmail.com

2024 2nd International Conference on Computer, Communication and Control (IC4) | 979-8-3503-8793-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/IC457434.2024.10486733

***Abstract*— Agriculture is the study and practice of raising plants and animals for human consumption. Approximately 60.45% of India's land is devoted to farming, making it the country's second largest economic sector. Agriculture and the related agro-industrial sector are crucial to India's economy. Soil components (such as nitrogen-phosphorus &potassium), crop rotation, soil-moisture, air and soil-temperatures, rainfall, and other environmental factors all play important roles in successful farming. One piece of current evidence in this area is a technique that utilizes machine-learning (ML) strategies like Random-Forest (RF), Decision-Tree (DT), and ANN to pick the best crop. In this study, we employ deep learning approaches to improve the suggested model, and together with crop prediction, we obtain clear information regarding the quantities of soil attributes requisite, along with their individual costs. It's more precise than the current model. By analyzing the data provided, it aids farmers in making informed business decisions. Predicting a proper yield requires thinking about the land's climate and soil. The goal is to introduce a Python-based system that makes use of strategic thinking to anticipate the most productive harvest under specific conditions while keeping costs to a minimum. While LSTM &RNN are employed as deep-learning((DL)-strategies in this work, SVM is used as the ML-strategy.**

***Keywords*— Machine-learning, crop-yield, agriculture, prediction, deep-learning**

I. INTRODUCTION

The agricultural sector plays a crucial role in the international market. Understanding overall agricultural output is crucial for determining food security challenges and limiting the consequences of atmosphere change as the populace continues to expand. Predicting crop yields is a major challenge in the farming industry. Both natural factors (such as rainfall and temperature) and artificial ones (such as pesticides) have significant effects on crop production. Decisions about farming risk management and yield anticipating benefit from an accurate understanding of past crop yields [1]. Predicting crop yields is difficult for policymakers on a global, national, and local scale. When deciding what and when to sow, farmers may utilize a consistent model for estimating crop yields. There are a number of methods for estimating crop yields [2]. Crop analysis and prediction using the Internet of Things and machine learning is depicted in Figure 1.

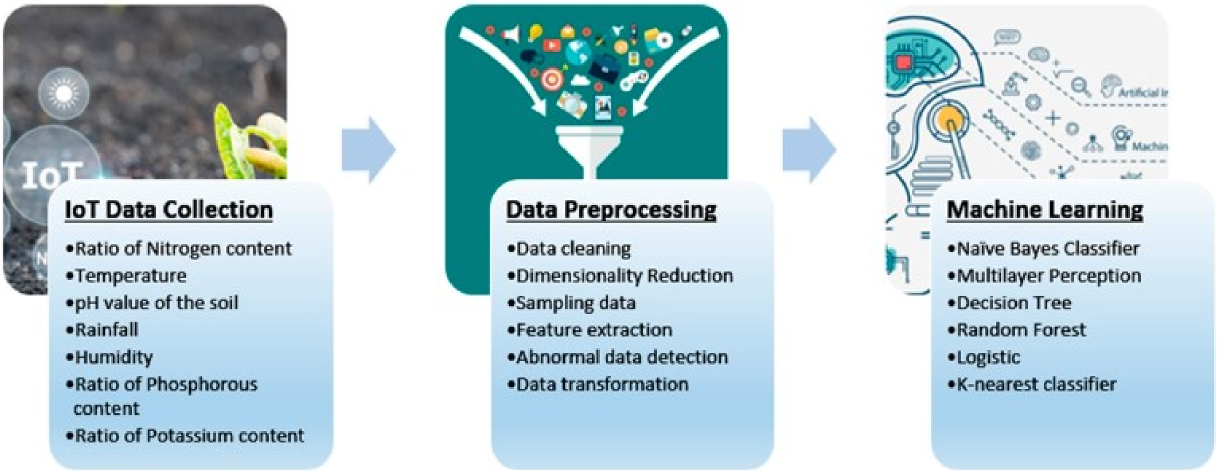

Fig. 1. Outline of the ML &IoT integrated crop yield anticipation process

ML is a practical approach that can improve yield anticipation across multiple parameters. It's a branch of AI that puts an emphasis on learning new things. By spotting trends and associations, machine learning (ML) can mine data for previously unknown insights. For the models to be effective, data sets depicting outcomes based on past experience must be used during training [3]. Machine learning has the potential to aid in agricultural production forecasting, crop selection, and management throughout the growing season. In order to improve the agricultural yield prediction study, several machine learning methods were implemented. Multivariate-regression, DTs, association-rule mining, and ANNs are only some of the ML-strategies that have recently been used to the prediction of crop yields [4]. Predicting crop yields using deep learning models has been increasingly popular in recent years. DL is a sort of ML that can anticipate outcomes from a diversity of raw data groupings. For instance, using ten years of field data, DL-strategies can construct a probability design that sheds light on crop yield under varying climatic conditions. Artificial neural networks (ANNs) are used in deep learning to simulate human cognitive processes. Multi-layered artificial neural networks are the engine behind deep learning. Multilayered networks called DNNs (Deep Neural Networks) are capable of complex processes like illustration and generalization to understand visuals, audio, and text. Similar to how neurons make up the human brain, neural networks consist of layers of nodes. Connecting the nodes of one layer to those of another. The greater the number of layers, the more complex the network. Signals are sent from node to node and given weights in an artificial neural network [5].

The deployment of ML in agriculture has a number of challenges, including a dearth of data foundations, the high cost of sensors and other technologies, and the requirement of specific skill in developing and maintaining the various solutions. However, the potential benefits of implementing ML in cultivation will become clearer as more farms adopt precision agriculture and collect data. It's important to note that the application of ML in farming is still in its infancy, and that more study is required before the full potential of this technology can be realized. The outcomes thus far are encouraging, suggesting that machine learning will gain prominence in the near future [6].

The authors examine how machine learning is changing several different sectors and provide a survey of the methods used in previous research. By employing ML techniques to build designs that accurately anticipate labels based on the input data, they highlight the importance of collecting and analyzing exact data. This research explores a number of classification techniques that can be used to construct such models, including SVM, LSTM, and RNN. Massive shifts in business models are anticipated as a result of the increased efficiency and productivity brought about by the widespread adoption of ML-focused systems &solutions, as predicted by the authors [7].

As for the rest of the paper, the outline looks like this: A overview of the relevant review is provided in Section 2. The technique provided in Section3 is followed by a debate of the results in Section4. Section 5 provides the final analysis.

II. RELATED WORKS

ML-strategies are applied in agricultural yield anticipation techniques to increase the eminence of the crop so that the farmer's turnover is exploited. As a result, the economy as a whole benefit from the better quality of agriculture. The literature provides extensive coverage of this topic. Machine learning strategies for estimating palm oil production are reviewed in [8]. The writers examined the existing literature and made observations about the strengths and weaknesses of the various approaches proposed. In addition to a review and discussion of relevant prior research, this article presents a novel architecture for predicting palm oil yield using machine learning techniques. Since soil qualities have such a profound impact on crop output, the authors of [9] decided to investigate the relationship between soil quality and crop prediction and yield.

In addition, the authors compared the accuracy of each of these methods. To aid farmers in underdeveloped nations that rely on antiquated methods and are unable to assess the true worth of their goods on the market, the authors of [10] suggest a crop production model that would manage the crop utilizing ML-strategies. The suggested system is built on three scenarios: determining which crops are appropriate for a certain region, assisting the farmer in preparing the soil for those crops, and connecting the farmer directly with consumers. The authors used accurate, real-world climate, weather, and soil data in conjunction with SVM, Voting-Regression &RF. K-Nearest Neighbor, SVM, RF, and ANN are all examples of supervised ML-strategies that were proposed by the authors of [11] to assist farmers in making informed decisions about crop selection and production in light of the global shortage of natural resources. The research focuses on tracking the development of chili and cotton plants through the use of mobile phone photos and machine learning methods. The authors of the study hoped that by identifying the most effective ML-strategy for improved crop prediction and analysis, particularly for chili and cotton crops, their work will contribute to the development of smart farming.

In [12], data on soil prediction was gathered from 32 districts in Tamil Nadu, India, using information from the Tamil Nadu Agricultural University (TNAU). The authors claim that the presented comparison results aid farmers in making the right option regarding crop selection and production based on the proportional study of several ML-strategies. In [13], the authors suggested a three-stage framework consisting of data preprocessing and feature-extraction, classification, and performance assessment for the 22 crop kinds to be used in the study. The authors concluded that NB, with an precision of 99.45%, was the most effective classifier for this case. If the authors classified and evaluated performance for an actual problem, this research might yield better results.

In [14], the author conducted a study with the primary aim of assessing the effectiveness of several researches is that attempted to foretell fungal diseases on the crops. The authors conducted an in-depth analysis of several popular machine learning techniques. Based on the available data, the authors found that SVM, variants of choice trees, and Nave Bayes have been the most extensively used and successful machine learning models for yearly crop disease prediction. Using ML &CNN technique, the authors of [15] suggested a system for the early anticipation of crop-diseases in plants. The dataset is trained and tested using data collected from a community. Classifiers are trained to evaluate multiple possible diagnoses and select the one with the highest accuracy. Farmers can use the offered model to anticipate plant illnesses and choose the most suitable crop for their needs. Because of the difficulties Indian farmers have in predicting agricultural yields, the writers of [3] gathered data from the internet. The dataset was clustered with K-Means Clustering for ease of study and analysis, and the NB was utilized to determine the most productive crop to grow. The presented analysis and results demonstrate the usefulness of the suggested approach for farmers, particularly in the areas of early crop production prediction and best crop planning.

Based on the aforementioned literature study, it can be seen that this subject has been examined in the review from many angles. Most of the research in this area has focused on improving agricultural yield and overall production through the use of ML-strategies to aid farmers with crop anticipation. However, many study analyses lacked an actual problem to classify and evaluate the effectiveness of. In addition, there was a lack of transparency from the authors regarding the reliability of the results. In [4], the authors offered investigation and performance appraisal with a 99.45% accuracy rate, albeit only for 22 carefully chosen crops and not using actual data.

III. Proposed Methodology

ML &DL-strategies are used in the proposed framework to provide accurate predictions about crop yields. The proposed model is put to the test using a dataset of crop data. The crop is selected in light of the climate and soil conditions currently in place, as well as the soil's inherent components. When there are multiple possible outcomes, deep learning is utilized to determine which one will yield the best crop. By utilizing this strategy, crops are forecasted properly. Figure 2 illustrates how the SVM algorithm is executed by ML, whereas LSTM &RNN are executed by deep learning.

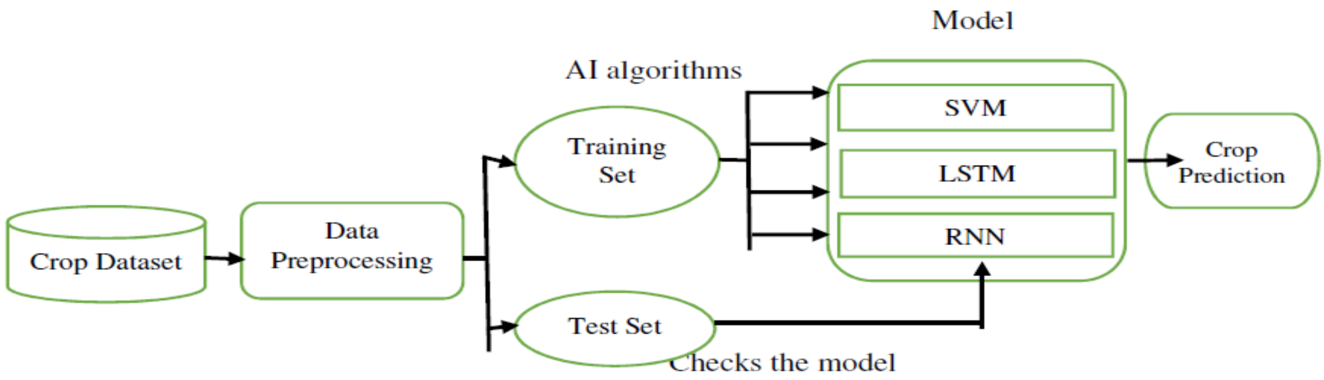

Fig. 2. Proposed Model workflow diagram

A. Dataset Description

The model is fed information from a massive generic crop dataset that includes agricultural parameters. Another dataset is taken as feature dataset. The data is gathered from a website called Kaggle. The total data set for crops is 7841 kilobytes in size. Temperature, precipitation, pH, RH, and landmass are among the variables used for forecasting in this data collection. Wheat, rice, maize, millet, pea, pigeon pea, sugarcane, green gram, etc. are just few of the crops sampled in this data collection. Each parameter used in a single-crop prediction has a range of possible values. Prediction parameters can be assigned any value from the dataset, for example, when wheat is used as the crop. All of the crops in the dataset have the same characteristics.

B. *Feature Selection*

There are several factors to think about when choosing features to use for classification. This is a serious problem because feature selection techniques require very large sample sizes. This study investigates the role of feature selection in inclassification.

Most categorization tasks necessitate supervised learning, which benefits prediction models. The number of redundant features in a large dataset can affect the running time of a learning algorithm, and reducing them improves the quality of the resulting classifier. Through the use of optimization techniques, the most computationally efficient possible combination of feature subsets can be determined. If you want to improve your classification accuracy, reduce the amount of features in your feature vector. Based on the outcomes of a decision model, a wide range of agricultural crops will be planted in the experimental field. There were 100 input variables that were grouped into 25 essential factors, such as land and water and the seasons. To aid farmers in crop selection across different agricultural regions, a decision-making model was constructed using data obtained via the extraction of characteristics. There were 100 different predictors in this data set. A Swarm of Weak Females25 out of the 100 input variables are chosen using an optimization method. To reduce training error, as suggested by the optimization, we use optimization to apply relative weight to features based on their relative information.

C. *Support Vector Machine (SVM)*

To classify data points in a clear way, the support vector machine algorithm seeks to find a hyperplane in N-dimensional space (where N is the number of features in the dataset). Different hyperplanes could be selected to separate the two types of data points. By analyzing the distance between data points in the two classes, one may determine which plane has the better edge. Increasing the edge distance provides some assistance in organizing future data points with greater certainty.

D. *Long short-term memory (LSTM)*

Long-term dependencies are possible in this variety of recurrent neural network (RNN). Neural networks are a set of approaches that attempt to discover patterns in data in the same way that the human brain does. RNNs are merely any feedforward neural network that also has memory built in [5]. Recurrent neural networks (RNNs), as its name suggests, are capable of performing the same function for several data inputs, but with severe dependencies on previous calculations. While most neural architectures use a feedforward strategy, LSTM incorporates feedback connections. In addition to individual data points like images, it can also cycle entire collections of data like discussions and videos. Time series problems, along with other complicated problems, are ideal candidates for this type of network's processing and prediction abilities. The vanishing gradient problem is a significant challenge for any RNN network, although LSTM, as a variant of RNN, helps address this issue. It uses back-propagation to train any model.

IV. RESULTS AND DISCUSSIONS

In order to obtain the outcome, the effectiveness of the suggested model is evaluated. The precision of the result depends on the application of specific equations. Here are the relevant equations:

$$Acc = \frac{Tr_{Positive} + Tr_{Negative}}{Tr_{Positive} + Tr_{Negative} + Fa_{Positive} + Fa_{Negative}} \quad \text{Eq. (1)}$$

$$Pe = \frac{Tr_{Positive}}{Tr_{Positive} + Fa_{Negative}} \quad \text{Eq. (2)}$$

$$Re = \frac{Tr_{Positive}}{Tr_{Positive} + Fa_{Postive}} \quad \text{Eq. (3)}$$

The research is put into action by first loading the collected dataset of crops. The process begins with the import of required libraries and packages and continues with the preprocessing of data. Trained data and test data are spit out. In the end, the necessary artificial intelligence algorithms are included into a model, and the results reveal the optimal crop for a given plot of land.

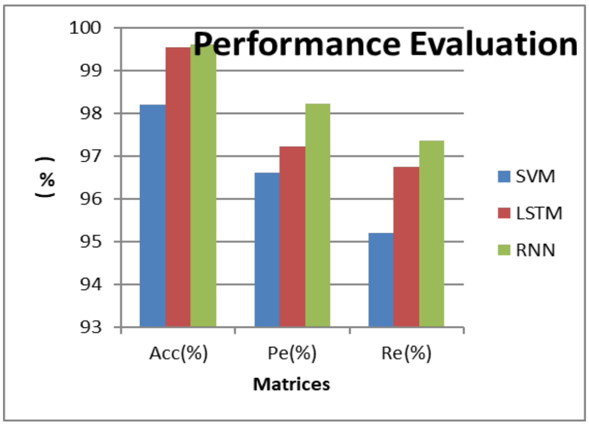

Fig.3. Proposed model performance assessment

The analysis of the dataset depends on a few key factors. The variables in question are, respectively, area, temperature, rainfall, soil location, and relative humidity. These characteristics are used to assess the dataset and make accurate predictions about the outcome. Some fields are only suitable for growing certain types of crops. Some examples of such crops are wheat, rice, corn, peas, pigeon peas, green gram, potatoes, sugarcane, soybeans, etc.

The proposed models' experiments are conducted on vast scale and performance is estimated in terms of accuracy (Acc), precision (Pe) and Recall (Re) are presented in table 1 and figure 3.

Table 1. Proposed models comparison

Model	Acc (%)	Pe (%)	Re (%)
SVM	98.21	96.61	95.21
LSTM	99.54	97.22	96.75
RNN	99.62	98.24	97.36

From, the figure 3 shows that, RNN model achieves better results compared to other models i.e. SVM, LSTM in terms of accuracy is 99.62%. The same is compared with state of art models and the results show that RNN attained superior accuracy than others. The Table 2 & figure 4 shows the comparison analysis.

Table 2. Comparative Analysis

Model	Acc (%)
RF	97.91
LR	97.28
CNN	98.67
Proposed (RNN)	99.62

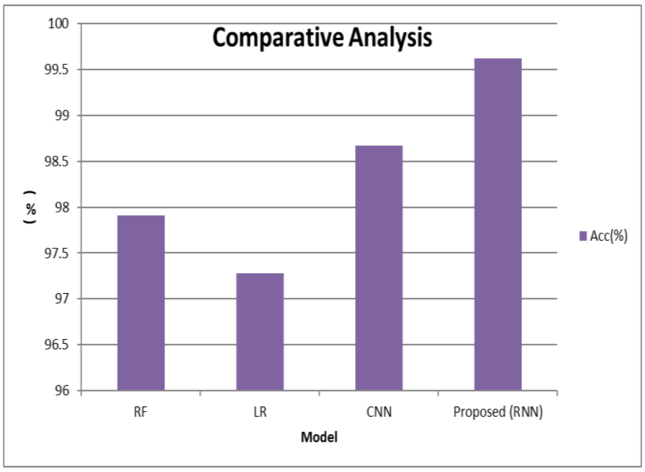

Fig.4. Comparative Analysis

V. CONCLUSIONS

To help farmers avoid losing money on their crops since they don't know how to grow them in a variety of soil and climatic circumstances, an AI-based model has been developed. The model is developed using a combination of supervised ML and unsupervised DL (LSTM &RNN). After assessing the prediction parameters, the model determines which of a variety of crops should be cultivated on a given plot of land in order to maximize profit while minimizing costs. No other published work employs the same methods for crop prediction, at least not that we could find. Thus, it is determined that the exactness of this investigation increases when equated to the current work that utilized other methodologies for crop estimate. It has a wide range of potential future applications and can be implemented and interfaced with a wide variety of skills and flexibility. Farmers need education, and thanks to mobile phones, they'll be able to easily access up-to-date data on how to maximize their harvests. There will be no future loss if the breeder is at home and the work needs to be handled at this very moment. Advances in agribusiness will be much appreciated, and they will ultimately aid farmers in agricultural output.

REFERENCES

[1] Patil, S.B., Rao, H.R., Chatrapathy, K., Kiran, A., Kumar, A.S. and Reddy, P.C.S., 2023, August. Ensemble Deep Learning Framework for Classification of Skin Lesions. In 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT) (pp. 1550-1555). IEEE.

[2] Reddy, P.C.S., Muller, P.S., Koka, S.N., Sharma, V., Sharma, N. and Mukherjee, S., 2023, November. Detection of Encrypted and Malicious Network Traffic using Deep Learning. In 2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE) (pp. 1-6). IEEE.

[3] Kumar, K., Pande, S.V., Kumar, T., Saini, P., Chaturvedi, A., Reddy, P.C.S. and Shah, K.B., 2023. Intelligent controller design and fault prediction using machine learning model. *International Transactions on Electrical Energy Systems*, *2023*.

[4] Baskar, S., Nandhini, I., Prasad, M.L., Katale, T., Sharma, N. and Reddy, P.C.S., 2023, November. An Accurate Prediction and Diagnosis of Alzheimer's Disease using Deep Learning. In 2023 IEEE North Karnataka Subsection Flagship International Conference (NKCon) (pp. 1-7). IEEE.

[5] Rao, C.N., Chatrapathy, K., Fathima, A.J., Sathish, G., Mukherjee, S. and Reddy, P.C.S., 2023, October. Intelligent Deep Learning Framework for Breast Cancer Prediction using Feature Ensemble Learning. In 2023 4th IEEE Global Conference for Advancement in Technology (GCAT) (pp. 1-5). IEEE.

[6] Joshi, K.P., Prasad, M.L., Natchadalingam, R., Reddy, P.C.S., Mukherjee, S. and Babu, G.C., 2023, November. An Accurate Prediction of Coronary Heart Disease Using Ensemble Algorithms. In 2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE) (pp. 1-5). IEEE.

[7] Mamatha, B., Rashmi, D., Tiwari, K.S., Sikrant, P.A., Jovith, A.A. and Reddy, P.C.S., 2023, August. Lung Cancer Prediction from CT Images and using Deep Learning Techniques. In 2023 Second International Conference on Trends in Electrical, Electronics, and Computer Engineering (TEECCON) (pp. 263-267). IEEE.

[8] Suneel, S., Balaram, A., Amina Begum, M., Umapathy, K., Reddy, P.C.S. and Talasila, V., 2024. Quantum mesh neural network model in precise image diagnosing. Optical and Quantum Electronics, 56(4), p.559.

[9] Sucharitha, Y., Reddy, P.C.S. and Suryanarayana, G., 2023. Network Intrusion Detection of Drones Using Recurrent Neural Networks. *Drone Technology: Future Trends and Practical Applications*, pp.375-392.

[10] Patil, S.B., Rao, H.R., Chatrapathy, K., Kiran, A., Kumar, A.S. and Reddy, P.C.S., 2023, August. Ensemble Deep Learning Framework for Classification of Skin Lesions. In *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)* (pp. 1550-1555). IEEE.

[11] Latha, S.B., Dastagiraiah, C., Kiran, A., Asif, S., Elangovan, D. and Reddy, P.C.S., 2023, August. An Adaptive Machine Learning model for Walmart sales prediction. In *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)* (pp. 988-992). IEEE.

[12] Sucharitha, Y., Reddy, P.C.S. and Chitti, T.N., 2023, July. Deep learning based framework for crop yield prediction. In *AIP Conference Proceedings* (Vol. 2548, No. 1). AIP Publishing.

[13] Kumar, G.R., Reddy, R.V., Jayarathna, M., Pughazendi, N., Vidyullatha, S. and Reddy, P.C.S., 2023, May. Web application based Diabetes prediction using Machine Learning. In *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (pp. 1-7). IEEE.

[14] Yadala, S., Pundru, C.S.R. and Solanki, V.K., 2023, March. A Novel Private Encryption Model in IoT Under Cloud Computing Domain. In *The International Conference on Intelligent Systems & Networks* (pp. 263-270). Singapore: Springer Nature Singapore.

[15] Deivasigamani, S., Rani, A.J.M., Natchadalingam, R., Vijayakarthik, P., Kumar, G.B.S. and Reddy, P.C.S., 2023, August. Crop Yield Prediction Using Deep Reinforcement Learning. In 2023 Second International Conference on Trends in Electrical, Electronics, and Computer Engineering (TEECCON) (pp. 137-142). IEEE.