Cultivating Prosperity: A Fusion of IoT Data with Machine Learning and Deep Learning for Precision Crop Recommendations

# Bhumika Fattepur

*School of CSE*

*KLE Technological University* Hubballi, India [bhumikavfattepur@gmail.com](mailto:bhumikavfattepur@gmail.com)

# Yash

*School of CSE*

*KLE Technological University* Hubballi, India [yashhalkati600@gmail.com](mailto:%20yashhalkati600@gmail.com)

# Sakshi A

*School of CSE*

*KLE Technological University* Hubballi, India [sakshiananthapur@gmail.com](mailto:sakshiananthapur@gmail.com)

# Sneha Varur

*School of CSE*

*KLE Technological University* Hubballi, India [sneha.varur@kletech.ac.in](mailto:sneha.varur@kletech.ac.in)

# Abhishek A

*School of CSE*

*KLE Technological University* Hubballi, India [asangadi45@gmail.com](mailto:asangadi45@gmail.com)

# Meena S M

*School of CSE*

*KLE Technological University* Hubli, India [msm@kletech.ac.in](mailto:msm@kletech.ac.in)

***Abstract*—The selection of crops in precision agriculture poses challenges, affecting food security and economic stability due to factors like sub-optimal yields and resource wastage stemming from variables such as soil composition and climate. These chal- lenges underscore the importance of holistic solutions.Utilizing an set of sensors, data collection becomes an indispensable cornerstone for ongoing research. This data gathering consti- tutes a pivotal element in facilitating exhaustive analysis and supporting informed decision-making within the agricultural domain.Addressing the problem at hand, the paper proposes a multi-layered approach that combines homogeneous and hetero- geneous ensemble learning models, alongside integrating deep learning architectures. It introduces powerful ensemble learn- ing strategies such as RandomForestClassifier, GradientBoost- ingClassifier, XGBClassifier, Support Vector Classifier (SVC), MLPClassifier, Logistic Regression, and a variety of other models. Through this ensemble approach, it leverages the strengths of these models to enhance the accuracy and robustness of crop recommendations. Additionally, the study delves into the realm of deep learning, employing Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and RNN-LSTM models, which are finely tuned to capture subtle data patterns essential for accurate crop yield prediction. Our study carefully evaluates the performance of these models through rigorous cross-validation and fusion techniques, highlighting their transformative potential to revolutionize crop recommendation practices.By providing data-driven insights to farmers, we aim to pave the way for sustainable agriculture and improved food security in an ever- evolving agricultural landscape.**

**Keywords-Support Vector Classifier (SVC), Long short-term memory (LSTM), Convolutional Neural Networks (CNN), Re- current Neural network (RNN),Multilayer perceptron (MLP)**

1. INTRODUCTION

Agriculture, a fundamental activity entrenched in human societies across history, [[7]](#_bookmark15) serves as a cornerstone that sustains civilizations. Its paramount role lies in providing essential food resources for a growing global population, intertwining with economic development to shape nations and influence communities’ well-being. As a primary livelihood source, agriculture plays a pivotal role in molding a country’s economic landscape, and influencing GDP, employment, and trade. Beyond its economic impact, [[8]](#_bookmark16) agriculture weaves into the social and cultural fabric of societies, underscoring its significance for sustained national development.The suc- cess of agriculture hinges on creating optimal environmental conditions, with factors like light, humidity, and soil quality pivotal for maximizing productivity. Strategic management of these conditions not only ensures efficient resource uti- lization but also contributes to sustainable farming practices, minimizing environmental impacts. [[8]](#_bookmark16) Crop selection, a key aspect, requires a delicate balance to avoid mismanagement leading to wastage, which could jeopardize food supplies and have severe economic repercussions. Addressing such challenges, agriculture faces the formidable threat of drought, necessitating innovative solutions beyond traditional practices. The advancement of IoT [[5]](#_bookmark13) technology provides real-time monitoring of crop health, environmental conditions, and soil quality, which is a major assistance to the Indian agriculture industry. This makes data-driven decision-making, precision farming, and more precise crop forecasts possible. IoT [[5]also](#_bookmark13) makes it possible to access market data, instructional materials, and remote control—all of which help farmers maximize

agricultural yields and enhance their financial stability while reducing resource waste.

Machine learning [[6]](#_bookmark14) models play a pivotal role in solving the aforementioned challenges. They can process vast amounts of historical data, encompassing various parameters such as weather conditions, soil attributes, and geographical factors. These models excel at recognizing intricate patterns and re- lationships within the data, enabling accurate predictions of crop yields. By leveraging machine learning, [[6]](#_bookmark14) we can min- imize uncertainties in crop selection and optimize agricultural practices.

In this research two objectives are set. Firstly, Collection of data utilizing a set of IoT sensors, ensuring a comprehen- sive and accurate dataset. Secondly, assessing the efficacy of various machine learning and deep learning methods through rigorous evaluation to determine their accuracy in predicting crop yields.

The research seeks to leverage the insights gained from these models to enhance agricultural sustainability. By offering data-driven insights and precise crop recommendations, the aim is to empower informed decision-making and promote sustainable agricultural practices.

Further, Section II introduces the Background Study, estab- lishing the context and importance of the research. In Section III, the Literature Survey reviews existing work, identifying gaps and emphasizing the research’s distinctive contribution. Section IV presents the Proposed Methodology, detailing the utilization of IoT sensors for data collection and the imple- mentation of machine learning and deep learning methods for predicting crop yields. Section V, Results and Discussions, analyzes the study’s findings, and Section VI, Conclusion and Future Scope, summarizes outcomes, underscoring potential areas for future research.

1. BACKGROUND STUDY
2. *Ensemble Learning:* Ensemble learning [[9]](#_bookmark17) is a powerful machine learning strategy that harnesses the collective wisdom of multiple models to enhance overall performance. Instead of relying on a single model, ensemble techniques leverage the diversity of various models to achieve superior prediction accuracy and resilience.
3. *Heterogeneous Ensemble Learning:* Heterogeneous ensemble learning. [[10]](#_bookmark18) combines a range of diverse machine learning algorithms into a unified model, aiming to boost overall predictive accuracy and resilience by capitalizing on the strengths of multiple models

Voting Classifier

The Voting Classifier is an ensemble approach that combines the predictions of multiple machine-learning models. Each model provides its prediction, and the final decision is made through either majority voting (for classification) or averaging (for regression). This technique is beneficial when aiming to utilize the diversity among models to enhance overall

predictive performance. Stacking Classifier

Stacking combines models by using a meta-model to optimize the combination of base models’ predictions, allowing for better performance when dealing with complex relationships. However, it can be more resource-intensive than simpler ensemble methods.

1. *Homogeneous Ensemble Learning:* Homogeneous ensemble learning [[10]](#_bookmark18) involves employing multiple instances of a single machine learning algorithm, such as the Random Forest Classifier (RFC), to construct a unified model. This approach is designed to harness the strengths inherent to a specific algorithm.

Bagging Classifier

Bagging is a technique that involves the generation of multiple data subsets, training of diverse models on these subsets, and combining their predictions to decrease variance and enhance accuracy. It is frequently applied in conjunction with decision trees.

AdaBoost Classifier

AdaBoost is an iterative approach that progressively gives greater importance to misclassified data points, subsequently training weak classifiers and merging them into a potent ensemble model. This method is especially valuable for enhancing the classification performance when working with weak base models.

1. *Recurrent neural networks:* Recurrent neural networks

[[11]](#_bookmark19)are a kind of neural networks that are optimized for processing sequential data. By preserving internal memory,see fig [1](#_bookmark0) RNNs can recognize patterns and dependencies over time. It might, however, be plagued by the vanishing gradient issue.

.

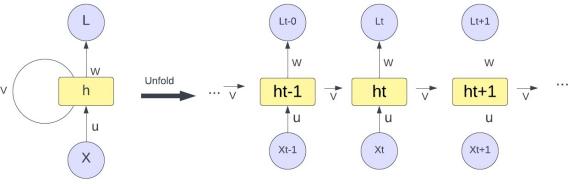


Fig. 1. Recurrent neural network

1. *Long Short-Term Memory:* Long Short-Term Memory

[[12]](#_bookmark20) is an RNN type that incorporates a memory cell to address the vanishing gradient problem and allow the network to selectively remember and forget information over extended sequences of time is called an LSTM. This makes it especially

useful for applications like natural language processing that call for the modeling of long-term dependencies,refer fig [2](#_bookmark1) .

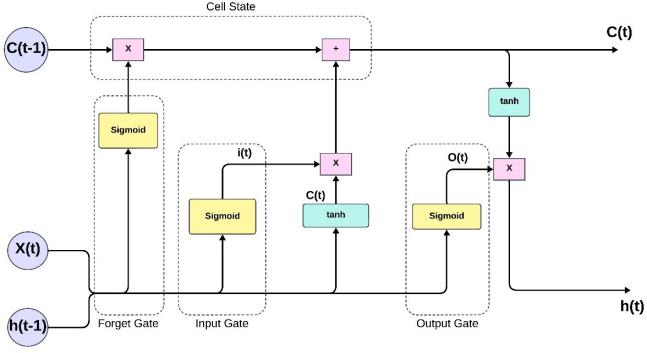


Fig. 2. Long-short term memory

1. *Convolutional Neural Network :* CNN [[13]](#_bookmark21) is a type of neural network architecture designed specifically for process- ing and image recognition. It employs convolutional layers to automatically learn spatial hierarchies of features, making it very successful for jobs involving visual data.
2. LITERATURE SURVEY

Rakesh et al. [[1]e](#_bookmark9)xplore the integration of machine learning, especially deep learning, to solve agricultural problems with a focus on crop optimization. The researchers proposed a three-stage system consisting of preliminary data, distribution, and performance evaluation, approving 22 products through correlation analysis, distribution analysis, co-sharing, and majority voting. Naive Bayes, in particular, appeared to be the most accurate, with an accuracy rate of 99.54 percent, exceeding the 98.52 percent accuracy achieved by the majority voter. This study demonstrates the potential of machine learning, providing insights and suggesting future research directions, including the integration of deep neural networks and in-flight applications to improve predictive capabilities.

Shahbaz et al. [[2]proposed](#_bookmark10) a new stitching method with 27 layers including LSTM model color blocking. The model outperformed existing methods by identifying the most suitable crop among 22 varieties with 97.26 percent accuracy,

96.9 percent sensitivity, 95.69 percent F1 score and 96.56 percent recall rate. Future plans include sharing business intelligence with business customers and building mobile applications and web interfaces to ensure trust.

Kusum et al. [[4]](#_bookmark12) have proposed the work to address the impact of global population growth on food security, highlighting the important role of agricultural technology in improving crop management. The researchers evaluated and compared machine learning methods, specifically support vector machine (SVM) and neural networks with short- term neural networks (RNN-LSTM), to obtain an accurate prediction of historical crop data from the Indian government website. It is worth noting that the accuracy of SVM is slightly

higher (78 percent) compared to RNN-LSTM (70 percent) and the results are based on the difficulty of using the real world. The article reveals the potential of machine learning to process crop data and highlights the impact of historical data on accurate predictions. The results demonstrate the feasibility of this machine learning technique in helping farmers with crop and yield predictions, paving the way for various SVM and RNN-LSTM hybrid models of the future.

Priyadharshini et al. [[3]](#_bookmark11) proposed ensemble learning using methods such as decision tree(accuracy 81 percent), KNN(accuracy 85 percent), linear regression(accuracy 88.26 percent), naive Bayes(accuracy 82 percent), neural networks(accuracy 89.88 percent), and SVM(accuracy 78 percent). Their research focuses on important issues such as soil geography and environmental degradation to provide Indian farmers with detailed information on crop management to reduce crop losses, failures, and failures. Future plans include launching a web interface and mobile app for use by farmers across the country.

1. PROPOSED WORK

The figure [3](#_bookmark2) depicts a crop recommendation system that utilizes a combination of RNN-LSTM, CNN, and ensemble models in two primary stages: training and testing. During the training phase, the system uses an existing dataset, while in the testing phase, data is collected using IoT sensors. Both stages involve similar preprocessing steps, and the trained models are then employed to generate crop recommendations. The system leverages ensemble techniques alongside RNN-LSTM and CNN models to enhance the accuracy and effectiveness of its recommendations.

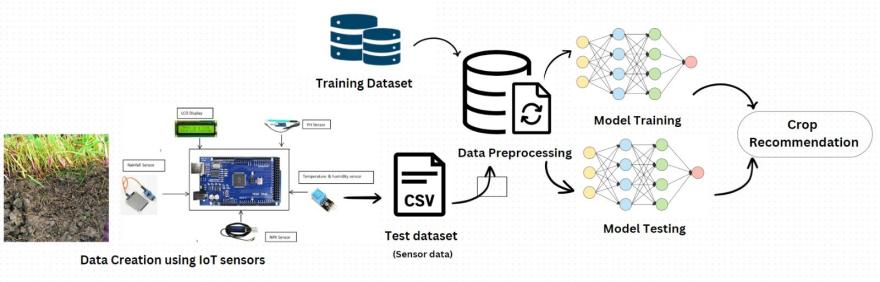


Fig. 3. Proposed model of crop recommendation system

1. *Data acquisition*

For the training phase, we utilized a dataset acquired from the Kaggle website. This dataset includes seven distinct features, covering rainfall, temperature, humidity, pH levels, as well as Nitrogen, phosphorus, and potassium content in the soil. Encompassing twenty-two different crop types like rice, maize, chickpeas, and others, this dataset serves as the foundation for analyzing environmental conditions and soil characteristics, forming the input for model training.

For the testing phase, the dataset with the same features was gathered manually by employing the set of IoT sensors, including pH sensors, rainfall sensors, DHT11 sensors to measure humidity and temperature, as well as NPK sensors to record nitrogen, phosphorus, and potassium values from the soil. This separate dataset is employed to evaluate and test our model’s performance. It is important to note that this data collection process occurred under a variety of weather conditions in farms located in the vicinity of the Hubbli- Dharwad region.

The dataset we gathered can exclusively recommend a specific set of crops due to the favorable soil and weather conditions in the local area. These recommended crops are rice, maize, black gram, banana, mango, papaya, coconut, cotton, lentils, pomegranate, muskmelon, and watermelon.

1. *Data Preprocessing*

The preprocessing of the training dataset involved address- ing missing values, verifying data completeness, and eliminat- ing outliers to improve the quality of the dataset. In the testing dataset, normalization was employed. This technique ensures consistent and accurate model evaluation by standardizing the scale of numerical characteristics across various data points.

1. *Approach-1-Ensemble Learning*
   1. *Heterogeneous Ensemble Learning:* In this approach, di- verse foundation models are selected, including MLP, Random Forest, Gradient Boosting, XGBoost, and linear SVM. Two ensemble models, Voting Classifier (Hard Voting) and Stack- ing Classifier, combine base model predictions for improved accuracy. Both ensembles are trained on the training set using the ”fit” approach, and predictions on the test data are made using the ”predict” approach. Evaluation is based on accuracy scores, measuring the percentage of correct predictions on the test data for the Voting and Stacking ensembles, refer [IV-C1](#_bookmark3)

.

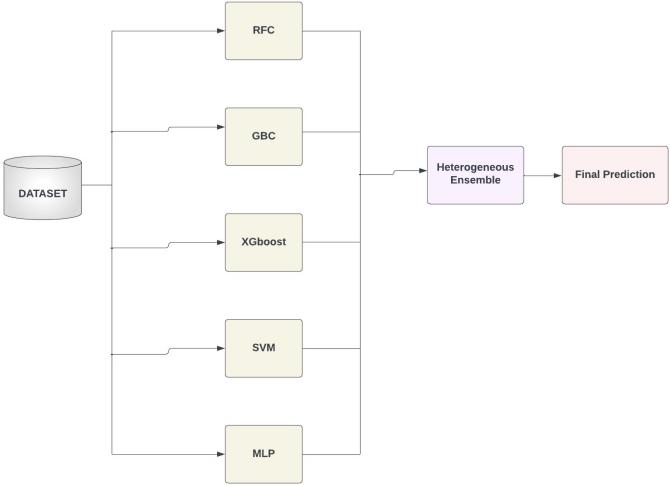


Fig. 4. Heterogeneous Ensemble Learning

* 1. *Homogeneous Ensemble Learning:* In this approach, the chosen foundation models share a common machine- learning algorithm but differ in subsets of data (estimators and depths). Both BaggingClassifier and AdaBoost utilize 100 base estimators. Bagging employs bootstrapping with Decision Tree Classifiers, while AdaBoost iteratively adjusts weights to emphasize misclassified samples. Both ensemble models are trained on (Xtrain, Ytrain) and predicted on Xtest. Evaluation involves accuracy scores, indicating the percentage of correct predictions on the test data for each ensemble, refer [5](#_bookmark4) .

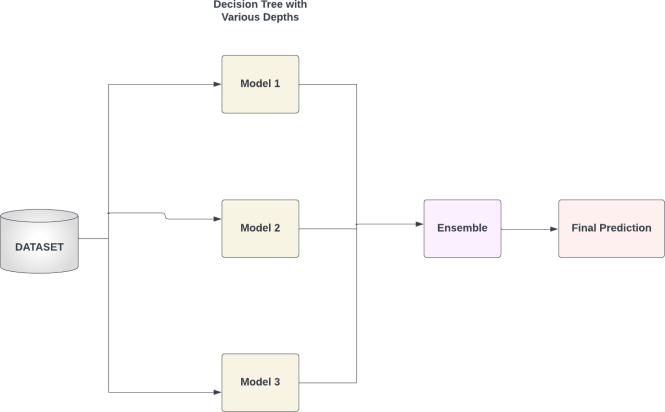


Fig. 5. Homogeneous Ensemble Learning

1. *Approach-2-RNN-LSTM*

In this approach, the model is developed using Keras and TensorFlow and is structured as a sequential model. It incorporates a Simple RNN layer and is followed by an LSTM layer, both with 50 units and tanh activation. The output layer consists of a Dense layer with softmax activation. The model is assembled using the Adam optimizer and categorical cross- entropy loss, with accuracy as the chosen evaluation metric. During training, a batch size of 32 is utilized for 100 epochs. Following training, the model is applied to the test set for predictions. The projected probabilities are converted to class labels using argmax, and the model’s accuracy is evaluated by comparing the predicted labels with the ground truth labels. The final step involves printing and reporting the accuracy of our combined RNN-LSTM model on the test set, refer [6](#_bookmark5) .

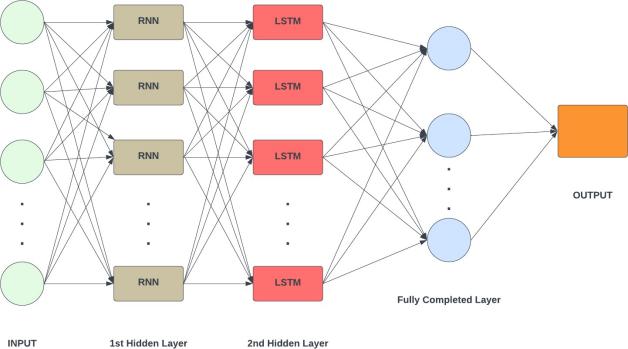


Fig. 6. RNN - LSTM

1. *Approach-3-Convolutional Neural Network*

In this approach, The proposed methodology employs a Convolutional Neural Network (CNN) to address outlier de- tection and class imbalance. Z-scores and the Interquartile Range (IQR) identify and remove outliers, while the Synthetic Minority Over-sampling Technique (SMOTE) tackles class imbalance. The CNN model, including Conv1D, MaxPool- ing1D, Flatten, Dense, and Dropout layers, utilizes softmax activation for multi-class classification. To prevent overfitting, an EarlyStopping callback is integrated during a training period limited to a maximum of 500 epochs. Validation against reshaped testing data yields the final accuracy, offering an effective solution for outlier detection, class imbalance, and multi-class classification, refer [7](#_bookmark6) .

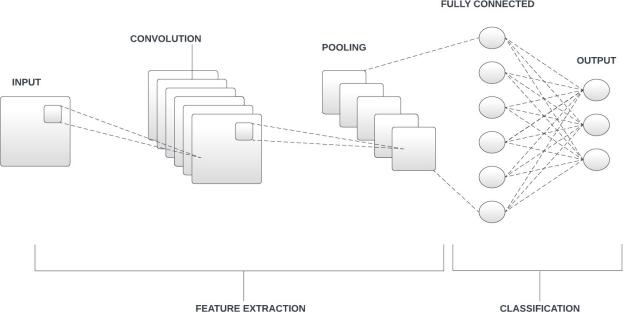


Fig. 7. CNN

1. RESULTS AND DISCUSSION

In the below results the accuracy of the model steadily grows over the epochs, peaking around 98.25 percent in approach 2, refer [8](#_bookmark7) and 97 percent in approach 3, refer [9.](#_bookmark8) This implies that the model is generalizing well to new data and learning efficiently.

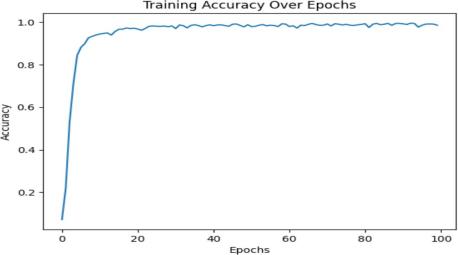


Fig. 8. Approach-2-Epoch v/s Accuracy

. .

The below table [V](#_bookmark8) showcases the training and testing accura- cies for all the models used for crop recommendation.

Tabel-1. Performance Summary Table

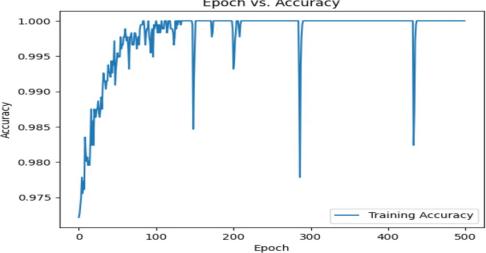


Fig. 9. Approach-3-Epoch v/s Accuracy

|  |  |  |
| --- | --- | --- |
| Model | Training Accuracy | Testing Accuracy |
| Voting Classifier | 99.54 | 94.56 |
| Stacking Classifier | 99.54 | 94.56 |
| Bagging Classifier | 98.63 | 92.8 |
| Boosting Classifier | 97.5 | 91.9 |
| RNN - LSTM | 97.7 | 90.2 |
| CNN | 97.7 | 90.23 |

1. CONCLUSION AND FUTURE SCOPE

The research has shown promising outcomes in the development of crop recommendation classifiers. Notably, the voting, stacking, and bagging classifiers achieved remarkably high training accuracy scores. However, it’s worth noting that the boosting classifier, RNN-LSTM, and CNN exhibited lower training accuracies, but these results still indicate their potential in crop recommendation systems.It’s important to emphasize that the testing accuracies were slightly lower than the training accuracies. This discrepancy can be attributed to the limited size and scope of the dataset, primarily focusing on crops cultivated within the Hubbli-Dharwad region. The dataset’s specificity may have hindered the models’ ability to generalize to a wider range of crops and regions.

This study can be used to identify potential future ad- vancements in crop recommendation systems. Integrating ad- vanced sensors, harnessing big data analytics and AI, and incorporating genetic and crop variety data offer promising avenues. Enhancing personalization, integrating weather data, automation technologies, and sustainability metrics are key areas for improvement. Emphasizing data security, blockchain traceability, and regulatory compliance is essential. Addition- ally, fostering community engagement and incorporating mar- ket predictions will refine crop recommendations, enhancing agricultural sustainability and profitability.

REFERENCES

1. R. K. Ray, S. K. Das and S. Chakravarty, ”Smart Crop Recommender System-A Machine Learning Approach,” 2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence), Noida, India, 2022, pp. 494-499, doi: 10.1109/Confluence52989.2022.9734173.
2. Sikandar, Shahbaz Mahum, Rabbia Aladhadh, Suliman. (2023). Automatic Crop Expert System Using Improved LSTM with Attention Block. Computer Systems Science and Engineering. 47. 2007-2025. 10.32604/csse.2023.037723.
3. P. A, S. Chakraborty, A. Kumar and O. R. Pooniwala, ”Intelligent Crop Recommendation System using Machine Learning,” 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 843-848, doi: 10.1109/ICCMC51019.2021.9418375..
4. https://cnre2023.ptu.ac.in/CNRE21/20.pdf .
5. Prasad, Ganapathi Vanathi, A. Devi, B.S.. (2023). A Review on IoT Applications in Smart Agriculture. 10.3233/ATDE221332. .
6. Sharma, Abhinav Jain, Arpit Gupta, Prateek Chowdary, Vinay. (2020). Machine Learning Applications for Precision Agriculture: A Comprehensive Review. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3048415. .
7. Ozor, Nicholas. (2008). Challenges and impacts of agricultural biotechnology on developing societies. African Journal of Biotechnology (ISSN: 1684-5315) Vol 7 Num 4. 7. .
8. Wreford, A., D. Moran and N. Adger (2010), Climate Change and Agriculture: Impacts, Adaptation and Mitigation, OECD Publishing, Paris, https://doi.org/10.1787/9789264086876-en. .
9. Ying, Xu. (2014). Ensemble Learning. .
10. Petrakova, Aleksandra Affenzeller, Michael Merkuryeva, Galina. (2015). Heterogeneous versus Homogeneous Machine Learning Ensembles. Information Technology and Management Science. 18. 10.1515/itms-2015-0021. .
11. Elsayed, Ebrahim. (2020). Recurrent Neural Networks. .
12. Hochreiter, Sepp Schmidhuber, Ju¨rgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735. .
13. O’Shea, Keiron Nash, Ryan. (2015). An Introduction to Convolutional Neural Networks. ArXiv e-prints. .
14. Ruchirawya, Thewahettige. (2020). Crop Recommendation System. .
15. Ruchirawya, Thewahettige. (2020). Crop Recommendation System. .
16. Pudumalar, S. Elangovan, Ramanujam Rajashree, R. Kavya, C. Kiruthika, T. Nisha, J.. (2017). Crop recommendation system for precision agriculture. 32-36. 10.1109/ICoAC.2017.7951740. .