A Data-Centric Approach to Weather Prediction: IoT and Machine Learning on Drone Platforms

Devendra Parihar*, Ajay Chaudhary‡, Shreehitha Peddoju† and Kavitha Kadarla§
*Helios Solutions, Vadodara, India, *‡Engineering College Bikaner, Bikaner, India
†National Institute of Technology Rourkela, India, §COER University, Roorkee, India
Email: *devendraparihar340@gmail.com, †shreehitha.p@gmail.com,‡ajaychaudhary@acm.org,§satkavi.23@gmail.com

***Abstract*—The Internet of Things (IoT) is increasingly influential in transforming various sectors, including agriculture, by providing cost-effective sensors and easily deployable infrastructure. This adoption is revolutionizing the management of agricultural activities, shifting from traditional practices to data-centric and automated decision-making processes. Machine learning (ML) algorithms are integral to this transformation, particularly in predicting and forecasting weather conditions critical for agriculture. However, ML algorithms' high computational and memory demands pose significant challenges for small IoT devices. The Emergence of Edge Devices with High Computation Capabilities may be a feasible solution. This paper proposes a drone-based weather prediction system designed to automate agriculture's weather prediction and decision-making process. The DL/ML model may be trained on historical weather at high-computation facilities, and this pre-trained model is deployed on the drone-mounted edge node. The edge node-based DL/ML pre-trained model, integrated with IoT sensors, processes real-time weather data to make informed actuation decisions. Our experimental results, focused on rainfall prediction, show that the system achieves high accuracy, with logistic regression performing best on the Bikaner dataset and XGBoost excelling on the Australian data. The precise weather predictions enabled by our system lead to more accurate agricultural decisions, thereby optimizing the automation of agricultural processes.**

***Index Terms*—Internet of Things,IoT, Edge Node, Drone, Weahter Predition, Weather Forecasting, Precision Farming, Smart Agriculture, Machine Learning.**

I. Introduction

The agriculture sector is one of the primary sources of income in most developing economies. It not only serves the financial aspect but also helps to provide food for the growing population. There is a dire need to enhance agriculture outputs; however, due to limited arable land, alternative solutions must be adopted to improve agricultural production. Some cutting-edge technologies that may enhance overall agri production growth in current circumstances are the Internet of Things (IoT), machine and deep learning algorithms, cloud computing infrastructure, and intermediate infrastructure such as edge, fog, and mist computing. These technologies can improve the overall resource and agri-decision management, such as optimal water cycles and other resource utilization. The IoT nodes sense the environmental variable and actuate decisions that help in the water & fertilizer cycles, pest & weed management, and other resource management. Predicting and forecasting are essential variables that immensely affect the outcome of agricultural products. A well-advanced prediction of weather, undesirable weeds, pests, and nutritional requirements may help improve end yield in many ways. Machine learning and deep learning algorithms-based models can predict and forecast weather variability and other factors that may be used for precision agriculture. The high computation and storage capabilities of cloud computing enable machine learning and deep learning algorithms to perform efficiently on large datasets. In a typical scenario, latency hampers the performance and decisions due to the long distance and limited bandwidth communication between the IoT nodes and cloud infrastructure. However, based on input data received through IoT sensor nodes, the edge node is immediately processed and delivered back to the IoT network using a trained model provided by a cloud computing facility[1]. The deployment of the drone-mounted edge node as a gateway node may enhance the capabilities of IoT devices in many folds by timely performance actuation based on delivered agri intelligence by the drone-mounted edge node. It helps in reducing overall latency and efficiently optimizes network traffic[2]–[4].

II. Background and Related Work

The IoT is a disrupted technology that enhanced the efficiency of almost all sectors by many folds[7]–[9]. The IoT is also widely adopted in the agricultural sector for precision farming and is one of the most common agricultural technologies used to optimize and automate operations with the availability of low-cost and off-shelf plug-and-play sensors [5], [6]. It helps in optimizing the overall activities and smartly monitors the key parameters. It automates activities such as remote monitoring, environmental data collection, actuation, and other activities such as pest and weed management to improve the overall yield of crops. The smartly deployed IoT sensor infrastructure may deliver services for an extended period by optimizing power consumption and energy harvestion[10]. The key challenge with IoT infrastructure is storing and processing the sensed data due to low-cost IoT hardware's limited storage and processing capabilities. As a result, to drive viable decisions using sensed data, the data must be stored and processed with a high computation facility. Recently, researchers have proposed innovative ideas for using intermediatory processing and storage hardware such as mist, fog, and edge computing devices based on the processing and storage capacity requirement. Integrating IoT and cloud computing through intermediate layers such as edge, fog, and mist, the data flow from the sensor to final consumption is highly streamlined [1], [11]. Deploying the drone-mounted edge, fog, and mist nodes as gateway nodes may enhance the capabilities of implementing actuation decisions for IoT devices. It helps in reducing overall latency and efficiently optimizes network traffic. It offers processing capacity locally closer to data generation and provides fast and almost real-time responses and decision-making using machine learning algorithms. The machine learning algorithm is a mathematical representation of an output function based on some input function. It can create a predictive or classification model based on the training data. Due to their role in analyzing historical data, building predictive models, and making accurate predictions, machine learning algorithms play a significant role in weather forecasting, using the capability of handling vast amounts of data and identifying underneath patterns and connections between various weather indicators. The Ensemble method, a set of combinations of many models, is deployed to enhance the system's overall accuracy. The machine learning model is updated based on input received in real-time monitoring data. The machine learning models' capabilities help determine the overall weather and uncertainty. The prediction is based on the current atmospheric variable and previous training data[12]–[14].

III. FRAMEWORK AND METHODOLOGY

The major hardware parts of the infrastructure are IoT sensors, actuator nodes, weather stations, drone-mounted edge nodes, cloud computing or high-capacity processing facilities, and storage facilities. The software part of the infrastructureate are (i). Data preprocessing, (ii). Model training on the historical data at a cloud computing facility or high data processing facility, (iii). Model deployed on a drone-mount edge node, and (iv). periodically, the model is retrained on environmental and historical data. The IoT devices play a vital role in precision farming; they are used for sensing and actuation purposes. The IoT sensor node and weather station may collect environmental and soil-related data. The weather station and IoT Senor nodes may have duplicate values depending on the sensing frequency. We had a high volume of sensed data if we kept a low sensing frequency. The system may miss vital data if we keep a low sensing frequency. The drone-mounted edge node is an intermediate device that acts as a gateway for IoT nodes. It helps reduce overall latency in data and decision delivery. The trained model is being deployed on it to take and deliver decisions to the IoT node in almost real-time without any latency. The model is trained using deep and machine learning algorithms such as Logistic Regression(LR), Decision Tree(DT), Random Forest, Gradient Boosting(GB), XGBoost, LightGBM, and CatBoost.

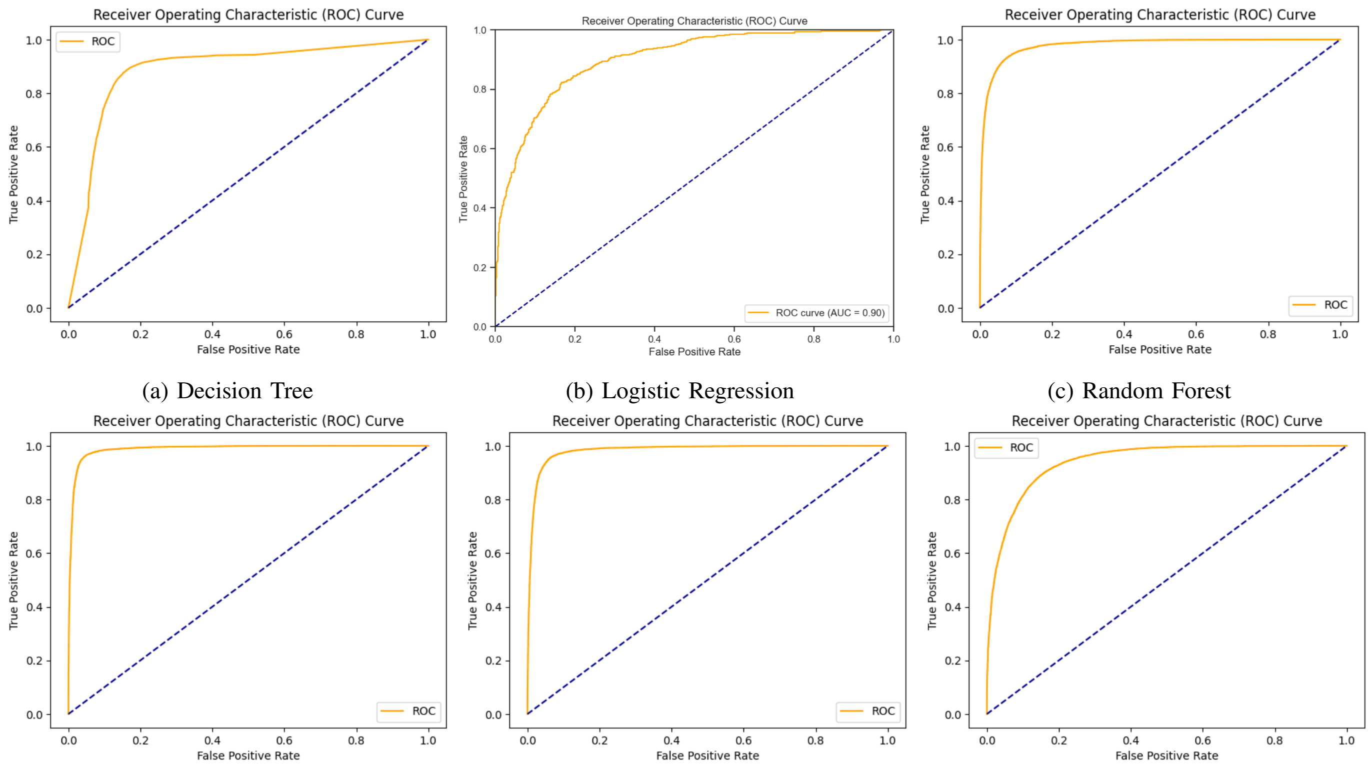
The data is captured through sensor nodes and weather stations. Sometimes, the data may have noise or faulty values. The sensor and weather station can generate data at a rapid pace, which often results in duplicate values. The data is cleaned, integrated, transformed, and reduced before being fed into the decision support system. We tested several models for weather forecasting and selected a few highly performed algorithms such as Logistic Regression(LR), Decision Tree(DT), Random Forest, Gradient Boosting(GB), XGBoost, LightGBM, and CatBoost. In the Logistic Regression[15], the probability of an instance belonging to a particular class is calculated using a sigmoid function. The weights are assigned to each feature, and the model combines them into the linear equation. However, the Decision Tree[16] is used in machine learning for regression and classification. The decision tree is trained for the best splitting criteria based on available information. In order to avoid overfitting, it may be pruned. Random Forest learning algorithm combines multiple decision trees to make more precise and accurate predictions. It combines random feature selection, multiple decision trees, and ensemble techniques[17]. The Gradient Boosting approach is a popular machine-learning approach for noisy data with complex dependencies, such as weather forecasts and recommendation systems. It combines several weak predictive decision tree-based predictors in order to create a powerful ensemble-based model[18]–[21]. XGBoost or eXtreme Gradient Boosting is another improved variant of gradient boosting[18]. Just like gradient boost, it also combines several weak decision tree-based predictive models to improve accuracy[20]. Light Gradient Boosting Machine(LightGBM) is another version of the Gradient boosting model[18] with a faster implimentation time. It has faster training time and is memory efficient compared to the Gradient boosting model[22]. Categorical boosting, or CatBoost, is a machine learning model based on a gradient boosting[18] approach focusing on categorical data. It uses greedy methods to limit the feature combination problem. The CatBoost can handle high-dimension data and missing data[19], [21].

IV. RESULTS AND DISCUSSION

In this section, we evaluate the performance of the proposed model against existing models in the literature deployed on the Australian Rainfall Weather Dataset(rain in Australia)[23]. We also deployed our model on weather datasets for Bikaner(India) between 1982 and 2021 [24], [25]. The preprocessed datasets are used to train and evaluate the machine learning algorithms to develop the model. We compared the selected machine learning algorithms (Decision Tree, Logistic Regression, Random Forest, XGBoost, Catboost, LightGBM)

TABLE I: Comparision of Various Machine Learning Models on Australian Rainfall Weather Data dataset

Measures	**LR**	**DT**	**RF**	**LGBM**	**Catboost**	**XGB**
TT(Sec)	1.93	0.474	20.38	1.627	129.1	49.78
Cohenś Kappa	0.582	0.720	0.851	0.728	0.877	0.898
Acc.(%)	79.52	86.12	92.65	86.62	93.92	94.96
prec (%)	79.47	86.36	92.76	86.61	94.17	95.15
recall(%)	79.52	86.12	92.66	86.62	93.92	94.96
F1-score	79.46	86.16	92.67	86.61	93.93	94.97



(a) Decision Tree (b) Logistic Regression (c) Random Forest

(d) XGBoost (e) Catboost (f) LightGBM

Fig. 1: ROC AUC Curves

for various performance measurement parameters such as Train Time(TT), ROC curve & AUC area, accuracy, precision, recall, and F1-score.

TABLE II: Comparision of Various Machine Learning Models on Bikaner, India dataset

Measures	LR	DT	RF	LGBM	Catboost	XGB
TT(Sec)	0.14	0.02	0.91	0.10	3.81	0.65
Cohenś Kappa	0.48	0.42	0.49	0.49	0.52	0.47
Acc(%)	91.29	88.28	90.77	90.83	91.24	90.31
prec(%)	90.29	88.14	89.84	89.92	90.42	89.41
recall(%)	91.29	88.28	90.77	90.83	91.24	90.31
F1-score	90.23	88.21	90.10	90.17	90.65	89.72

The comparison of algorithms for the Australian dataset is shown in Table I, while the Bikaner dataset is shown in Table II.From these tables, we can conclude that the system achieves high accuracy, with logistic regression(91.29%) performing best on the Bikaner dataset, while decision trees'(88.28%) performance is worst amongst the tested algorithms. However, XGBoost(94.96%) excels in the Australian data, and logistic regression(79.52%) had the lowest accuracy. For the Australian Dataset, the XGBoost has the highest precision(95.15%), recall(94.96%), and F1-Score(94.97%), While the logistic regression has the lowest precision(79.472%), recall(79.52%), and F1-Score(79.46%). For the Bikaner Dataset, the Catboost had the highest precision(90.42%) and F1-Score(90.65%), while logistic regression had the highest recall(91.29%). However, the decision tree had the lowest precision(88.14%), recall(88.28%), and F1-Score(88.21%). The decision tree is the fastest algorithm in training, while Catboost is the slowest. Cohen's Kappa is in the moderate agreement(0.41– 0.60) range for all algorithms.

The Receiver Operating Characteristic(ROC) curve and the Area Under the Curve(AUC) metric are used to evaluate the algorithm's performance. Figure 1 shows the ROC and AUC curves. The curve is more inclined towards Y - the axis and a higher value of AUC suggests that our proposed model had strong discriminatory power in predicting output and a high level of predictive performance. For the Australian weather datasets, the AUC value of the logistic regression is 78.95% as a result, it had lower accuracy amongst the selected model. The remaining machine learning algorithms, i.e., random forest, XGBoost, and Catboost, had AUC values greater than 90%, which is reflected in the accuracy of these algorithms.

A. *Comparision with existing Models*

The results of the proposed model are compared with existing works to evaluate the overall performance of the proposed model. We used accuracy as a prime parameter for comparison. Most of the existing works also deploy two or more two machine learning algorithms; for simplicity of comparison, we consider their best-performed machine learning algorithm or

TABLE III: Comparision with Exisitng Weather Prediction Algorithms

Authors	Dataset	Model	Accuracy
Oswal[13]	W_{Aus}	KNN	85%
Oswal[13]	W_{Aus}	KNN	85%
Deng [26]	W_{Aus}	LR	84.95%
Liu et al. [27]	W_{Aus}	SS AdaBoost	84.77%
Polishchuk et al. [12]	W_{Aus}	RF	85.9%
Sarasa-Cabezuelo [28]	W_{Aus}	NN	84.00%
Mahadware et al. [29]	W_{Aus}	CatBoost	81.37%
Zhao et al. [30]	W_{Aus}	LR,LSTM	85.00%
Baharisangari et al.[31]	W_{Aus}	ANN	84.73%
Dieber et al. [32]	W_{Aus}	XGBoost	85.00%
Yadav et al.[14]	W_{Aus}	XGBoost	86%
Goksu et al.[33]	W_{Aus}	EK-stars	87.15%
This Paper	W_{Aus}	XGBoost	94.96%
This Paper	W_{Bkn}	LR	91.29%

Weather Datasets:W_{Aus} *:Australian [23]* | W_{Bkn}*:Bikaner[24], [25]* |

model. The comparison is shown in the Table III. The proposed model with the XGBoost algorithm had the highest accuracy on the Australian Rainfall Weather Data(rain in Australia)[23] compared to other models or algorithms tested on the same dataset.

References

[1] A. Chaudhary, S. K. Peddoju, and V. Chouhan, “Secure authentication and reliable cloud storage scheme for iot-edge-cloud integration,” *Journal of Grid Computing*, vol. 21, no. 3, p. 35, Jun 2023. [Online]. Available: https://doi.org/10.1007/s10723-023-09672-z

[2] A. Mukherjee, N. Dey, and D. De, “Edgedrone: Qos aware mqtt middleware for mobile edge computing in opportunistic internet of drone things,” *Computer Communications*, vol. 152, pp. 93–108, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140366419315750

[3] A. Chaudhary, S. K. Peddoju, and K. Kadarla, “Study of internet-of-things messaging protocols used for exchanging data with external sources,” in *2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, 2017, pp. 666–671.

[4] I. Donevski, C. Raffelsberger, M. Sende, A. Fakhreddine, and J. J. Nielsen, “An experimental analysis on drone-mounted access points for improved latency-reliability,” in *Proceedings of the 7th Workshop on Micro Aerial Vehicle Networks, Systems, and Applications*, ser. Dronet ’21. New York, NY, USA: Association for Computing Machinery, 2021, p. 31–36. [Online]. Available: https://doi.org/10.1145/3469259.3470489

[5] M. S. Farooq, S. Riaz, A. Abid, K. Abid, and M. A. Naeem, “A survey on the role of iot in agriculture for the implementation of smart farming,” *Ieee Access*, vol. 7, pp. 156 237–156 271, 2019.

[6] A. Chaudhary, “A cluster based wireless sensor network deployment for precision agriculture in dried and arid states of india,” in *Proceedings of the 2014 International Conference on Information and Communication Technology for Competitive Strategies*, ser. ICTCS ’14. New York, NY, USA: Association for Computing Machinery, 2014. [Online]. Available: https://doi.org/10.1145/2677855.2677915

[7] N. Hossein Motlagh, M. Mohammadrezaei, J. Hunt, and B. Zakeri, “Internet of things (iot) and the energy sector,” *Energies*, vol. 13, no. 2, p. 494, 2020.

[8] A. Chaudhary and S. K. Peddoju, “The role of iot-based devices for the better world,” in *Information and Communication Technology*, D. K. Mishra, A. T. Azar, and A. Joshi, Eds. Singapore: Springer Singapore, 2018, pp. 299–309.

[9] N. Misra, Y. Dixit, A. Al-Mallahi, M. S. Bhullar, R. Upadhyay, and A. Martynenko, “Iot, big data, and artificial intelligence in agriculture and food industry,” *IEEE Internet of things Journal*, vol. 9, no. 9, pp. 6305–6324, 2020.

[10] A. Chaudhary and S. K. Peddoju, “Impact of clustering algorithms and energy harvesting scheme on iot/wsn infrastructures,” in *2021 IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*, 2021, pp. 603–608.

[11] S. Ketu and P. K. Mishra, “Cloud, fog and mist computing in iot: an indication of emerging opportunities,” *IETE Technical Review*, vol. 39, no. 3, pp. 713–724, 2022.

[12] B. Polishchuk, A. Berko, L. Chyrun, M. Bublyk, and V. Schuchmann, “The rain prediction in australia based big data analysis and machine learning technology,” in *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)*, vol. 1. IEEE, 2021, pp. 97–100.

[13] N. Oswal, “Predicting rainfall using machine learning techniques,” *arXiv preprint arXiv:1910.13827*, 2019.

[14] K. Yadav, A. Singh, and A. K. Tiwari, “Improvements in weather forecasting technique using cognitive internet of things.” *Computer Systems Science & Engineering*, vol. 46, no. 3, 2023.

[15] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.

[16] S. R. Safavian and D. Landgrebe, “A survey of decision tree classifier methodology,” *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[17] L. Breiman, “Random forests,” *Machine learning*, vol. 45, pp. 5–32, 2001.

[18] J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” *Annals of statistics*, pp. 1189–1232, 2001.

[19] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “Catboost: unbiased boosting with categorical features,” *Advances in neural information processing systems*, vol. 31, 2018.

[20] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[21] A. V. Dorogush, V. Ershov, and A. Gulin, “Catboost: gradient boosting with categorical features support,” *arXiv preprint arXiv:1810.11363*, 2018.

[22] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “Lightgbm: A highly efficient gradient boosting decision tree,” *Advances in neural information processing systems*, vol. 30, 2017.

[23] Kaggle. Rain in Australia. https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package , accessed on 15-Jun-2023,.

[24] NASA. NASA Power Project. https://power.larc.nasa.gov/, accessed on 15-Jun-2023,.

[25] imdlib. imdlib 0.1.17. https://pypi.org/project/imdlib/ , accessed on 15-Jun-2023,.

[26] F. Deng, “Research on the applicability of weather forecast model—based on logistic regression and decision tree,” in *Journal of Physics: Conference Series*, vol. 1678, no. 1. IOP Publishing, 2020, p. 012110.

[27] X. Liu, S. Luo, and L. Pan, “Robust boosting via self-sampling,” *Knowledge-Based Systems*, vol. 193, p. 105424, 2020.

[28] A. Sarasa-Cabezuelo, “Prediction of rainfall in australia using machine learning,” *Information*, vol. 13, no. 4, p. 163, 2022.

[29] A. Mahadware, A. Saigiridhari, A. Mishra, A. Tupe, and N. Marathe, “Rainfall prediction using different machine learning and deep learning algorithms,” in *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*. IEEE, 2022, pp. 1–8.

[30] Y. Zhao, H. Shi, Y. Ma, M. He, H. Deng, and Z. Tong, “Rain prediction based on machine learning,” in *2022 8th International Conference on Humanities and Social Science Research (ICHSSR 2022)*. Atlantis Press, 2022, pp. 2957–2970.

[31] N. Baharisangari, K. Hirota, R. Yan, A. Julius, and Z. Xu, “Weighted graph-based signal temporal logic inference using neural networks,” *IEEE Control Systems Letters*, vol. 6, pp. 2096–2101, 2021.

[32] J. Dieber and S. Kirrane, “A novel model usability evaluation framework (muse) for explainable artificial intelligence,” *Information Fusion*, vol. 81, pp. 143–153, 2022.

[33] G. Tuysuzoglu, K. U. Birant, and D. Birant, “Rainfall prediction using an ensemble machine learning model based on k-stars,” *Sustainability*, vol. 15, no. 7, p. 5889, 2023.