

# Crop Forecasting Method Based On Machine Learning

Ms. J.SASIKALA( Research Scholar)

Prof. Dr. Mr. A.VIJAYAN.

Master of Engineering Department of Computer Science  
T.J.S Engineering College, Tamil Nadu

Ms. Pavithra .V & Dr. Ms. Agnes .J Assistant Professor  
Master of Engineering Department of Computer Science  
T.J.S Engineering College, Tamil Nadu

**Abstract:** India is primarily an agrarian nation, where agriculture holds significant sway over both the country's economy and the livelihoods of its population. Most of the Indian farmers face a common challenge of not selecting the most appropriate crop for their land in accordance with the environmental requirements. As a result, they will see a major decline in their total level of productivity. So to overcome this problem crops are recommended based on soil, weather, humidity, rainfall, and other geographical factors which plays a major role in attaining quantity and quality of crops massively. So our project's aim is to analyse all these factors and match it with the required parameters of each crops and provide the best possible option to the farmers. This leads to a decrease in crop selection errors and an increase in yield. To achieve our aim we will use a recommendation system using an ensemble model with a majority voting technique that uses a Decision Tree, Logistic Regression, Naive Bayes, CNN, Random Forest (RF), and Extreme Gradient(XG) Boost as learners to recommend a crop for the site- specific parameters with high accuracy and efficiency.

**Keywords:** Recommendation System, XG Boost, Random forest, CNN, Decision Tree, Fuzzy Algorithm

## 1. Introduction

India is one of the world's oldest countries that still practices agriculture. However, due to globalization, agricultural practices have substantially changed in recent years. One of the major areas of focus is agriculture. They contribute significantly to society since they generate a huge amount of the food. In many countries, this is still the case. With an increasing population, many people are experiencing food shortages. Agriculture in today's world frequently employs sophisticated technologies. Farmers may increase profit, efficiency, safety, and environmental friendliness by using these improved agricultural technologies. For image processing, MATLAB is used as IDE. For training the model the algorithm of Hybrid NeuralNetwork is employed and also to improve the program's accuracy in calculating the soil's nutrient and pH levels. Machine learning is the branch of AI that doesn't require any explicit programming.

Our model will predict the output or result based on the input data which can be extracted from the past years. Here the aim of the project is to design a model with machine learning which can be used to get the information about the weather, soil, other geographical factors and so on which will be suitable for the cultivation of the individual crop. The envisioned model is set to be constructed employing machine learning algorithms such as Linear Regression, XG Boost, and Random Forest, the latter operating on decision trees. These sophisticated algorithms will enable the model to generate highly accurate predictions by meticulously analyzing input data. The predictive outcomes will be seamlessly presented to users through a web interface, implemented using the Python Django web framework. This integration of machine learning and web development ensures a robust and user-friendly platform, where individuals can readily access and interpret the predicted results for informed decision-making in agriculture.

## 2. Related Works

Research on crop yield prediction utilizes deep learning, facing challenges like indirect data-to-yield mapping and feature quality dependence. A novel solution integrates deep reinforcement learning. The proposed Deep Recurrent Q-Network, combining RNN and Q-Learning, achieves superior accuracy (93.7%) by preserving data distribution.

This paper focuses on collecting and analyzing environmental and agricultural data to enhance crop yield for farmers. The process involves data preprocessing in Python, utilizing the MapReduce framework for large-scale analysis,

<sup>1</sup> Assistant Professor, Department of Information Technology,

and applying k-means clustering for accuracy. The study explores correlations between crops, rainfall, temperature, soil, and seed types in two regions (Ahmednagar, Maharashtra, and Andaman and Nicobar Islands) using bar graphs and scatter plots. A self-designed recommender system predicts crops, displayed on a Flask-based Graphic User Interface. The scalable system can extend its crop recommendations to other states in the future.

Precision agriculture addresses this by considering soil characteristics, types, and crop yield data to suggest suitable crops, enhancing overall productivity. This method offers benefits such as optimal resource utilization and improved decision-making in farming. To implement precision agriculture, a recommendation system is proposed using an ensemble model with random tree, CHAID, K-Nearest Neighbor, and Naive Bayes as learners. This system ensures high accuracy and efficiency in recommending suitable crops based on soil parameters. The resulting classified image includes key data like weather, crop yield, and state and district-wise crops, enabling precise predictions of crop yields under specific weather conditions.

In India, we all know that Agriculture is the backbone of the country. This paper predicts the yield of almost all kinds of crops that are planted in India. This script makes novel by the usage of simple parameters like State, district, season, area and the user can predict the yield of the crop in which year he or she wants to.

### 3. Proposed System

The central objective of this initiative is to offer substantial support to end-users, particularly farmers, by presenting innovative ideas and addressing a critical issue in agriculture – profit and cost loss. The project's focal point involves the development of an advanced decision-making XG Boost model. By leveraging insights from two research papers documented in IEEE, the project aims to encapsulate cutting-edge strategies and methodologies. Ultimately, the goal is to empower farmers with a reliable and efficient tool that enhances decision-making processes, mitigates profit and cost losses, and contributes to overall agricultural sustainability. This project incorporates Weather API technology to get input of weather parameters from ‘WB-CPI: Weather Based Crop Prediction in India Using Big Data Analytics’ paper and soil input parameters gathering using the reference from ‘Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications’ paper. Combining both these ideologies helped us to develop a system that produces result with more accuracy and efficiency. The output of the project plays the

major role in crop selection for the field under study. This plays a major role in increasing the crop yield thereby increasing the profit.

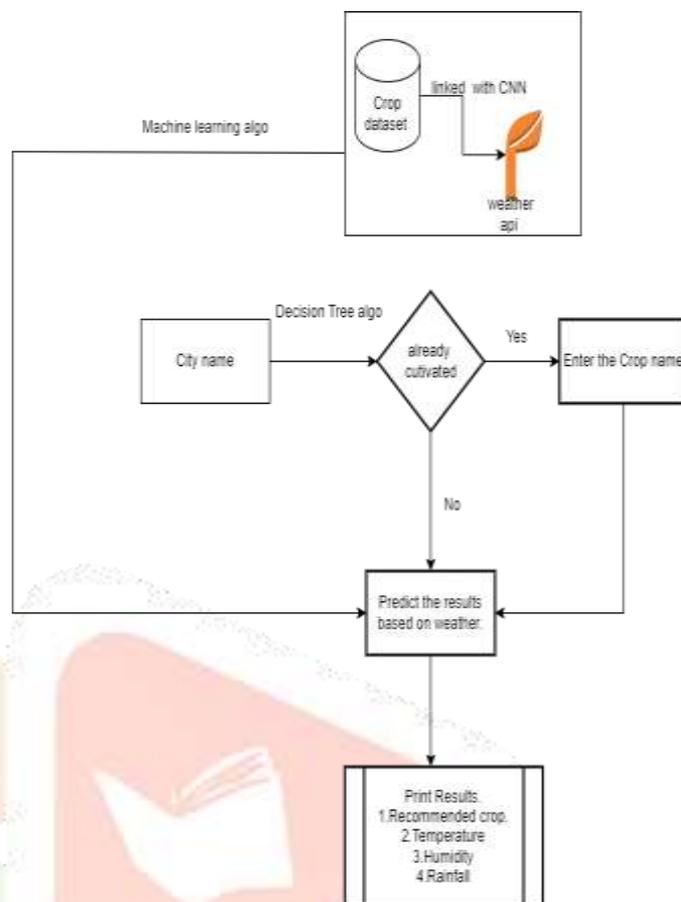


Fig. 1 Architecture Diagram

### XG Boost Model Training and Hyperparameter Tuning

- i. Split the dataset into testing sets and training sets.
- ii. Choose XG Boost algorithm for its advantages in handling complex relationships.
- iii. Fine-tune hyperparameters (learning rate, tree depth, regularization) through cross-validation.

$$f(x) \approx y(t-1) \quad \text{--->} \quad (1)$$

This represents the prediction from the previous iteration (t-1), denoted as y(t-1).

$$Z(q) \approx \sum \chi(x)^2 * (y(t-1) - f(x))^2 + \lambda * T * \text{sum}$$

$$(\sum \chi^2(x))^2 \quad \text{--->} \quad (2)$$

This formula represents the objective function that XG boost aims to minimize during training.

Using (1) and (2)

$$g(x) \approx y(t-1) + T * \sum (\Sigma \chi(x)^2 * (y(t-1) - f(x))^2)$$

This formula represents the update to the prediction for the current iteration.

T is a hyperparameter that controls the step size of the update.

These formulas describe the prediction, regularization, and update steps in the XG Boost algorithm. The algorithm seeks to find the optimal combination of weak learners by minimizing the objective function, which includes terms for prediction accuracy and regularization.

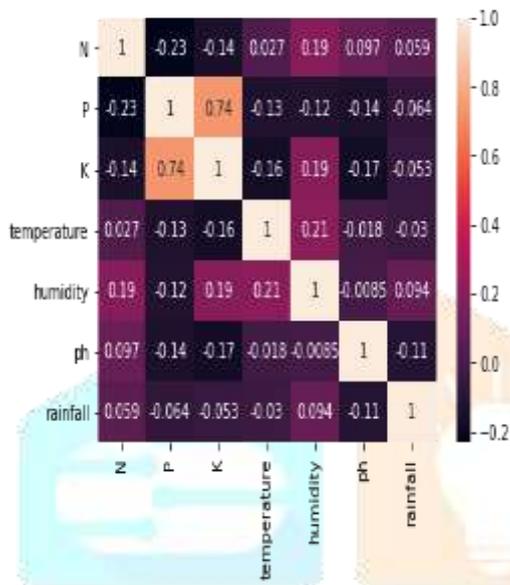


Figure 2. Shows how XG Boost model is trained with the collected data that is Temperature, Rainfall, Humidity and land details.

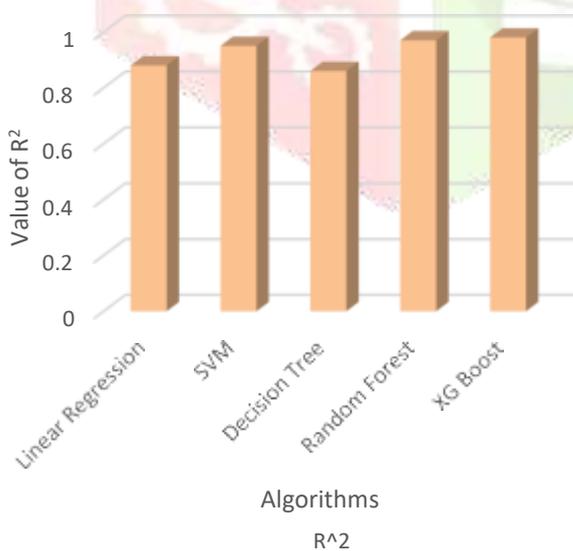


Figure 3. Shows the comparison of R<sup>2</sup> for XG Boost algorithm with other algorithms.

### Weather Integration

Real-time Weather Updates: Integrate APIs or services to provide real-time weather updates for accurate decision-making.



Figure 4. Weather Integration

## 4. Implementation and Results

The process comprises of five individual categories that include data input, preprocessing, data segmentation, classification, and the output. The Hybrid Neural Network is incorporated as the training model in the main motive to increase the system’s accuracy in the wayof providing the highly positively predicted values or levelsof the nutrients and the soil’s pH..

### 1. Data Collection

Soil nutrient Dataset required for this work was collected from Department of Agriculture at some of the districtslike Ariyalur, Salem, Trichy in Tamil Nadu. The dataset collected contains information like attributes, the values of soil samples that corresponds to the respective district fromwhich the soil was taken. There are around 12 attributes in the dataset and the total instance of 1676 soil samples are used for the proposed system. The attribute description of the dataset collected was depicted in the table provided below as Table 1.

### 2. Data Pre-Processing

In this step described about removing unwanted data from the dataset which helps to extract required result data from the dataset.

### 3. Data Conversion

If data mining uses MATLAB tool, the data must be in MAT format. All the data sheets converted into .MAT file “MATLAB file format”. This MAT file has sections those are Header and Data Information.

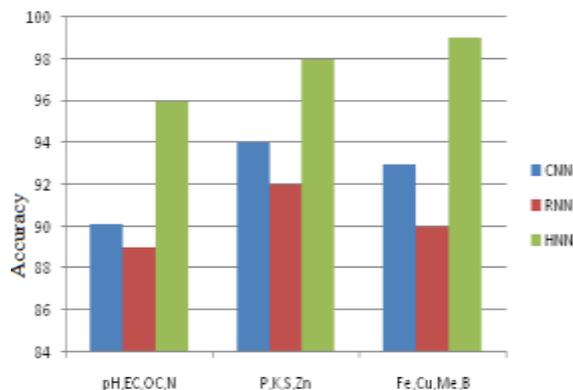
### 4. Classification

This technique of data mining is based on machine learning using concepts of algorithms. In this soil nutrient datasets are classified using XG Boost and Random Forest algorithms.

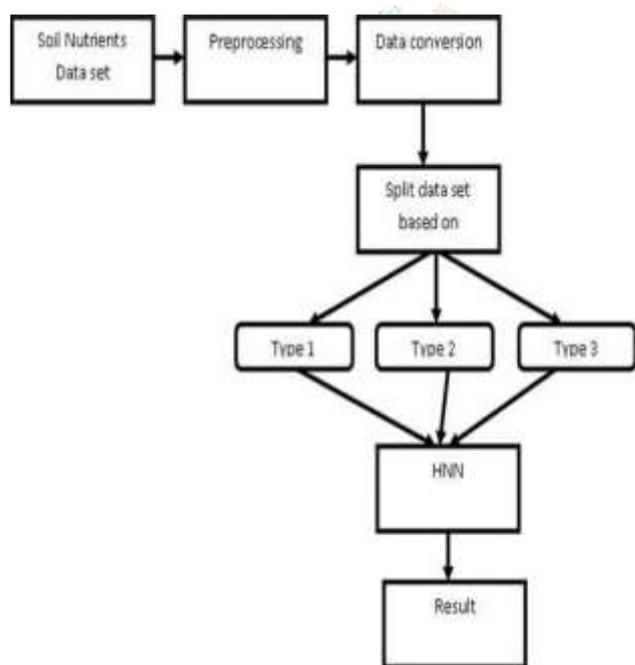
### 5. Prediction

The classification algorithm is noted for the accuracy and the performance analysis and it provides suggestion to farmer to choose the best crop for the soil.

in the rate of those algorithms with the help of the set of three different nutrients.



ACCURACY FOR EXECUTION TIME OF CLASSIFIERS



PROPOSED WORK FLOW

The dataset collected are pre-processed and the unwanted data was removed. After the step of pre-processing the complete set of data was divided into LOW, MEDIUM and the HIGH depending on the level of nutrients present.

Then the algorithms like CNN, RNN and the Hybrid Classification. Then based on these algorithms the nutrients in the soils are classified in the following levels Very High, High, Medium, Low and Very Low.

The Table:2 depicts the comparative analysis of classifiers. The following figures shows the time of execution of the various classification algorithms by grouping the three types of nutrients and also the accuracy

District	pH	EC	OC	N	P	K	S	Zn	Fe	Cu	Mn	B
Ariyalur	8.30	0.26	0.54	43.2	17.6	789	10.5	2.45	2.45	3.56	3.67	3.4
Ariyalur	5.10	0.17	0.40	60.60	16.25	60	17.50	0.53	5.23	0.17	3.38	1.0
Coimbatore	6.50	0.38	0.18	231.0	18.0	400	8.10	0.80	5.10	2.20	2.60	0.30
Coimbatore	7.80	0.18	0.18	173.0	13.6	284	17.24	1.0	5.30	1.40	2.60	0.41
Karur	7.90	0.6	0.35	80.00	2.70	189	8.10	0.84	1.40	0.02	1.08	1.10
Karur	8.10	1.50	0.15	45.00	2.70	253	0.11	0.10	2.60	0.03	1.70	1.10
Salem	7.90	0.10	0.28	162	7.50	315	29.2	0.23	10.5	0.72	7.23	2.30
Salem	7.80	0.10	0.11	189	10	378	33.20	0.32	15.5	0.74	0.32	1.0
Thanjavur	6.90	0.15	0.75	205	75	250	25	1.30	6.80	1.80	2.0	1.0
Thanjavur	7.70	0.17	0.69	189	75	300	20.60	1.00	7.40	0.90	2.0	1.0
Trichy	7.0	1.10	0.25	179	53.0	195.0	12.20	1.06	5.72	0.96	3.40	0.10
Trichy	6.80	0.40	0.19	147	25.0	185.0	12.60	0.95	6.27	0.91	5.76	0.10

Table 1 SAMPLE NUTRIENT DATA-SET

District	pH	EC	OC	N	P	K	S	Zn	Fe	Cu	Mn	B
Ariyalur	H	L	M	L	M	H	M	H	L	H	M	H
Ariyalur	L	L	L	L	M	L	H	L	M	L	M	M
Coimbatore	L	L	L	L	M	H	L	L	M	H	M	L
Coimbatore	H	L	L	L	M	H	H	L	M	M	M	L
Karur	H	L	L	L	L	M	L	L	L	L	L	H
Karur	H	M	L	L	L	M	L	L	L	L	L	H
Salem	H	L	L	L	L	H	H	L	H	L	H	H
Salem	H	L	L	L	L	H	H	L	H	L	L	M
Thanjavur	M	L	H	L	H	M	H	L	M	H	M	M
Thanjavur	H	L	M	L	H	H	H	M	M	L	M	M
Trichy	M	M	L	L	H	M	M	L	M	L	M	L
Trichy	M	L	L	L	H	M	M	L	M	L	H	L

Table 2 COMPARATIVE ANALYSIS OF CLASSIFIERS BASED ON NUTRIENTS

## 5. Conclusion

The need of maintaining the health of the soil resource base as a critical component of long-term development is Agriculture will be able to The soil nutrients are mostly responsible for yield production. The study of the nutrients presence in the soil helps is determining the types of crop, which can be used in the specific soil in order to get the good yield. The healthiest soils generate the healthiest and most plentiful food. We conclude that the Data Mining and Remote Sensing area provides many algorithms and prognosis strategies for classification of soil attributes databased on the literature. Thus, the collected data of temperature and humidity helped in the process of prediction and categorization system of climate using the algorithms like K-Nearest Neighbor, which helps in uncovering the hidden patterns. In order to extract water features from satellite photos, a variety of satellites are used. A few of the findings are talked over. And as the result an attempt was proposed by considering the existing issues as well as the futuristic possibilities of extraction from water body approaches to a conclusion.

## References

- [1] Elavarasan, D., & Vincent, P. D. (2020). Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. IEEE access, 8, 86886-86901.
- [2] Bhuyar, V. (2014). Comparative analysis of classification techniques on soil data to predict fertility rate for Aurangabad District. Int. J. Emerg. Trends Technol. Comput. Sci, 3(2), 200-203.
- [3] Baskar, S. S., Arockiam, L., & Charles, S. (2013). Applying data mining techniques on soil fertility prediction. International Journal of Computer Applications Technology and Research, 2(6), 660-662.
- [4] Gholap, J. (2012). Performance tuning of J48 Algorithm for prediction of soil fertility. arXiv preprint arXiv:1208.3943.
- [5] Gholap, J., Ingole, A., Gohil, J., Gargade, S., & Attar, V. (2012). Soil data analysis using classification techniques and soil attribute prediction. arXiv preprint arXiv:1206.1557.
- [6] Suman, B. B. N., & Naib, B. (2013). Soil classification and fertilizer recommendation using WEKA. IJCSMS International Journal of Computer Science & Management Studies, 13(05).
- [7] Hemaetha, N., & Nasira, G. M. (2016). Analysis of soil condition based on pH value using classification techniques. IOSR J ComputEng (IOSR-JCE), 18(6), 50-54.
- [8] Bhargavi, P., & Jyothi, S. (2011). Soil classification using data mining techniques: a comparative study. *International Journal of Engineering Trends and Technology*, 2(1), 55-59.
- [9] Rajeswari, V., & Arunesh, K. (2016). Analysing soil data using data mining classification techniques. *Indian journal of science and Technology*, 9(19), 1-4.
- [10] Noble, I., Bolin, B., Ravindranath, N., Verardo, D., & Dokken, D. (2000). Land use, land use change, and forestry. *Environmental Conservation*, 28(3), 284-293.
- [11] Henderson, F. M. (1995). Environmental factors and the detection of open surface water areas with X-band radar imagery. *International Journal of Remote Sensing*, 16(13), 2423-2437. (13), pp.2423-2437, September 1995.
- [12] Nath, R. K., & Deb, S. K. (2010). Water-body area extraction from high resolution satellite images-an introduction, review, and comparison. *International Journal of Image Processing (IJIP)*, 3(6), 265-384.
- [13] Prakash, A., & Gupta, R. P. (1998). Land-use mapping and change detection in a coal mining area- a case study in the Jharia coalfield, India. *International journal of remote sensing*, 19(3), 391-410.
- [14] Zhang, Q., Wang, C., & Shinohara, F. (2007, November). Automatic extraction of water body based on EOS/MODIS remotely sensed imagery. In *MIPPR 2007: Automatic Target Recognition and Image Analysis; and Multispectral Image Acquisition (Vol. 6786, pp. 1068-1074)*. SPIE..
- [15] Nath, R. K., & Deb, S. K. (2010). Water-body area extraction from high resolution satellite images-an introduction, review, and comparison. *International Journal of Image Processing (IJIP)*, 3(6), 265-384.
- [16] Li, S., Peng, S., Chen, W., & Lu, X. (2013). INCOME: Practical land monitoring in precision agriculture with sensor networks. *Computer Communications*, 36(4), 459-467.
- [17] Jones, A. D., Ngure, F. M., Pelto, G., & Young, S. L. (2013). What are we assessing when we measure food security? A compendium and review of current metrics. *Advances in nutrition*, 4(5), 481-505.
- [18] Ogutu, G. E., Franssen, W. H., Supit, I., Omondi, P., & Hutjes, R. W. (2018). Probabilistic maize yield prediction over East Africa using dynamic ensemble seasonal climate forecasts. *Agricultural and forest meteorology*, 250, 243-261.
- [19] Holzman, M. E., Carmona, F., Rivas, R., & Niclòs, R. (2018). Early assessment of crop yield from remotely sensed water stress and solar radiation data. *ISPRS journal of photogrammetry and remote sensing*, 145, 297-308.
- [20] Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in plant science*, 21(2), 110-124.

- [21] Raja, S. K. S., Balaji, V., & Vivekanandan, M. (2015). Autonomous mobile navigation robot for agricultural purpose. *International Journal of Applied Engineering Research*, 10(10), 27333-27341.
- [22] Srikanth, V. Natarajan, B. Jegajothi, S. S. L. Durai Arumugam and D. Nageswari, "Fruit fly Optimization with Deep Learning Based Reactive Power Optimization Model for Distributed Systems," 2022 International Conference on Electronics and Renewable Systems (ICEARS), 2022, pp. 319-324, doi: 10.1109/ICEARS53579.2022.9751908.
- [23] S.Kanaga Suba Raja, R. Rishi, E. Sundaresan and V. Srijit, "Demand based crop recommender system for farmers," 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), 2017, pp. 194-199, doi: 10.1109/TIAR.2017.8273714.
- [24] M.Vivekanandan, S. Kanaga Suba Raja, (2022) ' Virtex-II Pro FPGA Based Smart Agricultural System', Wireless Personal Communications, ISSN 0929-6212. Volume 125 Issue 1, pp 119–141. <https://doi.org/10.1007/s11277-022-09544-x>.
- [25] S. K. Suba Raja, D. A. S S L, R. P. Kumar and J. Selvakumar, "Recognition of Facial Stress System using Machine Learning with an Intelligent Alert System," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 1-4, doi: 10.1109/ICCMC56507.2023.10083566.
- [26] Ekaterina Katya, S.R. Rahman. (2020). Void Node Detection and Packet Re-routing in Underwater Wireless Sensor Network. *International Journal of New Practices in Management and Engineering*, 9(04), 01 - 10. Retrieved from <http://ijnpme.org/index.php/IJNPME/article/view/93>
- [27] Sindhvani, N., Anand, R., Vashisth, R., Chauhan, S., Talukdar, V., & Dhabliya, D. (2022). Thingspeak- based environmental monitoring system using IoT. Paper presented at the PDGC 2022 - 2022 7th International Conference on Parallel, Distributed and Grid Computing, 675-680. doi:10.1109/PDGC56933.2022.10053167 Retrieved from [www.scopus.com](http://www.scopus.com)

