International Journal of Reconfigurable and Embedded Systems (IJRES)
Vol. 12, No. 3, November 2023, pp. 503~508
ISSN: 2089-4864, DOI: 10.11591/ijres.v12.i3.pp503-508

Predicting yield of crop type and water requirement for a given plot of land using machine learning techniques

Nitin Padriya[1], Nimisha Patel[2]
[1]Faculty of Engineering and Technology, Sankalchand Patel University, Visnagar, India
[2]Gandhinagar Institute of Technology, Gandhinagar University, Moti Bhoyan, India

Article Info

Article history:

Received Jan 27, 2023
Revised Mar 27, 2023
Accepted Apr 27, 2023

Keywords:

Climate
Internet of things
Machine learning
Precision agriculture
Regression problem

ABSTRACT

Internet of things (IoT) smart technology enables new digital agriculture. Technology has become necessary to address today's challenges, and many sectors are automating their processes with the newest technologies. By maximizing fertiliser use to boost plant efficiency, smart agriculture, which is based on IoT technology, intends to assist producers and farmers in reducing waste while improving output. With IoT-based smart farming, farmers may better manage their animals, develop crops, save costs, and conserve resources. Climate monitoring, drought detection, agriculture and production, pollution distribution, and many more applications rely on the weather forecast. The accuracy of the forecast is determined by prior weather conditions across broad areas and over long periods. Machine learning algorithms can help us to build a model with proper accuracy. As a result, increasing the output on the limited acreage is important. IoT smart farming is a high-tech method that allows people to cultivate crops cleanly and sustainably. In agriculture, it is the use of current information and communication technologies.

Corresponding Author:

Nitin Padriya
Faculty of Engineering and Technology, Sankalchand Patel University
Visnagar, Gujarat, India
Email: nitinpadriya@gmail.com

1. INTRODUCTION

Agriculture plays a crucial role in global development, benefiting humans in various ways, making it a popular academic topic. Farmers, especially those cultivating non-traditional crops, are always eager for information. They seek guidance from sources like television, radio, newspapers, fellow farmers, government agricultural groups, farm suppliers, and merchants. Consequently, there is a need for a system that provides farmers with valuable information.

Accurate historical crop yield information is vital for making agricultural risk management decisions [1]. Challenges such as rising temperatures, changing precipitation patterns, and extreme weather events [2], as well as the variability of weather conditions and the complexity of factors influencing crop growth [3], pose significant obstacles. To address these challenges, machine learning has emerged as a widely adopted technology. Numerous machine learning techniques have been investigated and developed for agriculture. Given its effectiveness in resource efficiency, prediction, and management key aspects in agriculture researchers have examined the performance of various machine learning algorithms in this domain and others [4]. The prediction module employs machine learning algorithms to forecast crop yield, while the monitoring module uses these algorithms to identify potential problems like pests and diseases [5]. Machine learning refers to the ability of an electrical processing system to gather and utilize information. In crop management, machine learning has been applied to tasks such as yield prediction, disease detection, weed detection, crop quality assessment, and species recognition [6]. Internet of things (IoT) data can be leveraged to efficiently control irrigation systems, ensuring optimal watering for crops [7]. Additionally, quadcopter-based pesticide spraying mechanisms have been developed for accurate and efficient pesticide application [8]. This study specifically focuses on food crop cultivation and explores how machine learning can optimize land usage for maximizing crop output while conserving available resources. Crop production heavily relies on key land requirements, including topography, soil type, soil quality, moisture content, sunlight, and other factors that influence crop development on cultivable land.

2. RELATED WORK

Agricultural planning involves extensive forecasting of agricultural output to support company strategy development, risk evaluation, and goods storage. Two methods are commonly used for predicting future agricultural supply: statistical methods like autoregressive integrated moving average (ARIMA) and Holt-Winters, and machine learning methods [9]. The correlation between summer rainfall and agricultural product production is studied using statistical methods such as linear regression, correlation analysis, and homogeneity tests, to analyze changes in precipitation and temperature over a specific period [10].

In famine-prone areas, a machine learning technique is employed to model an early famine prediction application [11]. By utilizing historical data from Uganda during 2004 and 2005, machine learning techniques are tested to reduce societal vulnerability to famine risk. A novel approach utilizing genetic algorithms is proposed to optimize the parameters of random forest models. This approach focuses specifically on improving the accuracy of land cover classification, demonstrating the potential to enhance classification accuracy by up to 10% [12].

Agricultural productivity is influenced by climatic, geographical, biological, political, and economic factors [13]. Effectively processing large raw datasets is a primary challenge in agricultural data analysis. Accurate crop production forecasting using historical data is crucial for mitigating sensitive risks.

Random forests prove to be a valuable tool for predicting crop yield, especially when dealing with imbalanced data [14]. They can predict crop yield with significantly higher accuracy compared to multiple linear regression models. Crop yield data from various sources, including global wheat grain yield, maize grain yield in US counties, and potato tuber and maize silage yield in the northeastern seaboard region, are utilized [15].

Combining model-driven and data-driven approaches is considered the most promising strategy for crop yield prediction. Researchers also suggest that using multiple models and integrating different data sources can improve the accuracy of crop yield prediction [16]. An intelligent tool utilizing statistics and machine learning techniques is developed for rice yield prediction [17]. Classification and clustering processes are employed, with a support vector machine used for noisy data. Correlation analysis is utilized to assess the impact of different influencing factors on rice yield.

While machine learning techniques are frequently employed for crop yield forecasting, the lack of comprehensive evaluations of crops and techniques hinders appropriate decision-making [18]. Accuracy rates for various learning strategies are displayed based on the collected dataset. An IoT-based system for continuous measurement and monitoring of temperature, soil moisture, and relative humidity is developed using various sensors. Data is transmitted to a cloud server for storage and analysis. This system can be used to monitor environmental conditions in settings like greenhouses, farms, and industrial plants [19].

The production rate of crops in China is studied by dividing the country into six regions. Agricultural yields in Northwest and Southwest China are found to be positively associated with temperature change, while precipitation has little impact on the production of most crops in the Northwest. Positive relationships between precipitation, temperature change, and crop production are observed in East and Central-South China [20]. Hyperspectral wavebands in the range of 400 to 2,500 nm are utilized in a satellite crop classification technique to differentiate images of various crops in cultivated land in Egypt [21].

A two-tiered machine learning model approach is proposed for predicting crop yield. The first tier utilizes a decision tree to classify data into different groups based on data features, while the second tier employs a random forest to predict crop yield for each group [22]. The challenges of security and privacy in IoT-based agriculture systems are discussed, including the lack of trust between system entities, vulnerability of IoT devices to cyberattacks, and the difficulty of securing data in the cloud [23]. A plant nutrient management system is developed to correlate required plant nutrients with soil fertility, utilizing a supervised machine learning method called backpropagation neural network (BPN). Soil characteristics such as organic matter, micronutrients, and crucial plant nutrients that influence crop growth are considered [24].

3. PROPOSED METHOD
3.1. Supplies
Crop data sets: the data that fills this database contains the plant growth characteristics that were used to construct the individual decision trees in the random forest. The data sources come from a number of dependable databases. The machine learning approach is used to aid decision-making in the and is plant growth condition database.

3.2. Method
Machine learning was used to incorporate data that had been used to create parameters into a dataset. The parameters of several inputs that are crucial for plant development are contained in the dataset. In addition to offering an output solution, the machine learning approach creates a connection between these input characteristics and some internal prediction parameters.

3.3. System architecture
The development of machine learning subclasses required utilizing the random forest technique to construct subclasses based on different crop feature sets. Choosing a learning algorithm is a crucial factor to take into account for any machine learning challenge. There are numerous learning algorithms available right now, each with its own benefits and drawbacks. Before selecting a learning algorithm, it is important to consider factors like the requirement to explain capacity, the size of the data collection, different types of characteristics, and more, as shown in Figure 1.

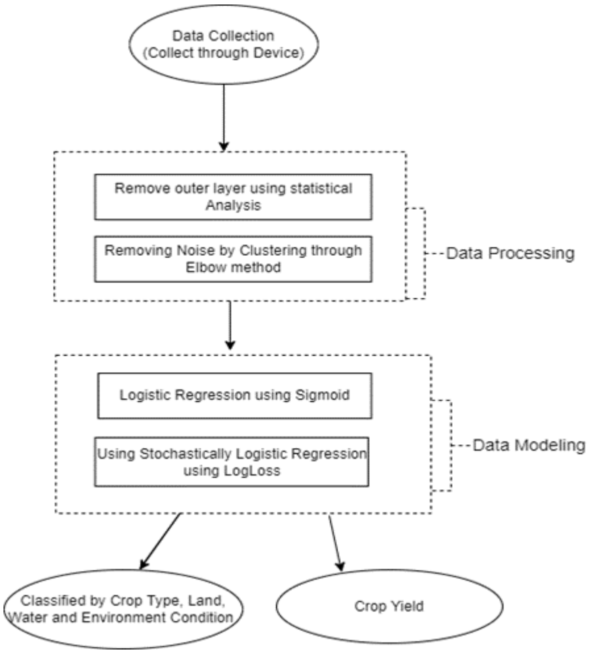

Figure 1. System architecture

We have gathered data from IoT sensors in the first step. Additionally, the data processing was divided into two stages. The outer layer is initially removed using static analysis, and then the noise is removed by clustering using the elbow approach. Following the removal of the noise, we conducted stochastically logistic regression using LogLoss and then logistic regression using sigmoid to identify the crop type, land, water, and environmental condition. It also provides crop yield.

4. IMPLEMENTATION
The suggested system has an input-gathering component that collects user input and runs the optimizer algorithm on it. Crop traits are divided into categories using the clustering method. The consumer is given access to the process data as comments. The fact that the growing crop requirements vary little from one region to the next is taken into account in this study.

When predicting future events based on historical data, predictive analytics uses data, statistical algorithms, and methods of machine learning. Some duplicate characteristics are discovered in the datasets during data cleaning. While making crop predictions, these features are not taken into consideration [25].

Figure 2 shows the distribution of agricultural conditions. The process of crop prediction starts with the import of external agricultural datasets. Pre-processing will proceed in stages after the dataset has been read, as explained in the data pre-processing section Chen trains the models in the training set using a decision tree classifier after the data has undergone pre-processing. We consider a number of factors, including temperature, humidity, soil parts hydrogen (pH), and expected rain, when estimating the crop. They serve as the system's input parameters, which may be manually entered or acquired by sensors.

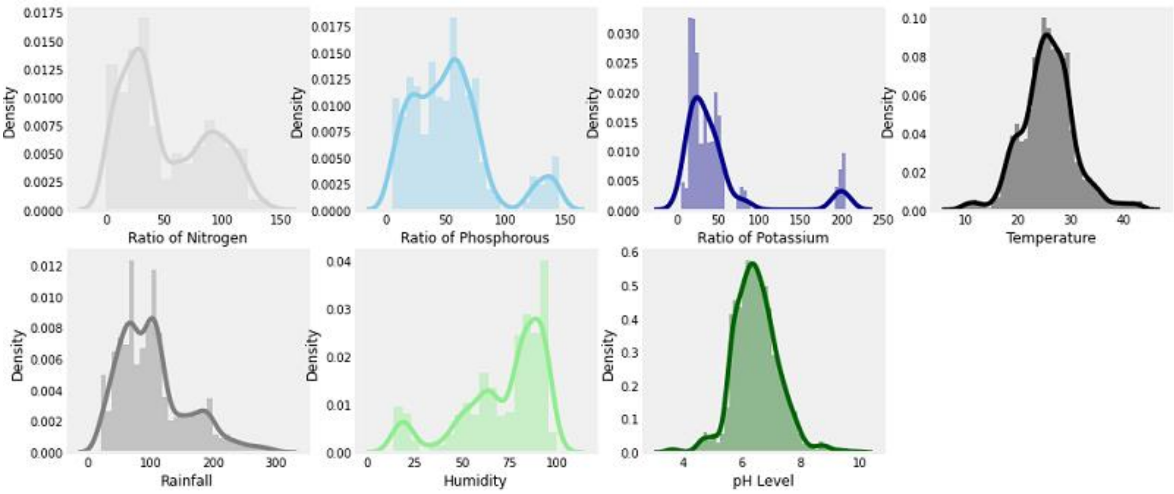

Figure 2. Distribution of agriculture conditions

Figure 3 impact of different conditions on crops in the logistic regression technique will forecast the crop based on the list data. Based on anticipated rainfall and soil composition, the system will choose the optimal crop for production. The recommended crop's seed requirements in kilograms per acre are also shown using this method.

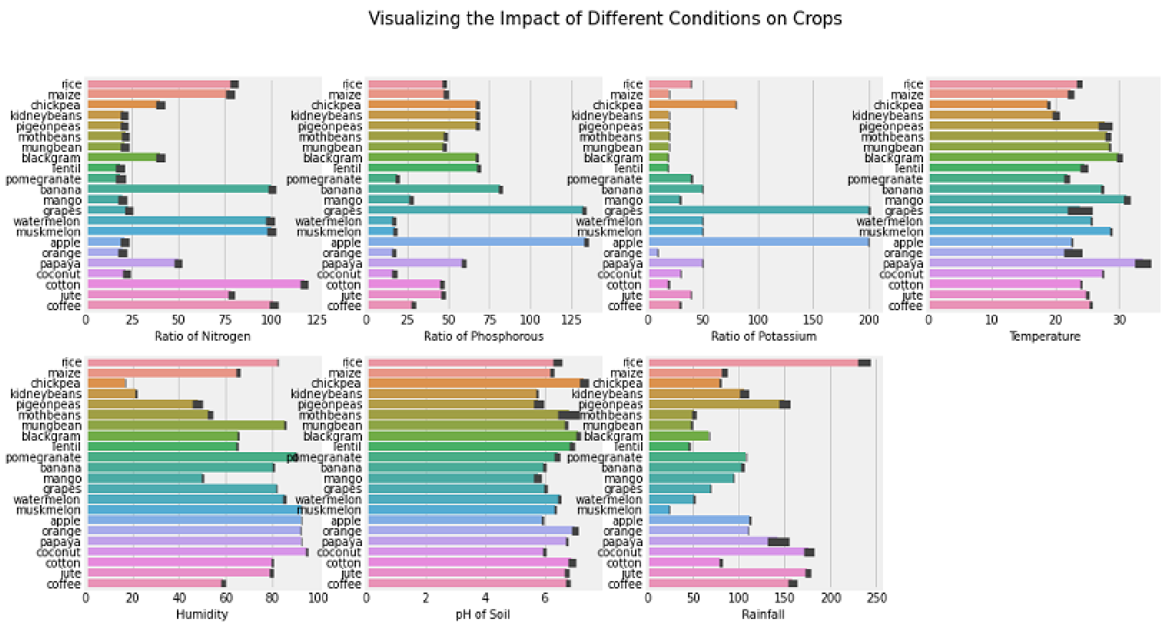

Figure 3. Impact of different conditions on crops

The suggested method recommends the best crop for a particular plot of land after taking annual rainfall, temperature, humidity, and soil pH into account. The system uses data from the prior year and the logistic regression technique to anticipate the year's rainfall; the user must supply the other elements. Table 1 shows the evaluation of model performance.

Table 1. Evaluation of model performance

Parameter	Precision	Recall	f1-score	Support
Apple	1.00	1.00	1.00	18
Banana	1.00	1.00	1.00	18
Blackgram	0.86	0.82	0.84	22
Chickpea	1.00	1.00	1.00	23
Coconut	1.00	1.00	1.00	15
Coffee	1.00	1.00	1.00	17
Cotton	0.89	1.00	0.94	16
Grapes	1.00	1.00	1.00	18
Jute	0.84	1.00	0.91	21
Kidney beans	1.00	1.00	1.00	20
Lentil	0.94	0.94	0.94	17
Maize	0.94	0.89	0.91	18
Mango	1.00	1.00	1.00	21
Mothbeans	0.88	0.92	0.90	25
Mungbean	1.00	1.00	1.00	17
Muskmelon	1.00	1.00	1.00	23
Orange	1.00	1.00	1.00	23
Papaya	1.00	0.95	0.98	21
Pigeonpeas	1.00	1.00	1.00	22
Pomegranate	1.00	1.00	1.00	23
Rice	1.00	0.84	0.91	25
Watermelon	1.00	1.00	1.00	17
Accuracy			0.97	440
Macro Avg.	0.97	0.97	0.97	440
Weighted Avg.	0.97	0.97	0.97	440

5. CONCLUSION

The crop selection method may improve the net yield rate of crops to be planted over the season. The Algorithm works to resolve the selection of crop (s) based on the prediction yield rate influenced by parameters (e.g., weather, soil type, water density, crop type), in our study, for crop-applied clustering and it is optimized using the elbow method. Further, this problem was solved by stochastic logistic regression and obtained an accuracy of 97% in crop yield. Farmers run the risk of choosing the wrong crop, which will reduce their income. We developed a system with a graphical user interface to predict which crop would be the greatest fit for a specific piece of land. Farmers are more likely to select the proper crop to cultivate, which leads to the agricultural business being improved.

REFERENCES

[1] Priya, Muthaiah, and Balamurugan, "Predicting yield of the crop using machine learning algorithms," *International journal of engineering sciences and reseach technology*, vol. 7, no. 4, pp. 1–7, 2018.

[2] R. Singh, T. Kumari, P. Verma, B. P. Singh, and A. S. Raghubanshi, "Compatible package-based agriculture systems: an urgent need for agro-ecological balance and climate change adaptation," *Soil Ecology Letters*, vol. 4, no. 3, pp. 187–212, Sep. 2022, doi: 10.1007/s42832-021-0087-1.

[3] R. K. Sidhu, R. Kumar, and P. S. Rana, "Machine learning based crop water demand forecasting using minimum climatological data," *Multimedia Tools and Applications*, vol. 79, pp. 13109-13124, 2020, doi: 10.1007/s11042-019-08533-w.

[4] C. Miller, V. N. Saroja, and C. Linder, "ICT uses for inclusive agricultural value chains," FAO, pp. 1–69, 2013, Accessed: May 09, 2023. [Online]. Available: http://www.fao.org/docrep/017/aq078e/aq078e.pdf.

[5] M. O. Adebiyi, R. O. Ogundokun, and A. A. Abokhai, "Machine learning-based predictive farmland optimization and crop monitoring system," *Scientifica*, vol. 2020, pp. 1–12, May 2020, doi: 10.1155/2020/9428281.

[6] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: a review," *Sensors (Switzerland)*, vol. 18, no. 8, 2018, doi: 10.3390/s18082674.

[7] F. Zhang, "Research on water-saving irrigation automatic control system based on internet of things," in *2011 International Conference on Electric Information and Control Engineering, ICEICE 2011 - Proceedings*, Apr. 2011, pp. 2541–2544, doi: 10.1109/ICEICE.2011.5778297.

[8] B. Sadhana, G. Naik, J. R. Mythri, P. G. Hedge, and B. S. K. Sharma, "Development of quad copter-based pesticide spraying mechanism for agricultural applications," *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, vol. 5, no. 2, pp. 121–123, 2017.

[9] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery, "Improving crop productivity through a crop recommendation system using ensembling technique," in *Proceedings 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions, CSITSS 2018*, Dec. 2018, pp. 114–119, doi: 10.1109/CSITSS.2018.8768790.

[10] K. K. Kumar, K. R. Kumar, R. G. Ashrit, N. R. Deshpande, and J. W. Hansen, "Climate impacts on Indian agriculture," *International Journal of Climatology*, vol. 24, no. 11, pp. 1375–1393, Sep. 2004, doi: 10.1002/joc.1081.
[11] W. Okori and J. Obua, "Machine learning classification technique for famine prediction," in *Proceedings of the World Congress on Engineering 2011, WCE 2011*, 2011, vol. 2, pp. 991–996.
[12] D. Ming, T. Zhou, M. Wang, and T. Tan, "Land cover classification using random forest with genetic algorithm-based parameter optimization," *Journal of Applied Remote Sensing*, vol. 10, no. 3, p. 35021, Sep. 2016, doi: 10.1117/1.jrs.10.035021.
[13] A. A. Raorane and R. V. Kulkarni, "Data mining: an effective tool for yield estimation in the agricultural sector," *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, vol. 1, no. 2, 2012.
[14] J. H. Jeong *et al.*, "Random forests for global and regional crop yield predictions," *PLoS ONE*, vol. 11, no. 6, p. e0156571, Jun. 2016, doi: 10.1371/journal.pone.0156571.
[15] A. Pang, M. W. L. Chang, and Y. Chen, "Evaluation of random forests (RF) for regional and local-scale wheat yield prediction in Southeast Australia," *Sensors*, vol. 22, no. 3, pp. 717, 2022, doi: 10.3390/s22030717.
[16] X. Chen and P.-H. Cournède, "Model-driven and data-driven approaches for crop yield prediction : analysis and comparison," *International Journal of Mathematical and Computational sciences*, vol. 11, no. 7, pp. 334–342, 2017.
[17] M. Sap, M. Noor, and A. M. Awan, "Development of an intelligent prediction tool for rice yield based on machine learning techniques," *Jurnal Teknologi Maklumat*, vol. 18, no. 2, pp. 73–74, 2006.
[18] Y. Zhang and L. Wu, "Crop classification by forward neural network with adaptive chaotic particle swarm optimization," *Sensors*, vol. 11, no. 5, pp. 4721–4743, May 2011, doi: 10.3390/s110504721.
[19] S. Nath, J. K. Nath, and K. C. Sarma, "IoT based system for continuous measurement and monitoring of temperature, soil moisture and relative humidity," *IAEME Publication*, vol. 9, no. 3, pp. 106–113, 2018.
[20] A. G.-Sanchez, J. F.-Solis, and W. O.-Bustamante, "Predictive ability of machine learning methods for massive crop yield prediction," *Spanish Journal of Agricultural Research*, vol. 12, no. 2, pp. 313–328, Apr. 2014, doi: 10.5424/sjar/2014122-4439.
[21] S. M. Arafat, M. A. Aboelghar, and E. F. Ahmed, "Crop discrimination using field hyper spectral remotely sensed data," *Advances in Remote Sensing*, vol. 02, no. 02, pp. 63–70, 2013, doi: 10.4236/ars.2013.22009.
[22] S. Shidnal, M. V. Latte, and A. Kapoor, "Crop yield prediction: two-tiered machine learning model approach," *International Journal of Information Technology*, vol. 13, no. 5, pp. 1983–1991, Oct. 2021, doi: 10.1007/s41870-019-00375-x.
[23] L. Ting, M. Khan, A. Sharma, and M. D. Ansari, "A secure framework for IoT-based smart climate agriculture system: toward blockchain and edge computing," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 221–236, 2022, doi: 10.1515/jisys-2022-0012.
[24] T. M. Shah, D. P. B. Nasika, and R. Otterpohl, "Plant and weed identifier robot as an agroecological tool using artificial neural networks for image identification," *Agriculture*, vol. 11, no. 3, p. 222, Mar. 2021, doi: 10.3390/agriculture11030222.
[25] I. Nitze, U. Schulthess, and H. Asche, "Comparison of machine learning algorithms random forest, artificial neuronal network and support vector machine to maximum likelihood for supervised crop type classification," in *Proceedings of the 4th Conference on GEographic Object-Based Image Analysis – GEOBIA 2012*, 2012, pp. 35–40.

BIOGRAPHIES OF AUTHORS

Nitin Padriya graduated from Gujarat University in 2007. He received an M.E. (Computer Engineering) degree from Gujarat Technological University in 2013, and he is currently pursuing a Ph.D. from Sankalchand Patel University, Visanagar, India. He can be contacted at email: nitinpadriya@gmail.com.

Dr. Nimisha Patel is currently working as a Professor and Head of CE/IT department at Gandhinagar Institute of Technology constituent Institute of Gandhinagar University since September 2022. She also worked as an Associate Professor and Head in Computer Engineering and Engineering department of Indus Institute of Technology and Engineering Constituent Institute of Indus University. Before this, she also worked as a Professor, Associate Professor, Assistant Professor and as a lecturer in the Sankalchand Patel College of Engineering, Visnagar, India. She has more than 18 years of teaching experience in the field of CE/CSE and Information Technology. She has also worked as a Dean, Faculty of Computer Science, Sankalchand Patel University, Visnagar and Associate Dean, R&D at Indus University. Dr. Nimisha Patel completed her Ph.D. (Computer Science and Engineering) with Cloud Computing as the major domain for research. She completed her M.Tech. (Computer Engineering) from S. V. National Institute of Technology (NIT), Surat and B.E. (Information Technology) from HNGU, Patan. She has published 50 research papers in various conferences and Journals of national and international repute with Scopus, eSCI, UGC, and indexing. She has delivered expert talks and also acted as session chair and reviewer for many conferences. Her area of domain is cloud computing, networking, information security, and parallel programming. She can be contacted at email: nimishapatelce@gmail.com.