Indonesian Journal of Electrical Engineering and Computer Science
Vol. 33, No. 3, March 2024, pp. 1748~1759
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v33.i3.pp1748-1759

Virtual analysis of machine learning models for diseases prediction in muskmelon

Deeba Kannan[1], Balakrishnan Amutha[1], Sattianadan Dasarathan[2], Daniel Rosy Salomi Victoria[3], Vikas Maheshkar[4], Ravindran Ramkumar[5], Dhandapani Karthikeyan[2]
[1]Department of CTECH, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai, India
[2]Department of Electrical and Electronics Engineering, SRM Institute of Science and Technology, kattankulathur, Chennai, India
[3]Department of Computer Science and Engineering, St. Joseph's College of Engineering, Chennai, India
[4]Department of Information Technology, Netaji Subhas University of Technology, New Delhi, India
[5]Department of Electrical and Electronics Engineering, School of Engineering and Technology, Dhanalakshmi Srinivasan University, Tiruchirapalli, India

Article Info

Article history:

Received Oct 24, 2023
Revised Nov 27, 2023
Accepted Jan 3, 2024

Keywords:

Back propagation network
Disease forecast
Machine learning models
Plant diseases
XGBoost

ABSTRACT

Muskmelon, a crop prized for its economic potential, has a relatively brief growth cycle. Disease susceptibility during this period can have a profound impact on yields, posing challenges for farmers. Environmental conditions are pivotal in disease occurrence. Unfavorable conditions reduce the likelihood of pathogens infecting vulnerable host plants as temperature and humidity influence pathogen behavior, including toxin synthesis, virulence protein production, and reproduction. Pathogens can lie dormant in the soil until suitable conditions activate them. When the right environment and host plants align, these dormant pathogens can cause outbreaks. Disease prediction becomes possible by analyzing environmental variables. Real-time data collected via strategically placed sensors focused on viral, fungal, and bacterial infections. Results indicated that the extreme gradient boosting (XGBoost) algorithm, with a maximum tree depth of 4 and 30 trees per iteration, achieved remarkable performance, yielding an accuracy of 97%. For comparison, the XGBoost model outperformed an 8-layer Backpropagation network with 7 nodes per layer, which achieved 95% accuracy. These findings underscore XGBoost's efficacy in forecasting and mitigating muskmelon plant diseases, offering the potential for improved crop yields and agricultural sustainability.

Corresponding Author:

Dhandapani Karthikeyan
Department of Electrical and Electronics Engineering, SRM Institute of Science and Technology
Kattankulathur, 603203, Chennai, Tamil Nadu, India
Email: karthikd@srmist.edu.in

1. INTRODUCTION

Muskmelon is a remunerative crop with a total cultivation time of around 65 to 75 days. The cultivation of muskmelon, however, faces significant challenges due to various diseases that can affect the crop, leading to substantial yield losses and economic implications for growers. Timely and accurate disease prediction and management are essential for maintaining crop health and ensuring a bountiful harvest. Here time is precious and the disease attack at this time duration will be a greater loss for the farmers [1]. Hence forecasting the possibility of a disease attack can alert the farmers to be prepared for any pathogen attack based on environmental conditions. Climatic changes have a direct impact on pathogen development in plants [2]. Major climatic factors considered are temperature, humidity, leaf wetness and soil moisture. Climate change is a predominant factor behind the spread of pests and pathogens. Climate change may have

an impact on the size of the population, the mortality rate, and the geographical spread of disease, and the incidence, growth, and geographical distribution of diseases [3]–[7]. The relationship between pathogen, host, environment and climatic factors is exhibited in the Figure 1.

In recent years, the application of machine learning (ML) techniques in agriculture has gained substantial attention as a promising approach to address these challenges. ML models can be trained to recognize patterns and predict the onset of diseases in muskmelon plants based on various input factors such as environmental conditions, growth stage, and historical disease data [8]–[12]. By harnessing the power of data analysis and predictive modeling, ML algorithms can potentially provide invaluable insights into disease outbreaks and assist farmers in implementing timely interventions to mitigate losses.

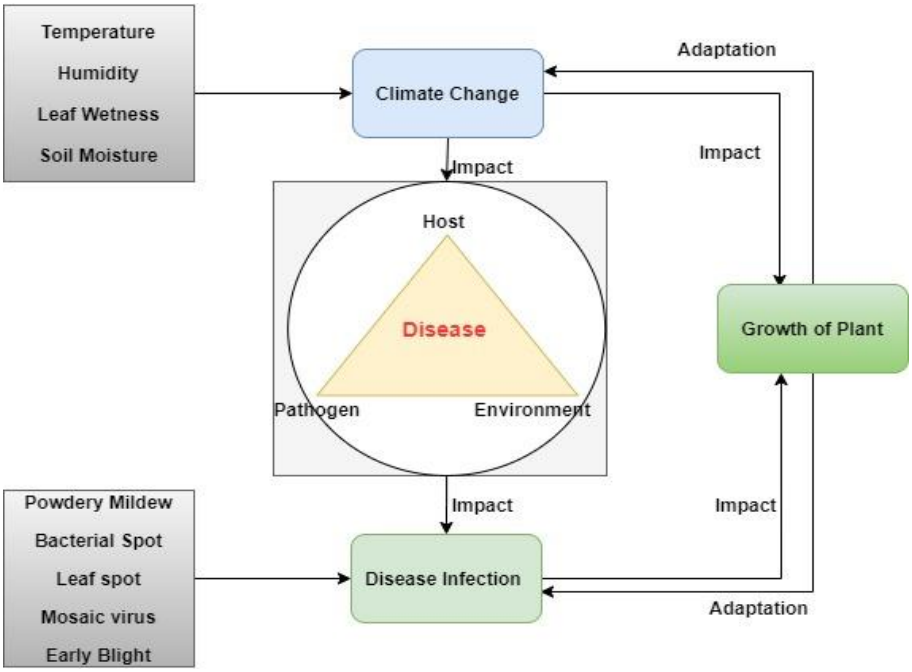

Figure 1. Relation between environmental factors and disease

This research paper explores the concept of "virtual analysis of machine learning models for diseases prediction in muskmelon." The term "virtual analysis" implies that the analysis and experimentation primarily occur in a digital environment using software tools and simulated data. In a virtual analysis setting, researchers can examine and optimize ML models without the need for physical experimentation, making it a cost-effective and efficient approach for agricultural research [13]–[19]. The primary objectives of this research are as follows:

- To collect and preprocess a dataset comprising relevant features related to muskmelon plants, including disease labels and environmental factors.
- To evaluate different ML algorithms for their effectiveness in predicting diseases in muskmelon plants.
- To conduct hyperparameter tuning and cross-validation to ensure the robustness and accuracy of the selected ML models.
- To use virtual analysis tools to visualize and interpret the results, aiding in the understanding of how these models make predictions.

This paper aims to provide insights into the application of ML for disease prediction in muskmelon, emphasizing the potential benefits of virtual analysis as a tool for improving agricultural practices. By harnessing the capabilities of ML, farmers and agricultural experts can make informed decisions, implement preventive measures, and effectively manage muskmelon diseases, ultimately enhancing crop yields and economic sustainability. In the subsequent sections of this paper, we will delve into the methodology, experiments, results, and discussions, presenting a comprehensive analysis of the virtual analysis of machine learning models for diseases prediction in muskmelon. Through this research, we aspire to contribute to the advancement of precision agriculture and the sustainable cultivation of muskmelon.

2. PROPOSED METHOD

The poposed method includes placing the sesors across the agrigultural filed. The real time climatic data like temperature, humidity, leaf wetness and soil moisture have been collected from the field. In this research, a comprehensive methodology has been devised to address the critical issue of disease prediction in agricultural fields, with a particular focus on muskmelon crops. The proposed approach incorporates a multidimensional strategy that integrates sensor-based data collection, real-time climatic monitoring, and advanced machine learning techniques to achieve accurate and timely predictions of leaf diseases in muskmelon plants.

The relationship between these parameters and pathogenic attack [8] regardless of these parameters need to be analyzed. The objective of this research is to develop a temporal prediction of leaf diseases upon vital features using multi class classification with imbalanced dataset based on environmental parameters. Steps involved in the process are as follows:

- Step 1: data set analysis–analyse the relationship between environmental parameters and diseases. Environmental parameters like temperature, soil moisture, humidity and leaf wetness has a greater effect on disease development. The foundation of this methodology lies in the strategic placement of sensors across the agricultural fields where muskmelon crops are cultivated. These sensors are equipped to capture real-time data related to crucial environmental parameters. By leveraging these sensors, a continuous stream of data is gathered from the field, ensuring the collection of vital information relevant to plant health and the surrounding conditions.
- Step 2: in this critical second step, a relational knowledge base is established. This knowledge base is enriched with environmental parameter values and their corresponding disease possibilities. Each entry in this database is a piece of the puzzle, capturing the intricate relationship between environmental conditions and the likelihood of disease occurrence in muskmelon crops. This comprehensive knowledge base forms the backbone of our disease prediction system, enabling the model to draw insights and predictions based on real-world observations.
- Step 3: the third step involves a meticulous analysis of the dataset. This process aims to unveil the underlying relationships between the attributes, representing the environmental parameters, and the class labels, which signify the various disease possibilities. By thoroughly understanding the dataset's structure and the connections between its elements, we equip ourselves with the knowledge required to make informed decisions and predictions [20], [26].
- Step 4: before the dataset is ready for the rigors of predictive modeling, data pre-processing is conducted. This step is essential for ensuring data quality and consistency [27]–[35]. It involves addressing missing values, a common issue in real-world datasets, and normalizing data to bring it to a standard scale. Data pre-processing lays the foundation for accurate and reliable model training and testing.
- Step 5: the dataset is divided into two distinct subsets in this step: a training set and a testing set. The training set encompasses 70% of the data, and the testing set contains the remaining 30%. This separation is crucial for model development and validation, enabling us to train our models on a substantial portion of the data while reserving a separate set for evaluating their performance.
- Step 6: at this stage, various classification algorithms are applied to the dataset. These algorithms, including support vector machines (SVM), decision trees, back propagation networks, XGBoost, and random forest [36]–[40] are carefully chosen for their ability to handle classification tasks effectively. This diversity allows us to explore a range of methods, maximizing the chances of finding the most suitable one for our specific problem.
- Step 8: once the appropriate classification algorithms have been selected, the models are trained and tested. This process involves feeding the training data into the models to teach them to recognize patterns and relationships between environmental parameters and disease possibilities [41]–[45]. Subsequently, the models are tested on the separate testing dataset to evaluate their performance and ensure they can make accurate predictions.
- Step 9: to assess the performance of the classification algorithms, quality of service (QoS) parameters like accuracy are chosen. Accuracy represents the model's ability to make correct predictions. These parameters act as benchmarks for evaluating and comparing the algorithms' performance [46]–[50]. Analysis of these parameters provides valuable insights into which algorithm is best suited for our specific dataset.
- Step 10: finally, the best-performing algorithm is identified based on its infusing capability. This selection process considers not only accuracy but also other relevant factors, such as the algorithm's ability to generalize and make predictions in real-world scenarios. The chosen algorithm will be the cornerstone of our disease prediction system, offering the capability to make informed and timely predictions based on environmental parameters.

In the initial step of our disease prediction methodology, we conduct a thorough analysis of the dataset, focusing on the relationship between environmental parameters and diseases affecting muskmelon crops. Sensors strategically placed across agricultural fields continuously capture real-time data on temperature, soil moisture, humidity, and leaf wetness. In the second step, a relational knowledge base is established, correlating environmental parameter values with disease possibilities. This comprehensive database forms the backbone of our prediction system.

The third step involves meticulous dataset analysis to unveil underlying relationships, while data pre-processing ensures quality and consistency [51]–[55]. The dataset is then divided into training and testing sets, allowing for robust model development. Multiple classification algorithms, such as SVM and random forest, are applied and rigorously tested.

QoS parameters, including accuracy, guide algorithm selection. The best-performing algorithm, chosen based on its infusing capability and real-world applicability, becomes the cornerstone of our disease prediction system. These steps encompass a systematic and data-driven approach to disease prediction in muskmelon crops. Through the creation of a relational knowledge base, dataset analysis, algorithm selection, and rigorous model training and testing, we aim to develop a robust system capable of providing valuable insights into disease possibilities based on environmental conditions.

2.1. Data collection

The initial process is to collect the real time images from agricultural field. The plant variety chosen under consideration is muskmelon (cucumis melo var. cantalupensis). The data are collected from 1 acre of agricultural land located in Sathappadi (Village), Attur (Tk), Salem (Dt), TamilNadu (DMS latitude:11° 35' 53.2176" N, DMS longitude 78° 35' 48.4872" E). The research mission planning to reconnaissance and clench a session is around 625. The plant data collection was executed during April to June and July to September of 2019, in two sessions and 4 times a day (6.30 am, 11.30 am, 3.30 pm, 7.30 pm). By encompassing the plant leaf images circumscribing the situational parameters of the agricultural field such as temperature, soil moisture, humidity, and pH level have been perceived for experimenting the research with timestamps.

2.2. Choosing appropriate classifier system

It is important to investigate the best classifier system for the type of imbalanced dataset, chosen. The dataset was trained on the best supervised machine learning algorithms as discussed in the introduction chapter. A comparative analysis is conducted on SVM, decision tree, back propagation network, XGBoost and random forest algorithm. The QoS parameter chosen for analysis of prediction is accuracy.

All the machine learning algorithms were executed for 10 iterations and an average accuracy is considered for comparison. A box plot graph is plotted based on the results obtained as shown in Figure 2. Among the algorithms compared are random forest, XGBoost and back propagation network perform better than the other ML models. Even though back propagation network gives maximum accuracy of 97%, XGBoost's average performance of 82% is better than the back propagation network is average performance of 79%. Hence back propagation net and XGBoost algorithms were chosen for our experimentation, based on average performance.

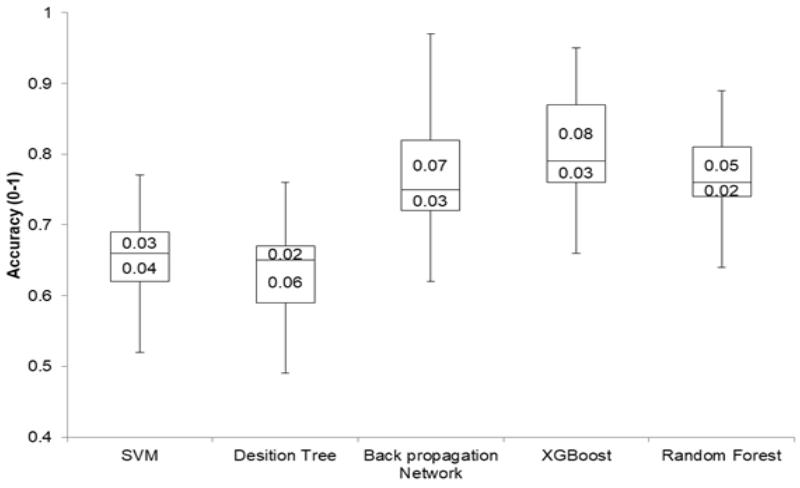

Figure 2. Comparative analysis of different classifier system based on accuracy

3. XGBOOST WITH VITAL FEATURE COMBINATION ALGORITHM

The gradient boosting framework is used by the decision tree-based ML method known as XGBoost [56], [57]. XGBoost was created by Tianqi Chen and was first kept up by the distributed (deep) machine learning community (DMLC). It is the most prevalent machine learning algorithm in use today, and its popularity has grown as a result of its success with both structured and tabular data. XGBoost is categorised as a boosting strategy in ensemble learning. To increase prediction accuracy, ensemble learning compiles various models into a group of predictors. By adding weights to the models in the boosting strategy, later models seek to correct the faults created by earlier models. XGBoost is a gradient boosted decision tree solution that improves speed and performance. Gain, weight, cover, total gain, and total cover are all feature significance kinds in XGBoost [58]. The default significance type is weight. The number of times a feature was utilized to divide the data across all trees is computed. The average gain across all splits where the feature was used is used to calculate gain.

3.1. XGBoost training

The training dataset in XGBoost is splitted into k-folds and the algorithm generates k-trees with k-fold cross validation. The structure of XGBoost algorithm with vital feature combination is displayed in Figure 3. The XGBoost algorithm steps will be explained after Figure 3.

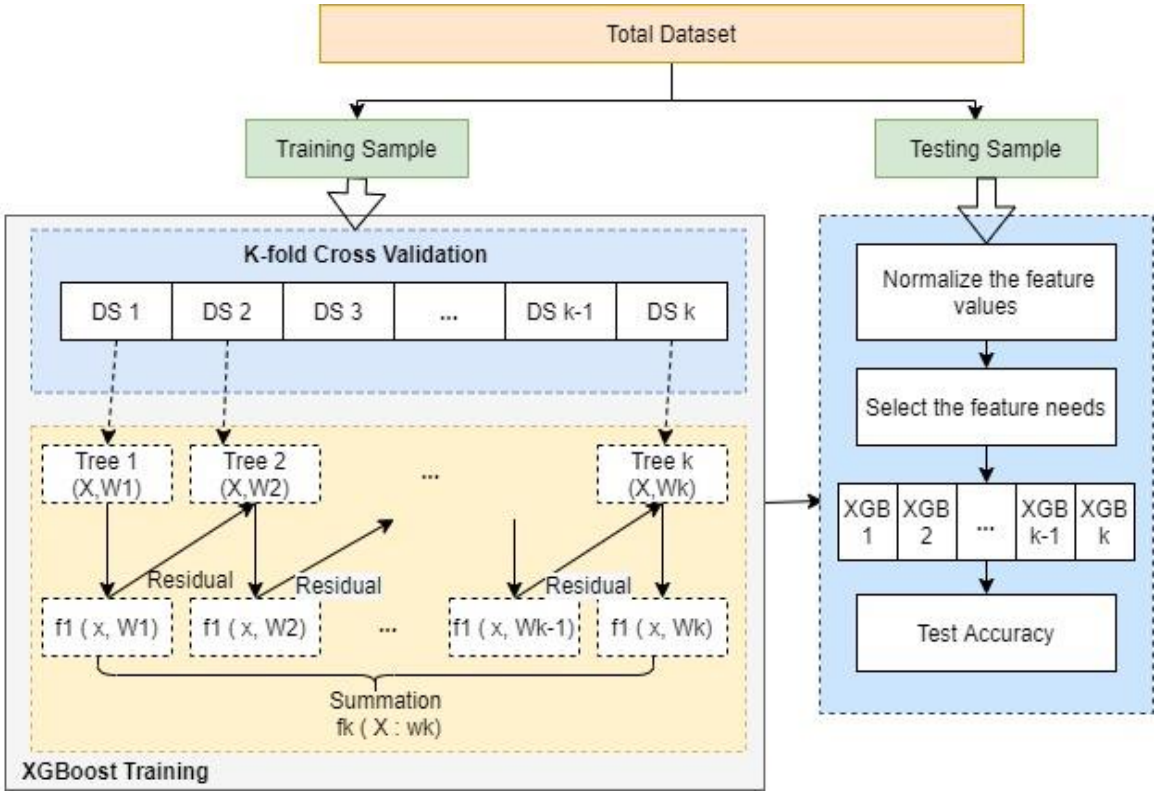

Figure 3. Working of XGBoost with vital feature combination as environmental parameters

The XGBoost algorithm steps:

- Initialization: initialize the model with a constant value, which is the global mean of the target variable for regression tasks or a balanced value for classification tasks [59]. Calculate the initial prediction, which is the model's prediction based on this constant value.
- Calculate the residuals: calculate the residuals by subtracting the initial prediction from the true target values. These residuals represent the errors the model needs to correct.
- Build a weak learner (decision tree): create a decision tree as a weak learner to capture the relationships between features and residuals. Train the decision tree to minimize the residuals by adjusting the tree's structure and feature splits.

- Update the model: add the newly created decision tree to the ensemble. Adjust the weights of the decision tree to minimize the loss function, taking into account both the residuals and a regularization term that prevents overfitting.
- Update the prediction: update the model's prediction by adding the contribution of the new decision tree. This updated prediction is the ensemble's current estimate of the target variable.
- Repeat: repeat steps 2 to 5 for a predefined number of iterations (trees) or until a stopping criterion, such as a maximum depth, is reached.

3.2. Back propagation network

To analyze the efficiency of XG-Boost, its accuracy is compared to that of a back propagation network [60], [61], the maximum accuracy is achieved by back propagation network. The basis of neural network training is utilized for back propagation network. Feed-forward back propagation network includes three layers namely input layer, hidden layers and output layer. The nodes that correspond to the variables at the input and output are present in both the input and output layers. There are 10 input neurons which corresponds to each environmental parameters. Similarly, the number of output neurons corresponds to the number of class labels. Totally four class labels are there for our research, one healthy class and three diseased class labels. The data travels between these nodes through the weighted links. The number of hidden layers and number of neurons in the hidden layers can be varied and based on the better combination the network model can be finalized.

4. PERFORMANCE ANALYSIS

The dataset used in our experimentation is collected from real time agricultural filed which consists of 2,510 entries with three class labels. Three main pathogen attack were addressed in this work namely viral, bacterial and fungal diseases along with one class label for healthy leaf condition represented as DS1, DS2, DS3, HLY. Ten environmental conditions have been considered as the combination in the feature set.

4.1. XGBoost performance

The performance of the XGBoost algorithm is measured by varying different parameters as the number of trees generated for each iteration and the depth of tree generated. The algorithm is iterated for 100 epochs and is executed for 10 times, to measure the average accuracy of prediction. Table 1 presents the parameters used for evaluating the performance of XGBoost. The performance of XGBoost is evaluated in Table 2 and Figure 4 by altering the number of trees created and maximum depth of each tree as 3 and 4. The experimental results inferred that XGBoost with maximum tree depth as 4 and 30 number of trees generated at each iteration performs well with a mean accuracy of 97%, than other combinations for a particular application.

Table 1. Attributes used for evaluating the performance of XGBoost

Attributes	Values
Disease class label	3 Diseases+1 healthy
Total number entries	2,510
Total attributes	10
Learning rate	α–0.15
Number of trees per iteration	5, 10, 15, 20, 25, 30
Tree depth	3.4
Total epoch	100
Total repetition	10
QoS parameters considered	Accuracy

Table 2. Accuracy of XGBoost with various number of trees generated

Number of trees generated	Max depth 3 (Mean accuracy)	Max depth 4 (Mean accuracy)
5	0.74	0.77
10	0.76	0.8
15	0.86	0.89
20	0.9	0.92
25	0.93	0.96
30	0.94	0.97

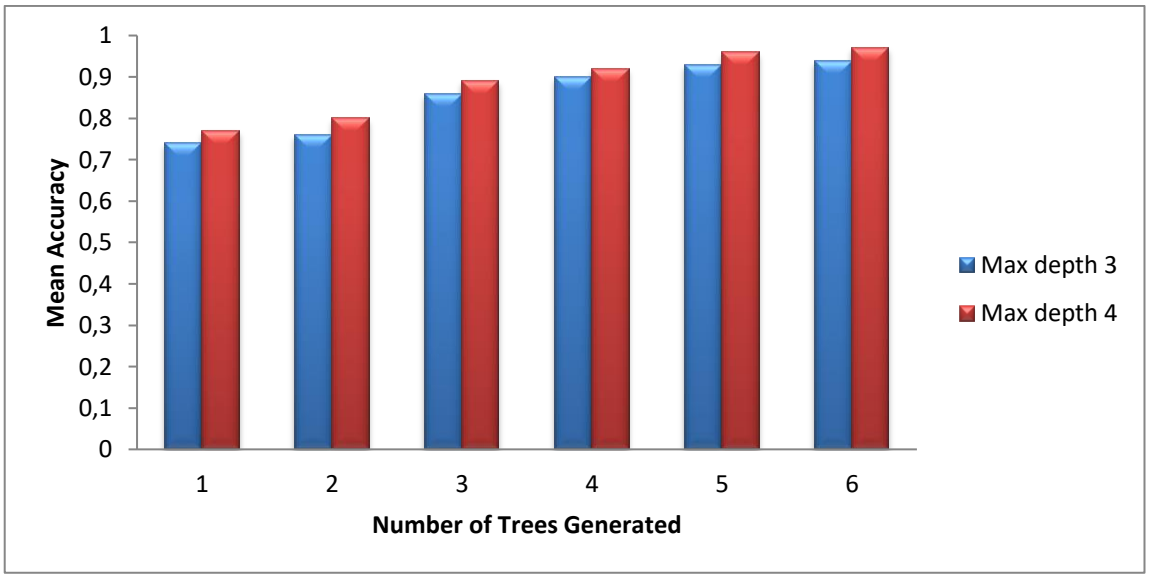

Figure 4. Accuracy of XGBoost with various number of trees generated

4.2. Back propagation net performance

The performance of the back propagation net is measured by varying the number of hidden layers and number of neurons in hidden layer. The algorithm is iterated for 100 epochs and is executed for 5 times, to measure the average accuracy of prediction. Table 3 presents the parameters used for evaluating the performance of back propagation net.

The network is analyzed with 40 different combinations of architecture by varying the number of hidden layers from 1 to 10 and number of neurons in hidden layers as 4, 5, 6, 7. Table 4 shows the accuracy of back propagation net with various hidden layers and hidden layer neurons and Figure 5 present the performance of back propagation net with accuracy as the mentioned QoS parameter [62]. Experimental results inferred that back propagation network with 8 hidden layers and each layer with 7 nodes perform better with an average accuracy of 95%.

Table 3. Parameters used for evaluating the performance of back propagation net

Parameters	Values
Disease class label	3 Diseases+1 healthy
Total number entries	2,510
Total attributes	10
Learning rate	α–0.1
Total epoch	100
Total repetition	5
QoS parameters considered	Accuracy
Analysis performed	Different hidden layers, different nodes of hidden layers

Table 4. Accuracy of back propagation net with various hidden layers and hidden layer neurons

Number of hidden layers	4 nodes	5 nodes	6 nodes	7 nodes
1	57.92	64.42	78.77	79.32
2	58.34	64.76	77.78	79.56
3	58.76	67.23	78.32	84.71
4	60.34	72.45	79.75	88.29
5	61.45	78.34	79.92	91.29
6	62.38	78.55	82.12	93.36
7	63.54	79.78	85.89	94.88
8	64.47	80.55	87.62	95.75
9	64.56	80.22	87.12	94.5
10	64.71	80.35	87.11	94.35

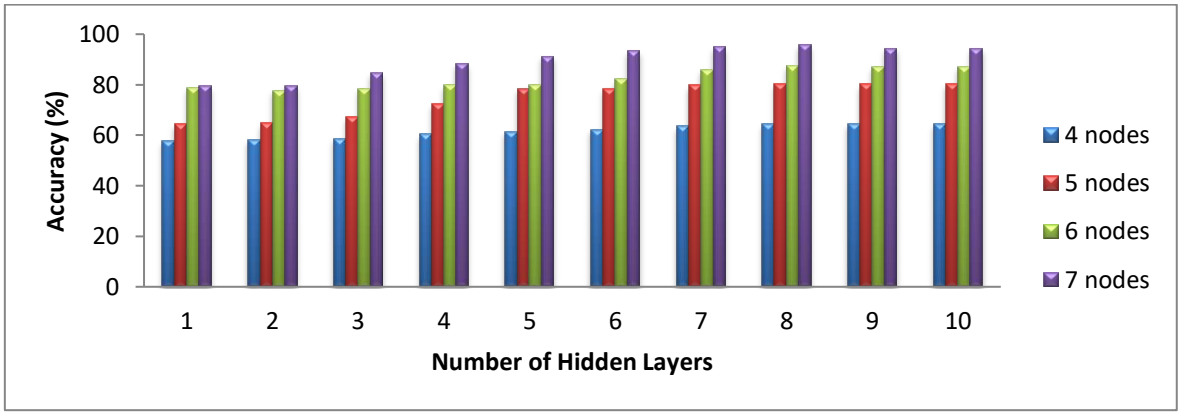

Figure 5. Accuracy of back propagation net with various hidden layers and hidden layer neurons

Our experimental findings reveal the remarkable performance of the XGBoost algorithm in the specific context of disease prediction. XGBoost, configured with a maximum tree depth of 4 and the generation of 30 trees at each iteration, consistently demonstrated its superior predictive power. Impressively, this configuration yielded an expected accuracy of 97%, signifying the algorithm's ability to make highly accurate forecasts. In stark contrast, an alternative approach was examined, utilizing a backpropagation network comprising 8 hidden layers, with each layer housing 7 nodes. This neural network architecture exhibited commendable performance, achieving an accuracy rate of 95%. While the backpropagation network displayed robust predictive capabilities, it fell slightly short of the XGBoost algorithm's exceptional accuracy. One notable highlight from our analysis is the XGBoost algorithm's proficiency in disease infection forecasting when compared to other models. The algorithm's ability to discern and capture intricate relationships between environmental factors and pathogen attacks is underscored by its superior performance. This finding underscores the practical significance of employing XGBoost as a predictive tool for disease control and prevention, offering agricultural experts and farmers a valuable means of safeguarding their muskmelon crops with the utmost accuracy.

5. CONCLUSION

Environmental conditions significantly influence plant diseases. When the environment is unfavorable for disease development, vulnerable hosts are less likely to be infected by aggressive pathogens. This study focused on three primary pathogen types: bacterial, fungal, and viral diseases. We conducted a comparative analysis of various ML algorithms using real-time datasets. The results of our experiments indicated that XGBoost, with a maximum tree depth of 4 and the generation of 30 trees at each iteration, outperformed other algorithms in this specific application, achieving an impressive expected accuracy of 97%. In contrast, a backpropagation network with 8 hidden layers, each containing 7 nodes, achieved an accuracy of 95%. Notably, XGBoost excelled in disease infection forecasting when compared to other models. Furthermore, there is potential for improvement by implementing a rule-based analyzer to select crucial features for each possible pathogen attack. This approach could enhance the model's ability to make accurate predictions and further contribute to disease control and prevention in muskmelon cultivation.

REFERENCES

[1] E. Stanaszek-Tomal, "Environmental factors causing the development of microorganisms on the surfaces of national cultural monuments made of mineral building materials—review," *Coatings*, vol. 10, no. 12, pp. 1–19, Dec. 2020, doi: 10.3390/coatings10121203.

[2] A. Pérez-García, D. Romero, D. FernÁndez-OrtuÑo, F. López-Ruiz, A. D. Vicente, and J. A. TorÉs, "The powdery mildew fungus Podosphaera fusca (synonym Podosphaera xanthii), a constant threat to cucurbits: Pathogen profile," *Molecular Plant Pathology*, vol. 10, no. 2, pp. 153–160, Feb. 2009, doi: 10.1111/j.1364-3703.2008.00527.x.

[3] L. Trecate, B. Sedláková, B. Mieslerová, V. Manstretta, V. Rossi, and A. Lebeda, "Effect of temperature on infection and development of powdery mildew on cucumber," *Plant Pathology*, vol. 68, no. 6, pp. 1165–1178, May 2019, doi: 10.1111/ppa.13038.

[4] X. F. Xin, B. Kvitko, and S. Y. He, "Pseudomonas syringae: what it takes to be a pathogen," *Nature Reviews Microbiology*, vol. 16, no. 5, pp. 316–328, Feb. 2018, doi: 10.1038/nrmicro.2018.17.

[5] C. Buttimer, O. McAuliffe, R. P. Ross, C. Hill, J. O'Mahony, and A. Coffey, "Bacteriophages and bacterial plant diseases," *Frontiers in Microbiology*, vol. 8, no. JAN, pp. 1–15, Jan. 2017, doi: 10.3389/fmicb.2017.00034.

[6] D. A. Cuppels and J. Elmhirst, "Disease development and changes in the natural pseudomonas syringae pv. tomato populations on field tomato plants," *Plant Disease*, vol. 83, no. 8, pp. 759–764, Aug. 1999, doi: 10.1094/PDIS.1999.83.8.759.

[7] R. A. C. Jones, "Global plant virus disease pandemics and epidemics," *Plants*, vol. 10, no. 2, pp. 1–41, Jan. 2021, doi: 10.3390/plants10020233.

[8] A. Giner *et al.*, "A mutation in the melon vacuolar protein sorting 41prevents systemic infection of cucumber mosaic virus," *Scientific Reports*, vol. 7, no. 1, pp. 1–12, Sep. 2017, doi: 10.1038/s41598-017-10783-3.

[9] Y. Jiang, G. Tong, H. Yin, and N. Xiong, "A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters," *IEEE Access*, vol. 7, pp. 118310–118321, 2019, doi: 10.1109/ACCESS.2019.2936454.

[10] X. Gu, Y. Han, and J. Yu, "A novel lane-changing decision model for autonomous vehicles based on deep autoencoder network and XGBoost," *IEEE Access*, vol. 8, pp. 9846–9863, 2020, doi: 10.1109/ACCESS.2020.2964294.

[11] A. Talwar, G. Lokhande, R. Jain, and S. Singh, "Estimation of aerodynamic parameters using cascade forward back propagation," in *2nd International Conference on Telecommunication and Networks, TEL-NET 2017*, Aug. 2017, vol. 2018-January, pp. 1–6, doi: 10.1109/TEL-NET.2017.8343508.

[12] N. A. Al-Sammarraie, Y. M. H. Al-Mayali, and Y. A. B. El-Ebiary, "Classification and diagnosis using back propagation artificial neural networks (ANN) algorithm," Jul. 2018, doi: 10.1109/ICSCEE.2018.8538383.

[13] M. M. Badža and M. C. Barjaktarović, "Classification of brain tumors from mri images using a convolutional neural network," *Applied Sciences (Switzerland)*, vol. 10, no. 6, p. 1999, Mar. 2020, doi: 10.3390/app10061999.

[14] V. K. G. Kalaiselvi, J. Ranjani, D. P. Rani, S. M. V. Kumar, and S. M. Priya, "Deep learning based pathogen infestation detection in plants," in *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, May 2021, pp. 1428–1433, doi: 10.1109/ICICCS51141.2021.9432106.

[15] W. T. Hay, S. P. McCormick, and M. M. Vaughan, "Effects of atmospheric CO_2 and temperature on wheat and corn susceptibility to fusarium graminearum and deoxynivalenol contamination," *Plants*, vol. 10, no. 12, p. 2582, Nov. 2021, doi: 10.3390/plants10122582.

[16] A. Das, M. N. Mohanty, P. K. Mallick, P. Tiwari, K. Muhammad, and H. Zhu, "Breast cancer detection using an ensemble deep learning method," *Biomedical Signal Processing and Control*, vol. 70, p. 103009, Sep. 2021, doi: 10.1016/j.bspc.2021.103009.

[17] A. S. Assiri, S. Nazir, and S. A. Velastin, "Breast tumor classification using an ensemble machine learning method," *Journal of Imaging*, vol. 6, no. 6, p. 39, May 2020, doi: 10.3390/JIMAGING6060039.

[18] S. Lee, M. Amgad, M. Masoud, R. Subramanian, D. Gutman, and L. Cooper, "An ensemble-based active learning for breast cancer classification," in *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, Nov. 2019, pp. 2549–2553, doi: 10.1109/BIBM47256.2019.8983317.

[19] D. G. P. Petrini, C. Shimizu, R. A. Roela, G. V. Valente, M. A. A. K. Folgueira, and H. Y. Kim, "Breast cancer diagnosis in two-view mammography using end-to-end trained efficientNet-based convolutional network," *IEEE Access*, vol. 10, pp. 77723–77731, 2022, doi: 10.1109/ACCESS.2022.3193250.

[20] M. Liu *et al.*, "A deep learning method for breast cancer classification in the pathology images," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 10, pp. 5025–5032, Oct. 2022, doi: 10.1109/JBHI.2022.3187765.

[21] X. Shu, L. Zhang, Z. Wang, Q. Lv, and Z. Yi, "Deep neural networks with region-based pooling structures for mammographic image classification," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2246–2255, Jun. 2020, doi: 10.1109/TMI.2020.2968397.

[22] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Scientific Reports*, vol. 9, no. 1, p. 12495, Aug. 2019, doi: 10.1038/s41598-019-48995-4.

[23] M. Karatayev, S. Khalyk, S. Adai, M.-H. Lee, and M. F. Demirci, "Breast cancer histopathology image classification using CNN," in *2021 16th International Conference on Electronics Computer and Computation (ICECCO)*, Nov. 2021, pp. 1–5, doi: 10.1109/ICECCO53203.2021.9663757.

[24] P. K. Samanta, N. K. Rout, and G. Panda, "A novel deep CNN model for improved breast cancer detection using ultrasound images," in *2023 International Conference on Communication, Circuits, and Systems (IC3S)*, May 2023, pp. 1–4, doi: 10.1109/IC3S57698.2023.10169383.

[25] S. Costanzo, A. Flores, and G. Buonanno, "Fast and accurate CNN-based machine learning approach for microwave medical imaging in cancer detection," *IEEE Access*, vol. 11, pp. 66063–66075, 2023, doi: 10.1109/ACCESS.2023.3291076.

[26] R. N. Patil, S. Rawandale, N. Rawandale, U. Rawandale, and S. Patil, "An efficient stacking based NSGA-II approach for predicting type 2 diabetes," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 1, pp. 1015–1023, Feb. 2023, doi: 10.11591/ijece.v13i1.pp1015-1023.

[27] J. Al Nahian, A. K. M. Masum, S. Abujar, and M. J. Mia, "Common human diseases prediction using machine learning based on survey data," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 6, pp. 3498–3508, Dec. 2022, doi: 10.11591/eei.v11i6.3405.

[28] A. S. Mustafa and W. R. Baiee, "Nutrition information estimation from food photos using machine learning based on multiple datasets," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 5, pp. 2922–2929, Oct. 2022, doi: 10.11591/eei.v11i5.4007.

[29] H. El Massari, N. Gherabi, S. Mhammedi, H. Ghandi, F. Qanouni, and M. Bahaj, "Integration of ontology with machine learning to predict the presence of covid-19 based on symptoms," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 5, pp. 2805–2816, Oct. 2022, doi: 10.11591/eei.v11i5.4392.

[30] S. Molla *et al.*, "A predictive analysis framework of heart disease using machine learning approaches," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 5, pp. 2705–2716, Oct. 2022, doi: 10.11591/eei.v11i5.3942.

[31] L. H. Poushy *et al.*, "Satisfaction prediction of online education in COVID-19 situation using data mining techniques: Bangladesh perspective," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 5, pp. 5553–5561, Oct. 2022, doi: 10.11591/ijece.v12i5.pp5553-5561.

[32] M. K. Alsmadi, "Modified SEIR and machine learning prediction of the trend of the epidemic of COVID-19 in Jordan under lockdowns impact," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 5, pp. 5455–5466, Oct. 2022, doi: 10.11591/ijece.v12i5.pp5455-5466.

[33] N. N. Moon *et al.*, "Predicting the mental health of rural Bangladeshi children in coronavirus disease 2019," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 5, pp. 5501–5510, Oct. 2022, doi: 10.11591/ijece.v12i5.pp5501-5510.

[34] T. A. Assegie, P. K. Rangarajan, N. K. Kumar, and D. Vigneswari, "An empirical study on machine learning algorithms for heart disease prediction," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 3, pp. 1066–1073, Sep. 2022, doi: 10.11591/ijai.v11.i3.pp1066-1073.

[35] D. Riswantini and E. Nugraheni, "Machine learning in handling disease outbreaks: a comprehensive review," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 4, pp. 2169–2186, Aug. 2022, doi: 10.11591/eei.v11i4.3612.
[36] M. Rahman, M. T. Ahmed, S. Nur, and A. Z. M. T. Islam, "The prediction of coronavirus disease 2019 outbreak on Bangladesh perspective using machine learning: a comparative study," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 4, pp. 4276–4287, Aug. 2022, doi: 10.11591/ijece.v12i4.pp4276-4287.
[37] A. A. Nayan, B. Kijsirikul, and Y. Iwahori, "Coronavirus disease situation analysis and prediction using machine learning: a study on Bangladeshi population," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 4, pp. 4217–4227, Aug. 2022, doi: 10.11591/ijece.v12i4.pp4217-4227.
[38] A. A. Jasim, A. A. Jalal, N. M. Abdulateef, and N. A. Talib, "Effectiveness evaluation of machine learning algorithms for breast cancer prediction," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 3, pp. 1516–1525, 2022, doi: 10.11591/eei.v11i3.3621.
[39] K. Jhajharia and P. Mathur, "A comprehensive review on machine learning in agriculture domain," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 2, pp. 753–763, Jun. 2022, doi: 10.11591/ijai.v11.i2.pp753-763.
[40] A. Bustamam, M. I. Sunggawa, and T. Siswantining, "Performance of multivariate mutual information and autocorrelation encoding methods for the prediction of protein-protein interactions," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 2, pp. 773–786, Jun. 2022, doi: 10.11591/ijai.v11.i2.pp773-786.
[41] A. H. Ahmed, M. N. A. Al-Hamadani, and I. A. Satam, "Prediction of COVID-19 disease severity using machine learning techniques," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 2, pp. 1069–1074, Apr. 2022, doi: 10.11591/eei.v11i2.3272.
[42] P. Vasavi, A. Punitha, and T. Venkat Narayana Rao, "Crop leaf disease detection and classification using machine learning and deep learning algorithms by visual symptoms: A review," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 2, pp. 2079–2086, Apr. 2022, doi: 10.11591/ijece.v12i2.pp2079-2086.
[43] T. Suresh, T. A. Assegie, S. Rajkumar, and N. K. Kumar, "A hybrid approach to medical decision-making: diagnosis of heart disease with machine-learning model," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 2, pp. 1831–1838, 2022, doi: 10.11591/ijece.v12i2.pp1831-1838.
[44] M. Panda, D. P. Mishra, S. M. Patro, and S. R. Salkuti, "Prediction of diabetes disease using machine learning algorithms," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 1, pp. 284–290, Mar. 2022, doi: 10.11591/ijai.v11.i1.pp284-290.
[45] M. El Midaoui, M. Qbadou, and K. Mansouri, "A fuzzy-based prediction approach for blood delivery using machine learning and genetic algorithm," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 1, pp. 1056–1068, Feb. 2022, doi: 10.11591/ijece.v12i1.pp1056-1068.
[46] R. Patil, S. Tamane, S. A. Rawandale, and K. Patil, "A modified mayfly-SVM approach for early detection of type 2 diabetes mellitus," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 1, pp. 524–533, Feb. 2022, doi: 10.11591/ijece.v12i1.pp524-533.
[47] S. Afrin *et al.*, "Supervised machine learning based liver disease prediction approach with LASSO feature selection," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 6, pp. 3369–3376, 2021, doi: 10.11591/eei.v10i6.3242.
[48] S. Tkatek, S. Amassmir, A. Belmzoukia, and J. Abouchabaka, "Predictive fertilization models for potato crops using machine learning techniques in Moroccan Gharb region," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 5, pp. 5942–5950, Oct. 2023, doi: 10.11591/ijece.v13i5.pp5942-5950.
[49] G. Emmanuel, A. Ramadhan, M. Zarlis, E. Abdurachman, and A. Trisetyarso, "Machine learning in drug supply chain management during disease outbreaks: a systematic review," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 5, pp. 5517–5533, Oct. 2023, doi: 10.11591/ijece.v13i5.pp5517-5533.
[50] M. Atif, F. Anwer, F. Talib, R. Alam, and F. Masood, "Analysis of machine learning classifiers for predicting diabetes mellitus in the preliminary stage," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 3, pp. 1302–1311, Sep. 2023, doi: 10.11591/ijai.v12.i3.pp1302-1311.
[51] L. Shrinivasan, R. Verma, and M. D. Nandeesh, "Early prediction of diabetes diagnosis using hybrid classification techniques," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 3, pp. 1139–1148, Sep. 2023, doi: 10.11591/ijai.v12.i3.pp1139-1148.
[52] S. Rajayyan and S. M. M. Mustafa, "Prediction of dementia using machine learning model and performance improvement with cuckoo algorithm," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 4, pp. 4623–4632, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4623-4632.
[53] M. K. Alsmadi *et al.*, "Susceptible exposed infectious recovered-machine learning for COVID-19 prediction in Saudi Arabia," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 4, pp. 4761–4776, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4761-4776.
[54] R. Rani, S. Kumar, R. R. Patil, and S. K. Pippal, "A machine learning model for predicting innovation effort of firms," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 4, pp. 4633–4639, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4633-4639.
[55] S. M. Hussain and S. Ghouse, "Detection and prediction of monkey pox disease by enhanced convolutional neural network approach," *International Journal of Public Health Science*, vol. 12, no. 2, pp. 673–681, Jun. 2023, doi: 10.11591/ijphs.v12i2.22310.
[56] N. A. Mashudi, N. Ahmad, and N. M. Noor, "Classification of adult autistic spectrum disorder using machine learning approach," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 3, pp. 743–751, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp743-751.
[57] N. Z. E. Zakariya and M. M. Rosli, "Physical activity prediction using fitness data: Challenges and issues," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 419–426, Feb. 2021, doi: 10.11591/eei.v10i1.2474.
[58] A. Srinivasulu and A. Pushpa, "Disease prediction in big data healthcare using extended convolutional neural network techniques," *International Journal of Advances in Applied Sciences*, vol. 9, no. 2, pp. 85–92, Jun. 2020, doi: 10.11591/ijaas.v9.i2.pp85-92.
[59] F. Abdali-Mohammadi, M. N. Meqdad, and S. Kadry, "Development of an IoT-based and cloud-based disease prediction and diagnosis system for healthcare using machine learning algorithms," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 4, pp. 766–771, Dec. 2020, doi: 10.11591/ijai.v9.i4.pp766-771.
[60] G. Saranya and A. Pravin, "A comprehensive study on disease risk predictions in machine learning," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 4, pp. 4217–4225, 2020, doi: 10.11591/ijece.v10i4.pp4217-4225.

[61] B. S. Shukur and M. M. Mijwil, "Involving machine learning techniques in heart disease diagnosis: a performance analysis," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, pp. 2177–2185, Apr. 2023, doi: 10.11591/ijece.v13i2.pp2177-2185.
[62] M. Yanto, Y. Wiyandra, and S. Defit, "Deep learning approach analysis model prediction and classification poverty status," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 1, pp. 459–468, Mar. 2023, doi: 10.11591/ijai.v12.i1.pp459-468.

BIOGRAPHIES OF AUTHORS

Deeba Kannan is an Assistant Professor at SRM Institute of Science and Technology, Chennai, Tamilnadu, India. She received her Ph.D. degree in Computer Science and Engineering, Chennai, Tamilnadu, India in 2021. Her research area is in IoT, deep learning and machine learning. She has 20 internatinal journal publication and also published two patents. She can be contacted at email: deebak@srmist.edu.in.

Balakrishnan Amutha is Professor at Department of Compute Science and Engineering, Chennai, Tamilnadu, India. She received her Ph.D. degree in Computer Science and Engineering, Chennai, Tamilnadu, India in 2012. Her main research area wireless networks, SDN and machine learning. She had more than 60 international journals and published five patents. She can be contacted at email: amuthab@srmist.edu.in.

Sattianadan Dasarathan was born in Pondicherry, India in 1976. He received his Bachelor of Engineering in Electrical and Electronics Engineering from Madras University in the year 2000, Master of Engineering in Power Systems from Anna University in the year 2005 and Ph.D. degree from SRM University in the year 2015. He currently holds the position of associate professor of the EEE Department at the SRM University, Chennai, India. His area of interest includes distributed generations, power system operation and control, FACTS and power system protection, DC microgrid. He is also a Lifetime Member of ISTE. He can be contacted at email: sattiand@srmist.edu.in.

Dr. Daniel Rosy Salomi Victoria is currently working an Associate Professor at St. Joseph's College of Engineering, Chennai. She has worked also in Karunya Institute of Technology, Coimbatore and Mepco Schlenk Engineering College, Sivakasi. She obtained her Ph.D. degree under Faculty of Information and Communication Engineering from ANNA University (2015), Chennai. She completed her M.E. degree in Computer Science and Engineering from Sathyabama University (2002), Chennai. She received her M.S. degree in Computer Science from BITS (1993), Pilani, Rajasthan. She completed her B.E. degree in CSE Department from Madurai Kamaraj University (1991). She has 32 years teaching experience. Her area of interests are computer networks, mobile computing, computer architecture, image processing, object oriented analysis and design and distributed systems. She has published papers in 7 SCI journals, 3 Scopus indexed journals, one UGC journal and other 8 international journals. She has presented papers in 4 IEEE international conferences, 6 national conferences and other 9 international conferences. She has published 3 Indian patents and one South African grant patent in her field. She is an author of one book chapter and 2 technical books. She obtained 8 online NPTEL Elite+silver certificates, 2 udemy certificates and many coursera certificates. She can be contacted at email: drosysalomi@gmail.com.

Dr. Vikas Maheshkar is assistant professor in the Department of Information Technology at Netaji Subhas University of Technology, Dwarka, New Delhi. His areas of interest include image processing and artificial intelligence. He has academic credentials with a Ph.D. from NIT Allahabad, M.Tech. (Computer Science and Engineering) from SATI, Vidisha and B.Tech. (Computer Technology) from Nagpur University. He can be contacted at email: vikas.maheshkar@nsut.ac.in.

Dr. Ravindran Ramkumar is currently working as an Assistant professor in the Department of Electrical and Electronics Engineering at Dhanalakshmi Srinivasan University, Trichy. He obtained his Ph.D. under Faculty of Electrical Engineering from Anna University (2022), Chennai. He completed his M.E. in Sethu Institute of Technology (2012), Madurai. He completed his B.E. in K.L.N. College of Information Technology (2008), Madurai. He has published more than 30 Scopus indexed journals and 6 SCI journals in his field. He has 11 years of teaching experience and 1-year industrial experience. His area of interest is power electronic converters, renewable energy and micro grid. He can be contacted at email: 2019ramkr@gmail.com.

Dhandapani Karthikeyan received a B.E. Degree in Electrical and Electronic Engineering from A.I.H.T College in Chennai, India (associated with Anna University in Chennai, India) in 2009, and an M.Tech. He received his bachelor's degree in Power Electronics and Drives from SRMIST (previously SRM University) in Kattankulathur, India, in 2013, and his Ph.D. in Multilevel Inverters in 2019. He is presently an assistant professor in the Department of Electrical Engineering at SRMIST (previously SRM University) in Kattankulathur and Chennai, India. His current research interests include power electronic multilayer inverters, alternating current drives, and direct current drives. He is a member of several professional organizations, including the IEEE, IET, IEI, and ISCA. He can be contacted at email: karthipncl@gmail.com.