WILEY | Hindawi

Research Article

Machine Learning- and Feature Selection-Enabled Framework for Accurate Crop Yield Prediction

Sandeep Gupta [iD],[1] **Angelina Geetha** [iD],[2] **K. Sakthidasan Sankaran** [iD],[3]
Abu Sarwar Zamani [iD],[4] **Mahyudin Ritonga** [iD],[5] **Roop Raj** [iD],[6] **Samrat Ray** [iD],[7]
and Hussien Sobahi Mohammed [iD][8]

[1]*Department of Computer Science and Engineering, JIMS Engineering Management Technical Campus, Greater Noida, UP 201308, India*
[2]*Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India*
[3]*Department of ECE, Hindustan Institute of Technology and Science, Chennai, India*
[4]*Department of Computer and Self Development, Preparatory Year Deanship Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia*
[5]*Universitas Muhammadiyah Sumatera Barat, Padang, Indonesia*
[6]*Education Department, Government of Haryana, Chandigarh, Haryana, India*
[7]*Sunstone Eduversity, Kolkata, India*
[8]*University of Gezira, Wad Madani, Sudan*

Correspondence should be addressed to Hussien Sobahi Mohammed; hussiensobahi@uofg.edu.sd

Agriculture is crucial for the existence of humankind. Agriculture provides a significant portion of the income for many people all around the world. Additionally, it provides a large number of work possibilities for the general public. Numerous farmers desire for a return to the old-fashioned techniques of farming, which provides little profit in today's market. Long-term economic growth and prosperity are dependent on the success of agriculture and associated companies in the United States. Agribusiness crop yields may be increased by carefully selecting the right crops and putting in place supportive infrastructure. Weather, soil fertility, water availability, water quality, crop pricing, and other factors are taken into consideration while making agricultural predictions. Machine learning is critical in crop production prediction because it can anticipate crop output based on factors such as location, meteorological conditions, and season. It is advantageous for policymakers and farmers alike to be able to precisely estimate crop yields throughout the growing season since it allows them to anticipate market prices, plan import and export operations, and limit the social cost of crop losses. The use of this tool assists farmers in making informed decisions about which crops to grow on their land. In this study, a machine learning framework for agricultural yield prediction is presented. Crop information is collected in an experiment's data set. Then, feature selection is performed using the Relief algorithm. Features are extracted using the linear discriminant analysis algorithm. Machine learning predictors, namely, particle swarm optimization-support vector machine (PSO-SVM), K-nearest neighbor, and random forest, are used for classification.

1. Introduction

Among contemporary farmers, precision agriculture (PA) [1] is a well-known and improved approach of farm management. Crop health and output are monitored via the application of agricultural and information technologies in precision agriculture. PA aims to lower agricultural input costs while retaining the quality of the final output. Bulk applications of chemical fertilizers and pesticides have long been the norm, with the whole field being treated as a single unit.

The UN Council has advocated increasing the global supply of high-quality food as a means of achieving this goal. As a result, new approaches are needed to deal with the

problem. One method of dealing with the issue is to forecast human population and agricultural yields.

It is advantageous for policymakers and farmers alike to be able to precisely estimate crop yields throughout the growing season since it allows them to anticipate market prices, plan import and export operations, and limit the social cost of crop losses. In addition to large agricultural companies and smallholders, agricultural enterprises and smallholders profit from such predictions since they are able to make educated decisions about the management and financing of their crops [2]. Because of the complexities of the data, crop production forecasting is a difficult assignment for policymakers to do successfully. Agricultural researchers and agroeconomists are always on the hunt for new mathematical strategies that might increase prediction accuracy while still making use of existing factors. In this field of study, the goal is to demonstrate how crop yields are related to the location of agriculture, while also taking into account environmental elements such as soil quality and irrigation systems. These models are built on the foundation of rule-based models with parameters. A solid understanding of the many linkages that may be established between agricultural methods and environmental circumstances is possessed by the professionals working in the project.

There is a problem with trying to build an empirical expert system using knowledge that cannot be characterized in such a way. Manual surveys and remote sensing data are used to forecast crop yield [3]. Mathematical studies based on previous years' observations and historical information may be useful for a certain region or country but cannot be applied universally. These problems have been handled by recent crop simulation model developments. Models of soil characteristics, climatic conditions, and crop management practices are used to simulate crop growth throughout the growing season in crop simulation. These modeling approaches need a massive data set in order to accurately estimate agricultural output [4] over large areas. It is common for researchers to use remote-sensing devices like satellites, planes, or even a simple camera.

Math, information theory, statistics, artificial intelligence, etc., all play a role in the field of machine learning because it is an interdisciplinary study. The primary goal of machine learning research is to develop fast and effective algorithms that can predict data. In data analytics, machine learning is a technique used to build predictive models of the data collected. Reinforcement, unsupervised, and supervised learning are the three main categories of machine learning tasks.

It is possible for machines to learn behavior based on the input they receive from encounters with the outside world through reinforcement learning. Data sets that have not been labeled using traditional methods such as cluster analysis can benefit greatly from the use of unsupervised machine learning. There must be labeled data for supervised machine learning to work. Each set of labeled training data has an input value and a desired target output value. You can use this inferred function as a basis for mapping new values into the training data using a supervised learning technique.

Reinforcement learning is preferred for solving decision-making problems, while unsupervised and supervised learning are both preferred for analyzing data from a data processing perspective.

In a study by Jasti et al. [5], ICT activities provided the most current information, advanced technology, and knowledge for farmers to improve their livelihoods and increase their productivity. The use of cutting-edge communication channels such as radio, mobile, and television for farmers' development and information evaluation is the primary focus of ICT's relevance. According to Hemamalini eta l. [6], farmers and individuals who have an interest in farming can benefit greatly from the use of information technology in the agricultural industry. Technology-enabled decision-making, increased productivity, and real-time communication are all critical for farmers. This is made possible by mobile communication tools used to provide market information, better weather forecasts, cross-market coordination, and a better understanding of agricultural market prices.

Individuals, businesses, and governments all rely on these data-driven models to make predictions. The food sector is presently developing machine learning approaches that can handle the complexity and unpredictability of input [7].

This article describes a machine learning-enabled methodology for agricultural yield prediction that is accurate and early in the season. The first input data set, which contains all of the crop-related details, is gathered. The Relief algorithm is then used to choose the characteristics that will be used. It is possible to achieve accurate findings through feature selection by categorizing relevant features that are connected to a certain real-world situation. The LDA technique is then used to extract the features from the data. The classification process is then carried out using machine learning techniques such as PSO-SVM, KNN, and random forest.

2. Literature Survey

According to Mohamed et al. [8], several studies have examined the difficulties connected with the deployment and optimization of big data applications in various machine learning algorithms employed by cloud data centers or their networks. Big data applications and cloud services were developed using a MapReduce programming methodology with an open-source platform and Hadoop. A variety of innovative analyses and computations were made possible by combining Hadoop with MapReduce. Commodity clusters in geographically dispersed data centers are typically used by these services to deliver elastic and cost-effective solutions. However, as the number of people using data centers to move, store, and analyze big data has risen, this has created a variety of new problems that highlight the necessity of finding ways to lease resources more cheaply and efficiently. As a result, companies that have a large number of tenants requesting big data services have been challenged by the requirement to optimize the leasing of resources in order to minimize excess or under-utilization. To give a

comprehensive overview of cloud computing's architecture, a new summary of big data programming paradigms and their applications was selected for this study. Any software-defined networking technology supporting big data systems and virtualization was included in this. The topologies and routing protocols as well as the traffic characteristics were also briefly reviewed to underline the consequences of big data, such as supporting networks and cloud data centers. A number of initiatives were undertaken to improve the performance and energy efficiency of big data systems for a variety of applications and measures of performance. It was the goal of the survey to compile all the relevant research and classify them by the level of data center, network, and application. There were also suggestions for future research.

Due to a lack of natural resources, information, knowledge, and data must be utilized to the fullest extent possible. The conversion of solar energy into chemical energy occurs, for example, during the process of photosynthesis. It is the soil's ability to store and distribute critical nutrients that will allow plants to thrive, and this process will be accountable for all forms of life on the planet. Because overexposure can lead to soil degradation, it is essential to use a fertilizer to preserve the quality of the soil. This makes soil analysis an excellent way to gauge the condition of the soil. If there are minimal or disorganized data, a soil analysis can assist provide a report by examining the soil in several laboratories. In order to provide fertilizer recommendations based on the existing composition of soil nutrition, many methods of machine learning analysis are being applied in this study. The results of soil testing at Tata's soil and water testing center were used in this investigation. The Hadoop Distributed File System (HDFS) was utilized to analyze the stochastic gradient descent (SGD) algorithm and the artificial neural network (ANN). Random forest (RF), SVM using RBF, K-nearest neighbors (KNNs), support vector machine (SVM) utilizing polynomial function, and the regression tree (RT) were used to assess performance. Overall, the experimental analysis was performed correctly. Receiver operating characteristics with AUC (ROC) curve, coefficient of determination (R2), root mean square error (RMSE), and mean absolute percentage error (MAPE) measures of validation were used. SGD was found to outperform all other techniques in a study of diverse solution classes. The results also backed up the choice of the remedy and its recommendation.

For big data, a Hadoop ecosystem with a pig hive or a machine learning component was required, as stated by Jankatti et al. [9]. McKinsey estimates that there will be a shortage of 15 million big data specialists by 2020. Apache Kafka, Apache Spark, and Apache Hadoop were only few of the technologies available to handle big data processing and storage issues. Utilizing 4 GB of data on the cloud laboratory and Hadoop MapReduce with various mappers and reducers using a Pig Script and Hive Queries, the processing speed was evaluated. Using Hadoop MapReduce Pig with Hive or Spark with Hive, the findings showed that machine learning with Hadoop increased the performance of processing with Spark. Machine learning's best performance was improved, thanks to the Pig, Hive, and Hadoop MapReduce jar.

Further, Chlingaryan et al. [10] presented some current breakthroughs in machine learning research. Several advancements in machine learning have been made during the past 15 years that give a cost-effective solution to the problem. Some sensor platforms and machine learning approaches have specific uses that combined many modalities. For precision agriculture (PA), the development of hybrid systems was combined with signal processing and machine learning approaches.

Predicting the yield of corn before harvesting offers vital information, and this forecast is given to the public in order to increase pricing efficiency for resolving knowledge asymmetry issues. In order to forecast maize yields, used long short-term memory (LSTM), a specialized recurrent neural network (RNN) approach. An hourly weather data cross-sectional time series helps to create sample areas appropriate for deep learning algorithms. Predicting time series with county-level data collected in Iowa using LSTM was an effective method that showed potential predictive ability.

In Parsabad-Moghan, Iran, Omid et al. [11] used ANNs to predict corn and seed yields. By conducting face-to-face interviews with 144 farms in 2011, data were gathered. These corn seeds and their grain corns required energy ratios of 0.89 and 2.65, respectively. MLP ANNs with six neurons in their input layer ranged from one to three layers, each having a distinct set of neurons for each layer. Fertilizers, biocides, human labor, diesel fuel, and machinery were some of the energy outputs. 6-4-8-1 and 6-3-9-1 topologies were most suited to this model's prediction of seed yields and corn yields. In this case, the model output value had a coefficient of determination (R2) of roughly 0.9998 and 0.9978 for the seed and grain corn, respectively. The R2 values for all comparable models of this regression were between 0.987 and 0.982. The PCA was used to decrease the number of dimensions in the data set for inputs. Tea production in Iran may be predicted by looking at how energy moves across the country. The PCA and ANN were shown to be effective in determining the best flow of energy, according to the results.

Data flow can now be obtained from a variety of sensors and security cameras, thanks to technological advances in the fields of automation and processing. The data supplied must be able to analyze and extract patterns from massive data sets in order to extract values from enormous volumes. Using data from Taiwan's electronic highway toll collection, Fan et al. [12] developed a novel machine learning algorithm and integrated it into their big data analytics platform. Models for highway travel time were constructed using historical and real-time data, as well as some modified information on the amount of time it takes to get from A to B. When it comes to controlling food and agricultural growth, yield forecasting is absolutely crucial. Crop yield forecasting, biomass change detection, and crop evapotranspiration analysis are just a few of the nonlinear issues that have been tackled with ANN-based models. When using the Landsat 8 Operational Land Imager (OLI)-based satellite data to derive spectral indices from field inventory data, Khan et al. (2020) examined the ANN approach of the multilayer perceptron (MLP) for the estimation of Mentha

crop biomass. Using field-measured biomass, the association between biomasses ($R2 = 0.762$, RMSSE = 2.74 t/ha) was calculated.

Machine learning algorithms can be used to effectively estimate crop yields, taking environmental factors into account. Fegade and Pawar employed the support vector machine (SVM) and ANN classifiers [13]. The actual quantity of rainfall, the maximum and lowest temperature, the kind of soil, the pH value of the soil, and humidity were all taken into account while making the crop predictions. The agriculture website of Maharashtra was used to get all the information. A total of nine agricultural zones were assigned to the data. Another interface was built for farmers to input the information they needed to anticipate the crop. An impressive 86.80% of the predictions made by the neural network were spot on.

It was in Andhra Pradesh's Visakhapatnam area that Khosla et al. [14] largely concentrated on forecasting kharif crops. The modular NN (MNN) was used to predict monsoon rainfall, which turned out to be the key determinant of kharif crop yield. After that, support vector regression (SVR) was used to calculate the real yield of kharif crops based on rainfall data and the region. Crop yields were increased by using effective agricultural tactics based on the MNN-SVR approach. This approach was proven to be superior to other methods for predicting the output of these kharif crops when compared with other algorithms. It was the aim of Tiwari and Shukla [15] to estimate agricultural yields using a variety of geographical information. Standard precipitation indexes are one kind, while normalized difference vegetation indices are another. An error standard, known as the BP-NN, was utilized for the aim of learning from the previous weather circumstances. This was accomplished by conducting training in a manner in which all features were employed in tandem with their yield value as their output. The entire experiment was conducted using a geospatial data set from the Indian state of Madhya Pradesh in order to improve the work's trustworthiness. When compared with previous methods, the suggested model was able to overcome a wide range of evaluation factors.

In recent years, machine learning has emerged as a strong tool. As far as artificial neural networks (ANNs) go, they are one of the most widely used machine learning techniques. However, until recently, such catalytic applications were not thoroughly researched. Catalysis research in the literature has been summarized by Li and his colleagues [16]. Using this strong approach, individuals were able to deal with difficult issues and speed up the advancement of the catalysis community even further. The review demonstrates the usefulness of ANNs for predicting catalysis and building novel catalysts, as well as for comprehending their structures, by observing from both an experimental and theoretical standpoint. There are dendrites and neurons in the neural network, and it has shown to be a powerful statistical analysis tool. The neural network is a mathematical representation of biological neurons. The structure presented by Mamatha et al. [17] may be learnt as data sets that are properly analyzed on key parameters to get the desired result. There are two kinds of outputs: those that
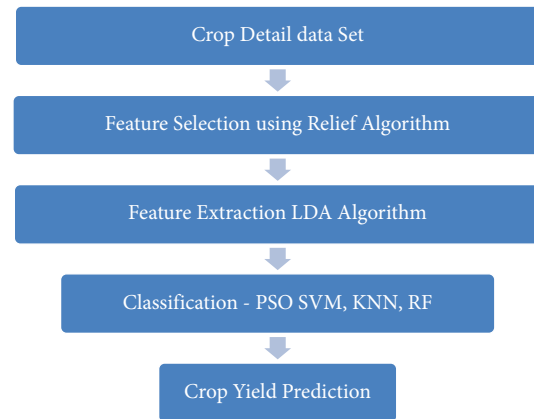

Figure 1: Machine learning-enabled framework for accurate crop yield prediction.

can be trusted and those that cannot. An Excel file or an Open Office document can be used to work with data sets. The MATLAB toolbox is used to train the neural network on the data. An additional assessment data set was used to determine whether or not the network has been trained. Trainers would be successful if their projected output matched their desired output, which this does.

3. Methodology

Crop yield prediction study necessitates a variety of production parameters and algorithms. Algorithms for finding the most predictive features are utilized by some, while others are used to find predictions. This section contains a machine learning-enabled framework for accurate and early crop yield production. First, the input data set containing all details related to crop is collected. Then, features are selected using the Relief algorithm. Feature selection helps in achieving accurate results by classifying important attributes related to a particular real-world problem. Then, features are extracted using the LDA algorithm. Then, machine learning algorithms, such as PSO-SVM, KNN, and random forest, are used for classification.

In 1992, Kira and Rendell [18] devised the Relief algorithm, an instance-based learning approach, to overcome binary classification challenges. Individuals may use a filtering technique to find feature-to-feature connections. When constructing feature statistics, nearest neighbors are used to take the interactions between variables that are in consideration. This approach, however, ignores any data with missing values or multiple classes.

Linear discriminant analysis (LDA) is used to decrease the number of dimensions in the original data matrix. PCA and LDA are also linear transformation techniques. In contrast to LDA, which is supervised, PCA is an unsupervised approach. While PCA finds the directions of highest variance, LDA aims to find a feature subspace that maximizes class reparability. By concentrating on class reparability, computational costs are minimized and overfitting is avoided [19]. The machine learning-enabled framework for accurate crop yield prediction is shown in Figure 1.

An easy approach to comprehend ML algorithms utilized in ML is through the KNN algorithm, which is the most commonly employed. It is a nonparametric supervised learning algorithm. The algorithm is also instance-based or "lazy" in its approach to learning. When a query is sent to the database, the algorithm consults the training data to determine how to respond. In comparison with other classifier algorithms, the training phase of KNN is quick because of this. It is slower in terms of time and memory; however, the testing phase gets faster. With KNNs, a new class of data points may be classified based on a data set with data points sorted into several categories. Each labeled data set consists of training observations, which implies that the algorithm creates the connection between x and y for each training data set (x, y). Processing is deferred until the KNN function has been identified. Neighborly contributions can be weighted in both classification and regression, making local neighbors more average than those further away. As the distance between neighbors increases, the 1/d weighting scheme is applied to each one [20].

Random forest is a popular strategy for creating predictive models, and it may be used to create predictive models. There are many uses of RF, including regression and classification. It is possible to find machine learning algorithms in this approach that can predict with high accuracy even when the data sets do not have appropriate parameter adjustment. When compared with other algorithms, this approach is incredibly user-friendly, as well as being extremely popular among people. RF stands for random forest, and as the name implies, this model creates random forests. When utilizing this strategy, a forest of decision trees is generated, each of which has been trained in a certain way. This method was used to generate the forest of all multiple choice trees that exists today. Consequently, they were integrated to provide even more exact projections [21] as a consequence of this.

With PSO-SVM (particle swarm optimization-support vector machine), nonprobabilistic binary linear classification is made simple. This method can be used to identify one or more target classes. Data are represented by a single dot (or point). It widens as a result of the distinct variations between the various groups. Based on where they lie in the gap, those new instances are assigned to distinct target classes. When the input data sets are unlabeled, nonlinear classification is conceivable. An unsupervised learning approach is used to categorize the data because there are no goal classes to assign to the instances. Once the function-based clusters have been established, further instances can be added. An effective recommendation system based on a nonlinear support vector machine has been presented. Most commonly, nonlinear support vector machine techniques are employed to handle unlabeled data [22].

4. Results Analysis

Data on 1200 crops examples have been gathered. For example, there are attributes like "year," "region," "crop," "season," "area," "production," "acres," "tonnes," "average temperature," "average rainfall," "solar radiation," "soil,"
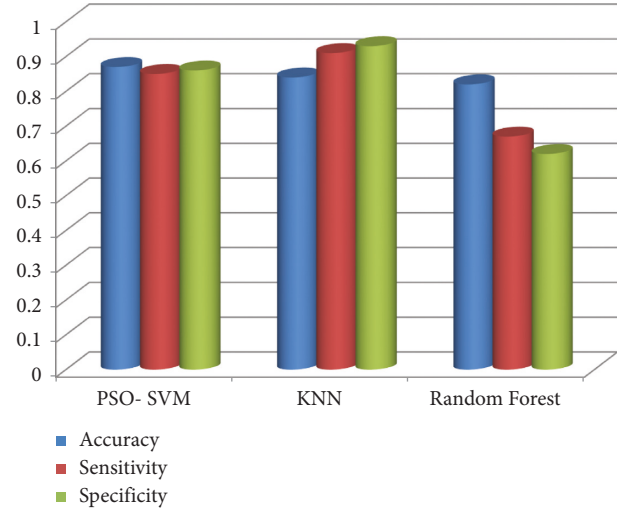

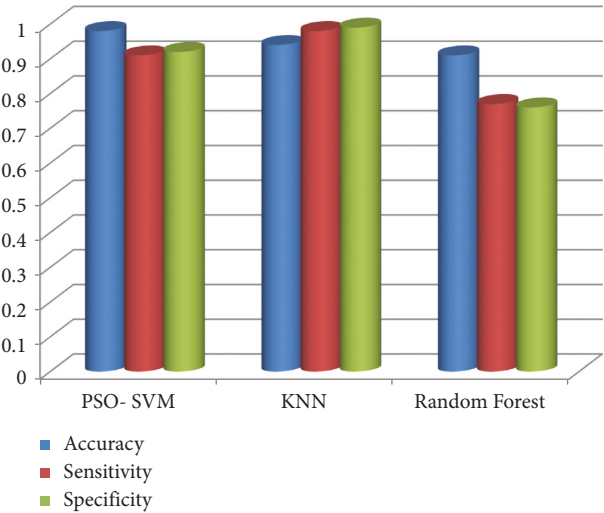FIGURE 2: Classifiers' performance without feature selection and feature extraction.

FIGURE 3: Classifiers' performance with feature selection and feature extraction.

"pH value," "soil type," and "major fertilisers" like "nitrogen," "phosphorus," and "potassium," as well as "minimum rainfall" and "minimum temperature," among others.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)},$$

$$\text{Sensitivity} = \frac{TP}{(TP + FN)}, \qquad (1)$$

$$\text{Specificity} = \frac{TN}{(TN + FP)},$$

where TP = true positive, TN = true negative, FP = false positive, and FN = false negative.

Figures 2 and 3 exhibit the accuracy, sensitivity, and specificity of PSO-SVM, K-nearest neighbor, and random forest for crop yield prediction with and without

preprocessing. It is clear that the feature selection and feature extraction has improved performance of machine learning algorithms. When compared with other classifiers, PSO-SVM has the best accuracy. The KNN method has a sensitivity and specificity superior to the rest of the classifiers.

5. Conclusion

Agribusiness crop yields may be increased by carefully selecting the right crops and putting in place supportive infrastructure. Weather, soil fertility, water availability, water quality, crop pricing, and other factors are taken into consideration while making agricultural predictions. Machine learning is critical in crop production prediction because it can anticipate crop output based on factors such as location, meteorological conditions, and season. The use of this tool assists farmers in making informed decisions about which crops to grow on their land. It is stated in this paper that a machine learning framework for agricultural yield prediction may be used. The data set for an experiment contains information about the crop. After that, the Relief algorithm is used to pick the features to be used. The PCA technique is used to extract the features. PSO-SVM, KNN, and random forest are examples of machine learning predictors that are used for classification. It is clear that the feature selection and feature extraction have improved the performance of machine learning algorithms. When compared with other classifiers, PSO-SVM has the best accuracy. The K-nearest neighbor method has a sensitivity and specificity superior to the rest of the classifiers.

Data Availability

The data shall be made available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] W. Haoxiang and S. Smys, "Big data analysis and perturbation using data mining algorithm," *Journal of Soft Computing Paradigm*, vol. 3, no. 1, pp. 19–28, 2021.

[2] M. Sivakami and P. Prabhu, "Classification of algorithms supported factual knowledge recovery from cardiac data set," *International Journal of Current Research and Review*, vol. 13, pp. 160–165, 2021.

[3] A. Bora, N. Vasantha Gowri, M. Naved, and P. S. Pandey, "An utilization of robot for irrigation using artificial intelligence," *International Journal of Future Generation Communication and Networking*, vol. 14, no. 1, 2021.

[4] A. Raghuvanshi, U. K. Singh, G. S. Sajja et al., "Intrusion detection using machine learning for risk mitigation in IoT-enabled Smart irrigation in Smart farming," *Journal of Food Quality*, vol. 2022, Article ID 3955514, 8 pages, 2022.

[5] V. D. P. Jasti, A. S. Zamani, K. Arumugam et al., "Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis," *Security and Communication Networks*, vol. 2022, Article ID 1918379, 7 pages, 2022.

[6] V. Hemamalini, S. Rajarajeswari, S. Nachiyappan et al., "Food quality inspection and grading using efficient image segmentation and machine learning-based system," *Journal of Food Quality*, vol. 2022, Article ID 5262294, 6 pages, 2022.

[7] A. Raghuvanshi, U. Singh, and C. Joshi, "A review of various security and privacy innovations for IoT applications in healthcare," *Advanced Healthcare Systems*, Wiley, Hoboken, NJ, USA, pp. 43–58, 2022.

[8] S. H. Mohamed, T. E. H. El-Gorashi, and J. M. Elmirghani, "A survey of big data machine learning applications optimization in cloud data centers and networks," 2019, https://arxiv.org/abs/1910.00731.

[9] S. Jankatti, B. K. Raghavendra, S. Raghavendra, and M. Meenakshi, "Performance evaluation of Map-reduce jar pig hive and spark with machine learning using big data," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 4, p. 3811, 2020.

[10] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review," *Computers and Electronics in Agriculture*, vol. 151, pp. 61–69, 2018.

[11] M. Omid, A. Farjam, A. Akram, and Z. F. Niari, "A neural network based modeling and sensitivity analysis of energy inputs for predicting seed and grain corn yields," *Journal of Agriculture, Science and Technology*, vol. 16, 2018.

[12] S. K. S. Fan, C. J. Su, H. T. Nien, P. F. Tsai, and C. Y. Cheng, "Using machine learning and big data approaches to predict travel time based on historical and real-time data from Taiwan electronic toll collection," *Soft Computing*, vol. 22, no. 17, pp. 5707–5718, 2018.

[13] T. K. Fegade and B. V. Pawar, "Crop prediction using artificial neural network and support vector machine," in *Data Management, Analytics and Innovation*, pp. 311–324, Springer, Berlin, Germany, 2020.

[14] E. Khosla, R. Dharavath, and R. Priya, "Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression," *Environment, Development and Sustainability*, vol. 22, pp. 1–22, 2019.

[15] P. Tiwari and P. K. Shukla, "A hybrid approach of TLBO and EBPNN for crop yield prediction using spatial feature vectors," *Journal of Artificial Intelligence*, vol. 1, no. 2, pp. 45–59, 2019.

[16] A. Li, S. Liang, A. Wang, and J. Qin, "Estimating crop yield from multi-temporal satellite data using multivariate regression and neural network techniques," *Photogrammetric Engineering and Remote Sensing*, vol. 73, no. 10, pp. 1149–1157, 2007.

[17] C. H. Mamatha, B. P. Reddy, R. Kumar, and S. Kumar, "Analysis of big data with neural network," *International Journal of Civil Engineering and Technology*, vol. 8, no. 12, pp. 211–215, 2017.

[18] K. Kira and L. A. Rendell, "Feature selection problem: traditional methods and a new algorithm," in *Proceedings of the Tenth National Conference on Artificial Intelligence*, pp. 129–134, San Jose, CA, USA, July 1992.

[19] Z. Lu and Z. Liang, "A complete subspace analysis of linear discriminant analysis and its robust implementation," *Journal of Electrical and Computer Engineering*, vol. 2016, Article ID 3919472, 10 pages, 2016.

[20] A. Gupta and L. K. Awasthi, *P4P: Ensuring Fault-Tolerance For Cycle-Stealing P2P Applications*, GCA, Ahmedabad, India, 2007.

[21] A. Gupta, "Performance insight 360: A cloud-based quality management framework for educational institutions in India," in *Proceedings of the IEEE 15th Conference on Business Informatics*, Vienna, Austria, July 2013.

[22] M. Rakhra, R. Singh, T. K. Lohani, and M. Shabaz, "Meta-heuristic and machine learning-based Smart engine for renting and sharing of agriculture equipment," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5561065, 13 pages, 2021.