Materials Today: Proceedings 51 (2022) 793–797

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Towards application of various machine learning techniques in agriculture

Santosh T. Jagtap [a,*], Khongdet Phasinam [b], Thanwamas Kassanuk [c], Subhesh Saurabh Jha [d], Tanmay Ghosh [e], Chetan M. Thakar [f]

[a] Department of Computer Science, Prof. Ramkrishna More College, Pradhikaran, Pune, Maharashtra, India
[b] Faculty of Food and Agricultural Technology, Pibulsongkram Rajabhat University, Phitsanulok, Thailand
[c] School of Agricultural and Food Engineering, Faculty of Food and Agricultural Technology, Pibulsongkram Rajabhat University, Phitsanulok, Thailand
[d] Department of Botany, Institute of Sciences, Banaras Hindu University, India
[e] Department of Microbiology, Dinabandhu Andrews College, Baishnabghata, South 24 Parganas, Kolkata – 700084, West Bengal, India
[f] Department of Mechanical Engineering, Government College of Engineering, Karad, Maharashtra, India

ARTICLE INFO

Article history:
Received 26 May 2021
Received in revised form 7 June 2021
Accepted 15 June 2021
Available online 26 June 2021

Keywords:
Precision Agriculture
Machine learning
Feature Extraction
ICT

ABSTRACT

Since the invention of the computer, all available information in every field has been digitized and made available to people who use computer resources. As a result, massive amounts of data are being generated in every domain at an alarming rate. Agriculture is one such area of interest for researchers. Machine learning is the process of extracting useful information from various types of data. The classification of objects is an important area within the field of data mining, and its application extends to a variety of areas, whether or not in the field of science. Although k-Nearest Neighbor classification is a simple and effective technique, it slows down the classification of each object. Furthermore, the classification's effectiveness suffers as a result of the uneven distribution of training data. The purpose of this paper is to look into the applicability of various machine learning techniques in agriculture.

Selection and peer-review under responsibility of the scientific committee of the 1st International Conference on Computations in Materials and Applied Engineering – 2021.

1. Introduction

Indian agriculture [1] is in the early stages of adopting Information and Communication Technology (ICT) techniques for farm management and output improvement. ICT has the potential to benefit all farmers, including small landholders, who are the most vulnerable to crop losses. Agriculture is the most important part of the Indian economy and is critical to India's growth, employing more than 40% of the Indian population (directly or indirectly). ICT techniques can help vulnerable farmers, particularly small stakeholders, take appropriate preventive/mitigating actions in the event of crop diseases, adverse weather, or even soil health.

Machine Learning (and its subset deep learning) and Artificial Intelligence [2] have contributed to an explosive increase in the application of computer science to previously thought-to-be-impossible complex science problems. The foundation of this thesis is machine learning, specifically Deep Learning and its subset, Convolutional Neural Networks (CNNs). Deep learning methods have been used successfully to solve problems that are simple for humans, such as game play or object recognition, but are difficult to describe mathematically or are computationally prohibitively expensive.

Image recognition, in particular, has undergone a paradigm shift, with use cases sprouting up all over the place. Machine Learning enables applications to predict outcomes with greater precision and accuracy without being explicitly programmed. It works by creating procedures that take input data and then predict an output based on statistical analysis of the data.

Machine learning [3] and data mining employ similar statistical analyses, both of which involve searching for patterns in data and updating outputs in response to new inputs. Furthermore, machine learning is powering virtual assistance technologies by combining several deep learning models to provide relevant context and to

* Corresponding author.
E-mail addresses: st.jagtap@gmail.com (S.T. Jagtap), phasinam@psru.ac.th (K. Phasinam), t.kassanuk@gmail.com (T. Kassanuk), subheshs.jha2@bhu.ac.in (S.S. Jha), tanmay.tanmay.ghosh780@gmail.com (T. Ghosh), cthakar12@gmail.com (C. M. Thakar).

https://doi.org/10.1016/j.matpr.2021.06.236
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the 1st International Conference on Computations in Materials and Applied Engineering – 2021.

interpret natural speech. Machine learning [4] is making it possible for us to live happier, healthier, and more productive lives.

Agriculture is critical to the country's economy since it feeds the whole population. In this way, it connects and interacts with all of the country's relevant companies. If a country has a reasonably big agricultural foundation, it is considered socially and economically wealthy. Agriculture is the principal source of employment in the majority of countries. Large farms frequently necessitate the hiring of additional personnel to assist with planting and farm animal care. The majority of these large farms have processing units nearby where their agricultural goods are finished and developed.

This article provides a comprehensive review of machine learning algorithms applicable in the area of agriculture. This will help future researchers to develop machine learning based solutions for agriculture sector to reduce agriculture waste, water irrigation etc.

2. Literature survey

This section contains literature review of various machine learning algorithms, which are applicable in agriculture domain for various tasks like- disease detection in crop, intelligent irrigation, and soil classification, monitoring and tracking.

2.1. Machine learning techniques

Machine learning (ML) is a new area of data mining that allows a computer program to become more accurate in predicting outcomes without being explicitly programmed. These ML algorithms are frequently classified as either supervised or unsupervised. For inference (classification, regression), supervised learning algorithms use labeled training data, whereas unsupervised learning algorithms use unlabeled data to discover hidden existing patterns (clustering).

Classification is the process of converting an input set of instances P into a unique set of attributes Q, also known as target attributes or labels. Various applications use classification techniques such as decision tree classifiers, bayesian classifiers, artificial neural networks, nearest neighbor classifiers, random forest, and support vector machines [5]. We'll talk about each of them briefly. Each technique operates on the basis of the learning algorithm it employs.

A decision tree is one of the most common and straightforward classifiers for solving classification problems. A decision tree is a graph in which instances are sorted based on their feature values to classify them. The decision tree is made up of nodes and branches, where each node represents a classification instance and each branch represents a value that the node can take on. In decision, instance classification begins at the root node, and instance sorting is based on their feature values.

In some applications, predicting the class label for a given set of input attributes is difficult. Furthermore, class variables are nondeterministic, even when using the given input attribute set values to match some of the attributes of the training data set. This is conceivable owing to the presence of some noisy data and puzzling aspects that are not taken into account during analysis. For example, predicting the likelihood of heart disease in a specific person based on the routine that person follows.

In this case, it is possible that most people who eat healthy foods and exercise regularly are at risk of developing heart disease due to other factors such as smoking, alcohol consumption, and possibly heredity. In such cases, the classification model is defined based on commonly known heart disease attributes, which cannot provide accurate information. There is a need to model probabilistic relationships between the attribute set and the class label in such applications, and the Bayesian classifier is all about justifying such tasks [6].

The concept of an artificial neural network (ANN) is inspired by biological neural networks, which are used to construct animal brains. Because it is made up of interconnected nodes and directed links, ANN is also known as a connectionist system. Each connected link is given a weight and is in charge of transmitting a signal from one node to another. When a node receives a signal, it processes it before transmitting it to another node.

The signal at the connection between artificial neurons in common ANN implementations is essentially a real number, and the output of each neuron is calculated by a non-linear function of the sum of all its inputs. The strength of the signal increases or decreases as learning progresses due to the weights of artificial neurons and the connections between them [7].

There are two strategies for making a model learned in ML classification. One of them is that as soon as the training set is available, the model begins learning; such models are known as eager learners. Another model observes all training examples but performs classification only if the attributes of the test instance exactly match any one of the training instances. Such students are referred to as lazy students [8].

Each example is treated as a data point in a d-dimensional space by the Nearest Neighbour (NN) classifier, where d is the number of attributes. The distance between the given test example and all data points in the training set is calculated. The k-Nearest Neighbors of data point X are the k points closest to the X.

The data point is then classified based on its neighbors' class labels. If a data point has more than one class labeled neighbor, the data point is assigned the class label with the greatest number of class labels. The value of k's nearest neighbors should be determined exactly. If the value of k is too small, it may misclassify due to the presence of noise in the training data. On the other hand, if the value of k is too large, there is a chance of misclassification because the set of nearest neighbors may contain data points that are located far away from the neighbourhood of the test attribute.

To begin, Random forest is a supervised machine learning algorithm that consists of a forest of decisions made by multiple decision trees generated using random vectors. This approach may be used to address classification difficulties as well as regression procedures. The result generated by the random forest is related to the number of trees it combines in the forest in such a way that as the number of trees in the forest increases, the possibility of obtaining greater accuracy increases. It is important to note that creating the forest is not the same as creating decision trees [8].

The primary distinction between decision trees and random forests is that in the case of random forest classification, finding the root node and splitting the feature nodes will occur at random. Because of its benefits, random forest classification is popular. One of them is that it can be used for classification as well as regression. Another advantage of this method is that if a sufficient number of trees are available, the problem of overfitting is avoided. In addition, a random forest classifier can handle missing values and can be modelled in the case of categorical values.

The random forest classifier has applications in medicine, banking, e-commerce, and the stock market. Random classifiers are used in banking to identify loyal and fraudulent customers. In medicine, Random Forest is used to identify the correct combination of medicines and to recognize the disease based on a patient's previous medical records. Random Forest classier is used in the stock market to observe a stock's behavior and then identify the loss and profit. In the context of e-commerce, Random Forest may be used to forecast customer product recommendations.

The supervised learning model used for classification is the Support Vector Machine (SVM). It has gotten a lot of attention in the classification field. In the SVM model, instances of the distinct cat-

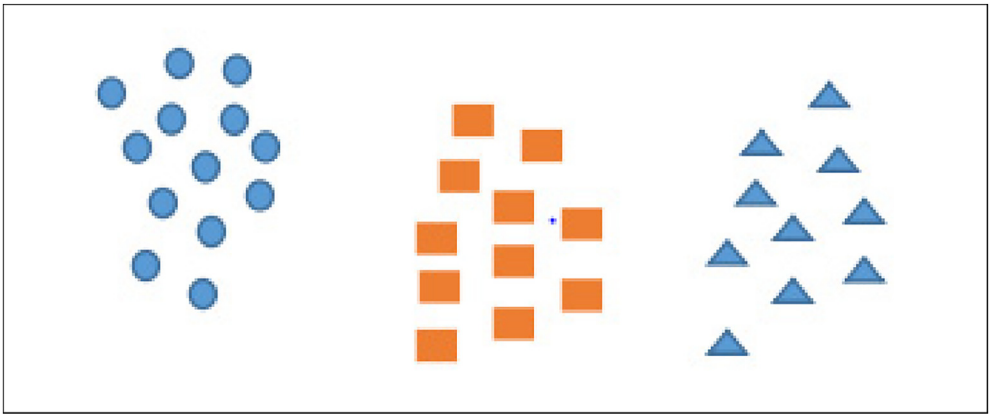

Fig. 1. Clustering.

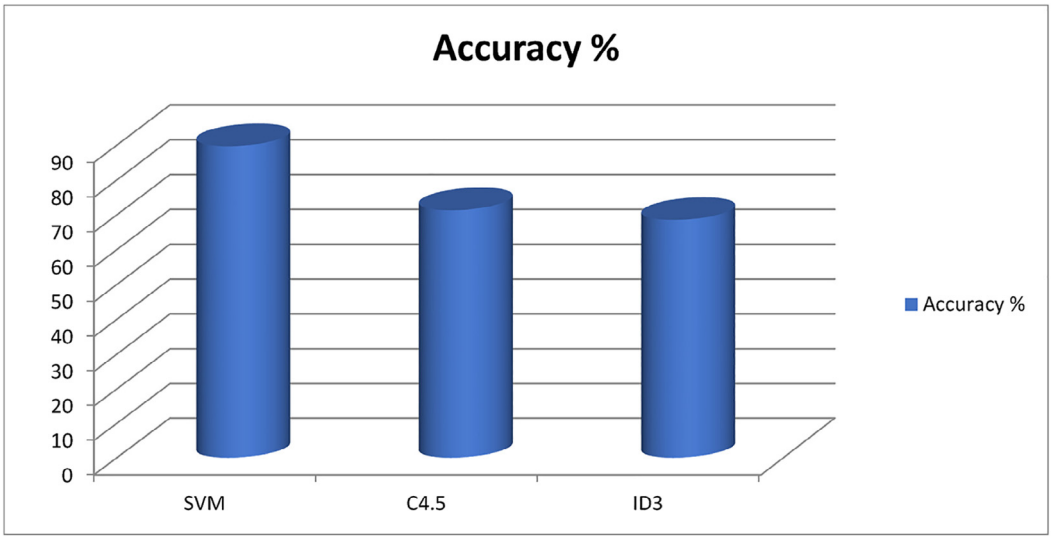

Fig. 2. Accuracy of Classifiers.

egories are separated in vector space by a visible gap. As soon as a new sample comes, it is mapped into the specific vector space, and its label is assigned to a category based on which side of the gap it falls [9]. Using the kernel trick, an SVM can perform non-linear classification efficiently.

Clustering, also known as cluster analysis, is the task of grouping a set of objects so that objects in one group are more similar to each other than objects in another, as illustrated in Fig. 1. The clustering would improve as the similarities between objects in one group and the dissimilarities between objects in different groups increased. Clustering is the foundation of data mining and can be applied in a variety of fields such as image processing, data compression, computer graphics, machine learning, and many others.

As shown in Fig. 1, clustering can be used in conjunction with other techniques for categorizing objects, such as classification, segmentation, and partitioning. When we compare Cluster analysis to classification, we can say that clustering is unsupervised learning. Cluster analysis differs from classification in that knowledge of classes is retained in classification, whereas knowledge of classes is not retained in clustering. Furthermore, in the case of classification, new samples are classified into known classes, whereas in the case of cluster analysis, groups are suggested based on data patterns [9].

Hierarchical clustering is popular for several reasons, including the following: 1) It does not require a specific value, as k-means clustering does. 2) The generated tree contains meaningful taxonomy. 3) To compute the hierarchical clustering, only the distance matrix is required. There are two types of algorithms available for hierarchical clustering: an agglomerative algorithm that uses a bottom-up approach, and a divisive algorithm that uses a top-down approach.

2.2. Feature Extraction

Decock et al. [10] used mutual similarity relations in the domain of discourse to represent linguistic hedges (e.g., cool, warm) as fuzzy rough approximations.

In determining existing practices for the management of soil nutrients, the extent of those practices and perception of soil fertility changes, Pilbeam et al. [11] used triangulated data. According to the survey, the use of farm manure and chemical fertilizer are the two most important soil fertility maintenance practices (mainly urea and di-ammonium phosphate).

The comparison of various sensory types and instruments, including field-based electronic sensors, spectral radiometers, vision machines, multispectral and hyperspectral remote sensing, satellite imaging, thermal imaging and olfactors systems, was carried out by Lee et al.[12]. Lee et al. The study also examined and discussed how precision farming and crop management, especially with regard to specialty crops, can take place in these sensing tech-

nologies. The work relating to that impact on aspects such as accuracy, reliability and time-consuming, has been the imprecise or ambiguous character of information in decision-making.

In order to determine the membership and non Member Function of FUZZY System Reliability, Garg [13] proposed the hybridized technique called Particle Swarm Optimisation Vague Optimisation. The system uses regular arithmetic operations rather than flimsy arithmetic operations and particular swarm optimisation to prevent uncertainty. The system builds membership functions.

As a common approach to the representation and rationalisation of uncertainty, Sturlaugson and Sheppard illustrated a Bayesian network. When the complexities of the network increase, Bayesian is used as an intractable determinant with a large number of nodes and states.

Mueller et al. [14] have developed studies on soil structure relevance and how soil quality overall is preserved. Their findings indicate that for visual soil structure assessment, soil with a concentration in clay greater than 30%, i.e. unfavorable soil structure, was not reliably detected.

Chu et al. [15] aimed for a method of solving problems with pattern recognition based on measures of similarity using intuitionist fuzzy sets from Atanassov. In addition to the [16,17,18,19] convenience of computing and ranking processes, a computer interface decision support system has been developed to help decision-makers more efficiently make diagnoses. This helped a lot in crop monitoring, tracking, disease detection and freshwater saving.

3. Result and discussion

A data set of 500 images of crops was created for experimental study. In experimental analysis, three classification algorithms namely, SVM- Support Vector Machine, C 4.5 and ID3 classifiers are used. These machine learning algorithms classified different crop images. This will help in identifying diseases prediction in crop. It will result in reduction in crop waste. It is shown in Fig. 2:

4. Conclusion

Agricultural research has benefited from technological advancements, particularly by incorporating industrial advances into a sustainable agriculture production system. By electrifying every farming procedure, technology has transformed farming into a viable business. This saves the farmer money and eliminates the middleman who buys low from farmers and sells high to end consumers. Recent applications of computational intelligence techniques (such as evolutionary algorithms, neural networks, and so on) provide solutions to site-specific decision modeling problems in agricultural systems.

Agriculture is important to the country's economy since it feeds the whole population. It links and interacts with all of the country's relevant enterprises in this way. A country is considered socially and economically prosperous if it has a sufficiently large agricultural basis. In the majority of countries, agriculture is the primary source of employment. Large farms usually require the hiring of extra workers to help with planting and farm animal care. The bulk of these huge farms have close processing plants where their agricultural products are processed and developed.

Machine learning's adaptability, promotion, and reduced costs helps in assessing the complicated link between the input and output of agricultural systems utilizing analytical approaches that are characterized by non-linearity, time variable features, and numerous unknown elements. This paper provides a review of various machine learning-based algorithms that can be used in agriculture for tasks such as crop disease detection, intelligent irrigation, soil classification, monitoring, and tracking.

CRediT authorship contribution statement

Santosh T. Jagtap: Conceptualization, Methodology. **Khongdet Phasinam:** Data curation. **Thanwamas Kassanuk:** Visualization, Investigation. **Subhesh Saurabh Jha:** Writing - review & editing. **Tanmay Ghosh:** Validation. **Chetan M. Thakar:** Writing - original draft, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] G. Gyarmati, T. Mizik, The present and future of the precision agriculture, in: 2020 IEEE 15th International Conference of System of Systems Engineering (SoSE), 2020, pp. 593–596. doi:10.1109/SoSE50414.2020.9130481.
[2] R. Katarya, A. Raturi, A. Mehndiratta, A. Thapper, Impact of Machine Learning Techniques in Precision Agriculture, in: 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), 2020, pp. 1–6. doi:10.1109/ICETCE48199.2020.9091741.
[3] A. Sharma, A. Jain, P. Gupta, V. Chowdary, Machine learning applications for precision agriculture: a comprehensive review, IEEE Access 9 (2021) 4843–4873, https://doi.org/10.1109/ACCESS.2020.3048415.
[4] M.M. Anghelof, G. Suciu, R. Craciunescu, C. Marghescu, Intelligent System for Precision Agriculture, in: 2020 13th International Conference on Communications (COMM), 2020, pp. 407–410. doi:10.1109/COMM48946.2020.9141981.
[5] I. Khan, X. Zhang, M. Rehman, R. Ali, A literature survey and empirical study of meta-learning for classifier selection, IEEE Access 8 (2020) 10262–10281, https://doi.org/10.1109/ACCESS.2020.2964726.
[6] S.P. Samuel, K. Malarvizhi, S. Karthik, S.G.M. Gowri, Machine Learning and Internet of Things based Smart Agriculture, in: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, pp. 1101–1106. doi:10.1109/ICACCS48705.2020.9074472.
[7] B. Sharma, J.K.P.S. Yadav, S. Yadav, Predict Crop Production in India Using Machine Learning Technique: A Survey, in: 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020, pp. 993–997. doi:10.1109/ICRITO48877.2020.9197953.
[8] A. Muniasamy, Machine Learning for Smart Farming: A Focus on Desert Agriculture, in: 2020 International Conference on Computing and Information Technology (ICCIT-1441), 2020, pp. 1–5, doi: 10.1109/ICCIT-144147971.2020.9213759.
[9] K.N. Bhanu, H.J. Jasmine, H.S. Mahadevaswamy, Machine learning Implementation in IoT based Intelligent System for Agriculture, in: 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1–5. doi: 10.1109/INCET49848.2020.9153978.
[10] M. Decock, A.M. Radzikowska, E.E. Kerre, A fuzzy-rough approach to the representation of linguistic hedges, in: Technologies for Constructing Intelligent Systems, Springer, Berlin, 2002, pp. 33–42.
[11] C.J. Pilbeam, S.B. Mathema, P.J. Gregory, P.B. Shakya, Soil fertility management in the Mid-Hills of Nepal: practices and perceptions, Agric. Hum. Values 22 (2) (2005) 243–258.
[12] W.S. Lee, V. Alchanatis, C. Yang, M. Hirafuji, D. Moshou, C. Li, Sensing technologies for precision specialty crop production, Comput. Electron. Agric. 74 (1) (2010) 2–33.
[13] H. Garg, An approach for analyzing fuzzy system reliability using particle swarm optimization and Intuitionistic fuzzy set theory, Multiple-Valued Logic and Soft Computing 21 (3) (2013) 335–354.
[14] L.E. Sturlaugson, J.W. Sheppard, Principal Component Analysis PreProcessing with Bayesian Networks for Battery Capacity Estimation. Conference in Instrumentation and Measurement Technology, IEEE, Minneapolis, 2013, pp. 98–101.
[15] C.H. Chu, K.C. Hung, P. Julian, Complete pattern recognition approach under Atanassov's Intuitionistic Fuzzy Sets, Knowl.-Based Syst. 66 (2014) 36–45.
[16] J. Gholap, A. Lngole, J. Gohil, Shailesh, V. Attar, Soil data analysis using classification techniques and soil attribute prediction, Int. J. Comput. Sci. 9 (3) (2012) 14.
[17] S. Ghosh, S. Koley, Machine learning for soil fertility and plant nutrient management using back propagation neural networks, Int. J. Recent Innov. Trends Comput. Commun. 2 (2) (2014) 292–297.

[18] S.S. Dahikar, V.S. Rode, Agricultural crop yield prediction using artificial neural network approach, Int. J. Innov. Res. Elect. Electron. Instrum. Control Eng. 2 (1) (2014) 683–686.

[19] M. Kaur, H. Gulati, H. Kundra, Data mining in agriculture on crop price prediction: techniques and applications, Int. J. Comput. Appl. 99 (12) (2014) 975 8887.