Chapter 12
Application of Machine Learning Algorithms in Agriculture: An Analysis

Kalpana Jain and Naveen Choudhary

Abstract Machine learning is being rapidly adopted in various industries: According to Research and Markets, the machine learning market is projected to grow to $8.81 billion by 2022, at a compound annual growth rate of 44.1%. One of the main reasons for its increasing use is that companies are collecting big data from which they need to obtain valuable information. Machine learning is an efficient way to make sense of that data. In the current situation, we are talking about the emerging concept of smart farming that makes farming more efficient and effective with the help of high-precision algorithms. The mechanism that drives it is machine learning, the scientific field that gives machines the ability to learn without being strictly scheduled. It has emerged alongside big data technologies and high-performance computing to create new opportunities to unravel, quantify, and understand data-intensive processes in agricultural operational environments. This paper reviews the exiting techniques and methods of machine learning applicable in the agriculture sector.

Keywords Agriculture · Machine learning · Training · Data · Algorithms

1 Introduction

Agriculture is essential to the world's sustainability. Humans profit in one direction or the other from agriculture, which makes agricultural practices a crucial field of research. Farmers often need details, particularly when growing crops that are not popular in their land or crop. The ordinary farmer has links to crude data outlets such as television, radio and journals, fellow growers, government departments, farmers, and merchants. Therefore, a framework is required, which enables farmers to access relevant details. Machine learning is one of the trends; thus, different strategies and applications function within the context of machine learning. In recent years, many machine learning systems have been validated and established in agriculture.

K. Jain (✉) · N. Choudhary
Department of Computer Science and Engineering, College of Technology & Engineering, MPUAT, Udaipur, India

R. Mathur et al. (eds.), *Emerging Trends in Data Driven Computing and Communications*, Studies in Autonomic, Data-driven and Industrial Computing,
https://doi.org/10.1007/978-981-16-3915-9_12

Research has also been carried out on the efficacy of multiple machine learning algorithms in agriculture and other applications since machine learning is an extremely effective method for efficient utilization of tools, forecasts, and management that is required for agriculture. Machine learning is the skill and the implementation of information through an electrical processing device.

The planet has to be more conscious of sustainable agriculture for some time now. According to the UN, by 2050, the population of the planet could reach 9.7 billion. Experts from the World Resource Institute predict that food demand could grow from 50 to 70% to feed these people. Agriculture can, however, have disastrous environmental consequences because of its wastewater use, transport of carbon dioxide, and the improper use of fertilizers that cause pollution. India has a long history as one of the world's biggest farmers. The nation has a comprehensive network of commercial farmers and small-scale growers; it is one of the major rice, cotton, sugar cane producer and exports large amounts of wheat, maize and produce other crops. India is therefore the ideal testing ground to transform its farming sector using revolutionary new technology. Intelligent farming supports farmers (large and small) for the value chain by the provision of materials, consultancy services and storage services, transport, and international exchange aid [3].

Now several companies are developing a software solution that guarantees sustainable agriculture both for the world and for farmers. They plan to create an open digital network for farmers and others to help grow more at less cost and with less environmental effects. This new technology uses innovative machine learning software, geospatial data mining, and cloud computing to provide farmers with advice and feedback in real time [33]. The challenge was to design a framework that could process data in real time from many sources, achieve visibility in all farming phases, and automatically provide farmers with recommendations. With satellite and drone geo-spatial data, the company may track ground conditions, which imply possible field productivity [19]. This data can be mixed with other outlets, including weather and market data, in order to consider farmers' real-time conditions and to make intelligent, real-time recommendations. The machine also gives a simpler way of identifying and avoiding accidents such as plague or water shortage and helps to avoid crop destruction. As the industry expands and evolves, different forms of machine learning are developing that can be explored in new applications. Many examples of machine learning implementations today, therefore, fall into two categories: supervised and unregulated learning.

1.1 Supervised Learning

A common method of machine learning is supervised learning that is often used to build training patterns to predict future happenings such as fraudulent credit card purchases in applications that use historical data. It is a method of machine learning that recognizes inputs and outcomes and uses tagged examples to train algorithms.

Supervised learning uses techniques for pattern analysis such as grouping, regression, estimation, and gradient change. These trends are then used to predict mark values of unlabeled results [10]. In drug research and development, this form of machine learning has already been used, with applications including target validation, biomarker identification, and automatic clinical pathology analysis. This means that machine learning promotes data-driven decision-making and facilitates the discovery and development process and improves success rates [12].

1.2 Unsupervised Learning

Unsupervised learning works without previous validation for data sets compared with controlled learning. It instead analyzes the collected data to describe the framework and trends. Unregulated machine learning is being used in factories for predictive maintenance. Machines can analyze and use data and algorithms that cause device failures to anticipate problems before they occur. This allows less needless downtime as factories order parts from the retailer of automation equipment to be replaced before a failure takes place. Research by Deloitte found that the use of machine learning technologies [18] in development decreased the unexpected downtimes by 15–30% and hence the costs of maintenance by 30%. Humans are not the only people able to think about themselves: machines like Google Duplex can also pass the Turing test today. Manufacturers may use machine learning to boost maintenance processes and make informed decisions in real time based on findings.

1.3 Problem Statement

It is well known that the prevention and timely diagnosis of any disease will bring us the strategic advantage over said disease and in agriculture it is no exception, because knowing what ails a crop or plant increases the chances of success in treatment [24]. In the developing world, more than 80% of agricultural production is generated by small farmers [1], and reports of yield losses of more than 50% due to pests and diseases are common [9]. Besides, the greater proportion of people with problems of poverty and famine (50%) live in these productive areas, which makes the small farmers particularly vulnerable group to disruptions in the food supply caused by pathogens. There are methods to determine the diseases of any plant, such as taking samples of vegetative tissue to a specialized laboratory or taking an expert agronomist to the cultivation site; in either of the two methods, the disadvantage lies in the time necessary to obtain the results. That is why they have been considered the use of artificial vision techniques and pattern recognition, as well as some classification algorithms that automatically determine the possible disease, facilitating the task of specialists to develop their work and that they can find a timely diagnosis for treatment. And as [7] says, the Tools for the automatic recognition of plant diseases have the potential to become a valuable source of information to aid decision-making in agriculture.

In the next section, the techniques used are shown, as well as the most significant research works that address the challenge of detecting a disease, based on the analysis of the characteristics present in the leaves of the crops.

2 Techniques and Algorithms

To arrive at the classification of diseases, the authors rely on proven methodologies, like data acquisition, pre-processing, feature extraction, and recognition, which are the steps or procedures followed to obtain results. For the acquisition of the images, digital cameras are used to capture the leaves or the parts where the damage caused by the disease is visible, as well as set so that the images available on the Web are made available to everyone. Public serves as the basis for training the model. For the images acquired by the camera, the shots were in controlled environments with acceptable resolutions, although it is important to mention that [27] used a mobile phone, to acquire the images and do an alternate experiment to see how much it affected the quality of the images.

Once the images are acquired, they go to pre-processing, where they will be treated such as scaling, noise elimination, color space transformation, histogram equalization, and everything else that can be done to maximize the characteristics. By the time you move on to the next process, the image is processing and noise removed. Therefore, when applying segmentation techniques, they will separate the points of interest with better precision, obtaining valuable data that will be more descriptive of the disease. Once you have the characteristics, then next thing is to do the classification, make use of the algorithms, obtain the results, and make a description of them. It is in this step when the effectiveness of the procedure is shown, if it has been classified according to expectations.

The next section shows the significant works related to the detection of diseases in various types of plants , using various machine learning algorithms as listed in Table 1.

3 Literature Review

Using different machine learning algorithms such as mathematical and statistical approaches, crop prediction can be carried out. Any of the approaches that are currently being tested are discussed here.

Table 1 Machine Learning Algorithms and their characteristics

Algorithm	Properties
Fuzzy logic	It is based on heuristic rules; it is used for processes highly not linear [15]. Easy implementation
SVM	Look for a hyperplane that works as a separator. In typical training and problems, it is very efficient [32]
Bayes	It is very efficient where this type of environment is used supervised learning. Large amounts of data are not required for their training [4]
KNN	Search for observations closest to the one you are trying to predict and classify the point of interest based on most training data [34]
ANN	Qualifying is very efficient, at the cost of training computationally expensive [37]
CNN	The performance of convolutional neural networks in the recognition object and image classification has made great progress in recent years. They tend to be more accurate at the cost of high computational cost. In training, it may require a considerable number of images to produce reliable results [5]

3.1 Artificial Neural Network

The Network of Artificial Neurons is the Artificial Neural Network (ANN). It is founded upon the biochemical functions of the human brain. This is one of the markers of controlled learning. Neural network has to be trained once, for example, after equivalent trends can be expected in future data, practical solutions to problems can be generated even though the input data is incorrect/incomplete. The accuracy of ANN continues to improve by including more and more details. ANNs are often willing to embrace their ambiguity without understanding the values behind them. For any method, ANN may extract the correlation between input and output, and compare the SVM and ANN algorithms to classify diseases in various crops that have been attacked by fungi, bacteria, nematodes, and nutrient deficiencies. They mention that the symptoms of plant diseases exhibit different properties such as color, shape, and texture, and based on this, the characteristics are obtained. They consider color as an important dimension, but applying dimensionality reduction they discover by experimentation that, out of 24, only 8 characteristics are significant for the classification of diseases. In the end, they put the two algorithms to the test and found that with SVM they obtained 92.17% precision and that with ANN only 87.48%, which is why they show that, for this case, the SVM is a better classifier. Another paper was presented by [19]. In Ireland, over 80% of farmland is grassland, a source of feed for the dairy and livestock industries. There have been very few studies worldwide that estimate small weeds like Ireland using remote sensing data. Certain computational models have been developed to estimate the amount of grassland biomass available in two intensively controlled grassland farming companies in Ireland such as the multilinear regression (MLR), the artificial neural network (ANN), and the adaptive neuro-fuzzy inference system (ANFIS). For 12 years (2001–2012), in situ weekly biomass measurements were used for model creation on the first test site (Moorepark) (2001–2012). In all three versions, 5 vegetation indices plus two spectral (red and close infrarouge) bands extracted from an 8-day MODIS product were fed. ANFIS has provided better biomass figures (RM2 Moorpark < 0.85, RMSE Moorepark = 11.07; RG range 2 > = 0.76, RMse grange = 15.35) than ANN and MLR. This project will provide a blueprint for the analysis and measurement of spatial data for the recovery of various biophysical variables. Another paper introducing the method [25] network integration is very useful for many vector-based algorithms and is now a hot subject of study in network analysis. A lot of studies have been conducted on network integration, but much of the material that is already present in the network was overlooked. In this article, we suggest a new method of network embedding using the semi-supervised SSKNMF kernel, from which previous knowledge can be implemented and more practical network features learned. By using the objective function based on the L2,1 standard, it can become more resilient against noise. These evolutionary principles are very useful in managing the following agent-based paradigm. Our proposed algorithm performs with thorough testing considerably better than current members. The author of [7] recommends that a deep-rooted network architecture be extended through the use of VGGNet [17] and AlexNetOWTBn [9] to automatically identify diseases arising in tomato leaves. Early plague, powdery mildew, and mildew were the pathogens to be classified. The photographs captured are preprocessed using techniques of image processing such as noise reduction, regression, and the improvement of the image processing to minimize expense and time. Later on, the data set characteristics were derived using convolution charts, in which the input data were imaged by healthy and contaminated leaves. While the architecture showed very accurate results; for this analysis, only AlexNetOWTBn architecture was used to achieve 32.23% and VGG 33.27%.

3.2 Information on Fuzzy Network

The remote sensing and other parameters for crop yield prediction were evaluated through aggressive neural network processing. They employed the versatile method of Neuro-fuzzy Inference (ANFIS). The inputs to ANFIS are the soil moisture content, field biomass, and the repository organ. It has only a single number, or another output node, i.e. yield, that is pursued. In that remote sensing data, the other problem in forecasting yield does not go far behind in time. In order to construct a design to estimate future values, every forecasting attempt is therefore compelled to add a very limited number of previous years. By leaving one year out and utilizing all the other material, the agreement is disciplined. In relation to the return of the year that is left out, they measure the variance of our calculation. The system is used for all years and the average efficacy of prediction has been given. A paper presented by [17] seeking a suitable motor imaging machine is a huge task. For positive discrimination, the collection of discriminatory characteristics is important. This research will include a controlled method to classify biased traits in the EEG signal MI classification. The function selection approach eliminates dimensionality. Any EMG measure is mapped to a continuum. The spatial characteristics are derived from each sub-band. A high-dimensional function vector is paired with several sub-bands. In order to allow an effective definition, the neighborhood component analysis-based filtering approach is used to select appropriate features. Least Squares is a supervised mathematical learning technique used in categorized regression data to improve accuracy. The chosen features are used to identify the SVM. In order to reduce the function factor, the irrelevant features are discarded. Evaluation of the strategy proposed is performed with the data sets 4a and IV 2b of the BCI Competition III. Both data sets can be used as benchmark data types to help test BCI machine learning. Simulations demonstrate the supremacy of the time series prediction system proposed a study that works to predict disease [36]. All the major factors influencing agricultural development are agricultural diseases and insect pests. Early detection and pestilence prediction can mitigate economic harm from plagues. This paper uses a neural network to automatically classify crop diseases. The statistics are taken from the 2018 public data sets of the AI competitor with 30 illnesses of 10 crops. In this post, we use the training model Inception-ResNet-v2. These are approaches widely used in prediction networks dependent on convolution. After the integration is complete, the ReLU function is enabled. The findings reveal that the average outcome rate of the model is 86.1%. After being educated in this approach, our students developed and introduced a smartphone application for the iPhone. And we carried out an actual examination to validate the theory. The findings revealed that the device would correctly predict crop diseases. Another concept is given by [6] that used a system of fuzzy logic and decision trees and with the help of a human expert, they were able to make a recognition of coffee diseases, where they obtained the characteristics of the symptoms that occur in the plant and thus be able to do your decision tree. The results they obtained are within 85% accuracy. Although this research does not deal with artificial vision techniques, it can be noted that a very important step has been taken for the detection of diseases, using expert systems applying decision trees with fuzzy logic.

3.3 Support Vector Machine

Cluster analysis or clustering is a method in which artifacts that are identical but distinct from individuals in other classes are defined. It is used mostly for data processing. In several areas, clustering is used, such as computer education, pattern detection, image processing, knowledge collection, and agriculture. There are different clustering algorithms like k-means and k-medoid set, but the popular and important clustering SVM is the algorithm. A paper presented the concept of the cluster [23]. The objective of this paper was to develop an effective method for harvesting automated cucumber harvesting. The proposed algorithm comprises several processing and data mining techniques aimed at classifying the image of carrots properly. Computer Vision project uses an SVM, a Euclidean distance transform, a bag-of-visual-words (BoW) classifier, and a watershed transform in order to classify images into seven broad categories. Several experiments were carried out to generate the data sets which will be used for training and validation of the classifiers. Detection was tested at both levels of pixel and cucumber by evaluating its result to the ground truth data. The high percentage accuracy at the pixel level and at the tissue level proves that the proposed algorithm is highly reliable in cucumber harvesting applications. Comparative analysis is done by [7]. According to the Food and Agriculture Organization, world production of date fruits is forecast at about 8.5 million tons and 1.3 million tons in 2018 in Saudi Arabia. The most popular varieties of dates are Badri, Khalas, Muneer, NabootSaif, and Suleh in Saudi Arabia. The co-dependencies in maturity are immature, Khalal, Khalal, Rutab, Pre-Tamar, and Tamar. The way fruit is processed has a big influence on earnings. This paper proposes a smart harvest method by means of computer vision and DL to decide the form, maturity, and weight of dates. The method contains three sub-systems: maturity estimate, form estimate, and date weight estimate (DWES). We used four architectures of Deep Neuro Networks, ResNet, VGG-19, Inception-V3, and NASNet for DWES as well as the SVM (regression and linear) support for DMES. We also focused on the Smart Robotics Research Center to test the device proposed. DTES attained a median performance of 99.175%, a 99.225% F1 score, average accuracy of 99.8%, and an average recall of 99.05% using further performance metrics. The DMES has achieved a high-quality precision, which averages 99,058, 99,34, 99,64, and 99,08%. DWES reached a maximum score of 84.27% with SVM Linear.

3.4 Convocational Neural Network

A CNN network was developed to determine the impact of climate change on potato development. They have also seen a network of beliefs integrating the volatility of potential climate change, taking into account the fluctuations of current weather parameters including temperature, radiation, rainfall, and potato production. They claimed their network would help politics Agricultural makers. They validate their model with synthetic weather conditions and equate the findings with the traditional mathematical model and conclude that the performance of the faith network is greater. A paper was presented by [17]. AI has been recently extended to a variety of sensing activities to anticipate, monitor, and/or understand. However, it is limited to embedded devices. We propose a low-power sensor with AI onboard to improve the application in agriculture. For this reason, we used a CNN to come up with a system which achieves 83% of average IoU and 97% of seeds recognition accuracy on the test set. The proposed solution would conduct seed identification and seed germination detection via spectral domain image processing. For the training of CNN, we collect data of the image of the seed germination process over the span of many stages. The whole framework would be measured by industrial regulations. The studies have demonstrated that the device can open up a massive amount of possibilities for smart applications. [21] write another paper**:** The argument over finding an optimal sample set for signals on graphs is an excellent one. The proposed sighter collection is based on a localization operator that uses both vertex and spectral domain positions. We show the relationships between the new procedure, selecting sensors in machine learning methods, and using graph frequency methods to select sensors. The method suggested does not have to measure the vectors of the variable operator while also considering (graph) information. By measuring performance metrics such as quality and memory, we evaluate the effectiveness of our approach. Another definition is the same [2]. Accurate data will greatly reinforce spatial growth and sustainable natural resource and farm management in Singapore. The research focuses on the detection of greenhouse photographs from SPOT-7, and MSI photographs from Sentinel-2. In chosen test areas, multiple classifiers are employed, such as k-nearest neighbor (KNN), random forest (RF), and support vector machine (SVM). The image is created by multi-resolution scanning. Secondly, for each image unit, spectral, texture, and remote sensing indications were obtained. The greenhouses were classified by classifiers. The classification accuracy of this task was evaluated using an ambiguity matrix. The study of the KNN and RF classification revealed a slightly higher mean accuracy (OA) and Kappa (β) of 0.69 and 0.85. The image of Sentinel-2 was rated according to KNN techniques. Another article has been published by [1]. Recently, deep learning methods are becoming extremely relevant to solve image processing problems. Among various algorithms, Convolutional Neural Network (CNN)–based and Recurrent Neural Network (RNN) based systems have achieved state-of-the-art results on satellite and aerial imagery in several applications. While these methods are part of research are for some, there is currently no functional or generic version at the consumer level for remote sensing culture. In this letter, we introduce a method to use distance learning strategies with remote sensing imagery and geographic details. This project is focused on the libraries Orfeo toolbox for image processing and TensorFlow for numerical computation. It is capable of processing a large amount of information without any limitation on image size or hardware configuration. Another paper is written by [3]. There is also a need for more refinement and specificity in the study of human behaviors. Deep education will soon have a significant effect as a study is carried out on security applications in the area of elderly tracking and the detection of individuals and objects left in the public room. Wearable devices for human activity have been developed but these technologies can cause users unwanted physical and mental distress. Researchers have concentrated on using image-based HAR and have placed it at the forefront of electronic consumer design. This essay explores an intelligent method to classify human behaviors through image analysis and profound learning methods. A skeleton-based approach can also produce detailed findings in a number of conditions and domain structures. This paper explores the creation of an efficient HAR embedded knowledge approach. Two public databases of people's routine behaviors, such as experimental outcomes, are used for tests to render the proposed method more reliable and successful on both data sets than most equivalent schemes [13]. As robotics and computer technologies evolve, the agriculture industry is increasingly evolving. Farmers shift to agriculture in which selective productivity is significant. This is called agricultural precision. To prevent weed growth in crops, it is important to correctly identify seed rows. We are proposing a new approach called CRowNet which uses CNN and Hough to detect crop rows in a drone image (UAV). It consists of a "SegNet" model made up of "SegNet" and a

"CNN-based Hough transform" (HoughCNet). The approach suggested could yield better results than traditional techniques. An accuracy rate of 93.58% was obtained for the identification of crop rows with IoUs over 70%. Furthermore, the trained model can detect crops on a given field of crops.

3.5 Deep Learning

Deep learning is an artificial intelligence (AI) feature that imitates the operation of the human brain in data processing and generates templates for decision-making use. Deep learning is a branch of artificial intelligence machine learning, which has networks that can learn unattended from unstructured or unscheduled results, often named profound neural learning or deep neural network. A concept is given by [14]. The Internet of Things would revolutionize health systems, agriculture, banking, electricity, and transportation. With the incorporation of software-defined networks (SDNs) and network feature virtualization (NFV) in the edge-cloud interplay, IoT platforms are significantly improving with developments. Deep learning has been getting its existence more influential due to enormous evidence of Internet of Things (IoT). DL algorithms may have privacy issues when applying to confidential data. These algorithms depend on conventional server-based approaches, that is, high server processing capacity is needed. We would like to suggest a new learning method by differentially private training algorithm called LATENT. The functionality of LATENT allows a data owner to introduce randomization until the data are exposed to possibly untrusted machine learning service. It divides the neural design into three layers: (1) convolutional module; (2) randomization module; and (3) completely linked module. Therefore, the Randomization module will effectively perform a privacy guard service for SDN-enabled NFV. There would be substantial variation in government policy-making owing to this aspect. Our empirical review on classification using convolutional deep neural networks indicates fine results in accuracy (e.g., 92–96%). Another concept is presented by [20] where they conducted 60 experiments using deep convolutional neural networks to identify 14 crop species and 26 diseases, using trained models such as AlexNet [16] and GoogleNet. They used the approach of [12] that demonstrates for the first time that end-to-end supervised training with a CNN architecture is a potential option for a large number of classes, going beyond the traditional approach of using hand-built features. Within the Plant Village data set convolutional with RelU activation function, each followed Max Pooling layer and the second part contain two dense layers to focus color and gray scale. In the end, they showed that working with color characteristics gives them 99.84% better results compared to grayscale with 95.54%. Another article [22] shows a reinforcement learning (RL) method of maximizing charging strategies in a public electric vehicle (EV) charging station. "Online" in the sense that this price mechanism takes choices on a real-time basis, and "model-free" in the sense that it does not use simulations of unknown occurrences. A challenge with RL is we need to optimize the total charging rates to satisfy customers' demands before departure time.

This feature-based method can further improve the performance of the proposed algorithm. Compared to representative benchmark algorithms; our algorithm achieved the highest charging-station profit. This paper presents [31] a novel Computational Intelligence vision sensing approach to analyze the content of nutrients in plant leaves. We propose developing the Deep Sparse Extreme Learning Machines (DSELM) framework for fusion and GA for image normalization as well as to reduce variation of images due to differing degrees of sunlight. Of course, we do relevant working image segmentation with DSELM. In this paper, four moments of color distribution of the leaves images (mean, variance, skewness, and kurtosis) are used as predictors in the estimation of nutrient profiles. Our machine learning model consists of a variety of DSELMs and combines them to predict nitrogen content in wheat leaves. Studies also show that the proposed approach outperforms other approaches in terms of efficiency and processing time. Another method is proposed by [24] that proposes algorithms to identify multiple plant diseases, based on color analysis and using a pair-wise classification algorithm. According to them, their methodology allows them to operate in uncontrolled conditions and thus be able to cover a large number of diseases. This method was tested with a large, unrestricted set of leaf images containing symptoms belonging to 74 diseases, 4 pests, and 4 abiotic disorders, affecting 12 different plant species. The results obtained were between 40 and 80% precision. [29] used an ANN, KNN, Naive Bayes, a hybrid of self-organizing maps (SOM), and radial base function (RBF); all this in order to determine the best algorithm to classify diseases of rust, coffee wilt (CWD), and the CFD that affects the coffee fruit. In their work, they state that they obtained 58.16% for KNN, for NaiveBayes they obtained 53.47%, for ANN 79.04%, and for the combination of RBF and SOM they obtain 90.07% accuracy, which shows a great improvement with respect to the previous algorithms, although they note that the latter takes them longer in training. [26] tell us that they found a feasible solution for the diagnosis and identification of four alfalfa diseases. They extracted 129 characteristics of texture, color, and shape from the 1651 images using the Relif F, 1R, and CFS methods. To classify the diseases, they used SVM, KNN, and Random Forest. The best classifier was SVM and the Relif method for obtaining the characteristic, since they achieved 97.64% in precision for the training set and 94.74% for the test set. [28], experiments have been done to recognize diseases or burns presented in the leaves. The crops examined were bananas, rice, citrus fruit, and roses. They suggest segmentation using genetic algorithms after analyzing the photos and clustering. They used the color rivalry technique to remove attributes and they think it easier to use the color picture than the standard grayscale. To identify them, they used MDCs achieving K-Mean 86.54%, MDCs with an algorithm they proposed, enhanced 93.63% and SVMs with a proposed algorithm, which received substantial improvements of 95.71%. Many of these percentages represent the total average of 54.306 photos from the four crops of 58 groups with 14 crop types and 26 diseases (or their absence), as shown by the highest precision of 99.35%. This target was accomplished. Therefore, without any feature engineering, the model correctly classifies crops and diseases from 38 possible classes in 993 out of 1000 images. The authors noted that training requires a lot of computational work but for classification it is less than a second; they believe it could be implemented on a cell phone [30].

Propose a web tool, where the farmers who grow the pomegranate will upload an image of the fruit to be analyzed, with the trained model to check whether or not the fruit is infected. The technique used is based on extracting the characteristics of the pomegranate images, such as color, morphology, and the color coherence vector, to later perform the classification using the SVM algorithm, giving them an accuracy of 85% with a 10-megapixel camera. Knowing that farmers will not always have the optimal capture means to upload the images to the system, the authors carried out significant tests with the cameras of mobile devices with resolutions of 5 and 3 megapixels, and obtained results of 82 and 79%, respectively. Tang et al. [30] carried out a study that consisted of analyzing 9,000 images of tomato leaves, to produce a model that could be used on smartphones, with the purpose of identifying 5 types of diseases. Your model would be based on a deep convolutional network, but it would be made up of two parts, the first part of the model (the extraction functions) was the same for full-color focus and grayscale focus, consisting of 4 layers. It is used by [34] and has been working with convolutional neural networks of various crops, looking for disease levels in plants, being able to classify healthy crops with 89% accuracy, slightly ill with 31%, moderately ill 87% and seriously ill with 94%. Although the results are satisfactory, he concludes that the classification of plant diseases from digital images is very difficult. On the other hand, the limitations of the data set in terms of quantity and variety of samples continue to prevent the emergence of truly comprehensive systems to perform disease classification functions [35]. Many deep learning models have recently been used to identify various forms of plant diseases, but relatively little work has been done in predicting whether or not these illnesses are dangerous. In addition, it is necessary to thoroughly monitor the severity of plant diseases because it facilitates plant protection decisions. We create a data set of images with the aid of Plant Village and crowd AI of contaminated citrus fruit leaves. Six separate algorithms for the magnitude of citrus HLB were then tested. The experimental findings show that, when dealing with severity detection, the model Inception v3 with epochs = 60 will obtain a higher precision rate of 74.38% than other models. We also decided to explore whether GAN data increase would help substantially enhance the learning output of the model. A new set of 14,056 leaves was eventually used in the V3 network preparation. We have provided more than 85% precision and very high performance.

Table 2 shows the summary of the results obtained by the exposed investigations and the comparative analysis of crop production prediction accuracy with different machine learning classification methods. It shows that crop prediction classification depends on the method, crop and it varies on different parameters, result shows.

4 Discussion and Results

One of the advantages that we can observe when [6] used the fuzzy logic classifier algorithm is how easy and fast it could be implemented, since it does not require a large collection of data for its training, and it presented more or less favorable

Table 2 Summary of the algorithms and their accuracy results

References	Culture	Method	Accuracy in percentage
[4]	Coffee	Fuzzy logic	85.00
[5]	Pomegranate	SVM	82.00
[6]	Coffee	KNN ANN N. BAYES RBF and SOM	58.16 79.04 53.47 90.07
[7]	Banana, Bean, Lemon, Rose	MDC + K-Means MDC + SVM	86.54 93.63
[8]	Bean Fruit Cassava Citrus Fruit Coconut tree Coffee Corn Cotton Grape Passion fruit Soy Wheat Sugar Cane Passion fruit Soy Wheat Sugar Cane	Proposal Classification by pairs	50.00 46.00 56.00 71.00 53.00 40.00 76.00 58.00 56.00 58.00 59.00 70.00
[9]	Alfalfa	SVM	94.74
[10]	Wheat, sunflower, grape, corn, cucumber, cotton, cabbage, tomato	ANN SVM	87.48 92.17
[11]	Strawberry	Fuzzy logic	97.00
[12]	Various	CNN	99.35
[13]	Tomato	CNN	99.84
[14]	Bean Cassava Naranjo Coconut Corn Coffee Cotton Cashew Grape Kale Passion Fruit Soybean Sugar cane Wheat		95.00 83.00 62.00 97.00 66.00 77.00 100 61.00 82.00
[15]	Tomato	CNN	33.27

results when wanting to classify. Now, if the target class is decreased, then it would have better results, as shown by the work of [26], which got better accuracy just by predicting just one disease. The proposal made by [37], by using RBF and SOM, seems quite good, since they are the ones that gave them the best results, compared to KNN, ANN, and Bayes, and this is because their approach to obtain the characteristics by their texture and color made it have better results, since the choice of the latter is the one that works best for these cases.

From the SVM algorithms, it can be concluded that it is another good alternative for solving these cases, although at first, they were not intended to solve this type of problem; with the passage of time, they have been adapted and presenting very good results, such as those of [3, 4].

When the classification is for few diseases, it could be said that it is relatively easy, but when you really want to cover more diseases or more crops, that is when the complexity increases, and a proposed solution is the one presented [24]. Classification by pairs it is based on the theory where plants present similar symptoms when attacked by the same disease and that the algorithm can be retrained for new diseases. Another proposal for the same plants and the same diseases was presented by [5] when using convolutional neural networks; he declares that it is the best way to do this type of study, and that the only limitation today is the availability of large amounts of data to do a good training of the net.

5 Conclusion

Nowadays, an increasing range of computer application Agricultural learning strategies are required for which a lot of accessible data from several resources can be examined to find the secret information. This is an advanced area of study and is anticipated to develop in the near future. The convergence of computer science with agriculture allows predicting crops. It must draw on objective methods for forecasting pre-harvest crops. The creation of an effective model would have some merits over the conventional approach of forecasting as it could be seen in this review of articles, the works that obtained the best results are convolutional neural networks; their use is increasingly biased for this type of problem, since diagnoses are obtained closer to what a human expert would determine. The only problem would be that the training is too computationally expensive and requires a lot of data to do it. If the study data were not enough, the recommendation would be decision trees with fuzzy logic, since it is the algorithm with the best results after convolutional networks. Although it is not very precise, we can trust that as technologies advance; more optimal values can be reached at a lower computational cost. Another important point that we can note is that with the use of new technologies such as machine learning and pattern recognition, diseases in crops can be detected and a timely diagnosis made, reducing the risk of agricultural and economic losses, which would bring a direct benefit to the farmers who implemented it.

References

1. Ai Y, Sun C, Tie J, Cai X (2020) Research on recognition model of crop diseases and insect pests based on deep learning in harsh environments. IEEE Access 14
2. Anand P, Singh Y, Selwal A, Alazab M, Tanwar S, Kumar N (2020) IoT Vulnerability assessment for sustainable computing: threats, current solutions, and open challenges. IEEE Access 8:168825–168853
3. Arachchige PCM, Bertok P, Khalil I, Liu D, Camtepe S, Atiquzzaman M (2020) A trustworthy privacy preserving framework for machine learning in industrial IoT systems. IEEE Trans Industr Inf 16(9):6092–6102
4. Chaudhury A (2009a) Machine vision system for 3D plant phenotyping. In: IEEE/ACM transactions on computational biology and bioinformatics, vol 16, No 6
5. Chen J, Huang YY, Li YS, Chang CY, Huang YM (2020) An AIoT based smart agricultural system for pests detection. IEEE Access 8:180750–180761
6. Du G, Wang Z, Li Z (2020) A new cloud robots training method using cooperative learning. IEEE Access 8:20838–20848
7. Elavarasan D, Vincent PMD (2020) Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. IEEE Access 8:86886–86901
8. Faisal M, Alsulaiman M, Arafah M, Mekhtiche MA (2020) IHDS: intelligent harvesting decision system for date fruit based on maturity stage using deep learning and computer vision. IEEE Access 8:167985–167997
9. Feng S, Zhao J, Liu T, Zhang H, Zhang Z, Guo X (2019) Crop type identification and mapping using machine learning algorithms and Sentinel-2 time series data. IEEE J Sel Top Appl Earth Obser Remote Sens 12(9):3295–3306
10. Fleming SW, Good body AG (2019) A machine learning metasystem for robust probabilistic nonlinear regression-based forecasting of seasonal water availability in the US West. IEEE Access 7:119943–119964
11. He Q, Zhang Y, Tang S, Liu H, Liu, (2019) Network embedding using semi-supervised kernel nonnegative matrix factorization. IEEE Access 7:92732–92744
12. Horng G-J, Liu M-X, Chen C-C (2020) The smart image recognition mechanism for crop harvesting system in intelligent agriculture. IEEE Sens J 20(5):2766–2781
13. Jin (2020) 'Clustering Life course to understand the heterogeneous effects of life events, gender, and generation on habitual travel modes. IEEE Access 8
14. Josef S, Degani A (2020) Deep reinforcement learning for safe local planning of a ground vehicle in unknown Rough Terrain. IEEE Robot Autom Lett 5(4):6748–6755
15. Khan TM, Robles-Kelly A (2020) Machine learning: quantum versus classical. IEEE Access 8:219275–219294
16. Lee W, Ham Y, Ban T-W, Jo O (2019) Analysis of growth performance in swine based on machine learning. IEEE Access 7:161716–161724
17. Lee W, Kim M, Cho D (2019) Deep learning based transmit power control in underlaid device-to-device communication. IEEE Syst J 13(3):2551–2554
18. LeVoir SJ, Farley PA, Sun T, andXu, C. (2020) High-Accuracy adaptive low-cost location sensing subsystems for autonomous rover in precision agriculture. IEEE Open J Ind Appl 1:74–94
19. Li W, Ni L, Li Z-L, Duan S-B, Wu H (2019) Evaluation of machine learning algorithms in spatial downscaling of MODISL and surface temperature. IEEE J Sel Top Appl Earth Obser Remote Sens 12(7):2299–2307
20. Meng B (2020) Modeling alpine grassl and above ground biomass based on remote sensing data and machine learning algorithm: a case study in east of the Tibetan Plateau, China. IEEE J Sel Top Appl Earth Obser Remote Sens 13:986–2995
21. Minh DL, Sadeghi-Niaraki A, Huy HD, Min K, Moon H (2018) Deep learningapproachforshort-termstocktrendspredictionbasedontwo-streamgated recurrent unit network. IEEE Access 6:55392–55404

22. Molla MKI, Shiam AA, Islam MR, Tanaka T (2020) Discriminative feature selection-based motor imagery classification using EEG signal. IEEE Access 8:98255–98265
23. Phyo CN, Zin TT, Tin P (2019) Deep learning for recognizing human activities using motions of skeletal joints. IEEE Trans Consum Electron 65(2):243–252
24. Ren A (2020) Machine learning driven approach towards the quality assessment of fresh fruits using non-invasive sensing. IEEE Sens J 20(4):2075–2083
25. Sakiyama A, Tanaka Y, Tanaka T, Ortega A (2019) Eigen decomposition-free sampling set selection for graph signals. IEEE Trans Signal Process 67(10):2679–2692
26. Shafi U (2020) AMulti-Modal approach for crop health mapping using low altitude remote sensing, internet of things (IoT) and machine learning. IEEE Access 8:112708–112724
27. Sharma A, Jain A, Gupta P, Chowdary V (2021) Machine learning applications for precision agriculture: a comprehensive review. IEEE Access 9:4843–4873
28. Sott MK, Furstenau LB, Kipper LM, Giraldo FD, Lopez-Robles JR, Cobo MJ, Zahid A, Abbasi QH, Imran MA (2020) Precision techniques and agriculture 4.0 technologies to promote sustainability in the coffee sector: state of the art, challenges and future trends. IEEE Access 8:149854–149867
29. Sun L (2020) Application of machine learning to stomatology: a comprehensive review. IEEE Access 8:184360–184374
30. Tang Z, Wang H, Li X, Li X, Cai W, Han C (2020) An object-based approach for mapping crop coverage using multiscale weighted and machine learning methods. IEEE J Sel Top Appl Earth Obser Remote Sens 13:1700–1713
31. Wang S, Bi S, Zhang YA (2021) Reinforcement learning for real-time pricing and scheduling control in EV charging stations. IEEE Trans Industr Inf 17:849–859
32. Wu T, Luo J, Dong W, Sun Y, Xia L, Zhang X (2019) Geo-Object-Based soil organic matter mapping using machine learning algorithms with multi-source geo-spatial data. IEEE J Sel Top Appl Earth Observ Remote Sens 12(4):1091–1106
33. Yu J (2020) A deep learning approach for multi-depth soil water content prediction in summer maize growth period. IEEE Access 8
34. Z (2020) Smart farming becomes even smarter with deep learning-a bibliographical analysis. IEEE Access 8:105587–105609
35. Zeng Q, Ma X, Cheng B, Zhou E, Pang W (2020) GANs-Based data augmentation for citrus disease severity detection using deep learning. IEEE Access 8:172882–172891
36. Zerrouki N, Harrou F, Sun Y, Hocini L (2019) A machine learning-based approach for land cover change detection using remote sensing and radiometric measurements. IEEE Sens J 19(14):5843–5850
37. Zhu J, Zhao X, Li H, Chen H, Wu G (2018) An effective machine learning approach for identifying the glyphosate poisoning status in rats using blood routine test. IEEE Access 6:15653–15662