Agriculture Automation Using Machine Learning: A Review

Randeep Kaur[1], Harmeet Singh[2]
[1]*Research Scholar, CSE Department*
CT University, Ludhiana, India
[2]*Assistant Professor, CSE Department*
CT University, Ludhiana, India

***Abstract*—Machine learning is a part of data science and software computing technologies that give us new opportunities for processing a large amount of data in various existing areas. Agriculture automation is one of the major parts of concerns for everyone related to this field. Because there exist many areas where many crucial problems related to agriculture arise. The problems are like plant diseases, storage management, control of pesticides, weed management, and water management. All these problems can be tackled by algorithms of computer vision. In this paper, we tried to represent research done by various experts related to agriculture automation using various algorithms of machine learning or image processing. This paper also gives a brief review of various algorithms that come under different categories in machine learning.**

Keywords: Machine Learning, CNN, Decision Tree, Automated Agriculture, Regression Techniques

I. Introduction

India is the sixth-largest developing country on the earth. Agriculture is the root of the providence of India. In India, most of the human resource is taking benefits through agriculture field. Agriculture has a remarkable impact on Indian GDP. Agriculture is rapidly adopting software computing techniques in field farming. Machine learning is becoming the most unrestrained technology in agriculture services because it can learn, understand, and give results to different problems in different factors. Farming is degrading due to a lack of knowledge for doing difficult decisions based on variables that are interconnected with each other. There exist various factors related to agriculture which include specifications of soil and its conditions, change of climate, and many more. So, Machine learning can be merged with other technologies like image processing to create a platform for smart agriculture and to help farmers with quick decision-making in agriculture environments.

II. Machine Learning

Machine learning is a subbranch of computer vision. It is a technique which makes the machines to work as a human without giving a command or programmed by a user. Self-driving cars, speech recognition, image recognition, and understanding of the human genome are some examples of applications made with help of machine learning techniques. Machine Learning is going to be used in various applications with increasing popularity. It can be used in various types of evaluative applications like Knowledge discovery in databases, computer-aided learning, voice recognition, and programmable object interfaces. Machine learning provides perspective solutions in all areas. There exist subcategories of Machine learning i.e., supervised, and unsupervised algorithms, and then further into the types of solutions they are trying to attain.

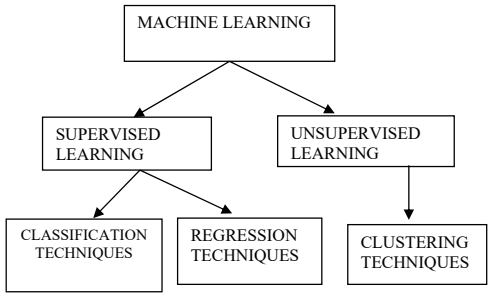

Fig. 1

Machine learning is everywhere throughout the whole expanding and gathering cycle. Software computing technologies do not take decisions based on historical data but use artificial intelligence combined with examination of weather to meet the aggressive demand of agriculture. Machine learning has a fundamental impact on the agriculture field. There exist various applications in agriculture where machine learning can be applied for farmers.

III. Types of Machine Learning Algorithms

A. Supervised Learning

Supervised learning is a subcategory of machine learning. It is a task of learning a function that maps the input of labeled datasets to the output to train the system or any model for the detection and classification of any

ISBN: 978-93-91355-11-1

image or video. It also detects an error while finding the expected output.

B. Unsupervised Learning

Unsupervised learning is also a subcategory of machine learning. The algorithms related to this category do not need any users. They learned on their own by learning patterns from unlabelled data and give results based on that pattern. In this type of learning a large amount of data is divided into different clusters.

C. Reinforcement Learning

Reinforcement learning is a part of machine learning-based in which an agent does work to get maximum reward. In this type of learning the model keeps on learning in the whole process.

IV. Different Categories of Algorithms

A. Regression Algorithms

Regression means analyzing the particular problem which prognostic. These types of algorithms use the relation between various types of variables. Linea, Logistic, Stepwise, and Ordinary Least Squares are some of the examples of regression algorithms.

B. Memory-Based Algorithms

A memory-based learning model develops an exact definition of target function by storing samples of training data. The solution to the problem can be determined by predicting the target function value with the help of stored instances. Some of the memory-based algorithms are Self-Organising Map, Locally Weighted Learning, K-Nearest Neighbour, Learning Vector Quantisation, etc.

C. Decision Tree Algorithms

A decision tree algorithm works exactly like a tree architecture which consists of optimized results of various problems. The solutions of the algorithm are based on certain restraints. It begins with a node called origin and then further part into several nodes making subtree until a prediction is made thus forming a tree-like architecture. This algorithm helps to formalize the problem. This technique will give potential results more accurately and makes the system faster and efficient. These types of algorithms are used to solve classification and regression problems.

D. Bayesian Algorithms

Machine learning algorithms that use Bayes' theorem to classify the part of the image are called Bayesian algorithms and examples of such algorithms are independent Bayes algorithm, Gaussian Naive Bayes, decision network, etc.

E. Support Vector Machine

Support Vector Machine is the most popular technique of machine learning. It takes a decision plane to create a decision boundary line between a cluster of data values recognized with different tags. It is a subpart of a supervised algorithm for classification. It can perform both linear and nonlinear classification which depends upon the system on which it is being applied.

F. Clustering Algorithms

Clustering algorithms use a fixed pattern of datasets to classify the data and labeling the data according to that pattern. K-medoids, K-Medians, EDA, Ward's minimum variance method, Agglomerative nesting, Gaussian Mixtures, are various types of clustering algorithms that can be used in agriculture.

G. Artificial Neural Network (ANN) Algorithms

An artificial neural network is a type of algorithm in machine learning which creates a model that is based on the function of actual neutronics of humans. These are considered as irregular models because it attempts to find out cumbersome relations between input and output data. It takes the data which is sampled from the entire set and thus helps in increasing accuracy and efficiency. Some of the examples of algorithms that come under this category are feed-forward artificial networks, recurrent networks, Radial Basis Function Network, etc.

V. Applications in Agriculture

A. Species Identification

Even professional farmers suffer from difficulty in recognizing two plants with similarity in some cases where there are only little variations in color or shape of part of plants. Machine learning helps in increasing the accuracy and fastness of recognition of species with the help of time, resources, and cost of farmers can be saved. Various types of machine learning algorithms can help to find out a particular vegetable type by training the model with the image of a vegetable. Species of plant identification is a classification problem that can be detected by supervised machine learning techniques. Many researchers have worked to find out the solutions to this problem by creating different algorithms. For the identification of species, the first step is to analyze the images of various plants. In the application phase, the model which has been trained is used to access new images and detect the species in a particular image and give accurate results. Images contain countless pixels with different colors and values which are interrelated to each other. The complexity of these images can be reduced by extracting the features. Many researchers have focused on various types of feature extraction

ISBN: 978-93-91355-11-1

methods which can help to decrease the complexity of several images and can make the classification process fast. They used generic algorithms to find out different interesting points and their description which can help to detect the species easily.

Wu et al. (2018) developed a technique in which they used the image thresholding method to detect the features like separate background, noise, and contour of the image. They extracted rectilinear derivations from the 12 leaf shape features of species. The approach was applied on 32 species and give accurate results with 90% value. The drawback of this technique is that it can only deal with species which are having different leaf shapes.

Jin et al. (2018) proposed an approach in which features of leaf teeth are extracted. In this method firstly, the binarization method is used to detect the features after that image is segmented to have the interesting point and then the contour detection approach is applied to the segmented image. The proposed method attained a classification rate of 76% for the eight species but it cannot be applied to species having significant appearances of leaf teeth.

Seeland et al. applied a combination of SURF detector and SIFT local shape descriptor for developing a classification model for the identification of plant species. It was concluded that this combination is superior to other detector–descriptor combinations. They have applied the Fisher Kernel encoding which proves to be a higher-ranking image descriptor for classification and had proven to be superior.

B. Selective Breeding

Reproduction for the particular features in plants is a very complex and time-dependent process. Many researchers are exploring machine learning applications to make this task easy and accurate manner to achieved resourceful information. Researchers have already developed many models that can properly pick out which plant category would be beneficiary under specific conditions of the environment. This task is possible only by sustaining data sets to machine learning algorithms. These findings can give new opportunities to the farmers in the agricultural sector as a whole. With these farmers can easily identify which quality of which plant would result in the most beneficial productivity of plants under particular environmental conditions.

Jansen et al. (2012), designed a platform that performs a controlled environment. They worked on two separate platforms. They took images of a single plant as a dataset. In the first method, they transfer individual plants themselves to an enclosed box where screening of plants done with help of multiple sensors. In the second method, sensors are to be placed near individual plants to take measurements. Their results showed that the second method gives good results and is mostly used for the applications of fluorescence imaging.

An et al. (2016) developed a model to calculate the quality of plants like length and area of the leaf. They took 1000 images of rosette plants by using cameras which are spread evenly on the top of fields. This system can help farmers to do easily plant phenotyping.

Minarine et al. (2017) presented a model which can be used for the segmentation of plant image automatically. They used different algorithms of machine learning for accuracy. They used an image analysis pipeline for the segmentation which is applied to the images of soybean plants. They took images of soybean at early growth stages. They used a 3-d point cloud for this purpose and took 2D images using an image-based high-throughput phenotyping platform in a greenhouse.

Jianfeng Zhou et al. (2018) developed a platform to evaluate powerful responses of soybean. For phenotyping, they estimated leaf chlorophyll and plant height. A single camera was used to take 400 images of soybeans by moving the camera in all directions. They used the structure from motion method for stitching the sequential images of soybean plants. They conclude that images of the larger number of plants can be achieved in a lesser amount of time by using this platform.

Jing Zhou et.al (2019) evaluated the performance of machine learning methods like Support Vector Machine and K-means clustering which can be used to segment the image. They used 75 images of soybean plants at early growth stages as a dataset. The dataset was collected using an image-based high-throughput phenotyping platform. They used the Structure from Motion (SfM) method for developing a 3D point cloud. They used Histogram of Oriented Gradient descriptor for the separation of overlapped plants. At last, a comparison was done between manual and automated segmentation and their result showed that K-means clustering is good for the background removal and the separation of overlapped plants, but SVM and HOG descriptors give better results with minimal error rates.

C. Agrochemical Production

Agrochemical products are very important in eco-revolution. Various synthetic products like defoliant, weedkiller, and toxicants have been used by farmers to resist the damage caused by unwanted insects and bacteria. But these chemicals have a mischievous effect on the environment and health of human beings and animals. Machine learning can help farmers with this problem. Many researchers have worked out this application for developing a model for the detection of weed and which will give accurate results.

Haug et al., (2014) developed a model for weed detection and removal. They used SVM for the

classification and also used shallow neural networks. They used Scale-Invariant Feature Transform for the feature extraction of plants. But there exists the main degraded feature of these weed recognition models is that they do not give results with noisy images. So, till now there exist several challenges such as point of view, flake and noncomparable variations, distortion of the image, closure of image, brightness conditions, and disordered background associated with image classification.

Lecun et al., (2015). used Convolutional Neural Networks for the detection of weeds in plants. This kind of technique is used for complex problems. It gives efficient results with a high rate of discrimination in results for the correct identification of plants to make the agricultural practices automatic. They conclude that many more architectures of CNN can be used to make the automatic systems for the identification of weeds in the field which can be used by farmers to make their work easy.

McCool et al., (2017) developed a convolution neural network model by compression of the fine-tuned network using a mixture of fast networks. Their results give less processing times with accuracy and high throughput without losing too much identification efficiency. Their results showed that by using CNNs we can directly estimate segmented images from the entire image and the result gives a pixel-wise segmentation of the complete image from which useful information can be extracted without doing step of preprocessing and feature extraction method.

D. Soil Analysis and Disease Detection

The characteristics of soil like wetness, inversion, and nitrogen levels play an important role in the quality of crops. Most farmers spread pesticides equally per square meter in the field. So, wasteful use of resources affects the budget of farmers and further increases the budget of farmers. In this case, Machine learning techniques can be used to examine the levels of soil erosion and the quality of individual crops. The result of machine learning algorithms can be used to identify on which exact location of the land they have to throw the pesticides. It helps farmers to use pesticides exactly wherever required.

Yujiana et al. (2019) developed a model for the detection of disease from images. They worked on fine-tuning and architecture of the deep convolutional neural network. They used different architectures of CNN. An experimental comparison of the deep learning architecture was done by them. They accessed diseased and healthy images of leaves of 14 plants that come under different classes. The data set was taken from plant Village. They concluded that Dense Nets are a type of network which can give accurate results even in degraded quality of image with noise and other effects. There will be no chance of deterioration of results. Moreover, while using Dense Nets small number of parameters are required and can achieve efficient performances in less time. It achieves a testing accuracy score of 97.75% to beat the rest of the architectures.

Ferentinos et al. (2018) developed a model using a Convolutional Neural Network for the detection of plant disease. They have taken more than 85000 images of leaves of healthy and diseased plants which are of different categories. Several model architectures were trained with images by dividing the dataset into training and testing. Their results showed the best performance of architecture with a success rate of 99.53% while identifying the diseases in a plant. Its high success rate concludes that the same architecture can be further expanded to identify the disease of integrated plants in real environment conditions.

Vijai et al. (2017) developed an automatic model for the detection and classification of diseases of plants using the images of their leaves. A genetic algorithm is used for this purpose. They also compared the performances of different classification techniques used to detect different types of diseases in plants. The proposed algorithm is being tested on various leaves of plants like Tomato, potato, lemon, Banana, Beans. The optimum results were obtained in very less computational time. Their results showed that by using images of leaves we can detect the diseases at an early stage and this model gives accurate results with less usage of resources and time.

Vidya et al. (2016) proposed an algorithm for creating a model for the detection and classification of diseases present in fruits. They used the K-means clustering algorithm for the segmentation image of the plant and after that feature are extracted from the segmented image and finally on the modified image a classification technique named Multi-class Support Vector Machine was applied. They concluded that the proposed model which was created by using this algorithm can give accurate results for the automatic detection of diseases of fruit in less time. They achieved accuracy up to 93%.

E. Weed Detection

Growth of darnel in fields of agriculture gives serious problem to crop farming. They grow very quickly with crop field and causes various plant diseases due to which production of crop decrease and reduce the profit of farmers. Farmers use many germicides to remove the weeds from the field, but this solution raises many environmental and economic problems or issues. Also, Manual weeding is an inefficient solution because it increases the cost of labor and also increases time consumption which further creates a problem for the farmers. So, machine learning can be used here for helping farmers to detect weed-penetrated regions.

ISBN: 978-93-91355-11-1

Perez et al. (2015) developed a model for the detection of weed in plants. They combined the OBIA method and support vector machine algorithm classifier. They concluded that the program-oriented analysis gives better results than a pel-based analysis. The OBIA method gave transcendent information which includes bell curve metrics, and also gives an ExG and information about the shape of the object and which further could be used by the SVM classifier. They used this model in the real fields of sunflower and achieve the classification accuracy of 96%.

Lopez et al. (2016) used the OBIA method for linear vegetation. They applied this method to objects that are in touch with rows of crops as it's very hard to classify that object. So, with this method, they can easily assign unclassified vegetation object to the class of weed and this is further dependent on the degree of insubstantial similarity in values of NDVI. The results of classification give the knowledge about the detection of weeds in the field.

Gao et al. (2018) suggested a hybrid method for the detection of inter-row weed. They used the combination of Hough transform-based algorithm and an OBIA method and after that applied a Random Forest classifier. With this approach, they proposed a semi-automatic method to automatically detect the objects which are in a row with vegetation objects that can be used as labeled data set for training the model.

Jafari, et al. (2018) tried to develop a procedure for detecting the species of weeds in sugar beet fields. They employed a support vector machine and artificial neural networks to enable the vision system in the detection of the weeds based on their pattern. They used the four species of common weeds in sugar beet fields as a dataset. They extracted the feature by using the Fourier descriptor. They achieved an accuracy of 92.92% with the ANN classifier and an accuracy of 95.00% with the SVM classifier which gives overall accuracy of 96.67%.

Table 1: Application of Machine Learning in Agriculture and Used Techniques

Year	Application	Problem Description	Techniques Used	Results
2020	Disease Detection in leaves of plants	A cucumber leaf disease detection using classification method	Sharif saliency-based (SHSB) method, SVM, VGG-19 & VGG-M	Give 98.08% accuracy
2019	Selective Breeding	Used different segmentation techniques to detect the quality of plant.	K- means clustering SVM SFM HOG descriptor	SVM & HOG descriptor gives better results
2019	Disease Detection using images	A model is developed for the detection of diseases using architectures of Deep Convolutional Neural Network	CNN Dense net architecture Deep CNN	Give accuracy of 97.75%
2018	Weed Detection	A hybrid method is used for the detection of inter-row weed.	Hough transform OBIA method Random forest classifier	Give accuracy of 96.4%
2018	Weed Detection	Develop a procedure for detecting the species of weeds in sugar beet fields.	SVM Classifier ANN classifier Fourier descriptor	Give accuracy of 92.92%, 95.00%, 96.67%
2018	Disease Detection	A model is developed using a Convolutional Neural Network for the detection of plant disease.	CNN	Give accuracy of 99.53%
2018	Species identification	Used the image thresholding method to detect the features like separate background, noise, and contour of the image.	Rectilinear derivations	Give accuracy of 90%
2017	Selective Breeding	A model is presented which can be used for the segmentation of plant image automatically.	Image analysis pipelining segmentation	Give accuracy of 90.01%
2017	Disease Detection	An automatic model is developed for the detection and classification of diseases of plants using the images of their leaves.	Genetic algorithm Classification techniques.	Optimal results are obtained with less computational time
2017	Agrochemical Production	Developed a convolution neural network model by compression of the fine-tuned network using a mixture of fast networks.	CNN	Give accurate results with less processing time.
2016	Disease Detection	A model is created for the detection and classification of diseases present in fruits.	Multiclass SVM K-means clustering	Give accuracy of 93%
2016	Weed Detection	For linear vegetation applied a method to objects that are in touch with rows of crops as it's very hard to classify that object.	OBIA METHOD	Give accurate results
2015	Weed detection	Used Convolutional Neural Networks for the detection of weeds in plants	CNN	Gives efficient results with high rate of discrimination.

ISBN: 978-93-91355-11-1

VI. Conclusion

This paper reviews various machine learning algorithms used in different agriculture practices for helping farmers to solve the problems they face during agriculture. Today each person is using machine learning knowingly or unknowingly. With the adventure of Computer vision, we can make models or systems which can work like think like the brain of humans and work accordingly. Computer Vision, Digital image processing, Artificial intelligence, Internet of things are the essential techniques for making agriculture smart. So, in this paper, we tried to throw light on various ideas and algorithms given by different researchers. In this survey paper, we have presented various practices of agriculture where machine learning techniques can be used for making it smart. Which can help farmers for efficient and accurate results with not many resources and reduce their cost. Machine learning has made the machines independent; it also helps farmers to reduce the constant vigilance on the fields and save their time. Many developments have been done by researchers in this field but still, there is a need to do more advancements like choosing of correct data set with which machine can be trained. The purpose of this survey was to create the interest of many more researchers in this field and to make furthermore advancements in making agriculture automatic and help the farmers. Therefore, in this paper, we have considered various algorithms that can be applied to smart agriculture. At last, we can say the machine learning algorithms can be used in boundless applications and it will persist as a dynamic field of research with extensive development options.

References

[01] The economic times. GDP data; 2018. Link <https://economictimes.indiatimes.com/news/economy/indicators/view-being-the-worlds-6th-largest-economy-means-littlefor-indias-future/articleshow/64966415.cms>.

[02] India at a glance. FAO in India. Food and Agriculture Organization of the United Nations; 2018.

[03] Singh A, Ganapathysubramanian B, Singh AK, Sarkar S. Machine learning for high-throughput stress phenotyping in plants. Trends Plant Sci 2016;21(2):110–24.

[04] Samuel, A.L. Some Studies in Machine Learning Using the Game of Checkers. IBM J. Res. Dev. 1959, 44, 206–226. [CrossRef]

[05] L. Rokach, O. Maimon, *"Top-Down Induction of Decision Trees Classifiers – A Survey"*, IEEE Transactions on Systems,

[06] Wäldchen J, Mäder P. Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review. Archives of Computational Methods in Engineering. 2017; 1–37.

[07] Cope JS, Corny D, Clark JY, Remagnino P, Wilkin P. Plant species identification using digital morphometrics: A review. Expert Systems with Applications. 2012;39(8):7562–7573.

[08] Wu SG, Bao FS, Xu EY, Wang YX, Chang YF, Xiang QL. A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network. In: Proceedings of the IEEE International Symposium on Signal Processing and Information Technology, 2007. pp. 11–16.

[09] Jin T, Hou X, Li P, Zhou F. A Novel Method of Automatic Plant Species Identification Using Sparse Representation of Leaf Tooth Features. Palos ONE. 2015;10(10): e0139482. pmid:26440281

[10] Seeland M, Ranney M, Alaqraa N, Wäldchen J, Mälder P. Plant species classification using flower images—A comparative study of local feature representations. PLOS ONE. 2017;12(2):1–29.

[11] Aggarwal K.K., Yogesh Singh, A. Kaur, O.P. Sangwan "A Neural Net-Based Approach to Test Oracle" ACM SIGSOFT Vol. 29 No. 4, May 2004.

[12] Singh A, Ganapathysubramanian B, Singh AK, Sarkar Machine learning for high-throughput stress phenotyping in plants. Trends Plant Sci 2016;21(2):110–24.

[13] M.M. Kamal et al., "Classification of Leaf Disease from Image Processing Technique," Indonesian Journal of Electrical Engineering and Computer Science. vol.10, pp. 191-200, 2018

[14] Akodagali J, Balaji S. Computer vision and image analysis-based techniques for automatic characterization of fruits a review. Biotechnol Bioeng 2013;38(9):1001–6.

[15] Zhou XB, Zhao JW, Li YX. Apples Shape Grading by Fourier Expansion and Genetic. Program Algorithm[M]. IEEE; 2008.

[16] Mavridou, E., et al., 2019. Machine vision Systems in Precision Agriculture for Crop Farming. J. Imag. 5 (12), 89.

[17] Kamilaris, A., and Prenafeta-Boldú, F. X. Deep learning in agriculture: A survey. Computers and electronics in agriculture 147 (2018), 70-90.

[18] Qin, J., et al. Classification of citrus peel diseases using color texture feature analysis. In: Food Processing Automation Conference Proceedings, 28-29 June 2008, Providence, Rhode Island. 2008. American Society of Agricultural and Biological Engineers.

[19] R. Sathya, Annamma Abraham, "Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification", (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 2, No. 2, 2013.

[20] Faith praise, F., Birch, P., Young, R., Obu, J., Faith praise, B., and Chat win, C. Automatic plant pest detection and recognition using k-means clustering algorithm and correspondence letters. International Journal of Advanced Biotechnology and Research 4, 2 (2013), 189-199.

ISBN: 978-93-91355-11-1

[21] N. Satya Priya, E. Nivetha and Rashmita Khilar, "Efficient Knowledge-Based System to Detect Diseases in Lemon Leaf" Imperial Journal of Interdisciplinary Research, Vol.2(5), pp.275-280, 2016.

[22] Kamilaris, A., and Prenafeta-Boldú, F. X. Deep learning in agriculture: A survey. Computers and electronics in agriculture 147 (2018), 70-90.

[23] Calabi-Floody M, Medina J, Rumpel C, et al. Smart fertilizers as a strategy for sustainable agriculture 2018;147: 119–57.

[24] Tuli, N. Hasteer, M. Sharma and A. Bansal, "Framework to leverage cloud for the modernization of the Indian agriculture system," IEEE International Conference on Electro/Information Technology, Milwaukee, WI, 2014, pp. 109-115.

[25] Bai, X., Li, X., Fu, Z., et al., 2017. A fuzzy clustering segmentation method based on neighborhood grayscale information for defining cucumber leaf spot disease images. Comput. Electron. Agric. 136, 157–165.

ISBN: 978-93-91355-11-1