Section A-Research paper
ISSN 2063-5346

Plant disease detection in agriculture using Machine Learning – a survey

1. **Sukhwinder Kaur**, Department of Computer Science and Applications, Khalsa College, Amritsar, India
E-mail: sukhwinderkaurasrpb@gmail.com

2. **Saurabh Sharma**, Assistant Professor, Sant Baba Bhag Singh University, Jalandhar, India

Abstract

The primary factor in a country's development is agriculture. To effectively manage plant diseases and increase crop production, it is necessary to accurately identify a disease when it first manifests. Machine learning has been used in a wide range of sectors, and it is now being used in agriculture. The diagnosis of illnesses from images of a plant is one use of machine learning in agricultural research. A variety of machine learning algorithms, such as random forest, clustering, Gaussian models, linear / logistic regression, decision trees, Naive Bayes, K-nearest neighbors, support vector machine and deep learning techniques can be used for disease diagnosis in the agricultural sector. The review paper's objective is to examine the various machine and Deep learning models techniques that are effectively used in the agriculture industry.

Keywords Machine learning, deep learning, plant disease classification, Agriculture

Introduction

Crop management, fruit counting, water management, weed identification, soil management, harvesting, seed categorization and disease detection are some of the applications of DL in agriculture. This article focuses on plant disease detection. Plants are affected by various types of diseases. In agriculture industry, there's demand to reduce the diseases in plants to expand the quantity and condition of crop. Diseases that are resulting from living organisms are known as biotic diseases which can be particularly a result of virus, fungi and bacteria. Diseases to nonliving surroundings are categorized as abiotic diseases for example hail robust winds and other weather conditions, spring frosts etc. Unlike biotic diseases, abiotic sicknesses are less dangerous due to their non-infectious nature. The process of manually identifying these diseases is challenging. Various image processing technologies may be used to evaluate the quality and health of the leaf which can help to reduce plant diseases. This is done by collecting its image and using various techniques such as disease detection algorithms. This allows for lower costs and results in increased production. Plant diseases symptoms are seen in plant leaf, stem, and fruit. Early detection of plant diseases will allow farmers to provide most reliable and effective crop yield [1]. The majority of plant illnesses may be diagnosed early on by checking the plant leaves. The pattern and color of the infected plant leaf will be different from that of the usual leaf when recognizing and diagnosing plant disease. There are numerous other features related to image shape, holes on the infected leaf through which plant disease can be detected.

Machine Learning Techniques

The processes shown in Figure 1—show how to identify plant diseases.

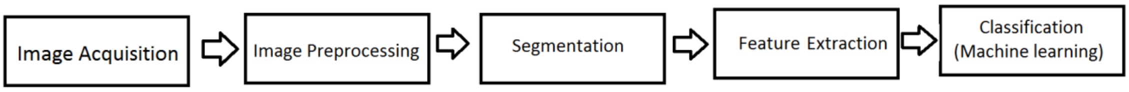

Figure 1

Image acquisition deals with Capturing of image under different environmental conditions or by using available datasets.

Image pre-processing aims to enhance the images beyond what can already be seen. In order to see the image's features, noise removal is done here most importantly. Color space conversion, cropping, smoothing, and enhancement are common pre-processing techniques. Image segmentation is a technique for dividing an image down into its individual objects or components in order to extract the necessary attributes from it. Leaf segments containing the diseased and unaffected areas are separated. A crucial technique for locating the features of interest is feature selection. The necessary features are removed from the processed data in the form of a table so that the model can detect them. The features that are chosen can change depending on the types of crops and diseases. Based on the features obtained, classifiers are used to distinguish between the various diseases that affect plant leaves. In order to identify plant diseases, some classifiers include Convolution Neural Network (CNN), Support Vector Machines (SVM), K nearest neighbors (KNN), and Artificial Neural Network (ANN), among others.

Numerous modeling techniques (including Deep Learning, Support Vector Machines, and Artificial Neural Networks) are created for the purpose of identification of plant diseases. Data is essential in machine learning. The methods which are used in Machine learning classify the data into classes. Unsupervised and supervised machine learning are the two main types. Other types include reinforcement learning and semi-supervised learning. Unsupervised learning is a learning technique which draws inferences from data sets without the training data and relies mainly on clustering. Supervised learning is regarded as learning by example. In supervised learning, training dataset is made and is fed into the system to obtain the meaningful outputs. This helps in decision making. The training data size influences the accuracy of machine learning classifications.

A discriminative classifier is a Support Vector Machine in which large amount of training dataset is required. It defines separating hyper plane which categorizes new dataset. This hyper plane divides a plane into two halves in two dimensions, with each class located on one side. Here original training datasets are mapped by a kernel function. After mapping the mapped outputs are linearly separable. SVMs are employed in both binary and multiclass classification. The majority of supervised machine learning applications use the probabilistic classifier known as Naive Bayes. It works by calculating the likelihood of certain data attributes belonging to a particular class. To develop a probabilistic data model, the Bayes theory of probability is used. It is straightforward and simple to grasp and used when size of training set is less. The Max Entropy classifier is a probabilistic classifier that belongs to the exponential model class; it takes into consideration the fact that the features are not conditionally independent of each other. Out of all the models that fit the training set of data, it selects the model with the maximum entropy. The decision tree (DT) classifier denotes the training data hierarchically. The tree is made up of nodes representing the input values, edges represent possible moves and path from node to leaf gives the target values from which classification can be created.

A collection of Decision Trees assembled into a Random Forest. The Random Forest algorithms are often used for classification and regression tasks and can handle categorical features very well. This method is capable of dealing with a large number of training instances as well as high-dimensional spaces. The random forest approach is used in issues where numerous trees are produced and trained by partitioning the training sets, and the results are derived by aggregating all trees.

The K-nearest Neighbor (KNN) technique searches the complete training set for the K most similar instances to continually segregate data as new data is received. KNN is based on the idea that an instance's classification will be close to that of its similar instances (neighbors).
ANNs are a type of mathematical model that are based on how the human brain functions. These are described as "neurons" systems that can be used to solve machine learning problems, and it can learn mapping from input to output. Neural networks clusters and classifies the data. Inputs and outputs are connected by a layer of interconnected perceptrons, and there are several layers in deep neural networks. A benefit of a DNN is its potential to automatically identify important low-level elements and connect them to high-level features in successive layers. A large amount of training data is necessary for recurrent neural networks (RNNs) and other supervised machine learning techniques.
CNN is the machine learning algorithm that is most frequently used for image recognition and visual analysis. a CNN transforms an input picture made up of raw pixels with use of different layers. The pooling layers lower the dimensionality of the input images while the convolutional layers are utilized to extract features. The fully connected layers act as classifiers by assigning input images to the appropriate class using high-level features. A convolution is a two-function process. the two functions are: a filter (or kernel) and input values (such as pixel values) at specific locations in the image. Each of these functions can be represented as an array of integers. An output is determined by the two functions' dot product. Based on the stride length, the filter is then moved to the following area of the image. After the method has been applied to the full image, it results in a feature (or activation) map .The Pooling layer decreases the amount of parameters that must be calculated and also the size of the image. The frequent used pooling technique is called max-pooling. In essence, max-pooling keeps only the greatest input value from a filter and rejects the others in order to total up the strongest activations throughout a neighborhood. some deep learning architectures including ResNet, LeNet, VGG16, VGG19, Inception Net, and MobileNet are trained on imagenet database.

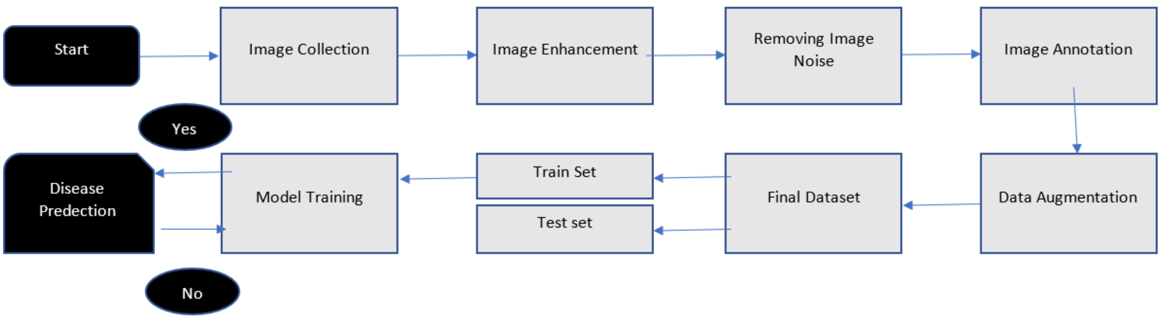

Figure 2

To overcome the problem of limited training data, Transfer learning [4] is used which is shown in Figure 3.

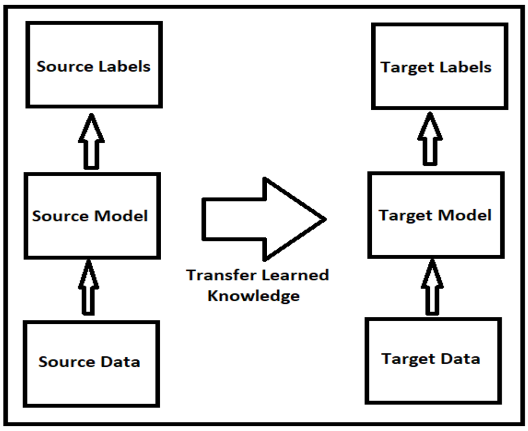

Figure 3

Literature review

Major crop losses can be prevented by using various methods to diagnose disease. The transfer learning of deep CNN was examined by the authors [5] to detect diseases in plant leaves. Images of samples connected to plant diseases were obtained and subsequently labeled in accordance with the expert's understanding of the subject. After the captured images were subjected to image-processing methods (image filtering, grey transformation, image sharpening and scaling, etc.) additional sample images were also produced to further improve the dataset utilizing data augmentation techniques. The experiment was conducted on a public plant village dataset using transfer learning. Validation accuracy of 91.83% after 30 epochs was achieved.

A method for identifying gray spots on tea leaves was put forth by combining algorithms for computer vision and image processing [6]. A Neural of Multi-Layer Perceptron, or MLP, was used; it was constructed using the features that could be extracted from the gray spotted tea leaves to identify them. To recognize the gray dots on tea leaves, the authors then used image processing methods, neural networks, and computer vision.

The authors [7] used the CNN in maize to identify diseases with accuracy rates of 99.9%), 91% and 87% for three diseases. The authors [8] trained a deep convolutional neural network to distinguish between five diseases using a public dataset of 9000 images of diseased and healthy tomato leaves, the trained model had a 99.84 percent accuracy rate.

2100 photos of tomato leaves from the internet, 500 photographs from local fields were used to create their dataset, and a convolution neural network were employed by the authors[9]. They applied Transfer learning on inception model to train CNN for tomato leaves and obtained 72% accuracy on splitting of training and testing dataset as 45-55%. They also demonstrated that more the accuracy is attained when more training data is provided and achieved the highest accuracy on 99%, on giving 90% image as training. Otsu's method/thresholding and SVM were used by Mangala et al. [10] to identify disease in paddy crops.

To classify rice diseases, the authors [11] made use of deep CNN's transfer learning. In addition to healthy rice plants, images of plant diseases were taken. The authors obtained an accuracy of 91.37% by using the SVM as a classifier and the pre-trained deep CNN as a feature extractor.

With 2,207 images in each category, the authors [12] classify four categories of wheat diseases: yellow rust, stem rust, normal and powdery. They achieved 84.54% accuracy when training the classifier with the Convolution Neural Network (CNN).

The authors [13], used images of six diseased and healthy tomato leaves (from PlantVillage dataset) to compare AlexNet and VGG16 net for detecting tomato leaf disease. They were able to achieve classification accuracy of 97.29% (VGG16 net) and 97.49% (for AlexNet).

The method which was proposed by the authors [14] applied a CNN which was focused on the Alex Net model. These four prevalent apple leaf diseases were all discovered. It demonstrated that convolutional neural networks achieve an overall accuracy of 97.62% by taking a dataset of 13,689 images of apple leaves.

The authors [15] first identified the disease affected portion with GLCM and LBP features and used classifiers k-nearest neighbors ensemble classifier and SVM classifier. Four different plant diseases were discovered and categorized by the researchers. With this method based on CNN for classification of the disease of strawberry plants to healthy plants and the diseased strawberry plants, the authors [16] classified images of strawberry leaves with 92% accuracy using deep learning for disease diagnosis. Based on a global gradient and local data, active contour model of a cotton leaf with disease was developed by the authors[17]. This model was successful in accurately extracting the contour curve of segmented cotton diseased leaves with weak edges and fuzzy edges, as well as the general cotton diseased leaves.

By utilizing the K-means clustering technique and the Otsu's classifier and collecting the shape and texture data, the authors [18] detected the diseased region of the leaves. Classification was carried out using a neural network based classifier. The authors [19] applied a CNN classifier to analyze leaf images from four different classes in the PlantVillage database in order to categorize Soybean plant diseases. With a 99.32% accuracy rate, their model demonstrates that it is possible to extract important features and categorize plant diseases using CNN.

The accuracy was 99.53% according to the author [20], who used an open database of 87848 photos of 25 distinct plants in 58 different classes of plant disease combinations after applied deep learning classification technique.

A deep model approach was put forth by the author [21] to develop a classifier for disease detection. They made use of a sizable dataset that included 14828 images and nine tomato diseases. Additionally, they proved that pre-trained Deep Models outperform those without pre-training while computing accuracy. The authors [22] described a method based on deep convolutional neural networks (CNNs) to recognize ten different types of rice disease. The CNN-based model, which outperformed the conventional machine learning model, had an accuracy of 95.48 %. The image retrieval approach was suggested by the authors [23] as a solution to the soybean disease detection issue. To classify the diseases affecting the soybean plants, they used image retrieval techniques. In addition to the two feature descriptors Color histogram (HIST) and Wavelet decomposed color histogram (WDH), they also employed a number of other color and texture related feature descriptors. Additionally, they tested the segmentation strategy on six different varieties of soybean plant diseases. The authors [24] implemented genetic algorithm to detect bakanae disease of rice plants. They classified the infected and healthy plants by Support vector machine (SVM) classifiers. The authors suggested a four-step image processing strategy for classifying apple diseases [25]. The contaminated area is initially retrieved using the K-means clustering segmentation technique. Combining features based on texture, color and shape results in more distinct characteristics. The diseases apple blotch, rot, and scab were investigated during the training and classification process using a Multiclass Support Vector Machine. The authors [26] have designed a method that uses genetic algorithms for picture segmentation to automatically identify and categorize diseases in plants. The training and test sets for leaves of four plants only included a small set of number of photos. The color co-occurrence approach has been used to extract characteristics that combine texture and color features. With accuracy rates of 86.54% and 95.71%, respectively, the diseases have been categorized with k-mean clustering and the SVM classifier using the Minimum Distance Criterion. When the genetic algorithm and the classifier using the Minimum Distance Criterion are combined, accuracy rises to 93.63%. In order to automatically determine whether a black pepper leaf is healthy or infected with diseases, the authors [27] proposed an algorithm. On black pepper leaves, the suggested method was tested, and the results showed 100% accuracy in identifying healthy or infected leaves. The Gabor wavelet transform technique was used by the authors [28] to extract relevant tomato leaf features. They detected leaf diseases Powdery mildew or early blight using Support Vector Machines (SVM) and applied background subtraction method. They evaluated the performance by taking a dataset containing 100 images for each type of tomato disease and attained 99.5 % accuracy. The authors [29] applied CNN to extract four types of rice disease spot and then applied SVM for classification and obtained accuracy of 96.8% by combination of CNN with SVM model. In order to accurately diagnose different cotton leaf diseases, authors [30] suggested a model based on meta Deep Learning. The dataset includes 2385 photos of healthy and diseased leaves collected from the field. The dataset was trained using meta deep learn model, Custom CNN, and VGG16 Transfer Learning after data augmentation. They achieved an accuracy of 98.53%.

Conclusion

It's important to identify plant diseases as soon as possible to prevent contamination of the entire crop and to improve the quality of agricultural outputs. ML models have been used in numerous agricultural applications. This paper discusses various methods for identifying and categorizing plant diseases. These methods are employed to examine both healthy and diseased plant leaves.

References

[1] Sharma, A.; Jain, A.; Gupta, P.; Chowdary, V. Machine learning applications for precision agriculture: A comprehensive review.IEEE Access 2020, 9, 4843–4873.

[2] Sharma, Rohit, Sachin S. Kamble, Angappa Gunasekaran, Vikas Kumar, and Anil Kumar, “A systematic literature review on machine learning applications for sustainable agriculture supply chain performance”, Computers & Operations Research 119 (2020): 104926.

[3] Sharada P. Mohanty, David P. Hughes and Marcel Salathé, “Using Deep Learning for Image Based Plant Disease Detection”, Frontiers in Plant Science |www.frontiersin.org, September 2016 |Volume7| Article1419.

[4] Tan, C.; Sun, F.; Kong, T.; Zhang, W.;Yang, C.; Liu, C.A survey on deep transfer learning. In International Conference on Artificial Neural Networks; Springer:Berlin/Heidelberg, Germany, 2018; pp.270–279

[5] Chen, J., Chen, J., Zhang, D., Sun, Y., &Nanehkaran, Y. A. (2020). Using deep transfer learning for image-based plant disease identification. Computers and Electronics in Agriculture, 173, 105393.

[6] Binh, P. T., Nhung, T. C., & Du, D. H. (2019, December). Detection and Diagnosis Gray Spots on Tea Leaves Using Computer Vision and Multi-layer Perceptron. In International Conference on Engineering Research and Applications (pp. 229-237). Springer, Cham.

[7] Sibiya, M., &Sumbwanyambe, M. (2019). A computational procedure for the recognition and classification of maize leaf diseases out of healthy leaves using

convolutional neural networks. AgriEngineering, 1(1), 119-131.

[8] Ashqar, Belal & Abu-Naser, Samy. (2019). Image-Based Tomato Leaves Diseases Detection Using Deep Learning. International Journal of Engineering Research. 2. 10-16.

[9] Hasan M, Tanawala B, Patel KJ. March. Deep learning precision farming: Tomato leaf disease detection by transfer learning. In Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE); 2019.

[10] Mangala N, Raj PB, Hegde SG, Pooja R. 2019. Paddy leaf disease detection using image processing and machine learning. Int J Innov Res Elec Electron Instrument Control Eng. 7(2):97–99.

[11] Shrivastava, V. K., Pradhan, M. K., Minz, S., & Thakur, M. P. (2019). Rice Plant Disease Classification Using Transfer Learning of Deep Convolution Neural Network. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42(3/W6).

[12] Hussain A, Ahmad M, Mughal IA, Ali H. Automatic disease detection in wheat crop using convolution neural network. In The 4 th International Conference on Next Generation Computing; 2018.

[13] Rangarajan AK, Purushothaman R, Ramesh A., Tomato crop disease classification using pre-trained deep learning algorithm. Procedia Computer Science. 2018;133:1040-1047.

[14] Bin, Liu & Zhang, Yun & He, DongJian& Li, Yuxiang . Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks. Symmetry 2018, 10, 11.

[15] Min Oo, Y., & Chi Htun, N. (2018). Plant Leaf Disease Detection and Classification using Image Processing. International Journal Of Research And Engineering, 5(9), 516-523. doi:10.21276/ijre.2018.5.9.4

[16] Park H., Jeesook E., & Kim S.H., "Crops Disease Diagnosing Using Image Based Deep Learning Mechanism," 2018 International Conference on Computing and Network Communications (CoCoNet), 2018.

[17] ZHANG Jian-hua, KONG Fan-tao, WU Jian-zhai, HAN Shu-qing, ZHAI Zhi-fen Automatic image segmentation method for cotton leaves with disease under natural environment Journal of integrative agriculture 2018 17(8)1800-1814

[18] Saradhambal, G., Dhivya, R., Latha, S., Rajesh, R., 'Plant Disease Detection and its Solution using Image Classification', International Journal of Pure and Applied Mathematics, Volume 119, Issue 14, pp. 879-884, 2018

[19] Wallelign, S., Polceanu, M., &Buche, C. (2018, May). Soybean plant disease identification using convolutional neural network. In The thirty-first international flairs conference.

[20] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311-318.

[21] Brahimi, M., Boukhalfa, K., & Moussaoui, A. (2017). Deep learning for tomato

diseases: classification and symptoms visualization. Applied Artificial Intelligence, 31(4), 299-315.

[22] Lu, Y., Yi, S., Zeng, N., Liu, Y., & Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks. Neurocomputing, 267, 378-384.

[23] Shrivastava, S., Singh, S. K., &Hooda, D. S. (2017). Soybean plant foliar disease detection using image retrieval approaches. Multimedia Tools and Applications, 76(24), 26647-26674.

[24] Chung, C.L., Huang, K.J.; Chen, S.Y.; Lai, M.H.; Chen, Y.C.; Kuo, Y.F. Detecting Bakanae disease in rice seedlings by machine vision. Comput. Electron. Agric. 2016, 121, 404–411.

[25] Dubey, S. R., & Jalal, A. S. (2016). Apple disease classification using color, texture and shape features from images. Signal, Image and Video Processing, 10(5), 819-826.

[26] Singh, V., Misra, A.K., 'Detection of Plant Leaf Diseases Using Image Segmentation and Soft Computing Techniques', Information Processing in Agriculture, Volume 8, pp. 252-277, 2016

[27] Thamizharasi A, Caroline Mary and Dr. N.V. Radhakrishnan (2016) "Detection and Assessment of leaf spot disease in black pepper leaf." International Journal of Innovative Research in science engineering and technology 5(14): 10-21

[28] Mokhtar, U., Ali, M. A., Hassenian, A. E., &Hefny, H. (2015). Tomato leaves diseases detection approach based on support vector machines. In 2015 11th International Computer Engineering Conference (ICENCO) (pp. 246-250). IEEE.

[29] F. Jiang, Y. Lu, Y. Chen, D. Cai and G. Li, "Image recognition of four rice leaf diseases based on deep learning and support vector machine", Comput. Electron. Agricult., vol. 179, Dec. 2020.

[30] Memon, M.S.; Kumar, P.; Iqbal, R. Meta Deep Learn Leaf Disease Identification Model for Cotton Crop. Computers 2022, 11, 102. https://doi.org/10.3390/computers11070102

[31] Zhang, S., Wu, X., You, Z., & Zhang, L. (2017). Leaf image based cucumber disease recognition using sparse representation classification. Computers and electronics in agriculture, 134, 135-141.

[32] Ali, H., Lali, M.I., Nawaz, M.Z., Sharif, M., Saleem, B.A., 'Symptom based automated detection of citrus diseases using color histogram and textural descriptors', Computers and Electronics in Agriculture, Volume 138, pp. 92-104, 2017

[33] Dey, A. K., Sharma, M., &Meshram, M. R. (2016). Image processing based leaf rot disease, detection of betel vine (Piper BetleL.). Procedia Computer Science, 85, 748-754.

[34] Jha, S., Jain, U., Kende, A., Venkatesan, M.: Disease detection in tea leaves using image processing. Int J Pharm Bio Sci Volume 7 Issue 3, 2016 (July - September), Pages:165-171

[35] Shanmugam, L., Adline, A. A., Aishwarya, N., &Krithika, G. (2017, April). Disease

detection in crops using remote sensing images. In 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp. 112-115). IEEE.

[36] Halder, M., Sarkar, A., &Bahar, H. (2019). Plant Disease Detection by Image Processing: A Literature Review. image, 1, 3.

[37] Masazhar, A. N. I., & Kamal, M. M. (2017, November). Digital image processing technique for palm oil leaf disease detection using multiclass SVM classifier. In 2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA) (pp. 1-6). IEEE.

[38] Zhou, W., Li, H., &Tian, Q. (2017). Recent advance in content-based image retrieval: A literature survey. arXiv preprint arXiv:1706.06064.

[39] Ker, J., Wang, L., Rao, J., & Lim, T. (2018). Deep learning applications in medical im age analysis. IEEE Access, 6, 9375-9389.

[40] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org.

[41] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 [cs], Sep 2014. arXiv: 1409.1556.

[42] Mathew, Amitha, P. Amudha, and S. Sivakumari. "Deep Learning Techniques: An Overview." International Conference on Advanced Machine Learning Technologies and Applications. Springer,Singapore., 2020,pp.599-608.