



# Classification of Weeds Using Neural Network Algorithms and Image Classifiers

Rakesh Joshi<sup>(✉)</sup>, Garima Sharma, Vikas Tripathi, and Ankita Nainwal

Department of Computer Science and Engineering, Graphic Era Deemed to Be University,  
Dehradun, Uttarakhand, India  
rakeshjoshi21021369@gmail.com

**Abstract.** In agriculture it is very important to differentiate between crop seedling and weed from the farms, the traditional means to classify were mostly relies on manual methods which is not time and cost efficient. Automated Deep learning-based Weed classification plays crucial role to solve this in agriculture sector for effective crop yield and weed management. Also this study could help for precision agriculture, cost reduction and time efficiency in the field. In this research paper, we present several deep learning-based approach for automated classification between weed and crop seedlings. For the classification task we used popular image classification algorithms, which includes Convolution Neural Network (CNN), Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighbors (KNN). The study utilizes a diverse dataset between crop seedling and weed image data. We also demonstrated the results of deep learning based algorithm and compare them to find the best for the classification task. The work highlights the capability of deep learning-based model for efficient weed management, aiding farmers in identifying and controlling weeds while growing crop seedlings, and hence leads to improved productivity.

**Keywords:** Weed classification · image classification algorithm · crop seedlings · Deep learning · CNN · ANN · SVM · KNN · Comparative analysis

## 1 Introduction

In traditional agriculture, weed classification which is fundamental task required manual labour [1] on the field, which is costly for the farm and lead to loss for crop yield. In this work we have presented the Weed Classification from the image dataset. We pre-processed the images by resizing, data augmentation and then used them to classify between crop seedlings and weeds.

In farms the early stage of a growing plant is called plant or crop seedlings which are seems similar as weeds. So, it became a tough task to differentiate between them and remove the weeds from the farm [2]. This activity requires focussed manual labour on field which is time consuming and costly. To automate the recognition of weed in the farm we can use technology like deep learning to differentiate the image [3] between weed and plant seedlings. Using technology like [4] Deep Learning will make automation

recognition task easy and can deal with large image dataset in shorter amount of time. With improved model architecture the classification task between crop seedlings and weed will be easier and faster.

The Dataset [5] used in the classification work is divided into two classes weed and crop seedlings. But the manual dataset was unbalanced because images of crop seedling were very few. So, new images were added from a crop seedlings dataset [6], which helped to balance the dataset to fulfil the requirement for classification. The types of weed in weed class are like field pennycress (*Thlaspi arvense*), shepherd's purse (*Capsella bursa-pastoris*), field chamomile (*Matricaria perforata*), field pansy (*Viola arvensis*). Similarly the crop seedlings in the dataset are like pumpkin (*Cucurbita pepo*), radish (*Raphanus sativus* var. *sativus*) black radish (*Raphanus sativus* var. *niger*).

With the help of basic pipeline for image classification deep learning concepts [7], various models were prepared. For achieving the required model goal, the procedure we followed started with Data collection, in simple terms it means collecting the required dataset for the image classification task. Dataset pre-processing which includes steps like images resizing, cropping or normalization techniques further if required data augmentation techniques are also employed to get a more diverse and variable dataset. After splitting the dataset into train, test and validation set, using train set model architecture is trained on image classifier algorithm and neural networks. To tune up the model validation set is used. Then prediction on test set is performed from which further model evaluation is done like accuracy, F1 score, Recall, Confusion matrix and Precision – recall graph. For image classification Neural Network algorithm like Convolutional Neural Network (CNN) [8, 9] and Artificial Neural Network (ANN) [10] are used to train the model. CNNs are specifically designed for processing images through grid by leveraging the concept of convolution. It consists of multiple layers, including convolutional layers, pooling layers and fully connected layers. ANNs are basic form of neural networks used for general-purpose machine learning tasks. It consists of multiple layers, including input, hidden, and output layers, with each layer interconnected nodes called neurons. Model based on image classification algorithm like K-Nearest Neighbors (KNN) [11] and Support Vector Machine (SVM) [12].

For image classification KNN is used as baseline algorithm by representing each image as a feature vector and comparing it to the feature vectors of labelled images in the training set. SVM is a powerful and versatile algorithm for image classification, it can be used to learn a decision boundary that separates different classes in a high-dimensional feature space. It aims to find a hyperplane that maximally separates the data points of different classes, with a margin that maximizes the distance between the hyperplane and the nearest data points.

We have derived the result after fitting every image classifier and neural network in the model and results have shown which classifier or neural network based model works best for the classification task between crop seedling and weed.

## 2 Literature Survey

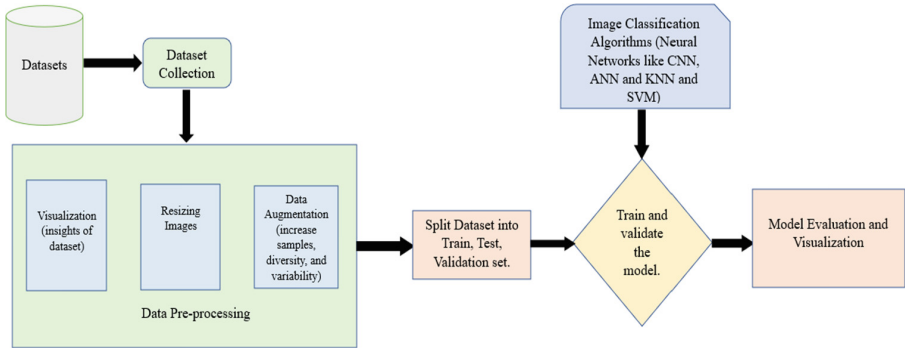
Weed management is being shifted from conventional agricultural practices to technology-friendly practice [13] that employs Machine Learning, Deep Learning, big data and other modern technology in past decade. Many authors have used different approach to solve this problem.

Author in [14] worked on a dataset that provided weed plant of different species and also suggested a benchmark scale to readers for easy comparisons of classification outcomes. Classification using effective convolutional neural network [15] demonstrated the unsupervised feature representation of 44 different plant species with high metrics. Study over the recognition of plant diseases used image classification, the development of plant disease recognition model, based on leaf image classification [16] using deep convolutional networks. This developed model was able to recognize 13 different plant disease. Study on classification of plant seedling using CNN [17] where 12 different species are classified, and model achieved an accuracy of 99% which seems promising for the agriculture sector. Work on Feature extraction for disease leaf [18], showed the efficiency of algorithms like KNN and SVM for the image classification and furthered recognizing crop disease based on extracted features. Study on autonomous robotic weed control [19] showed how robotic technology may also provide a means to do the task like hand weed control. Not only deep learning, Machine Learning based technology [20] are also contributing to the automation of many manual task. Work on vegetable plant leaf classification through Machine Learning[21] based models where classifiers like Decision Tree, Linear Regression, Naïve Bayes, MLP are used for the image classification task and result showed that MLP acquired an accuracy of 90% for the task. It is also been seen that Machine Learning can also play role in environmental sustainability [22], a model using Regression Kriging is used to classify the radiative energy flux at the earth surface. Comparison of various deep learning techniques has also been done by many researchers keeping varied application into view[23]. Recent IOT-enabled model for weed seedling classification [24] have shown that the Weed-ConvNet model with color segmented weed images gives accuracy around 0.978 for the classification task.

Our work has conducted a comparative analysis between different neural networks like CNN, ANN and image classifiers models like KNN, SVM and did the metric study on the performance of these deep learning based models. From the metric result and precision-recall graph the study concluded that neural network like CNN work best for the weed classification task and image classifier like KNN and SVM also perform good whereas ANN based model performance for classification seems to be low.

## 3 Methodology

The procedure for classification task between weed and crop seedling is widely based on Deep Learning method, below Fig. 1 shows the pipeline flow of the work. Starting from Data collection and visualisation of dataset and getting some insights, then Dataset Pre-processing through which we resize the image and by Data augmentation we make the dataset more balanced, diverse and variable while preserving the original classes.

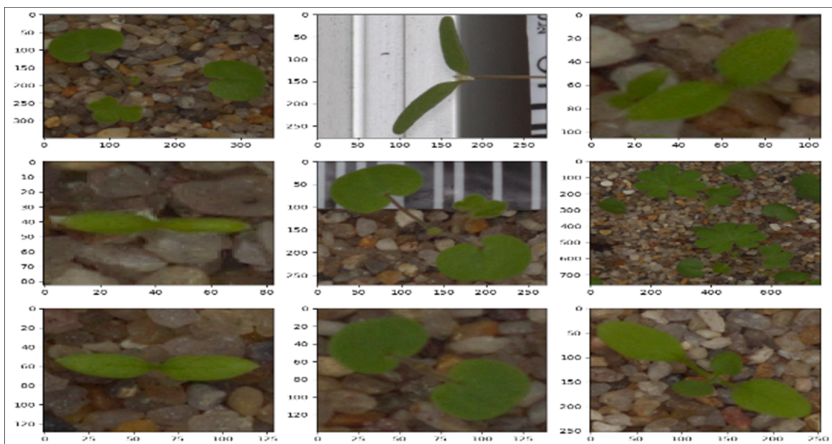


**Fig. 1.** Proposed methodology for weed classification system

Next step is to split the dataset into train, test and validation set and then applying the various image classification mode architectures like CNN (Convolutional Neural Network), ANN (Artificial Neural Network), KNN (K-Nearest Network) and SVM (Support Vector Machine). After training and validating the models we have evaluated the metrics of every model and tried to find out the comparative best model for the Classification.

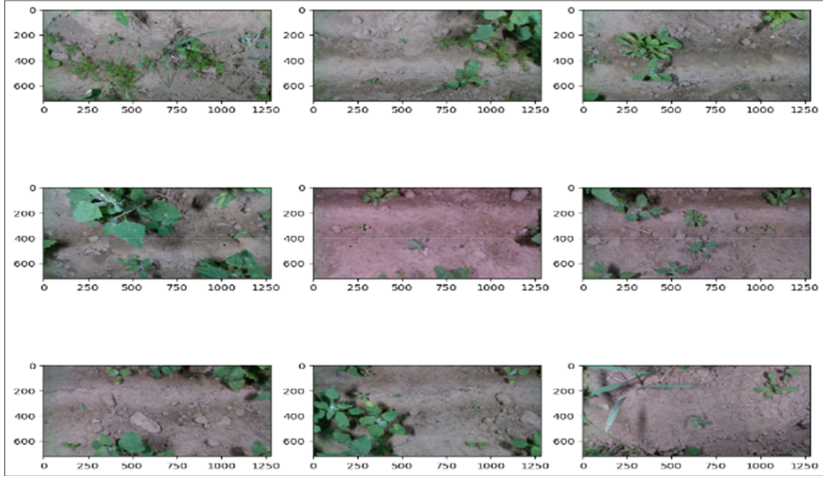
### 3.1 Data Collection

The collected Weed image dataset used for classification have around 2,047 images of weed and crop seedlings. In the record 931 are under crop seedling class and 1116 are weed images. Types of weeds in the dataset are goosefoot (*Chenopodium album*), catchweed (*Gallium aparine*), field pennycress (*Thlaspiarvense*), shepherd's purse (*Capsella bursa-pastoris*), field chamomile (*Matricariaperforata*), field pansy (*viola arvensis*) and others.



**Fig. 2.** Weed image representation on grid

Several types of plant seedlings are beetroot (*Beta vulgaris*), carrot (*Daucus carota* var. *Sativus*), zucchini (*Cucurbita pepo* subsp. *Pepo*), pumpkin (*cucurbita pepo*), radish (*Raphanus sativus* var. *sativus*) black radish (*Raphanus sativus* var. *niger*) and other seedling image data collected. Figures 2 and 3 show the grid plotting of the weed and plant seedling image respectively.

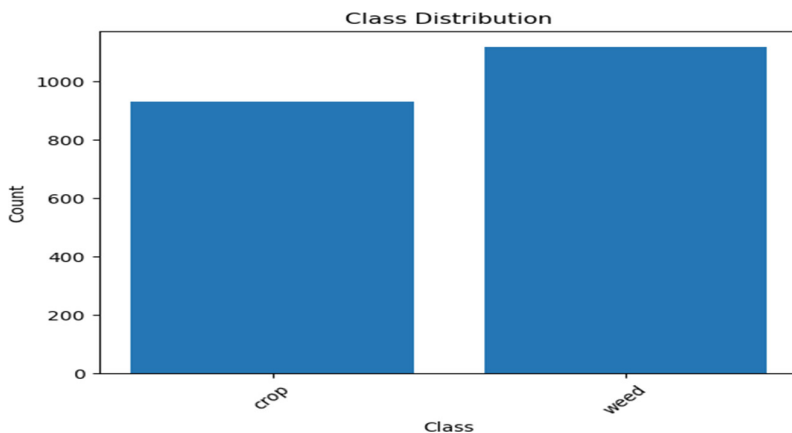


**Fig. 3.** Plant seedling image representation on grid

### 3.2 Data Pre-processing

This includes understanding the dataset through visualization and further pre-processing it, so that model can be applied on the dataset. Below Bar graph representation of Fig. 4 shows the image distribution of class weed and crop seedling in the dataset we acquired or collected.

As the raw image dataset have unequal size of images, and to perform any Deep Learning algorithm efficiently we perform Resizing of the images. Upon going through various trails for best resizing, we find the dimensions of 360 pixels width and 257 pixels height as the most fitting for the algorithms. Next, we performed Data Augmentation to increase the diversity and variability of dataset, techniques which we used here for data augmentation are random rotation of images, width shift, zoom range, horizontal and vertical flip which were applied randomly to images. After this new augmented dataset have around 6141 images, where Crop seedlings are 2793 and weed images were 3348. The very next step is to split the dataset into Train, Test and Validation set which will be further used in Model Architecture of different classification algorithms.

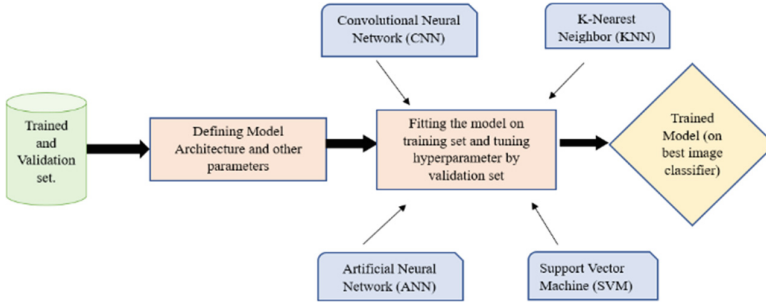


**Fig. 4.** Class representation in dataset

### 3.3 Classification Models Architectures and Algorithms

Image classification Deep Learning Algorithms are future applied through model in the dataset to achieve the classification task. First, we have used CNN (Convolutional Neural Network) model architecture using the Sequential model from keras. It consists multiple convolutional layers, a flatten layer, max-pooling layers and dense layers with dropout for regularization, more hyperparameter which were defined for CNN were Adam optimizer, 3 Convolutional layers of 32, 64, 128 filters, batch size of 32 and number of epochs were 7. After giving the architecture to the model, using fit method on Train set we trained the model, then we have compiled the model by specifying the optimizer, loss function and accuracy for evaluation matrix. Second classification algorithm we used is ANN (Artificial Neural Network), similar to CNN we defined the model architecture consists of flatten, dense, SoftMax layers and dropout further more hyperparameters were defined as Adam optimizer, dropout rate of 0.5, batch size of 32 and number of epochs were 6. Flatten is responsible for target image size, Dense layers are for learning patterns and to make predictions and dropout layers introduce regularization to reduce overfitting. Then we fit the model on Train set and after the compiled model we specify the evaluation matrix.

For SVM (Support Vector Machine) we extracted the features from the training set and make a train features 2-D array, after features vectors are normalized SVM classifier is trained and now features are extracted from validation set called valid features and the SVM classifier is used to evaluate the performance through valid features. Similarly, for KNN (K-Nearest Neighbors) first the train images are converted into RGB format, then converted to NumPy array. Then while creating and training of KNN classifier we set the number of neighbor 'k' to 3, create an instance and fit the classifier on training images and labels. Validation set is used for prediction and for tuning up the model, then performance is evaluated. Every prediction of the above neural network algorithms or classifier algorithms are compared with the prediction on test set and the result performance of test is evaluated through which we get the metric of each image



**Fig. 5.** Execution flow of the image classification models

classifier for weed Classification. This will help us to choose the best image classifier or neural network model to differentiate between weed and crop seedling.

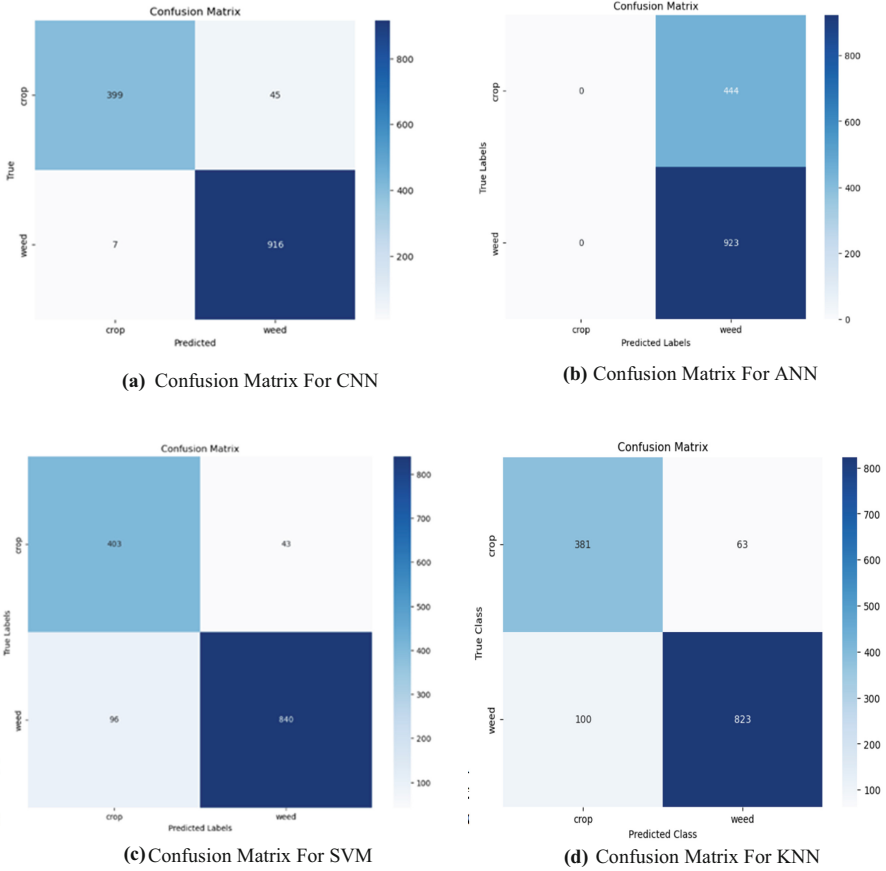
## 4 Results

To evaluate the prediction result, we have calculated various metrics like Accuracy, F1 score and Recall. The evaluation displays the performance of pre trained model CNN, ANN, KNN and SVM on the classification task of test set between weed and crop seedling which is shown in Table 1.

**Table 1.** Comparative study of the accuracy of various algorithms

S. no.	Classifier	Accuracy (%)	F1 score	Recall
1	Convolution Neural Network (CNN)	96.1	96	90
2	Artificial Neural Network (ANN)	67.5	49.3	32.47
3	Support Vector Machine (SVM)	89.9	89	90
4	K-Nearest Neighbors (KNN)	88	87	87

To visualize the performance of the model's confusion matrix of the neural network algorithm and for KNN and SVM is displayed in Fig. 6.

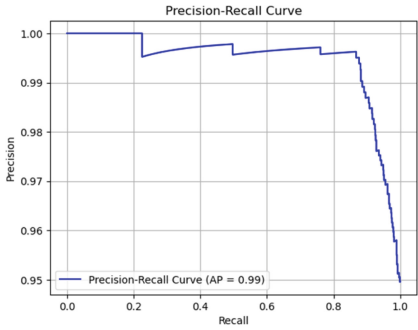


**Fig. 6.** Confusion matrix for various classifiers

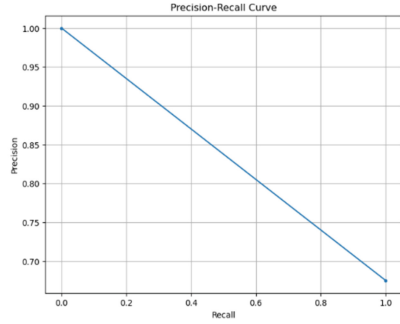
Confusion matrix consists of four terms which are True Positive, True Negative, False Positives, False Negative which future determine the evaluation metrics of any model. From metric table and confusion matrix it is clear that CNN based model architecture perform the classification task most precisely whereas ANN based model accuracy is not quite good. SVM and KNN based classify model also perform well. For further evaluation, precision-recall curve is used as shown in Fig. 7.

Precision Recall is a graphical representation between precision and recall for different classification thresholds or decision boundaries. In general, a model with a higher precision-recall curve closer to the top-right corner indicates better performance as it achieves high precision and high recall, which we can see in CNN, SVM and KNN but not in ANN. Through all the evaluation of the models we can see the classification model like Convolutional Neural Network (CNN), Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) perform very well for the classification task between crop seedling and weed.

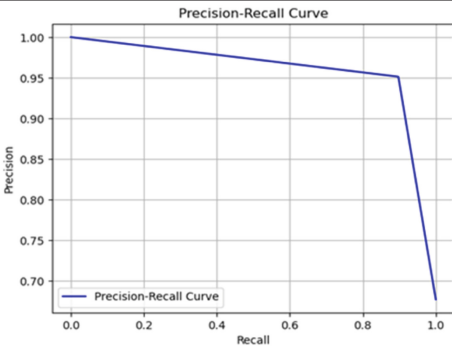




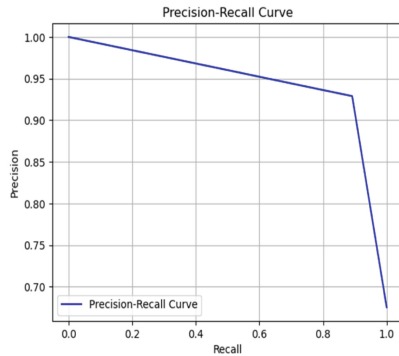
(a) Precision Recall Curve For CNN



(b) Precision Recall Curve For ANN



(c) Precision Recall Curve For SVM



(d) Precision Recall Curve For KNN

Fig. 7. Precision recall curves for various classifiers

## 5 Future Scope and Conclusion

The study shows that, Deep Learning based Neural networks and classifier used here helps to achieve the goal of classification between weed and crop on the image dataset. It provides us with the higher identification rate with more efficiency and less computation time. The strength of deep learning approach for Weed classification is its simplicity, accuracy, easy implementation for the required classification task. We have seen that using deep learning neural networks like CNN and classifier like KNN, SVM gives good results for the weed classification work in a farm. As in CNN based model under the defined hyperparameters gave accuracy around 96.1% and similarly in ANN based model under hyperparameters were gave accuracy around 67.5% which in comparison to CNN is very low, KNN and SVM have accuracy around 88% and 89.9% which is moderate. Although in this predefined dataset the computation time of all the models were very less but increase in dataset will increase the computation time and value of epochs can be increased to get the better result also. As the model will be applied more on real time data it would improve and eventually can be used for real time weed management in farm for agriculture which will be faster and more effective than the

traditional ways. However, these models will also face limitation to new undiscovered data. For Future work, it is worth trying to apply more pre-processing in dataset to achieve more refined images suitable for new and improved neural networks or image classifiers. More experiments on bigger real time dataset will take these models to achieve more real-life approach to solve classification problems of weed in a agricultural farm. Future research and study in this field using deep learning and machine learning will provide better opportunities to farmers in the agriculture sector. Finally, based on the result we believed that the image classification algorithm like K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and neural network like Convolutional Neural Network (CNN) works best for the classify weed and crop seedling. More work in them using technologies like deep learning and machine learning will help the farming sector both in time and cost in the following year.

## References

1. Fennimore, S.A.: Weed control, management, ecology, and minutia. The role of hand weeding in weed management. UC Weed Science (2014)
2. Brown, R.B., Noble, S.D.: Site-specific weed management, sensing requirements—what do we need to see? Cambridge Core, Published online by Cambridge University Press.(2017)
3. Rajagopal, A., et al.: A deep learning model based on multi-objective particle swarm optimization for scene classification in unmanned aerial vehicles. IEEE Access, 8, pp. 135383–135393 (2020)
4. Kamilaris, A., Prenafeta-Boldú, F.X.: Deep learning in agriculture and survey. Comput. Electron. Agric. **147**, 70–90 (2018)
5. Vvatsalgupta: Weed detection image dataset contains food crop and plant seedling. Kaggle (n.d.)
6. Giselsson, T.M., Jørgensen, R.N., Jensen, P.K., Dyrmann, M., Midtiby, H.S: A Public Image Database for Benchmark of Plant Seedling Classification Algorithms. preprint [arXiv:1711.05458](https://arxiv.org/abs/1711.05458) (2017)
7. Affonso, C., Rossi, A.L.D., Vieira, F.H.A., Ponce de Leon Ferreira de Carvalho, A.C.: Deep learning for biological image classification. Expert Syst. Appl. **85**, 114–122 (2017)
8. Sharma, N., Jain, V., Mishra, A.: An analysis of convolutional neural networks for image classification. Procedia Comput. Sci. **132**, 377–384 (2018)
9. Khari, M., Garg, A.K., Crespo, R.G., Verdú, E.: Gesture recognition of RGB and RGB-D static images using convolutional neural networks. Int. J. Interact. Multim. Artif. Intell. **5**(7), 22–27 (2019)
10. Bala, R., Kumar, D.: Classification using ANN: a review. Int. J. Comput. Intell. Res. **13**, 1811–1820 (2017). ISSN 0973-1873
11. Amato, G., Falchi, F.: kNN based image classification relying on local feature similarity. In: Proceedings of the Third International Conference on Similarity Search and Applications, SISAP '10. pp. 101–108 (2010)
12. Chandra, M.A., Bedi, S.S.: Survey on SVM and their application in image classification. Int. J. Inf. Technol. **13**(5), 1–11 (2021)
13. Shaner, D.L., Beckie, H.J.: The future for weed control and technology. Pest Manag. Sci. **70**(9), 1329–1339 (2013)
14. T. Giselsson, R. Jørgensen, P. Jensen, M.Dyrmann, and H. Midtiby: A public image database for benchmark of plant seedling classification algorithms, CoRR, [arXiv:1711.05458](https://arxiv.org/abs/1711.05458) (2017)
15. Sun, Y., Liu, Y., Wang, G., Zhang, H.: Deep learning for plant identification in natural environment. Comput. Intell. Neurosci. **2017**(4), 1–6 (2017)

16. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D.: Deep neural networks based recognition of plant diseases by leaf image classification. **2016**, Article ID 3289801 (2016)
17. Ashqar, B.A.M., Abu-Nasser, B.S., Abu-Naser, S.S.: Plant seedlings classification using deep learning. *Int. J. Acad. Inf. Syst. Res. (IJASIR)* **3**(1), 7–14 (2019)
18. Nandhini, N., Bhavani, R.: Feature extraction for diseased leaf image classification using machine learning. In: 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, pp. 1–4, ICCCI48352.2020.9104203 (2020)
19. Slaughter, D.C., Giles, D.K., Downey, D.: Autonomous robotic weed control systems: a review. *Comput. Electron. Agric.* **61**(1), 63–78 (2008)
20. Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D.: Machine learning in agriculture: a review. *Sensors* **18**(8), 2674 (2018)
21. Kumar, C., Kumar, V.: Vegetable plant leaf image classification using machine learning models. *LNNS*, **612** (2023)
22. Asha, P., et al.: Role of machine learning in attaining environmental sustainability. *Energy Rep.* **8**, 863–871 (2022)
23. Verma, P., Tripathi, V., Pant, B.: Comparison of different optimizers implemented on the deep learning architectures for COVID-19 classification. *Mater. Today* **46**, 11098–11102 (2021)
24. Tiwari, S., et al.: IOT-enabled model for weed seedling classification - an application for smart agriculture. *AgriEngineering* **5**(1), 257–272 (2023)