CLASSIFICATION AND DETECTION OF WEEDS AFFECTING SOYBEAN CROPS USING A 4-LAYERED CONVOLUTIONAL NEURAL NETWORK

Shreya Agrawal[1*], Sharon Christa[2], Sarishma Dangi[3]

[1] Dept. of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India
[2] Dept. of Computer Science and Engineering, School of Computing, MIT ADT University, Pune, India
[3] Dept. of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India
Email: [1] shreyaag252004@gmail.com

Abstract
Over the years, weed growth has evolved into a significant component that influences overall agricultural productivity and mostly incur to losses. Hence, timely weed detection and control can have a significant favorable impact on the productivity of crops and can provide valuable insights for precision farming. By using the concepts and techniques of machine learning, weeds can be easily identified and categorized. Convolutional Neural Network, one of the deep learning techniques, when experimented with led to better results than the majority of the others. In this work, we develop a CNN model that can be applied to weed pictures to recognize as well as categorize them depending upon their type. By using this model, one could considerably raise the effectiveness and efficiency of weed control in soybean crops.
Keywords: *Weed detection, CNN, Soybean Crops, Precision Farming, Deep Learning*

DOI: 10.48047/ecb/2023.12.si12.130

1. Introduction

Precision farming is a farming technique that assists farmers to ensure good crop yields through the season of sowing, irrigating and harvesting. It is a concept that helps to boost the production, reduce labor requirements, and effectively utilize fertilizers and other farming processes to the best possible level.
Precision farming makes use of modern technologies such as satellite and drone imagery, artificial intelligence and machine learning techniques, fog computing, edge computing etc. to improve the crop yield and also reduce labor costs. The images generated from drones are used to study crucial data about the crops such as crop status, soil conditions, environmental changes, yield etc. Based on this information, farmers can control and maneuver the course of their agricultural practices to ultimately enhance the crop productivity.
Modern day computational technologies used in precision farming include the use of Global Positioning System (GPS), Geographical Information System (GIS), remote sensing via satellite, big data, deep learning etc. The traditional methods of farming require more labor-intensive work and in most of the cases, lead to over-utilization of resources with a minimal impact on the crop productivity. As it is known that different soil types and crops have different

needs, but methods of traditional farming use resources such as water, fertilizers, etc. without considering the crop needs. Whereas precision farming divides a farm into separate areas, which allows diversifying management decisions for individual field parts, thus saving the resources and providing the crops with the right amount of nutrients along with the right set of agricultural processes.

Weeds are unwanted crops that serve as the biggest biotic constraint to agricultural production and yield. They are undesirable, tenacious, and harmful plants that grow along with other crops thus affecting their growth and development. As per [1], "*Weeds have become responsible for around 45% of the agriculture industry's crop losses mainly due to the competition with crops*". Weeds are plants that compete with other cultivated crops in terms of sunlight, water, nutrients, etc. Crops need nutrients and water to thrive, and as weeds proliferate in the field, they will compete with crops for the scarcely available resources, especially the seedlings of crops, causing them to wither or produce less as a result of a lack of nutrients or moisture. As per [2], *"By the mechanism of precision farming, we can identify weeds accurately at an early stage, and targeted controlling mechanisms can be applied to them, improving agricultural weeding efficiency".*

In India, soybean accounts for approximately 10 million hectare area of land under its cultivation, thus making it a significant crop. As per [3], *"Soybean is the world's most important seed legume, contributing 25% of the global edible oil and about two-thirds of the world's protein for livestock feeding"*. Weeds are regarded as the most significant issue in all significant soybean-producing nations. Summer annual weeds like Kochia (*Kochia scoparia*), giant ragweed (*Ambrosia trifida*), and common ragweed (*Ambrosia artemisiifolia*) mainly germinate in the early spring before the soybean planting. Even with cutting-edge technology, producers report significant losses because of weed interference with the crops throughout the season. According to estimates, weeds alone typically lower soybean output by 37%, while other fungi diseases and pests in agriculture are to blame for 22% of losses [4]. Thus, weed detection and management is significantly important to produce a high yield of soybean crops.

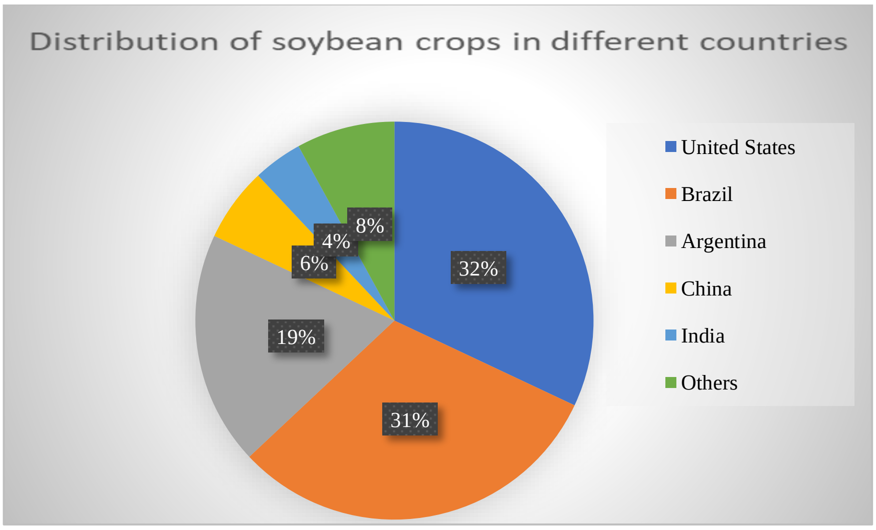

Fig 1. Distribution of soybean crops across the World

By using artificial intelligence-based techniques, these weeds can be timely identified, removed and thereby enhance the productivity of fields. In this work, we propose a deep learning technique called Convolutional Neural Network (CNN), and also give a comparative-based study on different techniques and their accuracy in detecting weeds in soybean crops. The main contributions of this work include:

- Implementation of a deep learning-based model i.e. CNN for timely and accurate weed detection and classification
- To check the accuracy of the trained model
- To present a comparative study between different models including MobileNetV2, ResNet50, fully convolutional network and our 4-layered Convolutional Neural Network.
- To present the future work in the area of weed detection via artificial intelligence-based techniques

The goal of this study is to provide a more efficient method for detecting weeds in a Soybean crop. The paper is divided into 5 sections, where Section I provides the necessary introduction to the area, Section II discusses the related work or literature in the area of weed detection via artificial intelligence-based techniques. Section III provides the proposed 4 layered CNN model and presents a detailed discussion for the same. Section IV analyses the model and compares it with the existing models including MobileNetV2, ResNet50, fully convolutional network and 4-layered Convolutional Neural Network. The discussion on future research directions in the aforementioned area is provided in Section V followed by the conclusion of this work.

2. Literature Survey

Traditional farming is a primitive practice of farming that involves the use of traditional tools like axe or hoes or sticks, intensive human labor, and the age-old beliefs of farmers. Agroforestry, intercropping, crop rotation, cover crops, traditional organic composting, integrated crop-animal farming, shifting cultivation, and slash-and-burn farming are examples of common traditional agricultural techniques. Although there are certain advantages associated with these types of practices such as limited use of fertilizers, resource utilization, biodiversity maintenance, and environment protection, there are certain negative impacts of traditional methods of farming which include degradation of soil due to very minimal or no use of fertilizers, crops getting infected by soil diseases, less efficient production or yield of crops. Before planting a crop that meet the requirements, farmers have to depend on the suitability of the land, the influence of the climate, and the availability of an adequate water supply[5][6]. The introduction of technology has completely changed the entire process of farming. The use of Artificial Intelligence (AI) means less human involvement, and it also analyses real-time data and helps agriculturists to predict soil conditions, crop health, and environmental conditions accurately. AI-trained machines can help study the data and generate autonomous irrigation and weed detection systems, that can contribute in increasing overall crop yield.
Different agriculture automation systems include field machinery, irrigation systems, greenhouse automation, automation fruit production systems, modern supply chain

management systems by using sensors, seedlings, and weeding robots and drones. In addition to fluid level sensors, which are used to measure the level of substances such as liquids, powders, and granular solids, soil sensors also keep an eye on the moisture content of the soil [7]. Robots that sow the seeds can be programmed to target particular agricultural farming areas. By doing this, labor and tedious farm work during sowing could be easily reduced. Drones, when paired with modern technologies, can be used for irrigation purposes and also for capturing images of the farms which can be used to analyze real-time data. Artificial Intelligence has several more applications in the field of agriculture, as represented in Figure 2. A farmer may accomplish more with fewer resources with the support of AI-powered farming solutions, that also improve quality and ensure speedy GTM (go-to-market) strategies for crops [8].

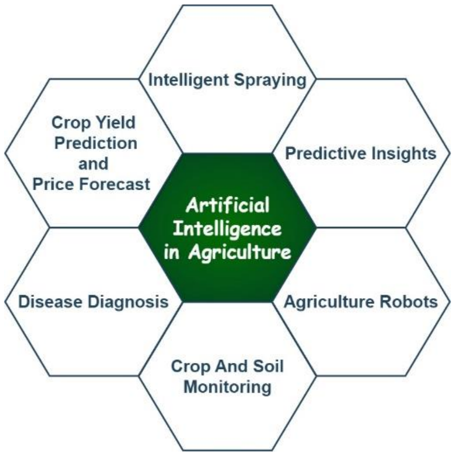

Fig 2. Various uses of Artificial Intelligence in Agriculture

In the agricultural field, there are two primary technologies deployed: namely proximity sensing and remote sensing. The data obtained from these technologies are used in testing the quality of the soil. The use of robotics and data-collection software in hardware solutions like Rowbot has already started to create the best fertilizer for crop cultivation in order to maximize crop yield [8]. Many researchers have reviewed and thoroughly examined IoT-based modern farming techniques. They claimed that micromanagement concepts for Statistical Analysis System (SAS) have begun to advance swiftly as a result of the development of intelligent low-powered wireless sensor technology that enables deployment in high densities. Since it was possible to closely monitor soil conditions, water-saving measures like smart irrigation technologies were created and put into use [7].

Although there is a growth of technology in the agricultural field, it has been observed that there are some issues such as farmers cannot use expensive technological facilities due to the lack of financial assistance. Also, due to the lack of smart and technological skills, farmers are reluctant to use such technical machinery [9]. The idea of smart farming and precision agriculture is to improve agricultural society, and it appears that the adoption of these technologies by the majority of farmers will be necessary for agricultural expansion to be viable.

3. Methodology

In this section, we present the research methodology for weed detection. The model is trained using a 4-layered CNN technique which has different types of hidden layers and a non-linear activation function called Rectified Linear Unit (ReLU). The process of weed detection by using artificial intelligence using deep learning in soybean crops is illustrated in Figure 3.

Data Collection

The dataset used for the research is a large dataset containing 15336 images, which are used for training and testing the algorithm. The dataset is 1.2GB in size and is further segmented into 4 classes: Soybean (7376 images), Soil (3249 images), Grass (3520 images), and Broadleaf (1191 images). Soybean contains the images of soybean crops, which is the main crop. Soil and Grass classes contain the images of soil and grass respectively, which are present just close to the main crops, and the Broadleaf class contains images of Broadleaf, which is a kind of weed that is often a problem as they compete with the crops for resources such as sunlight, water, and nutrients. The target of the algorithm is weed identification in soybean concerning soil, grass, and broadleaf. The constant management of crops is challenging due to the yearly, biannual, or perennial emergence of broadleaf weeds. They can multiply themselves easily and is hard to eradicate. “Popular broadleaf weeds are Chickweed, Clover, Dandelion, Wild Geranium, Ivy, Milkweed, Plantain, and Thistle” [10]. The difference in the structure and plant growth of Broadleaf weeds makes them easy to distinguish from narrow-leaved weedy grasses. Grassy weeds are different from broadleaf weeds as they initially appear like desirable grass, but later have harmful effects on the crops, thus they must be classified along with weeds.

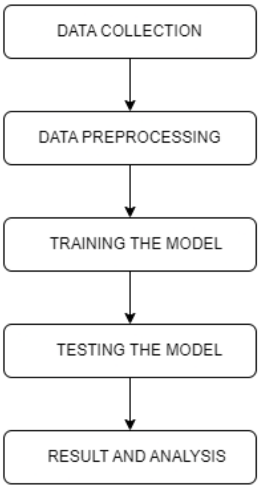

Fig 3. Process of detecting weeds in Soybean Crops

Fig 4. The images of (a) Soybean, (b) Soil, (c) Grass, and (d) Broadleaf

Data Preprocessing

To efficiently start the training process for the model, it is necessary to pre-process the images. Pre-processing of data means transforming raw data into a format that can be used by a machine learning or deep learning algorithm. The model imports *TensorFlow*[11] and *Keras ImageDataGenerator*[12] modules to perform image data preprocessing. The preprocessing techniques include rescaling the pixel values of the images by dividing them by 255 to normalize the data, randomly rotating the images by a maximum of 40 degrees, shifting the images horizontally and vertically by a maximum of 20% of the image size, applying shear transformations to the images, zooming in on the images by up to 20%, filling the images horizontally, and filling in any pixels in the image with the nearest pixel value. This is done to reduce the size of the dataset by removing irrelevant data, thus creating a high-quality dataset.

Training the model

The dataset used for training the model comprises about 10835 images. The model was trained on a machine with the configuration: 11th Gen Intel(R) Core (TM) i7-1165G7 @ 2.80GHz with a 2.80 GHz processor and 8 GB RAM with a 64-bit Operating System. The model classifies each image according to the class to which it belongs and is trained on a similar dataset 25 times and on images that are separated into a “training set” and a “validation set”. Convolutional Neural Networks (CNNs), a kind of deep learning neural network that is frequently employed in image and video recognition tasks, is the technique on which this model was trained.

Convolutional Neural Network

Neural network techniques, as their name implies, instruct the computer to process data in a manner modeled after the structure of the human brain, which has multiple layers and a large number of interconnected neurons or nodes. Neural network techniques can be categorized according to how the data flows from the input layer to the output layer. Some neural network techniques include Artificial Neural Networks (ANN), Convolution Neural Networks (CNN), and Recurrent Neural Networks (RNN).

The model is trained on the CNN technique, which has one input layer, multiple hidden layers, and an output layer. The hidden layers are composed of convolutional layers, pooling layers, and non-linear activation layers [13], [14]. The convolutional layers apply a set of filters to the input data, which are used to detect specific features in the image. The pooling layers then reduce the dimensionality of the data, while the non-linear activation layers introduce non-linearity into the model, allowing it to learn more complex representations of the data. The output layer of the CNN is a fully-connected layer, which produces a classification of the image according to its classes. The model is a 4-layered CNN model consisting of a stack of the following layers:

- *4 Conv2D layers, each of which applies a convolution operation to the input, with 64, 64, 128, and 128 filters respectively with the size of 3x3 and the activation function 'relu'.*
- *4 MaxPooling2D layers, each of which applies a max pooling operation to the output of the corresponding convolution layer with the pool size of 2x2.*

- *1 Flatten layer, which reshapes the output of the last pooling layer into a one-dimensional array so that it can be fed into the dense layers.*
- *1 Dropout layer, which randomly sets a fraction of input units to 0 at each update during training time, which helps prevent overfitting.*
- *1 Dense layer with 512 neurons, which applies a fully connected operation to the output of the dropout layer with the 'relu' activation function.*
- *1 Dense layer with 4 neurons, which applies a fully connected operation to the output of the previous dense layer with the 'softmax' activation function.*

The output of this layer will be a probability distribution over the 4 possible classes, with each element representing the probability that the input image belongs to that class.

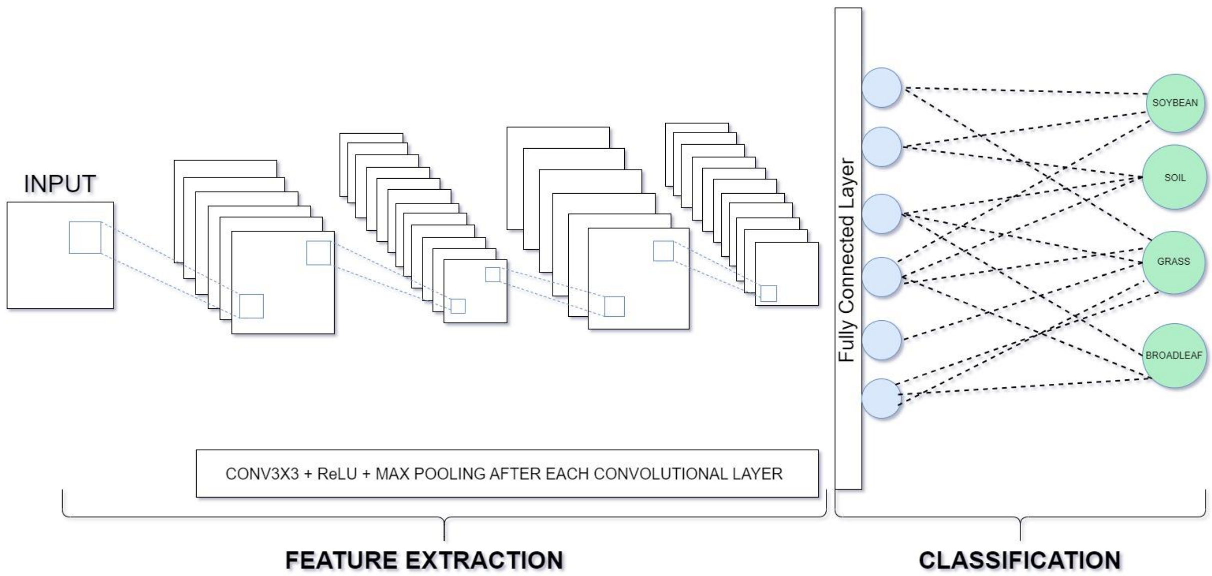

Fig 5. Trained CNN model

Testing the model

The trained model is later tested to generate an image's actual and predicted class. The model is also tested on approximately 4000 images to check the accuracy of the model. The images are classified according to their classes and the output displays the image, its actual class, and the predicted class. The testing of the model is done to check how the model is trained and whether it is not overfitted or under fitted on the training set. Thus, the tested model can now be deployed on an IoT device and used as a real-time invention, hence helping farmers to closely monitor their crops and plantation and detect weeds, and take effective methods to remove weeds from their plantation before the weeds destroy the plantation. Hence, farmers can get a better yield of crops.

4. Result and Analysis

The correctness of the model is evaluated by utilizing a hold-out for a validation set. The model was trained on a 70%-30% split between the training and validation datasets. The training accuracy of the model is approximately 98.8% and the validation accuracy is about 96.6%. This specifies that the 4-layered CNN-trained model correctly classifies most of the images, and can be used as a technique to detect weeds from a soybean crop field. A 4-layered CNN technique and a non-linear activation function called ReLU are used to train the model, which results in greater accuracy when tested. The table below lists some of the models that researchers have already worked with along with the percentage of validation for each model. Many researchers have also used techniques like MobileNetV2, ResNet50, and also CNN, but their accuracy is lower. As presented in Table 1, we can see that this model is more reliable and has better performance in terms of the prediction of weed seedlings from soybean crops.

Table 1. Comparison of proposed models with existing models

REF	MODEL	VALIDATION
[1], [2]	MobileNetV2	30.7%
[1]	ResNet50	82.3%
[15], [16]	Fully convolutional network	92.7%
This work	4-layered Convolutional Neural Network	96.6%

MobileNetV2 has the lowest accuracy in terms of weed detection and specifies that the model is not well-trained and produces the wrong output. ResNet50 has performed much better than MobileNetV2 as it has learned the distribution well. The key causes are twofold: first, it has more convolutional layers than MobileNetV2, and second, its classifier portion incorporates dropout, which encourages the network to learn a variety of gradient pathways to draw inferences [1]. Hence it detects weeds better than the MobileNetV2 technique. Because there are no completely connected layers in Fully Convolutional Networks (FCNs), the number of network parameters is drastically reduced. FCNs can accelerate and simplify network learning and inference while also greatly simplifying learning problems. Thus, they produce better accuracy than above both methods. The 4-layered Convolutional Neural Network, presented in this paper, classified the majority of soybean crops correctly and produced an accuracy of 98.8%, which is higher than all the other techniques.

In the model implemented in this paper, the case of overfitting and underfitting is also taken into consideration. An overfitted model is a model which has high training accuracy but comparatively low validation accuracy. An under-fitted model is just the opposite where it has low training accuracy and hence lowered validation accuracy. The number of epochs is just accurate in this training model, hence the problem of overfitting does not arise in the model [17].

The 4-layered CNN model consists of convolutional, pooling, and a fully-connected layer at the output which produces the correct classification of the image according to its classes. The model presented in this paper performs better than a normal CNN because it has more layers with a non-linear activation function ReLU, which enables it to collect input data with more complicated and abstract properties. Also, it enables the model to learn more intricate and non-linear correlations between input and output data. The ability to train and refine 4-layered CNN models has increased thanks to improvements in hardware and computing capacity, which may improve performance. A standard CNN only has two to three levels that don't have these properties. As a result, this model is superior to other deep learning techniques since it offers a greater accuracy rate and is also simpler to deploy.

5. Future Works and Conclusion

In conclusion, the research on weed detection in soybean crops using deep learning was successful in studying and developing a model that can accurately detect weeds in soybean crops. The model was trained on a dataset of images containing both soybean plants and various types of weeds. The use of deep learning in weed detection can greatly assist farmers in reducing labor costs and increasing crop yields. This highlights the potential for deep learning in agriculture and the importance of continued research in this field.

There are several areas for future work in the research on weed detection in soybean crops using deep learning. Some potential areas include:

- *Expanding the dataset*: The current dataset used for training and testing the model is limited in terms of the number of images and the variety of weeds included. Expanding the dataset to include more images and a wider range of weed species would help to improve the model's performance and make it more robust.
- *Incorporating other forms of data*: The current model is based on image data alone. Incorporating other forms of data such as weather data, soil data, and plant growth data could help to improve the model's accuracy and provide additional insights into crop growth.
- *Real-time implementation*: The current model is trained and tested on a dataset of images. A future direction would be to implement the model in real-time, for example, by using cameras mounted on drones or tractors to detect weeds in fields.
- *Developing a mobile application*: Developing a mobile application that utilizes the weed detection model could be a valuable tool for farmers, enabling them to quickly and easily identify weeds in their fields.

Overall, this technology has the potential to significantly improve the efficiency and effectiveness of weed management in soybean crops, and future works could help to further improve the model's performance and make it more widely applicable.

References

[1] N. Razfar, J. True, R. Bassiouny, V. Venkatesh, and R. Kashef, "Weed detection in soybean crops using custom lightweight deep learning models," *J Agric Food Res*, vol. 8, p. 100308, Jun. 2022, doi: 10.1016/J.JAFR.2022.100308.

[2] M. Dou, Z. Hong, and M. Shi, "An Improved Efficient Convolutional Neural Network for Weed Seedlings Detection," in *Proceedings - 2021 International Conference on Culture-Oriented Science and Technology, ICCST 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 285–289. doi: 10.1109/ICCST53801.2021.00067.

[3] D. K. Agarwal, S. D. Billore, A. N. Sharma, B. U. Dupare, and S. K. Srivastava, "Soybean: Introduction, Improvement, and Utilization in India-Problems and Prospects," *Agricultural Research*, vol. 2, no. 4, pp. 293–300, Dec. 2013, doi: 10.1007/s40003-013-0088-0.

[4] R. Vivian, A. Reis, P. A., L. Vargas, A. C. Camara Ferreira, and F. Mariani, "Weed Management in Soybean — Issues and Practices," in *Soybean - Pest Resistance*, InTech, 2013. doi: 10.5772/54595.

[5] A. Alam, A. H. Rizvi, K. Verma, and C. Gautam, "The Changing Scenario in Indian Agriculture: A Review," *International Journal of Scientific Research in Agricultural Sciences*, vol. 2, no. 7, pp. 118–127, Nov. 2014, doi: 10.12983/ijsras-2014-p0118-0127.

[6] D. K. Sreekantha, "AUTOMATION IN AGRICULTURE : A STUDY," *International Journal of Engineering Science Invention Research & Development*, 2016, [Online]. Available: www.ijesird.com

[7] S. Qazi, B. A. Khawaja, and Q. U. Farooq, "IoT-Equipped and AI-Enabled Next Generation Smart Agriculture: A Critical Review, Current Challenges and Future Trends," *IEEE Access*, vol. 10, pp. 21219–21235, 2022, doi: 10.1109/ACCESS.2022.3152544.

[8] V. Dharmaraj and C. Vijayanand, "Artificial Intelligence (AI) in Agriculture," *Int J Curr Microbiol Appl Sci*, vol. 7, no. 12, pp. 2122–2128, Dec. 2018, doi: 10.20546/ijcmas.2018.712.241.

[9] N. Aggarwal and D. Singh, "Technology assisted farming: Implications of IoT and AI," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, Jan. 2021. doi: 10.1088/1757-899X/1022/1/012080.

[10] "Broadleaf Weeds — Department of Plant Science." https://plantscience.psu.edu/outreach/plant-id/broadleaf (accessed Feb. 20, 2023).

[11] "TensorFlow." https://www.tensorflow.org/ (accessed Jul. 02, 2023).

[12] "Keras: Deep Learning for humans." https://keras.io/ (accessed Jul. 02, 2023).

[13] S. Bhadula*, S. Sharma, P. Juyal, and C. Kulshrestha, "Machine Learning Algorithms based Skin Disease Detection," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 2, pp. 4044–4049, Dec. 2019, doi: 10.35940/ijitee.B7686.129219.

[14] L. Singh, P. Chattopadhyay, J. Kaur, K. Chattopadhyay, and N. Mishra, "Machine Learning Based Prediction and Impact Analysis of Various Lockdown Stages of COVID-19 Outbreak – A Case Study of India," *Journal of Graphic Era University*, Jun. 2021, doi: 10.13052/jgeu0975-1416.922.

[15] Z. Wu, Y. Chen, B. Zhao, X. Kang, and Y. Ding, "Review of weed detection methods based on computer vision," *Sensors*, vol. 21, no. 11. MDPI AG, Jun. 01, 2021. doi: 10.3390/s21113647.

[16] P. Lottes, J. Behley, A. Milioto, and C. Stachniss, "Fully convolutional networks with sequential information for robust crop and weed detection in precision farming," *IEEE Robot Autom Lett*, vol. 3, no. 4, pp. 2870–2877, Oct. 2018, doi: 10.1109/LRA.2018.2846289.

[17] A. Ghildiyal, K. Singh, and S. Sharma, "Music Genre Classification using Machine Learning," in *Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 1368–1372. doi: 10.1109/ICECA49313.2020.9297444.