iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VI Month of publication: June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35173

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Weed Classification on Images using Machine Learning

Rafath Samrin[1], Oruganti Nikhitha[2], B. Praveena[3], B. Durga Prasad[4]
[1]Associate Professor, [2, 3, 4]Student, Department of CSE, CMR Technical Campus, Hyderabad, India

Abstract: Classify the plant species images so that the machine distinguishes between crop seedlings and weed. The images include some stones in the background, we don't want the machine to learn the stones pixels and base its prediction on all the components of the seedlings + stones. Color-based segmentation: Keep just the green color pixels keep just the seedlings pixels. Clean plant images without stones in the background. Ready to be fed to the model. Weed control is considered as a major obstacle for the farmers in farming. The plant productivity decreases due to the poor control of the weed in farming. So inorder to increase the plant productivity we use the concept of weed classification.
Keywords: weed, seedlings ,stones ,color ,images.

I. INTRODUCTION

Weeds are the most dangerous ones that destroy the yield of the crop and as well as the quality of it. As the yield of the crop decreases, there is an increase in the weed i.e it is indirectly proportional to each other. Image segmentation technique was used. We start our image pre-processing by importing a Small-flowered Cranesbill image as a test image. We will try to process this image in order to get rid of the unnecessary pixels. Feeding this picture to the model directly will result in a target leakage. The classifier will learn the stones and the container's pixels and build its prediction upon these pixels as well. This is something we would like to avoid.

We start our pre-processing with simply segment an object from an image based on color using Open CV. A popular computer vision library written in C/C++ with bindings for Python, we visualize the small-flowered Cranesbill image we opened above in RGB space to see the distributions of the color pixels.

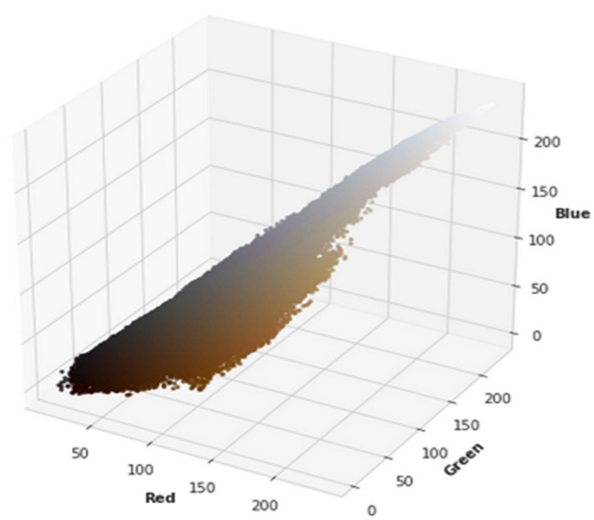
RGB is considered an “additive” color space, and colors can be imagined as being produced from shining quantities of red, blue, and green light onto a black background.

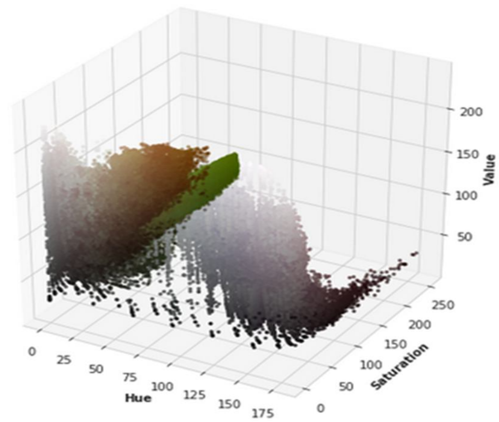
HSV stands for Hue, Saturation, and Value (or brightness), and is a cylindrical color space. The colors, or hues, are modeled as an angular dimension rotating around a central, vertical axis, which represents the value channel. Values go from dark (0 at the bottom) to light at the top. The third axis, saturation, defines the shades of hue from least saturated, at the vertical axis, to most saturated furthest away from the Centre.

To pick out a range color, we can use a color picking app online]. I found the green color ranges in a forum:

- minimum green (H=36, S=25, V=25)
- maximum green (H=70, S=255, V=255)

We will use the Open CV function cv2.inRange() to try to threshold our plant image with the minimum and maximum green values





A. Image Segmentation

It is a process of partitioning the digital image into image objects (also known as image pixels). The main aim of image segmentation is to change its representation into an image which could be easily understood and easily analized. It is used to locate objects, lines, curves, pixels etc.in images. The image segmentation is a set of pixels and segments that cover the entire image. Image segementation is mainly used for object recognition and recognizing the boundaries of the objects.

We use image segmentation to reduce the complexity of the object and it helps in easily analyzing the image.

Types of image segmentation

Threshold

Edge Based

Region based

Cluster

Water Shed

PDE

ANN

Each type of segmentation technique is used based on the project we develop. But here we will be using the

Thresholding Segmentation method

Edge Based Segmentation method

Region Based Segmentation method

Image Sharpening refers to any enhancement technique that highlights Edges and fine details in an image. It consists of adding to the original image a single that is proportional to a high pass filtered version of the original image.

B. Clustering

It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different.

Methods of Clustering

1) *Kmeans Algorithm:* Kmeans algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. The way kmeans algorithm works is as follows:

- Specify number of clusters K.
- Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
- Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.

2) *Agglomerative Clustering:* This is the most common type of hierarchical clustering used to group objects in clusters based on their similarity.

Steps to agglomerative hierarchical clustering

- Preparing the data
- Computing(dis)similarity information between every pair of objects in the data set.

- Using linkage function to group objects into hierarchical cluster tree, based on the distance information generated Objects/clusters that are in close proximity are linked together using the linkage function.

- Determining where to cut the hierarchical tree into clusters. This creates a partition of data.

C. Gradient Descent

Gradient descent is simply used to find the values of a function's parameters (coefficients) that minimize a cost function as far as possible.

Types Of Gradient Descent

Batch Gradient Descent

Stochastic Gradient Descent

Mini – Batch Gradient Descent

D. Root-Mean-Square Error (RMSE)

Mean square error (MSE) indicates the average difference of the pixels throughout the image. A higher MSE indicates a greater difference between the original and processed image. The root-mean-square error (RMSE) is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed.

II. LITERATURE SURVEY

Muhammad Hamza, states that Accurate mapping of weeds is a pre-requisite for weed density estimations and variable rate herbicide prescription. Semantic segmentation using deep learning is a promising technique for this purpose. Bottleneck for employing semantic segmentation is unavailability of labelled agriculture images at pixel level, which has been addressed in this paper. After acquiring high-resolution RGB images from canola field, background is segmented as a first labelling step, and then minority class pixel are manually labelled in background-segmented image. The trained model on this dataset maps only weeds and combines crop pixels along with background pixels. Our methodology has better results when we compare it to some of the most recent works involving semantic segmentation and weed detection. It has shown MIOU value of 0.8288 and FWIOU value of 0.9869 for ResNet-50 based SegNet model. Also, the paper makes comparison between UNET and SegNet meta-architectures for ResNet-50 and VGG16 base models. It is found that SegNet has better performance than UNET on this dataset. For base model feature extractors, ResNet-50 has superior performance than VGG16. In future work, soil properties will be included in the study for investigating a relationship between weed densities and soil characteristics with the purpose to facilitate variable herbicide prescription for different soil zones. Maintaining the Integrity of the Specifications.

Wenhua Mao, states that Weed detection is a key problem of spot spraying that could reduce the herbicide usage. Spectral information of plants is very useful to detect weeds in real-time for the fast response time. However, the cost of an imaging spectrograph-based weed detection system is too high. Therefore, the main objective of this study was to explore a method to classify crop and weed plants using the spectral information in the visible light captured by a CCD camera. One approach to weed classification was to directly use of G and R component of RGB color space. Another was to utilize the spectral information among the green band that hue was regarded as wavelength, and saturation was represented as reflectance. The result of statistic analysis showed that both of them using the G-R and H-S optimized segmentation line of crop and weeds could be used to detect weed (lixweed tansymnustard) from wheat fields. Moreover, the method of using the H-S optimized model could avoid the affect of lighting.

J.F.Thompson, States that Weeds are frequently distributed non-uniformly within fields. Herbicide use could be reduced either by applying it only to weed-infested areas or by applying a low dose rate to the whole field and normal dose rate to weed patches. A 4 necessary prerequisite is an effective method of weed detection. Tractor-mounted detectors do not appear to be feasible. Thus real-time weed detection and sprayer control are not currently possible. It is shown, however, that spatially variable herbicide application using a sprayer location system and a field map of weed location has potential. The data map would be constructed from a number of weed-location techniques based on image analysis, such as tractor-mounted video cameras, aerial photographs and manual field observation.

A.olsen, states that this work introduces the first, large, multiclass weed species image dataset collected . We anticipate that the dataset and our classification results will inspire further research into the classification of rangeland weeds under realistic conditions. Future work in this area includes: improving the accuracy and robustness of classifying this dataset, field implementation of our learning models as the detection system for a prototype weed control robot and investigating the use of NIR spectroscopy and hyperspectral imaging for weed species classification. The great lengths taken to collect a dataset including the real life complexity of the rangeland environment should allow for strong in-field performance.

Y.Gharde, states that in his study two MCSs were designed and assessed for the classification of paddy crops and weeds from the digital images. The approach was first to create digital images of paddy crops and weeds from paddy fields using digital cameras, and the Raspberry Pi camera was fixed at different heights from the ground to make the method device-independent. The soil and water background was removed. Texture, color, and shape features were extracted. Two selection-based MCSs were designed, one with calibrated random forest and calibrated SVM classifiers called as MCS-1 and another MCS with uncalibrated random forest classifier and uncalibrated SVM classifier called as MCS-2. MCS-1 and MCS-2 outperformed the single classifier systems. In addition, it was found that the MCS designed with calibrated classifiers performed slightly better than the MCS designed with uncalibrated classifiers. The features extraction and classification process proposed in this research work were applied on a publicly available paddy crop and weed dataset and results obtained are very promising. The study showed that the extracted features are good enough to classify paddy crops and weeds. This work could be used to recommend suitable herbicide for a particular type of weed based on the classification results to avoid 5 broadcast application of the herbicides. This could lead to the reduction of herbicide-resistant weeds, contamination of the groundwater, and other ill effects of overuse of the herbicides.

Syed Moazzam, states that Weeds are major cause due to which farmers get poor harvest of crops. Many algorithms are developed to classify weeds from crops to autonomously destroy weeds. Color-based, threshold-based and learning-based techniques are deployed in the past. From all techniques, deep-learning-based techniques stand out by showing the best performances. In this paper, deep learning-based techniques are reviewed in the case where these are applied for weed detection in agricultural crops. Sunflower, carrot, soybean, sugar beet and maize are reviewed with respect to the weeds present in them. Deep learning structures and parameters are presented, and research Gaps are identified for further research.

II. ALGORITHMS AND TECHNIQUES

A. CNN

We will use Convolutional Neural Network (CNN) to build our model. CNN are similar to regular neural networks: they are made up of neurons and have learnable weights and biases. The main difference is: the CNN takes advantage of image and accepts an image as a image with width, height and depth while the regular neural nets flattens the input. So, it is not good in identifying some features of an image while CNN easily can. The main building blocks of a CNN architecture are explained in the methodology section. The classifier I built is a Convolutional Neural Network, which is a state-of-the-art algorithm for image classification problems. It needs a large amount of data. Fortunately, we were having enough data to train our model. This algorithm takes an image as an input, passes through a sequence of layers and returns the probability for each class.

B. VGG16

VGG16 (also called Oxford Net) is a convolutional neural network architecture named after the Visual Geometry Group from Oxford, who developed it. It was used to win the ILSVR (Image Net). To this day is it still considered to be an excellent vision model, although it has been somewhat outperformed by more recent advances such as Inception and ResNet.

C. Encoding Targets and Data Split

After preprocessing the images, I encoded the class names using to categorical() from Kera's utils. Later, the data is split into training and validation sets in the ratio of 85:15 using train_test_split() from scikit-learn.

D. SVM

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N the number of features) that distinctly classifies the data points. 11 To separate the two classes of data points, there are many possible hyper planes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e. the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

III. PROCEDURE

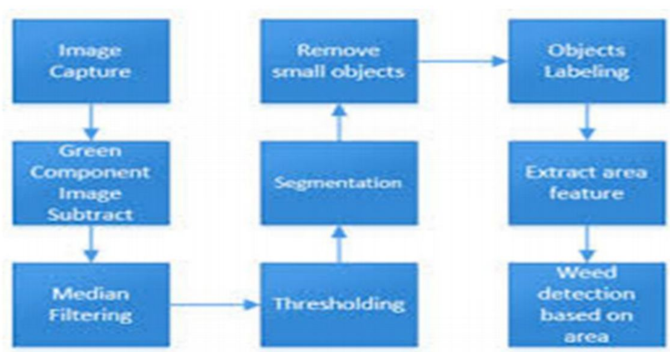


Figure Flow chart of Weed Detection

A. Data Exploration

The Plant Seedlings Dataset contains images of approximately 960 unique plants belonging to 12 species at several growth stages. It comprises annotated RGB images with a physical resolution of roughly 10 pixels per mm. The datasets are available for download in different versions i.e., Raw images, Cropped Images and Segmented Images. Here we use cropped versions of the dataset. A sample of raw type of images of dataset is show below.

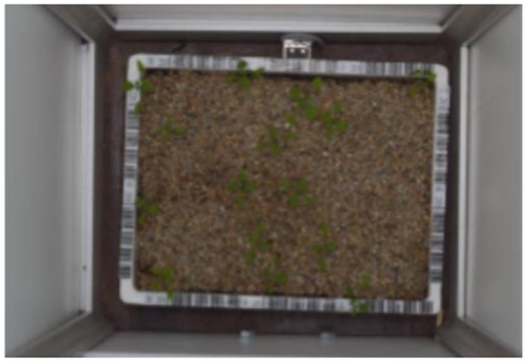
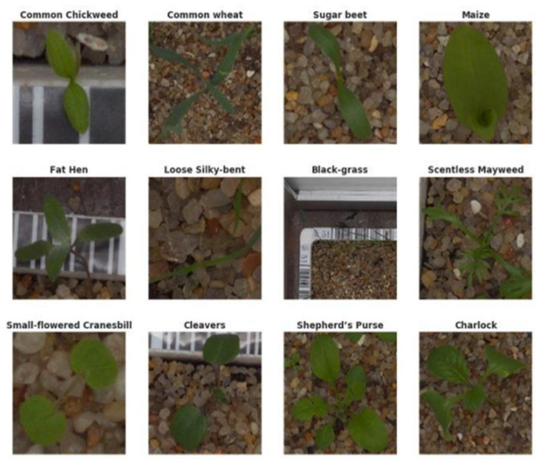


Figure Raw Chick Weed Image Sample

Since it is not possible to include too many species in the database, only a subset of high importance to the Danish agricultural industry are chosen to make the database. There are 12 species of images in the database:



After the data species are collected we apply the techniques of segmentation and classification so that we can easily distinguish the plants and the weeds. Based on HSV we can detect the plants easily and remove the weeds.

STEPS:

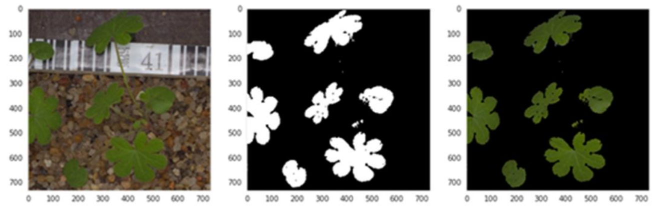
- Image Pre-processing.
- Visualize the plant image in RGB color Space.
- Visualize the plant image in RGB color Space.
- Segmentation.
- Visualize the plant image result of Segmentation.
- Build A computer Vision Model.
- Model Evaluation.

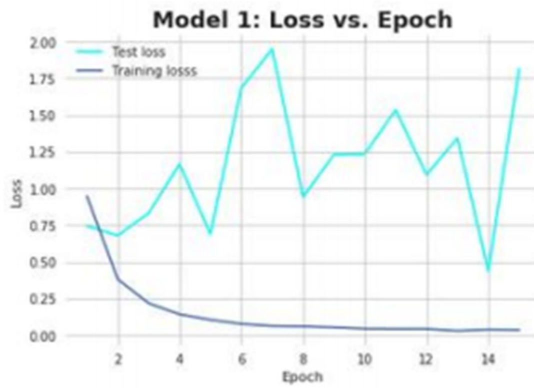
IV. RESULTS

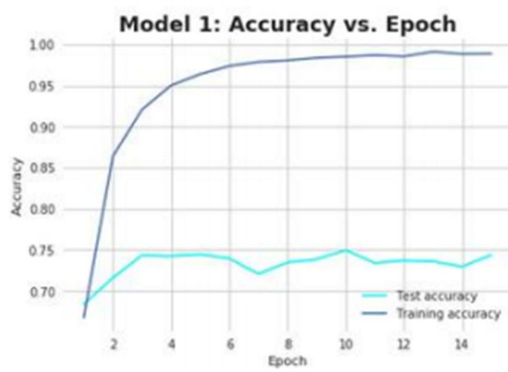
During the development of the model, a training set is used to train and a validation set is used to validate the model. The model we built might be sensitive to outliers because, the images that are present in the data set have intensities in range of 0-150 only. If a new image apart from these intensities is given for prediction, it may perform poorly.

- We used VGG16 Model in CNN to predict output.
- We also used SVM as a other machine learning algorithm .

- By using SVM we got the accuracy of 74%.





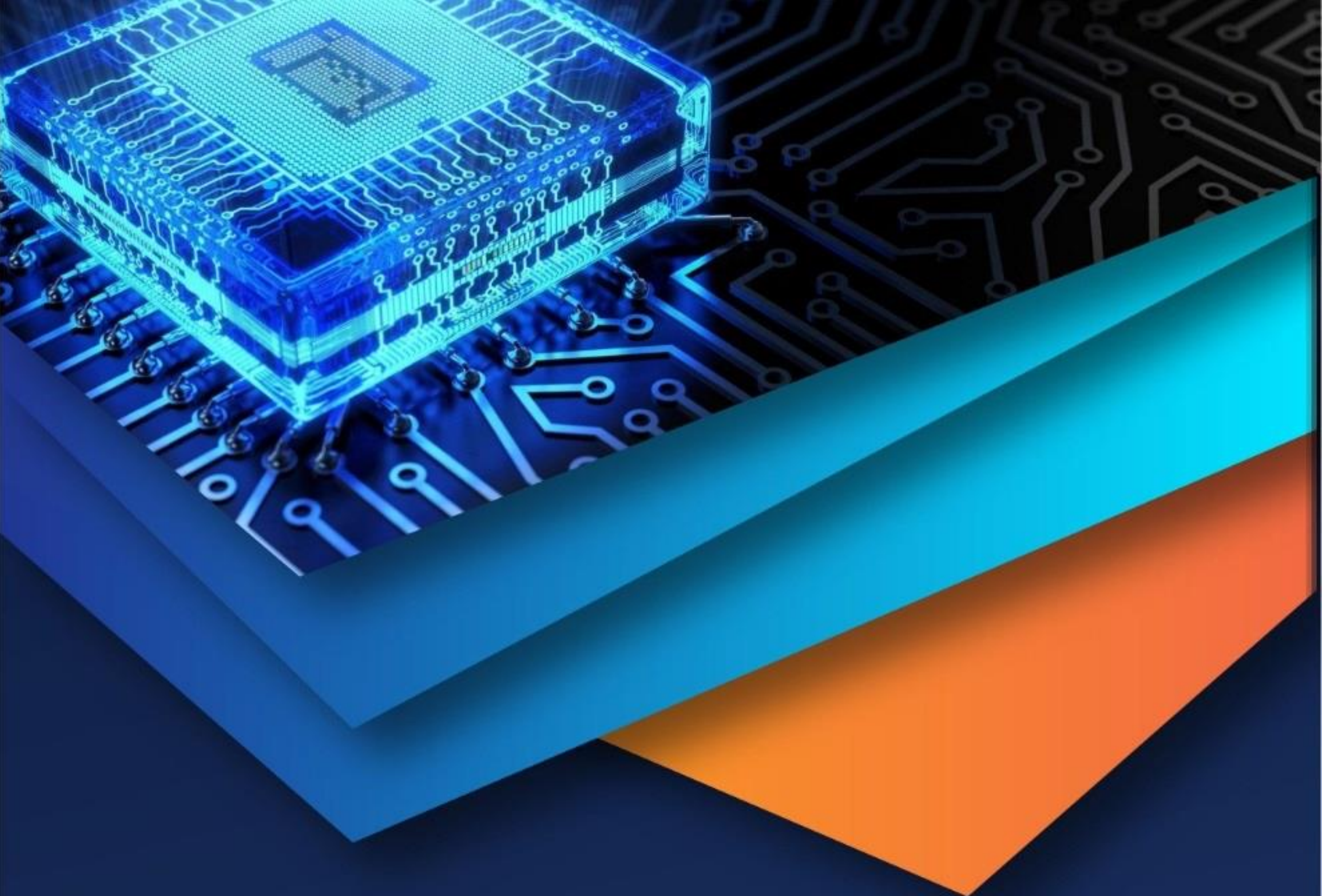


V. CONCLUSION

The model we built was with the aim of identifying weed plants in initial stages. The input plant images will be clustered into 2 groups namely Main Crop and the weed Plant. Only the green Component of the plant will be visualized due to we can group the images into their respective Clusters. Finally, we can detect the weed plant from the input images Model did find hard time in identifying Black-Grass from Loose Silky Bent. Model did very good job in detecting common chickweed, Loose Silky Bent, and identified Small-Flowered Cranesbill very accurately with f1-score of 0.99.

REFERENCES

1. F. Ahmed et al., "Automated weed classification with local pattern-based texture descriptors", *Int. Arab J. Inf. Technol.*, vol. 11, no. 1, pp. 87-94, 2014.
2. W. Wong et al., "Genetic Algorithm Tuned SVM Classifier for Weed Species Recognition".
3. G. Cohen et al., "One-class support vector machines with a conformal kernel. a case study in handling class imbalance" in Structural Syntactic and Statistical Pattern Recognition, Springer, pp. 850-858, 2004.
4. A. J. Ishak et al., "Weed Feature Extraction Using Gabor Wavelet and Fast Fourier Transform. M2USIC06-249", *Proc. Inter. Symposium on Information & Communications Technologies*, 2006.
5. A. Juraiza Ishak et al., "Weed detection system using support vector machine", *Information TheItsApplications 2008. ISITA 2008. International Symposium on*, 2008.
6. S. Frandina et al., "On-line laplacian one-class support vector machines" in Artificial Neural Networks and Machine Learning-ICANN 2013, Springer, pp. 186-193.
7. S. Tshewang, B. M. Sindel, M. Ghimiray and B.S. Chauhan, "Weed management challenges in rice (Oryza sativa L.) for food security in Bhutan: A review[J]", *Crop Protection*, vol. 90, pp. 117-124, 2016.
8. A. Dass, K. Shekhawat, A. K. Choudhary, S. Sepat, S. S. Rathore, G. Mahajan, et al., "Weed management in rice using crop competition-a review[J]", *Crop Protection*, vol. 95, pp. 45-52, 2017.
9. B. Vijayalakshmi and V. Mohan, "Kernel-based PSO and FRVM: An automatic plant leaf type detection using texture shape and color features[J]", *Computers and Electronics in Agriculture*, vol. 125, pp. 99-772, 2016.
10. T. Rumpfa, C. Römer, M. Weis, M. Sökefeld, R. Gerhards and L. Plümer, "Sequential support vector machine Classification for small-grain weeds discrimination with special regard to Cirsiumarvense and Galiumaparine[J]", *Computers and Electronics in Agriculture*, vol. 80, pp. 89-96, 2012.
11. Y. Le Cun, Y. Bengio and G. Hinton, "Deep learning[J]", *Nature*, vol. 521, pp. 436-444, 2015.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)