Wireless Personal Communications (2023) 133:2379–2410
https://doi.org/10.1007/s11277-024-10873-2

Deep Learning Techniques in Leaf Image Segmentation and Leaf Species Classification: A Survey

Anuj Kumar[1] · **Silky Sachar**[1]

Accepted: 20 January 2024 / Published online: 22 February 2024

Abstract
Plants have elemental importance for all life forms. The research areas in the field of plant sciences for botanists and agriculturists include the identification of plant species, classification of weeds from crops, detection of various diseases that hamper the growth of plant, and monitoring the growth and its semantic interpretation. Trained botanists can easily identify plant species based on the leaf shape, texture, structure or arrangement of leaves, however, the recent trend in smart agriculture demands the use of intelligent systems for the same task. Last decade has seen an enormous rise in the use of deep learning in the field of automatic plant species recognition based on the leaf images. In this work, we have surveyed various state-of-the-art deep learning techniques (Convolutional Neural Networks, Mask RCNN, Recurrent Neural Networks, Generative Adversarial Networks) that have been applied in the field of leaf image segmentation (separation of leaf from the whole image) and classification of leaves into various species. This contribution will help the new researchers in the field to get a foundation on the trends being employed in deep learning for generation of synthetic leaf images, segmentation and classification of leaves into various species. Various difficulties and future scope have also been presented.

Keywords Leaf identification · Deep learning · Segmentation · Image processing · Classification

1 Introduction

Plants serve the ecological balance in nature and also provide the human beings and animals with food, shelter, oxygen among other innumerable benefits. Most species now also face the risk of extinction. With the increase in population, there is a need to be aware of our natural environment and the quality of crops being consumed. Since the advent of computer vision techniques and artificial intelligence methods, assessing the genetic information of plants, their classification into different species, identifying different diseases

✉ Silky Sachar
silky.sachar@gmail.com

Anuj Kumar
anuj_gupta@pu.ac.in

[1] Department of Computer Science and Applications, Punjab University, Chandigarh 160014, India

in plants, separating weeds from the crop etc. have become convenient for the breeders and agricultural scientists as the results can be achieved in less time with less complexity and without human bias [54]. Automated plant classification systems can efficiently distinguish between therapeutic and harmful plant species with the assistance of programmed databases. Observing the latest trend in image processing, deep learning has given exemplary results in the field of automatic identification of plants based on leaf images [20, 22, 47]. Deep learning has also been a go-to tool for agriculturists in solving various research problems like identification of plant diseases[24, 31, 56], leaf counting [3, 41], weed classification [12], phenotypic analysis [5]. Deep learning consists of artificial neural networks that may have multiple hidden layers between input and output layers for automatic feature extraction from the input [26]. Extensive training of deep networks is accomplished with the help of Graphical Processing Unit (GPU). Deep learning finds many applications in image processing as it eliminates the need of hand-crafted feature extraction techniques that were traditionally employed in Machine Learning.

Figure 1 shows various building blocks employed in creation of deep learning model for the purpose of identification of plant species by inputting a leaf image. The leaf image datasets available publicly are of two types, one type consists of Unsegmented leaf images which contain background information like soil, pot, branch along with the leaf. Other type of leaf datasets contains Segmented leaf images which only have leaf against white/black background. Datasets are discussed in forthcoming Sect. 4. To increase the performance of the model, the input image dataset goes through basic pre-processing operations including noise removal, contrast enhancement, data augmentation. The dataset is then divided into training and testing datasets. The training dataset is expected to contain 70–80% images and the rest go to testing set. The purpose of training set is to train the network's millions of parameters to understand the patterns of input images and then its performance is evaluated on the test set. The performance metrics determine how the model performed while solving the task.

Contributions of different deep learning models by researchers, motivated us to pursue this work. In the literature we found different review articles enlisting different image processing techniques in identification of plant species. A systematic survey [51] presented an analysis of machine learning techniques used in identification of plants based on different

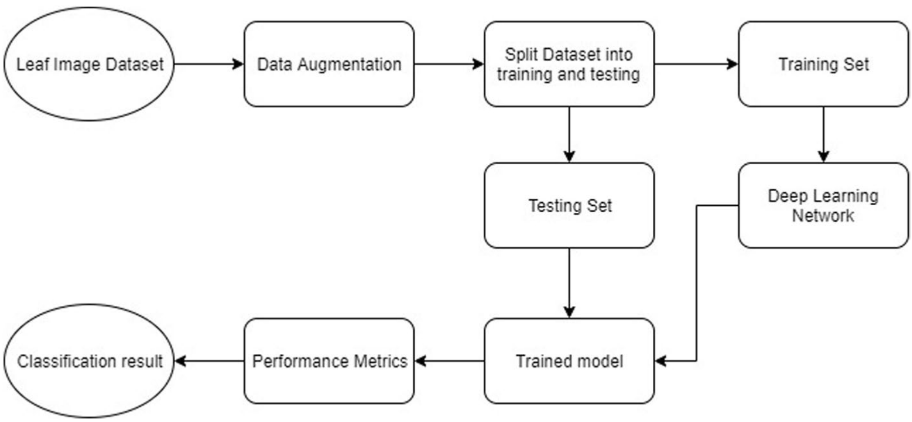

Fig. 1 Block diagram of deep learning-based model for automatic identification of plant using leaf image

parts like fruit, flower, leaves, branches etc. Another survey of techniques adopted in pre-processing, feature extraction and classification has been depicted in (Sachar and Kumar [43]). A review of deep learning techniques adopted for phenotypic plant recognition [57] also discusses the use of CNN for extraction of phenotypic traits of plants. Through this work, we have elaborated types of deep learning networks. Our contribution is two-fold. First, we have discussed different deep learning networks used in automatic segmentation of the leaf from the background. Second, we have discussed the deep learning networks used for the classification of leaf images into their species.

The rest of the paper is organized as follows: Methodology adopted for the selection of papers for this review is presented in Sect. 2. Section 3 discusses the use and potency of deep learning-based models in image processing with applications in image segmentation and image classification. Section 4 gives information of various datasets used in species identification using leaf images. Contribution by various researchers using deep learning in the field of image segmentation is presented in Sect. 5. Section 6 discusses various state-of-the-art work in the field of leaf image classification using deep learning. Section 7 discusses various challenges faced by researchers while employing deep learning and future scope in the field and Sect. 8 concludes the work.

2 Methodology

The step-by-step approach adopted for the selection of papers addressing the application and use of deep learning in two main research areas i.e. Leaf Segmentation (classifying the leaf pixels from non-leaf pixels) and Classification of leaf species into their respective classes is depicted in Fig. 2. Main highlights of the process are as follows:

- Firstly, about 150 research papers based on image processing in the field of agriculture were selected emphasizing mainly on the quality and recently published work in scholar databases like ScienceDirect, IEEE, and Google Scholar. The keywords used for selection were "Leaf", "Plant","Machine Learning", "Deep learning", "Segmentation".
- Highly cited papers published in the years 2018–2022 were selected as shown in Fig. 3.
- For further filtering of papers, Two criteras were used. One: the papers based on Leaf Segmentation and Two: the papers based on Plant identification/recognition using the leaf image were selected.
- From the above list, all the papers that used conventional machine learning approaches were deleted.
- The selected papers were the most recent papers proposing deep learning based solution and were taken into account.

3 Why Deep Learning?

Deep learning has been an effective technique to solve many problems pertaining to image recognition in various field like recognition of plants using flower species [14], identification of leaf diseases [21], detection of histopathological cancer diagnosis [30], cancer prognosis prediction [61] among many more. Mostly used Deep learning

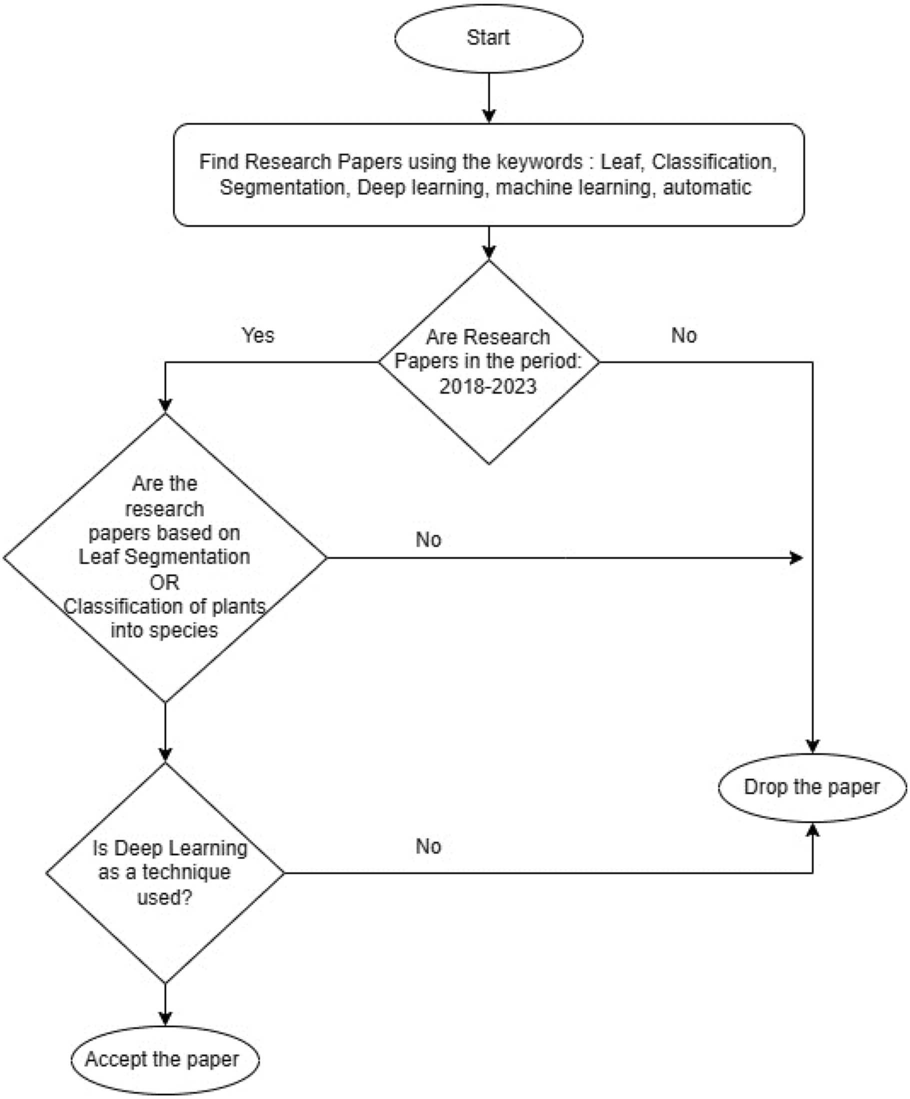

Fig. 2 Depiction of methodology adopted for paper selection

concepts are highlighted in Fig. 4. Deep learning techniques are part of supervised learning techniques with Convolutional Neural networks (CNN) beings used for all image classification tasks. The architectures like Mask RCNN (Region-based Convolutional Neural Networks), RCNN and faster RCNN are mostly used to solve segmentation problems. Hybrid CNN like Recurrent neural Networks (RNN) and Long Short Term Memory(LSTM) Networks being used in identification of leaves in complicated background are formed using hybrid CNNs. Transfer Learning using the pre-trained models like VGG-16, ResNet, Inception, GoogleNet etc. are widely used for classification

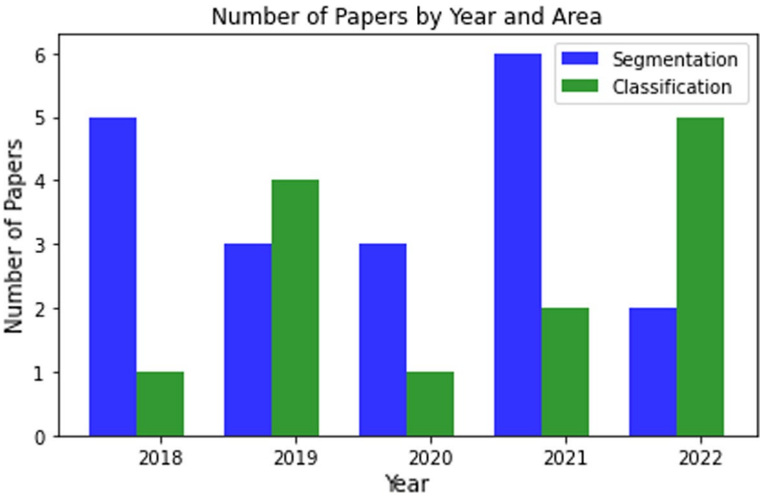

Fig. 3 Trends in research publications: segmentation vs. classification (2018–2022)

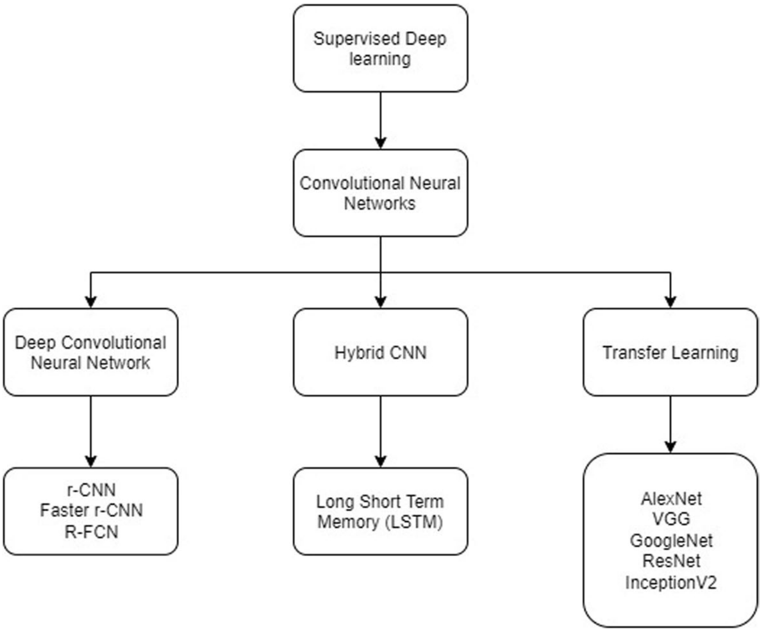

Fig. 4 Deep learning techniques

when the computing environments are constrained. Following are the most commonly observed advantages of deep learning:

- The deep learning models eliminate the need of hand-crafted feature extraction methods which helps to automate the image identification process.

- The approach allows to develop end-to-end solutions which simplifies the over-all architecture.
- Transfer learning, which allows the users to optimize the pre-trained CNN models, for their own problems has made the development work easy for many researchers. The deep models pre-trained on large datasets for weeks on GPUs can be modified to solve similar problems with less number of resources.
- Availability of open-source and easy to learn languages (Python) which can be employed to develop deep learning models has led to an added advantage.
- Different datasets are freely available online, which can be used by research communities to come up with better solutions for problems at hand.
- Google colab provides online access to GPUs like Tesla K80. 12 GB RAM is provided for a 12 h session.

Commonly faced problem with deep learning is the need of huge amount of training data which allows the model to better generalize the patterns of the data for its accurate classification. Not only this, but there should be sufficient number of samples in each class (balanced dataset) [26]. Another issue is the amount of training time required even by the use of specialized computing resources (GPUs) [39]. These shortcomings can be overcome by using transfer learning technique. Deep networks that have been pre-trained on large datasets for weeks on high-end GPUs can use the knowledge to solve similar types of problems on smaller datasets, thus saving a lot of time, effort and resources.

3.1 Convolutional Neural Networks (CNN)

Deep networks contain many layers between input layer and output layer [18]. Most commonly used network for image recognition based applications is the Convolutional Neural Network (CNN). During the last decade, customized CNNs have found applications in many domains due to availability of free development platforms and graphical computing hardware. CNNs consist of convolutional layers, non-linear Rectified Linear Unit (ReLU) and pooling layers. Convolutional layers perform convolution operations on input images. ReLU layers are responsible for introducing non-linear layer to get rid of negative values without affecting other layers. The mathematical equation is shown below:

$$f(x) = \max(0, x) \tag{1}$$

Equation (1) shows the result when an input x is fed into the non-linear function.

Pooling layer can be programmed to find the average or maximum value in a given portion of an image to overcome the problem of overfitting or translation variance. Pooling and ReLU layers have no parameters to be learnt during training. The last layer(s) consists of conventional fully connected neural network. Fully connected means that every output from 1 layer is connected to each neuron in the next layer. This leads to learning a large number of parameters in this layer. In the following Sect. 4 we have discussed how different researchers have used the CNNs in different ways to solve image identification problems.

4 Leaf Image Datasets

- Aberystwyth leaf Evaluation Dataset (ALED) [2]: This dataset contains about thousand images of Arabidopsis thaliana plant, taken in "top-view". The images were acquired from the National Plant Phenomics Centre using the Photon System Instruments PlantScreen plant scanner. The dataset also contains ground-truth, hand annotations.
- Pl@ntLeaves Database [38] This database contains tree leaf images in real world. Division of dataset into 3 categories represent 3 conditions under which images are taken. The scan category presents flattened leaf images in white background, scan-like category has images in uniform background. The photograph category contains images taken directly from tree leaves in natural environments. The photograph category images have non-uniform background, color variations, overlapping, shadows etc.
- SunFlower leaf dataset: This dataset contains about 200 sunflower leaf images, 50 images per variety for 4 varieties along with their ground truth labels. Size of images is 1024×1024, captured in cultivation fields, using UAVs.
- Computer Vision Problems in Plant Phenotyping dataset [40]: It is developed to compare the performance of the state-of-the-art segmentation methodologies. It is further divided into 3 datasets namely A1, A2 and A3 where A1 contains RGB images of Arabidopsis plants of wild-type, A2 contains RGB images of 4 varieties of Arabidopsis with different shaped and sized leaves, A3 contains tobacco plant images. The number of training set images in A1, A2 and A3 are 128,31, 27 and testing set are 33,9 and 56 respectively. The background appearance and composition is different for all 3 datasets. The dataset also addresses various complexities for segmentation like occlusion, out of focus scenes, leaf color variation etc.which makes the dataset more challenging and motivating for researchers to develop new segmentation algorithms.
- Flavia [55]: This dataset contains segmented leaf images. There are 32 species of leaves and each species has nearly 50–70 images. The images are taken against white background. All images are taken by scanner or digital cameras, of resolution 800×600 and stored in jpeg format.
- Swedish [45]: This dataset also contains leaf images taken from Swedish trees against white background. The dataset contains 15 species of leaves with 75 leaves in each class.
- LeafSnap dataset [25]: This dataset consists of 184 leaf species taken from trees in Northeastern United States.The authors also proposed a leaf classification system and a mobile application based on the same.
- Segmented Plant Seedlings dataset [13]. This dataset contains segmented RGB images of plant seedlings grown in greenhouse setting. All images have different spatial resolutions.
- MalayaKew Dataset [27]: This dataset contains a segmented and annotated dataset of leaf images against black background of 44 species of plants taken from Royal Botanical gardens, Kew, England.

5 Deep Learning in Segmentation

The pipeline adopted in Machine learning starts with the very important phase of image segmentation. The purpose of segmentation is to divide each image into characteristic regions and extract the target region. In this case, the image segmentation is performed to separate the leaf/green area from the non-green/non-leaf areas. The leaf or the green area is referred to as Foreground and the non-leaf area like soil, stem, branches, pot are termed as the Background. The whole process of segmentation significantly reduces the processing and inference time. Types of Segmentation can be semantic and instance segmentation. While achieving Semantic segmentation, the goal is to associate each pixel of an image with a class label example: leaf, pot, soil. Multiple leaves in a plant will be treated as single entity. To perform leaf counting for example, Instance segmentation will be taken into consideration as it will treat multiple objects of 1 class as different instances.

After the leaves have been segmented, the phenotypic traits of the plant could be extracted. Plant phenotyping helps to increase the yield of the crop. It can also help to determine the growth of the plant, the shape, surface etc. The main goal of the application of imaging techniques is to increase the output of plant phenotyping in non-destructive manner. In another direction, after segmenting the leaves from the background, further machine learning techniques could be applied for the extraction of features like shape, color, texture and venation. The feature vector obtained could be input into the classifier for detection of leaf diseases or detection of species. These applications lay stress on the fact that the accuracy of the final output dependents on the input supplied. If the quality of the segmented images will be fine, the expected output will be obtained and vice-versa.

The main aim of segmentation is to classify the image pixels as the foreground or background. The foreground pixels in the case of leaf image detection refer to the ROI (Region of Interest) leaf area which includes the leaf and the background pixels refer to the parts of the image other than the leaf, referring to the soil, pot, branches among others. The segmentation can be applied to the image in 2D and in 3D where the depth gives additional information to visualize data. The techniques are discussed in detail below:

5.1 Early Developments (Semantic and Instance Segmentation)

Major challenges that arise in the field of leaf segmentation are due to the variability in leaf shapes, partial overlap and occlusion by other leaves. Authors [53] used dataset from the Leaf Segmentation Challenge (LSC) which contains images of *Arabidopsis* and tobacco leaves with segmentation labels at the pixel level. To apply machine learning approaches for leaf segmentation, there is a need of large amount of manually annotated training data. Authors propose leaf instance segmentation framework by generating synthetic data to augment the real plant datasets. After creating an inspiration leaf and applying random deformation a textured and segmented leaf was obtained. As the arrangement of *Arabidopsis* leaves is a rosette pattern, uniform distributions were used to provide a variety of leaf positions. The combination of images of *Arabidopsis* plants were used to develop the state-of-the-art segmentation technique based on Mask-RCNN. Mask RCNN consists of a feature extractor providing the output to Region Proposal Network (RPN) and then into further producing box classification, box regression and object mask. ResNet 101 backbone was used with Feature Pyramid Network. 256 ROI per image were used during the training. The authors achieved 90% symmetric best dice score on the A1 test set.

5.2 Advancements in Deep Learning and Complex Segmentation

Mask RCNN and its applications were observed in the field of leaf instance segmentation and counting [59]. A1, A2 and A3 datasets from the LSC CVPPP challenge were used for training and testing. Base Conv Net ResNet 50 was used to obtain the feature maps. Then RPN recommends ROI which may include the instance base. ROI features are re-aligned through ROI Allign layer. Based on the result of leaf instance segmentation, leaf counting was also performed. To evaluate the results, SBD, DiC and mod of DiC were adopted as metrics. As Mask RCNN combines object detection, location and segmentation, it is robust to identify small leaves even in overlapping conditions. Contributing towards the segmentation of leaves in their natural habitats, authors in [34] created a labelled Dense-leaves dataset and also a CNN that can segment the overlapping leaves effectively from their backgrounds. The dataset contains 108 images of overlapping leaves from trees, bushes in and around Michigan State University. 20 leaves per image were outlined manually so that overlapped leaves could not be left out. It was made sure that the adjacent segments had no empty pixels between them and the background (branches, trunk) were labelled as 0. Fully convolutional- Pyramid network was used as binary classification problem to detect the boundary. After the boundary has been detected, an initial segment creation based on watershed algorithm was followed by segment merging. Dice, precision and recall of 0.915, 0.983 and 0.856 was reported which outperformed the state-of-the-art.

To perform automatic segmentation of leaf images in mobile applications from the complicated backgrounds, authors proposed a fully connected CNN [37]. The experiments were performed on the tomato leaf images captured under challenging field conditions. A comparative study was carried out to study the performance of 28 U-Net and 35 SegNet Convolutional Network Architectures architecture and a CNN named KijaniNet was proposed. The KijaniNet model was based on the multi-scale feature extraction from the encoder stages of the U-Net. The result was validated using 10 Fold cross validation and then the performance was also compared to the active polygon model and the Grab cut techniques. It was concluded that the SegNet was suitable for use in mobile applications and KijaniNet was suitable for the cases when the inference was made on the server.

The segmentation and classification of leaves in complicated backgrounds using Deep learning [59]. Due to lack of freely available datasets of leaves with complicated backgrounds, authors collected their own dataset of 15 species of leaf images associated with the South China Normal University. For the purpose of Annotating the dataset, MIT's Labelme was used. Segmentation was performed by using Mask R-CNN. Mask R-CNN consists of ResNet Feature Pyramid Network which forms the backbone of the algorithm. It extracts the features and present the feature map at different scales. The second module in the same, consists of Region Proposal Network (RPN) which uses sliding window to select the areas that contain different objects from the feature map. Next, the Region of Interest (ROI) is aligned by using bilinear interpolation and the same is passed on to the ROI classifier and bounding box regressor. The ROI classifier is also a CNN. The Classification was done by using transfer learning with the help of VGGnet classifier. The average Misclassification Error(ME) was 1.15% which is very less compares to the Otsu segmentation algorithm and Grabcut algorithm. The average classification accuracy was 91.5%.

5.3 Leveraging Synthetic Data and GANs

In the field of phenotyping, GAN-based model was proposed to create artificial images of Arabidopsis plants, focussing mainly on improving leaf counting [60]. Since the segmentation mask was used for conditioning, the model was able to produce high texture quality images. It was also able to produce realistic background for artificial images. The technique was evaluated using Mask RCNN model for leaf counting and it was observed that the average leaf counting error was reduced to 16.67% after augmentation using proposed technique. Another application of GAN was proposed in [42] where the authors produced synthetic images from real images of MalayaKew dataset. The dataset has 44 species of leaves having 3 types of shapes (lanceolate, lyrate and runcinate). The DCGAN (Deep Convolutional GAN) model was trained to segment the leaves into 3 shapes, using the combination of L1 and L2 regularizers also called Elastic Net.

5.4 Specialised Approach to Address Data Limitation

Since Deep learning models require a large amount of images for it to learn features of the data, a noteworthy solution has been provided by authors in [16]. In their work, the authors have proposed “Few-Leaf learning” which proves how only few images can be used to train a model to segment weeds from the background. The synthetic data was generated using image composition by pasting weed images on images with weed-free backgrounds. To predict the weed mask, PSPNet architecture with ResNet18 backbone was trained with synthetic images. Predicted mask was refined using the GrabCut technique. For effective evaluation, the technique was tested on 2 types of weeds *Rumex obtusifolius, Cirisium vulgare* with images taken from different fields and it was observed that the algorithm was able to generalized well.

5.5 Recent Innovations and Cutting Edge Techniques Based on Mask RCNN

In the field of plant phenotyping providing high throughput image processing, authors have proposed plant region segmentation based on the orthogonal transform coefficients and leaf counting based on Deep convolutional neural networks [40]. The algorithm was trained and tested on the Rosette plants on CVPPP dataset. After converting the image to YC_bC_r color space the plant region was extracted in the Cr plane. To remove noise, Wiener filter was applied followed by thresholding over the Lab and CMYK color space. The segmented region was hence sent as input to the fine-tuned DCNN models (AlexNet and VGG-19) and number of leaves were counted with the help of transfer learning. As a result, the FBD of 94.7% was attained on datasets (A1, A2 and A3) and also the Dice score of 93.72% was attained for tray plant images when compared to the existing methods. Mask RCNN technique for segmentation and 3D re-construction of 4 domestically managed Arabidopsis plants with leaves varying from 6 to 10, was proposed in [7]. The training set was created using the hand-marked outlines on the leaf images. Mask RCNN was used to create the 3D segmentation effect from the 2D images. The testing set was the point cloud top-view. 3D depth sampling in Stereoscopic vision was used to perform 3D reconstruction. Number of leaves identified were 81.5% and leaf coverage rate was 72.8%. To measure and study the morphological traits of

Digitized Herbarium Specimen (DHS) image samples [49] the images were reduced in size in the pre-processing stage. Mask RCNN was used as the base and then the residual block was re-designed to focus on the identity shortcut connections in the ResNet. The activation function chosen was Mish which outperformed the ReLu due to its' better information propagation ability. The segmentation results were evaluated using the Mean Intersection Over Union which reached 90.5% in cases of perfect, imperfect and leaf with missing parts.

5.6 Innovative Approaches for Specific Challenges Based on U-Net Architecture

To accurately determine the growth of plants, authors in [50] segmented the plant leaves using a deep learning architecture based on "U-Net". Four datasets from the Leaf Segmentation Challenge were taken for experimentation. All the images were resized to a standard shape and converted to RGB as expected by the CNN. To remove shadows and illumination effects, the images were first converted to HSV color space to apply CLAHE (Contrast limited Adaptive Histogram). The RGB images obtained were then input into the U-Net CNN. Dice score of 95.63% on the A1 dataset, 91.21% on the A2 dataset and 79.90% on the A3 dataset. Authors in [54] extracted the phenotypic features of plant such as leaf count, coverage and size. Based on the drone images, a bottom-up approach was used to perform simultaneous instantaneous segmentation of individual crop leaves in agricultural fields to extract the phenotypic traits. A clustering-based approach was proposed using full covariance matrix to cluster crop leaves and plant instances. The segmentation technique was also compared to the state-of-the-art techniques like Mask RCNN and Harmonic Embeddings. The proposed technique not only outperforms both but also performs simultaneous instance segmentation on both crop leaves and plants.

The advantages of Deep learning can be applied to achieve high throughput image processing and to realize the same, authors proposed a modified version of U-Net Architecture [5].. The experiments were performed on the CVPPP datasets. The modifications to the U-net were made in terms of convolution layers, activation and loss functions to attain high training speed and accuracy. Post segmentation, contour based, region-based shape descriptors and color information was obtained. The morphological parameters resulting from the segmented images were highly correlated with the measured values showing the effectiveness of the edge detection and segmentation.

Another modification of the U-Net architecture was proposed in [23] for the segmentation of plant leaves in complex backgrounds. For the segmentation prediction, image features were first extracted using the ResNets in the bottom-up path followed by up-sampling of the semantic feature maps in the top-down path. Then a feature fusion module was introduced to merge the up sampled results with the feature maps generated from the bottom-up path. The algorithm was compared with various other segmentation algorithms and F-measure, S-measure and MAE of 0.9686, 0.9596 and 0.0185 was obtained. While modifying the U-Net architecture, authors introduced a cross-layer feature future, Basic blocks to extract features, and patch learning in the training stage to accelerate the learning [46]. As segmentation is a pixel-level classification task, it tends to lose information if the input image size is small. The cross-layer feature fusion was developed to learn the semantic information from the shallow to deep layers. Basic blocks of ResNet were used to reduce overfitting. Patch learning method slices the input images into equal sized image blocks and processing the labels in the same way. To enhance the leaf segmentation and counting for the purpose of plant phenotyping, and to overcome the challenges like occlusion, varying brightness, shadows and blur, authors in [3] proposed an encoder-decoder model based on Eff-Unet++. For the purpose of feature extraction, EfficientNetB7 was used as the encoder and Unet++architecture with re-designed skip connections was used as the decoder. The model was trained and tested for 3 datasets namely: KOMATSUNA (plant images captured indoor), MSU-PID(multi-modality plant image dataset) and CVPPP(benchmark dataset for leaf segmentation containing Arabidopsis and Tobacco images). When compared with the state-of-the art Deep learning architectures(InceptionResV2-UNet, DeepLabV3,ResUnet, UNet and UNet++), the proposed model outperformed them evaluated using FBD, Best Dice, Intersection Over Union and Difference in Count.

5.7 3D Segmentation and High Throughput Techniques

Shi et.al advocate for the fact that 3D image of a plant provides more traits than its 2D counterpart and makes segmentation more effective [44]. They used 10 cameras to obtain multi-view set up of 62 tomato seedlings in early stage of development having 2 leaves each. To achieve better results in the segmentation, gray-scale cameras were used as they give a better transition of background from the foreground. After acquiring the images, they performed 2D semantic segmentation using FCN based on VGG16. Ground truth dataset was created using Label Me. Target classes were background, leaf, stem and node. Further, instance segmentation was performed using Mask-RCNN. To enhance the segmentation results, the 2D segmented images from different viewpoints were integrated into a 3D environment using voting strategy. Rviz cloud annotation tool was used to create the ground truth dataset. After determining the 2D to 3D correspondences, the 3D point cloud was obtained using shape-from-silhouette method. The performance of the system was evaluated using precision, recall and F1-score. Welch's t-test was used to compare the performances of 2D and 3D multi-view segmentation. Significant improvement in the performance was observed in 3D method compared to the 2D method. Authors propose DeepSegV3Maize, a pipeline developed to obtain high quality 3D point clouds of maize shoots and perform its automatic segmentation [29]. For High-throughput data acquisition, MVS-Pheno was used to capture multi-view images of maize in different growth stages and point clouds were reconstructed. The original point clouds were down-sampled and normalized. Label3DMaize was used for data annotation. After the creation of the dataset, PointNet deep learning model was used for stem-leaf and organ instance segmentation. Following the segmentation, phenotypes namely leaf length, width, inclination, leaf growth height, plant height and stem height were extracted. The Mean precision and F1-score for stem-leaf and organ instance segmentation was reported as 0.91,0.85 and 0.94,0.93 respectively.

In summary, the field of leaf segmentation has evolved from basic binary classification of pixels to sophisticated 3D modelling and synthetic data generation. Each study mentioned not only addressed the limitations of its predecessors but also paved the way for future innovations. This evolution underscores the dynamic and progressive nature of machine learning in plant phenotyping, where each development has been a stepping stone towards more accurate, efficient, and versatile segmentation techniques (Table 1).

Table 1 Leaf Image Segmentation techniques

Ref	Dataset	Segmentation Technique	Pros	Cons	Metrics
Ward et al. [53]	Dataset from CVPPP Leaf segmentation Challenge- Arabidopsis plant	ResNet backbone with Feature Pyramid for Mask RCNN	Achieved good results on 4 datasets of CVPPP leaf seg challenge	Texture of synthetic data need to be improved	Symmetric best dice (SBD) score = 90%
Xu et al. [58]	Leaf segmentation challenge Dataset. A1, A2, A3 datasets	Mask RCNN for segmentation and counting	Works for small-shaped and overlapping leaves. The ROI proposition takes anchors at multiple scales which makes small and big leaves visible to the model	Segmentation accuracy decreased slightly when no,of leaves in an image became 19	AP (Average precision), SBD(Symmetric Best Dice), DiC(Difference in Count) and |DiC|
Morris [34]	Dense leaves dataset created. It consists of 108 images at resolution 1024 × 768, each of dense foliage from trees, vines and bushes on or near the Michigan State University campus	Background detection- Pyramid CNN. Segment Building- Segment creation (watershed algo), Segment merging	Segment dense leaves in overlapping condition	Not dealt with leaves with weak boundary information	Dice, precision, recall
Ngugi et al. [37]	Tomato leaf dataset. Captured under all weather conditions	SegNet and U-Net, Kijani Net	No user interaction required. No constraints on orientation, shape or illumination of target leaf. Designed specially for mobile devices	Not tested on other varieties other than tomato	KijaniNet scores 0.9766 mwIoU and 0.9439 mBFScore on the test set
Yang et al. [59]	Own dataset-15 species. Images in complex BG	Mask RCNN for segmentation and VGG16 for classification	ResNet with FPN can extract features at multiple scales	Deep learning requires large amount of data	Misclassification error = 1.15% and acc = 91.5%

Table 1 (continued)

Ref	Dataset	Segmentation Technique	Pros	Cons	Metrics
Praveen Kumar and Dominic [40]	CVPPP	The RGB plant image is converted to YCbCr color space	Independent of device and preprocessing software, robustness of the scheme, better performance even in the presence of shadow region, it does not require ground-truth images or segmentation masks and applicable in various environments and phenotyping platforms with minimal modification	Computationally expensive	FBD, precision, Recall, F1-score, DiC, Absolute DiC. MSE
Guldenring et al. [16]	Rumex obtusifolius (broad-leaved dock) and Cirsium vulgare (spear thistle)	Generate synthetic images using image composition. PSPNet + RESNet for semantic segmentation. Grabu cut to refine predictions	Few leaves used	Can overfit	Gacc- Global accuracy(Pixel acc over the dataset). IoU for foreground objects (region similarity). Boundary F1 score(contour similarity)
Trivedi and Gupta [50]	Benchmark dataset of LSC-810 images	U-Net deep learning architecture	Lightweight and less complex	Requires re-training if more images need to be segmented	Dice score

Table 1 (continued)

Ref	Dataset	Segmentation Technique	Pros	Cons	Metrics
Cao et al. [4]	CVPPP	U-Net	The improved model re-duces the number of training weights, and the training speed is high. The result is high in accuracy. The proposed model can realize and batch image segmentation while avoiding a lot of human resources	Not suitable when the weed is too near to plant and is same in color to the plant	IoU, Recall
Kan et al. [23]	1000 leaf images from ImageNet dataset	U-Shaped CNN model	Avoids the concept of over/under splitting	Tested for few samples of scenes	Precision-Recall Curve, F-measure, Mean absolute error, Structural Measure
Sun et al. [46]	CVPPP 2017	U-shaped network based on encoder and decoder	Reduction of GPU usage by dividing the original image into image blocks	Tendency to overfit	Precision, Recall
Triki et al. [49]	The herbar- ium Haussknecht of Friedrich-Schiller-Universität Jena, Germany, has digitized more than 30,000 type specimens	Mask RCNN	Different sizes, shapes, textures, and colours are collected from dif- ferent specimen images. These herbarium images contain mature, immature, and damaged leaves with different shapes and sizes	Not able to measure the trait without scale. Challenging in different adhesion levels and overlapping conditions	Precision, recall, MIoU, and k-fold cross-validation

Table 1 (continued)

Ref	Dataset	Segmentation Technique	Pros	Cons	Metrics
Bhagat et al. [3]	KOMATSUNA dataset (contains Komatsuna plant images collected in the indoor environment, Multi-Modality Plant Imagery Dataset(MSU-PID)(Images of 16 Arabidopsis and 5 Bean plants	Deep learning efficient Net B7 as encoder and Unet + + as decoder	Convergence improved by addressing the vanishing gradient problem, improvement in semantic information degradation and reduction in computational complexity, enhancement of the multi-scale feature fusion	Tendency to overfit	1. Foreground Background Dice (FgBgDice) is dice 2. Best Dice (BestDice) 3. Intersection-Over-Union (IOU) 4. The difference in count
Weyler et al. [54]	RGB images of sugar beet fields. The dataset contains 1316 images with a size of 1024 px × 512 px	Bottom up approach-Encoder decoder architecture based on ERFNet	Compute valid covariance matrices for each instance, less prone to confuse crop leaves and plants with leaves or plants of weeds which are commonly present on real agricultural fields	Computationally expensive	Precision, Recall, Absolute difference in count, SBD, FBD
Shi et al. [44]	Own datatset. 62 tomato seedlings. 2 leaves each	FCN (Fully connected Neural Network) + RCNN based on 3D point cloud 2D instance seg and 2D semanticv seg = > 3D model	3D point cloud provides more detail than 2D. No downsampling required. It can deal with large and variable size point cloud	Precision in case of stem detection was less in 3D as it is a thinner structure. 3D method has less number of points in the BG as they are carved out in shape-of-silhouette technique. Inability to deal with abnormal views of the leaf	

Table 1 (continued)

Ref	Dataset	Segmentation Technique	Pros	Cons	Metrics
Li et al. [29]	Beijing Academy of Agriculture and Forestry) High-throughput data acquisition was performed using MVS-Pheno.a semi-automated multi-view image acquisition platform for individual plants	PointNet was introduced to implement stem-leaf and organ instance segmentation, and six phenotypes were extracted	High throughput data acquisition and data diversity. Fully automatic and up-scalable, which is particularly important for achieving high-throughput phenotype extraction	More resources used in implementing Deep learning algorithms	The mean precision and F1-Score of stem-leaf segmentation were 0.91 and 0.85, respectively. The mean precision and F1-Score for organ instance segmentation were 0.94 and 0.93, respectively
Chen et al. [7]	4 domestically managed Arabidopsis plants. Each plant had 6–10 leaves	Mask RCNN and 3D reconstruction	Effectively counts number of leaves and coverage area	Different scenarios like flowering, fruiting etc. were not considered	Recognition rate and leaf coverage rate. 81.5% and 72.8%
Zhu et al. [60]	CVPPP Arabidopsis plant images	cGAN	Proved to effectively generate synthetic Arabidopsis images from real images	Not tested on any other dataset	Average leaf counting error reduced by 16.67%
Purbaya et al. [42]	MalayaKew Dataset	DCGAN-based regularized model		GAN based models are sensitive to parameter changes	Error rate of 0.105% and 20.95%

5.7.1 Metrics Used for the Evaluation of the Segmentation Algorithms

To effectively evaluate the segmentation accuracy, the technique used should keep into account the effects such as Over-segmentation(when a reference region is represented by two or more regions), Under-segmentation(When two or more regions of reference are represented by a single region), Inaccurate boundary localization(When ground truth is produced by humans that segment at different granularities), and Different number of segments (Compare 2 segmentations when they have different number of segments). The following measure have been adopted by the researchers in the literature presented in this paper.

- PSNR: Peak signal to noise ratio. M and N refer to number of rows and columns. I is the standard image and I_1 is the segmented image. The formula measures how close the segmented image is to the standard image.

$$PSNR = 10lg\frac{255^2}{\frac{1}{MxN}\sum_{i-0}^{M-1}\sum_{j=0}^{N-1}\left(I(i,j) - I_1(i,j)\right)^2} \tag{2}$$

- Entropy: It is a statistical measure to calculate the number of bits needed to code the image data which in turn depicts the randomness.

$$H(x) = \sum_{i=1}^{k} p(i)log_2\frac{1}{p(i)} = -\sum_{i-1}^{k} p(i)log_2 p(i) \tag{3}$$

- Variation of information: it measures the sum of information loss and information gain between two clustering to explain the extent to which 1 clustering can explain the other.

$$VI(X, Y) : H(X) = H(Y) = 2I(X;Y) \tag{4}$$

  where, the clusters $X_1X_2X_3 \ldots X_K$ are represented by random variable X of range {1…K}. H(X,Y) represents the entropy of X and I(X,Y) represents mutual information between X and Y.
- Energy: Energy is a measure of information when an operation is formulated under a probability framework. The Energy has largest value of 1 when number of gray levels are constant. Smaller energy value corresponds to higher number of gray levels as in a complex image.

$$E(x) = \sum_{i=1}^{x} p(x) \tag{5}$$

- Terminology defined in following context:

$$TruePositive(TP) : \text{Correctly classified leaf are pixel} \tag{6}$$

$$FalsePositive(FN) : \text{Non} - \text{target leaf pixels are classified as leaf} \tag{7}$$

$$TrueNegative(TN)\text{:Non target leaf pixel is classified as non} - \text{target leaf} \tag{8}$$

$$FalseNegative(FN)\text{:When a target leaf pixel gets misclassified as non} - \text{target pixel} \tag{9}$$

- Specificity: It defines the ability of the algorithm to segment the normal regions of the input image.

$$Specificity = \frac{TN}{TN + FP} \tag{10}$$

- Sensitivity (Recall): It provides the information about the object and also defines proper segmentation of the input image.

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

- Segmentation Accuracy: It is based on the pixel classification as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$= \frac{FN}{TP + FN} \tag{13}$$

- False Positive Rate (FPR)

$$FalseNegativeRate(FNR) = \frac{FP}{FP + TN} \tag{14}$$

$$ErrorRate(ER) = \frac{FP + FN}{TP + FP + TN + FN} \tag{15}$$

$$Dice(F - measure) = 2 \times \frac{Precison \times Recall}{Precision + Recall} \tag{16}$$

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$Jaccard = \frac{TP}{TP + FP + FN} \tag{18}$$

$$Manhattan = \frac{TP + TN}{TP + FP + TN + FN} \tag{19}$$

- Best Dice:

$$BD\left(L^a, L^b\right) = \frac{1}{M} \sum_{i=1}^{M} \max_{1 \le j \le N} \frac{2|L_i^a \cap L_j^b|}{|L_i^a| + |L_j^b|} \tag{20}$$

where |.| denoted the number of leaf pixels and L_i^a for $1 \le$ i $\le$ M and $L_j^b for 1 \le j \le N represent the sets of leaf object segments from the leaf segmentations L^a and L^b respectively.$

- Symmetric Best Dice:

$$SBD\left(L^{ar}L^{gt}\right) = \min\left\{BD\left(L^{ar}, L^{gt}\right), BD\left(L^{gt}, L^{lr}\right)\right\} \tag{21}$$

where $L^{ar} representthealgorithmicresultand, L^{gt} representthegroundtruth$

- Foreground Background Dice:

$$\begin{array}{l}\text{Considering the foreground mask which constitutes the union of all leaf}^{\text{ÃćâĆňâĎć}}\text{s labels,}\\ \text{the dice score is used to evaluate the delineation of plant from the background}\\ \text{obtained algorithmically with respect to the ground truth}\end{array} \tag{22}$$

- Absolute difference in count(|DiC|):

$$|Dic| = \left|No.\ of\ leaves\ in\ L^{ar} - No.\ of\ leaves\ in\ L^{gt}\right| \tag{23}$$

- Misclassification Error (ME):

$$ME = 1 - \frac{|B_O \cap B_T| + |F_O \cap F_T|}{M \times N} \tag{24}$$

where B_O *represents the number of background pixels in the ground truth image*

B_T *represents the number of background pixels segmented by the algorithm*

F_O *represents the foreground pixels of the ground truth images*

F_T *represents the number of foreground pixels segmented by the algorithm*

- Mean Weighted Intersection over Union: To calculate this value, the Jaccard index or the intersection Over Union is calculated as:

$$IoU_{ln} = \frac{TP}{TP + FP + FN} \tag{25}$$

The mean weighted IoU is then obtained as:

$$mwIoU = \frac{1}{LN} * \sum_{l=1}^{L} p_l^{-1} \sum_{n=1}^{N} IoU_{ln}$$

where mwIoU is the average of the IoU for all classes weighted by the number of pixels in that class p_l

for all images in the test set

6 Deep Learning in Image Classification

The following section of the papers deals with the contribution of various researchers in the field of leaf image classification into their respective species. When an input image is input into the deep learning network, the layers are responsible for learning the patterns in the images and predict the final species. Initial efforts based on transfer learning could be observed in a comparison of performance of pre-trained models namely: VGG16, VGG19, ResNet50, InceptionV3, Inception-ResnetV2, and Xception was analyzed on a Bangladeshi leaf dataset collected by the authors[17]. In individual experiments, these models were initialized with their own weights and then fine-tuned to learn the features of the new dataset. Though, all the models listed are CNNs however each has its own peculiar features. VGG 16 and VGG19 have 16 and 19 layers respectively and VGG16 gives a better overall average accuracy, proving that adding more layers does not lead to increase in accuracy. ResNet contains a residual block with identity functions. Inception reduces complexity by reducing convolution filter size from 5×5 to 3×3. Xception introduces the concept of point-wise convolutions followed by depth-wise convolutions. The VGG16 model also proved to be invariant to rotation while providing maximum classification accuracy.

6.1 Advancements in Deep learning Models and Techniques

(a) *Innovations in loss function and mobile applications*: Authors proved that instead of using softmax loss function which is employed by most other researchers, advanced loss functions such as Additive Angular Margin loss and Large Margin Cosine Loss [48] lead to better discrimination of bean cutlivars. Experiments were performed on three levels (classifying species, cultivars from same species, cultivars from different species). The datasets consist of images from frontside and backsides of leaves. Transfer learning via fine-tuning of VGG16 layers was used and accuracies of 95.86%, 91.37% and 86.87% were reported on three levels.

A mobile app based on Deep learning was proposed in [35] to classify Malaysian herbal leaves. After converting images from RGB(color) to grayscale, shape features were extracted by utilizing Zernike and Hu moments and 5 texture features were extracted by Gray-Level Co-occurrence matrix (GLCM). These features were then input into Deep learning Neural network(DLNN). The performance of DLNN classifier was compared to Support Vector Machine (SVM) with RBF kernel and it was observed that DLNN gives better accuracy even when color of background was changed.

(b) *Utilizing RNN for temporal data*: Recurrent Neural Networks(RNN) help to model the spatial information in the most accurate manner [36]. In this contribution, the deep neural networks were used to consider the temporal correlation of data. RNNs can help to classify the time-series data. The potential of high spatial and temporal resolution Sentinel-1 remote sensing data to map agriculture land covers and to assess deep learning technique was evaluated. 2 RNN units- Long-Short Term memory (LSTM) and Gated Recurrent Unit(GRU) were analysed which differ by the number of learnable parameters. The CNN layers employed in these units extract non-linear time dependencies present in remote-sensing time series data. The softmax on top then predicts the species. GRU unit provided the maximum accuracy of 89% as it is an improvement over LSTM unit.

An enhanced RNN was proposed [8] which classifies leaf images into their respective species. Authors have used deep learning for segmentation as well as classification. The input images were pre-processed using median filter to remove noise. U-Net was used for segmentation. The shape, texture and color features extracted. Classification result was obtained using Enhanced RNN. The authors have proposed hybrid technique called Crow-Electric Fish optimization which optimizes the U-net's hidden node count and epoch as well as optimal hidden neurons of Enhanced RNN.

6.2 Specialized Approaches for application specific models

(a) *Attention mechanism and multi branch CNN:* To address the issue of plant recognition in natural environment, authors proposed an attention block and a multi-branch CNN [28]. The attention block reduces the effect of background on input image. This block consists of Convolutional, pooling and deconvolutional layers. The output from this block goes into a 3-branch CNN where each branch has same structure but different kernel sizes and number of parameters. The 3 branches extract the features of trunk, branches and leaves. The features are then fused and used for classification on BJFU100 dataset. This dataset has 100 species of plants and 100 images in each species.

(b) *SWP Leaf net model:* To model the botanist's leaf identification process in the lab, authors proposed a SWP LeafNet model [1]. The S-leafNet model developed from scratch determines the species based on segmented image where only margin and shape is visible. If the model can determine the species, it is output to the user and if not, then the information is retained and the colorful image is input into W-LeafNet model(developed from scratch). The knowledge from both steps can help determine the species, if there is still a doubt then to run a microscopic analysis, the image patch is sent into P-LeafNet model(based on MobileNetV2 petrained model). The model was tested on well-known datasets Flavia and MalayaKew and results are listed in Table 2.

6.3 Leveraging Synthetic Data and Optimization Techniques

Advancement in the field of deep learning has led to another major contribution of generating synthetic data from the real data. As known that deep learning models require large amount of training data, Generative adversarial networks (GAN) were introduced [15]. Leaves display a property of low inter-class and high intra-class variance visually. Authors in [32] proposed a GAN approach to generate artificial pictures of seedlings. The objective was to generate images as close as possible to real images, while maintaining high variance among various species. GAN models consist of discriminator network (D) and generator network (G). Discriminator network (D) is responsible for distinguishing between real and synthetic images and generator network (G) is responsible for producing synthetic images which look so realistic that D cannot distinguish between real and synthetic images. Authors proposed a combined network based on WGAN-GP and ACGAN and named in WacGAN which provides a supervised conditioning scheme for the model to produce visually distinct samples for multiple classes while maintaining high variability in each sample. To evaluate the proposed model, transfer learning was used to pre-train

Table 2 Comparison of Deep learning-based models for plant classification using leaf image datasets

Ref	Dataset	Technique	Models	Pros	Cons	Accuracy
Malik et al. [33]	PlantCLEF,UBD Botanical	Transfer learning	EfficientNet-B1			
Beikmohammadi et al. [1]	Flavia, MalayaKew	CNNs from scratch + transfer learning	MobileNetV2	Shallow network, fewer parameters	Complexity in training for the developers	99.6%, 99.8%
Habiba et al. [17]	Own dataset	Transfer learning	VGG16,VGG19,ResNet50, InceptionV3, InceptionResNetV2, Xception	Rotation Invariant		96%
Tavakoli et al. [48]	BeanLeafFS, BeanLeafBS	Transfer learning	Fine-tuned VGG16 and loss functions: Additive Angular Margin and Large Margin Cosine	The loss functions enforce intraclass denseness and intra-class variance	Not so effective for species level classification	95.87%
Muneer and Fati [35]	Botanical Garden UPM agriculture department and the Malaysian peninsular forestry department	Extracted Shape and texture input into the Deep Learning Neural Network and integrated into android mobile app	Number of features reduced	Does not provide a good accuracy if background is floral	93%	
Li et al. [28]	BJFU100	Attention block and multi-branch CNN developed	Influence of background is reduced and multi-view features extracted	Computationally expensive	97.89%	
Dudi and Rajesh [9]	Swedish, Mendeley dataset	The neurons and activation functions of CNN optimized using Shark Smell-Whale optimization algorithm	Works well on untrained dataset	Needs to be trained in case the number of images is lower than the required number of images	Better than others specified in the paper	
Carranza-Rojas et al. [6]	PlantCLEF 2022	2-level hierarchical softmax	ResNet, Efficient Net	Small models need lesser time and less computation resources	Other backbone models need to be tested	–

Table 2 (continued)

Ref	Dataset	Technique	Models	Pros	Cons	Accuracy
Wang and Wang [52]	Flavia, Swedish, LeafSnap	Siamese network as feature extractor KNN as classifier	Shows good results when size of dataset is small	Could be tested on more datasets to check the generalization capaibility	95.32%, 91.37%. 91.75%	
Ndikumana et al. [36]	Sentinel-1 data over an area in Camargue, France	Recurrent Neural Network (GRU and LSTM)		Occupies memory due to more number of parameters	89%	
Dudi and Rajesh [8]	Swedish and D-Leaf	U-Net for segmentation and RNN for classification. (Both optimized using C-EFO)	Optimization algorithm helps for feature selection	Specificity measure could be improved	97.1%	
Madsen et al. [32]	Plant Seedlings Dataset	Generative adversarial Networks to generate synthetic data from real data. Transfer learning using ResNet101 to evaluate the classification accuracy on artificial samples	Solves the problem of few samples in training data	Improvements to achieve more realistic images	58% on artificial samples	
Espejo-Garcia et al. [11]	Weed dataset	DCGAN based models for generating synthetic images and Xception for identification	Synthetic images in complex environment were generated	Texture and Shape information was missing	93% on noisy dataset	

ResNet101 classifier on artificial samples from the WacGAN and fine-tune on real samples. An increase in classification accuracy and convergence rate was observed. An average recognition accuracy of 59% was achieved on artificial samples. For automatic weed identification, authors combined GAN with agricultural transfer learning [11]. For generation and identification of synthetic real environment images taken in complex backgrounds, authors proposed a combination of DCGAN and transfer learning. Experiments were based on weed dataset containing 202 tomato and 130 black nightshade images at early growth stages [10]. Two DCGAN based models were proposed to generate images from 2 species and were evaluated using Frechet Inception Distance [19]. Pre-trained Neural networks like Inception, InceptionResNet and Xception were fine-tuned weed identification. Xception gave an accuracy of 93.23% on noisy dataset.

In order to get maximum classification accuracy on unseen data authors proposed [9] a CNN where the activation function was optimized by hybrid Shark Smell-based Whale Optimization algorithm(SS-WOA). The input images are pre-processed by conversion to grayscale, histogram equalization and median filtering to enhance the input quality. The threshold of the optimization algorithm was fixed by trial and error initially. The proposed CNN was unique as the hidden neurons and activation functions were optimized by the hybrid SS-WOA. The threshold for the CNN was optimized using the classification score. This technique made the classification accurate even for unseen dataset. This method also proved better when other optimization algorithms (like Particle-Swarm, Grey Wolf optimization) were applied.

6.4 Recent Innovations and Achievements

To tackle the problem of lost gradients and overfitting by the use of deep learning in smaller datasets, authors proposed a few-shot learning technique [52]. It was observed that CNN could be used as feature extractor module. The technique was based on Siamese Network framework where features of 2 different images could be extracted by using parallel 2-way CNN(inspired from GoogleNet) with weight sharing. The network used loss function to learn the metric space where similar samples were closes and dis-similar samples were farther from each other. Authors proposed spatial structure optimizer (SSO) to construct the metric space. K-nearest neighbors was used as classifier. The method when tested on Flavia, Swedish and LeafSnap datasets provided 95.32%, 91.37% and 91.75% accuracies on datasets with 20 samples each which shows its capability to work on smaller datasets. To solve problems in constrained resources authors won 4th position in PlantCLEF 2022 ranking [6]. The authors proposed a 2-level hierarchical softmax instead of the vanilla softmax layer. The performance of both types of softmax layers was compared on both ResNet and EfficientNet models. Other techniques that create constrained environment were automatic mixed precision, batch accumulation and gradient clipping. An android based automated real-time mobile application [33] was developed to identify medicinal plant species around Borneo region. Transfer learning was adopted and the CNN model EfficientNet-B1 was trained on popular datasets namely: PlantCLEF and UBD Botanical Garden and their own collected images. The training was done on Google Cloud Platform. The problem of im-balanced data was handled by using two-cost sensitive learning methods: focal loss and computation of class weights for each class. The mobile application also helped for geo-tagging of species and use crowdsourcing for creation of dynamic database.

From various contributions in the field of automatic leaf identification, it was observed that the capabilities of pre-trained models like VGG16 and ResNet50 can be leveraged, emphasizing their unique strengths and limiting their weaknesses. The use of GANs to generate synthetic images further expands the scope of training data used for deep learning. Moving towards more application-specific innovations, developments in the field of mobile applications for real time leaf classifications demonstrates the practical utility of these technologies in field settings. The problem of smaller datasets encountered in the field of deep learning was tackled using the advanced loss functions and exploration of few-shot learning techniques. The incorporation of sophisticated optimization algorithms such as Shark Smell-based Whale Optimization and attention mechanisms further enhanced the accuracy and efficiency of CNNs. The evolution of Deep learning by the integration of RNN and application of hierarchical softmax layers reflects deepening complexity and refinement in the model architectures.

7 Common Problems and Future Outlook of Deep Learning

1. Deep learning-based models are based on optimization of number of layers, learning algorithm to optimize weights and biases. The process of deep learning also relies on huge amount of data. The more variance in data, the better it can generalize the patterns and make predictions. The image data in case of plant studies also relies on the expert knowledge. The model needs to be customized to identify/segment different kinds of plants. In the future, more models need to be developed and tested to extract maximum feature information from the input image and achieve an optimal model accuracy. As observed from this survey, that emerging techniques such as Generative adversarial Networks(GAN) can have broader applications in leaf image identification. Supervised route in deep learning provides more stability to authors while unsupervised methods can lead to disorder.
2. While training Deep networks, it is important to fix hyperparameters like number of epochs so that the model does not overfit. Training speed is another major factor. Higher number of epochs, eventually increase the accuracy but the model loses the capacity to generalize and starts over-fitting the training data. Therefore, the relationship between the network scale, training speed should be adjusted comprehensively. The experiments also show that correct choice of classifier can lead to better results.
3. Image acquisition also plays a very important role for the identification of plant species. The factors into play while acquiring the image, also determine how well the image can be segmented or classified into its species. Weather conditions, occlusions by other objects, scaling and other uncertainties affect the output of the model. Collection of data is also determined by regional restrictions, plant varieties growing, healthy or diseased leaf. There needs to be a standard database and a set of unified standards for fair comparison of deep learning models.
4. There is similarity between leaf shapes of different species of plants. The researchers are also working on extracting new features. But the question remains, what if the features change after the model is trained. Depending upon the environment, the features like color and texture and shape tend to change, therefore the selection of appropriate feature selection technique should also be taken into consideration for future scope.

8 Conclusion

Despite the fact that deep learning has been a popular choice among researchers to solve plant identification problems, there prevail some open challenges like the training of models on the real-world data, automatic background removal, integrating the heavy models on hand-held devices to train the complex neural networks on millions of parameters. From the studies it was observed that the most common technique adopted in the field of deep learning was transfer learning. Transfer learning allows the researchers to use the deep CNNs like VGG16, VGG19, ResNet, GoogleNet etc. trained on the huge databases like ImageNet (1000 classes), to apply the knowledge on the target problem and provide a solution. These pre-trained networks have been trained for weeks on high-end GPUs in high-tech research labs. The weights learned by these networks, can be provided to the researchers working on constrained memory or processing environment and still gain the benefits. This is a great contribution to the research community coming from developing countries. As demonstrated, transfer learning- based methods have provided greater accuracies in segmentation as well as classification tasks compared to training networks from scratch.

Different datasets used in the segmentation and classification study have been described. Most authors have also considered their own grown/collected leaf/plant images. In cases when number of images were less, data augmentation techniques have been applied. Data augmentation techniques like zooming, rotation, cropping, adding noise, increase the variance in the dataset, thus allowing the model to better understand the patterns and make predictions. Data augmentation also helps to increase the number of images in the dataset to ease out pattern recognition for the model.

Comparison of deep learning models give an insight on which model to use in which scenario. Most studies also vouch for the fact that the deep learning models give better performance than the conventional machine learning algorithms for automatic plant identification. Deep learning-based models eliminate the need of a feature extractor module, thus making the process automatic. A large-set of papers have also provided a comparison of performance of their proposed technique to that of the pre-trained models. To tackle the problem of limited amount of data to efficiently train deep learning models, GANs have been proposed. These models allow the researchers to generate artificial images resembling the real images from the given dataset. These techniques also help to artificially create backgrounds in the given images. The pros and cons of all deep learning classification techniques and the metrics used for evaluation in each case have been tabulated in Table 2.

Mask-RCNN has been a widely used technique in automatic leaf image segmentation. The 2D imaging techniques, do not acquire areal and volume information as they lack the third dimension. Deep learning techniques could be applied after acquiring the 3D point cloud of the images. The pros and cons of all techniques and the metrics used for evaluation in each case have been tabulated in Table 1.

Based on the research papers reviewed, some future directions for the new researchers can be listed: (1) More deep learning algorithms need to be explored in the field of automatic leaf segmentation. (2) Performance of deep learning models vary for different datasets, hence there could be a scenario where a universal deep learning model could be proposed to identify plant species with higher recognition accuracy. (3) As the CNNs are continuously improving, another challenge for researchers is to find optimal parameters and layers to achieve convergence, thus solving the problems related to ensemble learning techniques could lead to future work.

This paper has been created as an attempt to provide latest research work carried out in the field of automatic leaf segmentation and classification using deep learning.

Author Contribution All the authors contributed to the conception and design of the manuscript. The data collection, material preparation and analysis were performed by Anuj Kumar and Silky Sachar. The first draft was prepared was prepared by Silky Sachar and reviewed and edited by Anuj Kumar.

Funding The authors declare that no funds or grants were received during the preparation of the manuscript.

Data Availability The datasets that have been studied by authors mentioned in the manuscript have been cited.

Code Availability The code used to generate figures can be made available on request.

Declarations

Conflict of interest The authors have no conflict of interest to disclose.

References

1. Beikmohammadi, A., Faez, K., & Motallebi, A. (2022). SWP-LeafNET: A novel multistage approach for plant leaf identification based on deep CNN. *Expert Systems with Applications, 202*, 117470. https://doi.org/10.1016/j.eswa.2022.117470
2. Bell, J., Dee, H. M. (2019). Leaf segmentation through the classification of edges.
3. Bhagat, S., Kokare, M., Haswani, V., et al. (2022). Ecological Informatics Eff-UNet ++: A novel architecture for plant leaf segmentation and counting. *Ecological Informatics, 68*, 101583. https://doi.org/10.1016/j.ecoinf.2022.101583
4. Cao, L., Li, H., Yu, H., et al. (2021). Plant leaf segmentation and phenotypic analysis based on fully convolutional neural network. *Applied Engineering in Agriculture, 37*, 929–940. https://doi.org/10.13031/AEA.14495
5. Cao, L., Li, H., Yu, H., et al. (2021). Plant leaf segmentation and phenotypic analysis based on fully convolutional neural network. *Applied Engineering in Agriculture, 37*, 929–940. https://doi.org/10.13031/AEA.14495
6. Carranza-Rojas, J., Gonzalez-Villanueva, R., Jimenez-Morales, K., et al. (2022). Extreme automatic plant identification under constrained resources. In *CEUR workshop proceeding* (vol. 3180, pp.2014–2024).
7. Chen, D. H., Zhang, Z. W., Chang, T. R. (2019). Leaf segmentation and 3D reconstruction of ARAFIDOPSIS based on MASK R-CNN. In *Proceedings–2019 8th international congress on advanced applied informatics IIAI-AAI* (pp. 1073–1074). https://doi.org/10.1109/IIAI-AAI.2019.00232
8. Dudi, B., & Rajesh, V. (2022). A computer aided plant leaf classification based on optimal feature selection and enhanced recurrent neural network. *Journal of Experimental & Theoretical Artificial Intelligence, 35*, 1–35. https://doi.org/10.1080/0952813X.2022.2046178
9. Dudi, B., & Rajesh, V. (2022). Optimized threshold-based convolutional neural network for plant leaf classification: A challenge towards untrained data. *Journal of Combinatorial Optimization, 43*, 312–349. https://doi.org/10.1007/s10878-021-00770-w
10. Espejo-Garcia, B., Mylonas, N., Athanasakos, L., et al. (2020). Towards weeds identification assistance through transfer learning. *Computers and Electronics in Agriculture, 171*, 105306. https://doi.org/10.1016/j.compag.2020.105306

11. Espejo-Garcia, B., Mylonas, N., Athanasakos, L., et al. (2021). Combining generative adversarial networks and agricultural transfer learning for weeds identification. *Biosystems Engineering, 204*, 79–89. https://doi.org/10.1016/j.biosystemseng.2021.01.014
12. Fawakherji, M., Youssef, A., Bloisi, D., et al. (2019). Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation. In *2019 3rd IEEE international conference on robotic computing* (pp. 146–152). https://doi.org/10.1109/IRC.2019.00029
13. Giselsson, T. M., Jørgensen, R. N., Jensen, P. K., et al. (2017). A public image database for benchmark of plant seedling classification algorithms.
14. Gogul, I., Kumar, V. S. (2017). Flower species recognition system using convolution neural networks and transfer learning. In 2017 4th *international conference on signal processing, communication and networking (ICSCN)* (pp. 1–6). https://doi.org/10.1109/ICSCN.2017.8085675
15. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. (2020). Generative adversarial networks. *Communications of the ACM, 63*, 139–144. https://doi.org/10.1145/3422622
16. Guldenring, R., Boukas, E., Ravn, O., Nalpantidis, L. (2021). Few-leaf Learning: Weed segmentation in grasslands (pp. 3248–3254). https://doi.org/10.1109/iros51168.2021.9636770
17. Habiba, S. U., Islam, M. K., Ahsan, S. M. M. (2019). Bangladeshi plant recognition using deep learning based leaf classification. In *5th International conference on computer, communication, chemical, materials and electronic engineering IC4ME2 2019* (pp. 11–12). https://doi.org/10.1109/IC4ME247184.2019.9036515
18. He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*.
19. Heusel, M., Ramsauer, H., Unterthiner, T., et al. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *Advances in neural information processing systems, 2017*, 6627–6638.
20. Hu, J., Chen, Z., Yang, M., et al. (2018). A multiscale fusion convolutional neural network for plant leaf recognition. *IEEE Signal Processing Letters, 25*, 853–857. https://doi.org/10.1109/LSP.2018.2809688
21. Hu, G., Yang, X., Zhang, Y., & Wan, M. (2019). Identification of tea leaf diseases by using an improved deep convolutional neural network. *Sustainable Computing: Informatics and Systems, 24*, 100353. https://doi.org/10.1016/j.suscom.2019.100353
22. Jasitha, P., DIleep, M. R., DIvya, M. (2019). Venation based plant leaves classification using GoogLeNet and VGG. In *2019 4th IEEE international conference on recent trends on electronics, information, communication & technology RTEICT 2019–Proceeding* (pp. 715–719). https://doi.org/10.1109/RTEICT46194.2019.9016966
23. Kan, J., Gu, Z., Ma, C., Wang, Q. (2021). Leaf segmentation algorithm based on improved U-shaped network under complex background. In *IMCEC 2021–IEEE 4th advanced information management, communicates, electronic and automation control conference 2021* (pp. 87–92). https://doi.org/10.1109/IMCEC51613.2021.9482382
24. Kumar, V., Arora, H., Sisodia, J. (2020). ResNet-based approach for detection and classification of plant leaf diseases. In *Proceedings international conference on electronics and sustainable communication systems ICESC 2020* (pp. 495–502). https://doi.org/10.1109/ICESC48915.2020.9155585
25. Kumar, N., Belhumeur, P. N., Biswas, A., et al. (2012) Leafsnap: A Computer vision system for automatic plant species identification BT–computer vision. In *ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part II* (vol. 7573, pp. 502–516). https://doi.org/10.1007/978-3-642-33709-3_36
26. Lecun, Y., Aurelio, M., Google, R. (2013). Deep learning tutorial. In *ICML*.
27. Lee, S. H., Chan, C. S., Wilkin, P., et al. (2015). *Deep-plant: Plant identification with convolutional neural networks* (pp. 452–456). Kingston University.
28. Li, P., Gong, X., Hu, X., et al. (2019). Plant identification based on multi-branch convolutional neural network with attention. *Communications in Computer and Information Science, 1043*, 472–481. https://doi.org/10.1007/978-981-13-9917-6_45
29. Li, Y., Wen, W., Miao, T., et al. (2022). Automatic organ-level point cloud segmentation of maize shoots by integrating high-throughput data acquisition and deep learning. *Computers and Electronics in Agriculture, 193*, 106702. https://doi.org/10.1016/j.compag.2022.106702
30. Litjens, G., Sánchez, C. I., Timofeeva, N., et al. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Science and Reports, 6*, 1–11. https://doi.org/10.1038/srep26286

31. Liu, B., Tan, C., Li, S., et al. (2020). A data augmentation method based on generative adversarial networks for grape leaf disease identification. *IEEE Access, 8*, 102188–102198. https://doi.org/10.1109/ACCESS.2020.2998839
32. Madsen, S. L., Dyrmann, M., Jørgensen, R. N., & Karstoft, H. (2019). Generating artificial images of plant seedlings using generative adversarial networks. *Biosystems Engineering, 187*, 147–159. https://doi.org/10.1016/j.biosystemseng.2019.09.005
33. Malik, O. A., Ismail, N., Hussein, B. R., & Yahya, U. (2022). Automated real-time identification of medicinal plants species in natural environment using deep learning models—a case study from borneo region. *Plants, 11*, 1952. https://doi.org/10.3390/plants11151952
34. Morris, D. (2018). A pyramid CNN for dense-leaves segmentation. In *Proceedings–2018 15th conference on computer and robot vision CRV 2018* (pp. 238–245). https://doi.org/10.1109/CRV.2018.00041
35. Muneer, A., & Fati, S. M. (2020). Efficient and automated herbs classification approach based on shape and texture features using deep learning. *IEEE Access, 8*, 196747–196764. https://doi.org/10.1109/ACCESS.2020.3034033
36. Ndikumana, E., Minh, D. H. T., Baghdadi, N., et al. (2018). Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing, 10*, 1–16. https://doi.org/10.3390/rs10081217
37. Ngugi, L. C., Abelwahab, M., & Abo-Zahhad, M. (2020). Tomato leaf segmentation algorithms for mobile phone applications using deep learning. *Computers and Electronics in Agriculture, 178*, 105788. https://doi.org/10.1016/j.compag.2020.105788
38. Nikbakhsh, N., Baleghi, Y., & Agahi, H. (2021). A novel approach for unsupervised image segmentation fusion of plant leaves based on G-mutual information. *Machine Vision and Applications, 32*, 1–12. https://doi.org/10.1007/s00138-020-01130-0
39. Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*, 1345–1359. https://doi.org/10.1109/TKDE.2009.191
40. Praveen Kumar, J., & Dominic, S. (2020). Rosette plant segmentation with leaf count using orthogonal transform and deep convolutional neural network. *Machine Vision and Applications, 31*, 1–14. https://doi.org/10.1007/s00138-019-01056-2
41. Praveen Kumar, J., & Domnic, S. (2019). Image based leaf segmentation and counting in rosette plants. *Information Processing in Agriculture, 6*, 233–246. https://doi.org/10.1016/j.inpa.2018.09.005
42. Purbaya, M. E., Setiawan, N. A., Adji, T. B. (2018). Leaves image synthesis using generative adversarial networks with regularization improvement. In *2018 International conference on information and communications technology ICOIACT 2018-Janua* (pp. 360–365). https://doi.org/10.1109/ICOIACT.2018.8350780
43. Sachar, S., & Kumar, A. (2021). Survey of feature extraction and classification techniques to identify plant through leaves. *Expert Systems with Applications, 167*, 114181.
44. Shi, W., van de Zedde, R., Jiang, H., & Kootstra, G. (2019). Plant-part segmentation using deep learning and multi-view vision. *Biosystems Engineering, 187*, 81–95. https://doi.org/10.1016/j.biosystemseng.2019.08.014
45. Söderkvist, O. J. O. (2001). Computer vision classification of leaves from swedish trees (pp. 74).
46. Sun, J., Zhao, J., Ding, Z. (2021). ULeaf-Net: Leaf segmentation network based on U-shaped symmetric encoder-decoder architecture (pp. 109–113). https://doi.org/10.1109/iscsic54682.2021.00030
47. Tan, J. W., Chang, S. W., Abdul-Kareem, S., et al. (2020). Deep learning for plant species classification using leaf vein morphometric. *IEEE/ACM Trans Comput Biol Bioinforma, 17*, 82–90. https://doi.org/10.1109/TCBB.2018.2848653
48. Tavakoli, H., Alirezazadeh, P., Hedayatipour, A., et al. (2021). Leaf image-based classification of some common bean cultivars using discriminative convolutional neural networks. *Computers and Electronics in Agriculture, 181*, 105935. https://doi.org/10.1016/j.compag.2020.105935
49. Triki, A., Bouaziz, B., Gaikwad, J., & Mahdi, W. (2021). Deep eaf: Mask R-CNN based leaf detection and segmentation from digitized herbarium specimen images. *Pattern Recognition Letters, 150*, 76–83. https://doi.org/10.1016/j.patrec.2021.07.003

50. Trivedi, M., & Gupta, A. (2021). Automatic monitoring of the growth of plants using deep learning-based leaf segmentation. *International Journal of Applied Science and Engineering, 18*, 1–9. https://doi.org/10.6703/IJASE.202106_18(2).003
51. Wäldchen, J., & Mäder, P. (2018). Plant species identification using computer vision techniques: A systematic literature review. *Archives of Computational Methods in Engineering, 25*, 507–543. https://doi.org/10.1007/s11831-016-9206-z
52. Wang, B., & Wang, D. (2019). Plant leaves classification: A few-shot learning method based on siamese network. *IEEE Access, 7*, 151754–151763. https://doi.org/10.1109/ACCESS.2019.2947510
53. Ward, D., Moghadam, P., Hudson, N. (2019). Deep leaf segmentation using synthetic data.
54. Weyler, J., Magistri, F., Seitz, P., et al. (2022). In-field phenotyping based on crop leaf and plant instance segmentation (pp. 2725–2734).
55. Wu, S. G., Bao, F. S., Xu, E. Y., et al. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. In *ISSPIT 2007–IEEE international symposium on signal processing and information technology* (pp. 11–16). https://doi.org/10.1109/ISSPIT.2007.4458016
56. Wu, Q., Chen, Y., & Meng, J. (2020). Dcgan-based data augmentation for tomato leaf disease identification. *IEEE Access, 8*, 98716–98728. https://doi.org/10.1109/ACCESS.2020.2997001
57. Xiong, J., Yu, D., Liu, S., et al. (2021). A review of plant phenotypic image recognition technology based on deep learning. *Electron, 10*, 1–19. https://doi.org/10.3390/electronics10010081
58. Xu, L., Li, Y., Sun, Y., et al. (2018). Leaf instance segmentation and counting based on deep object detection and segmentation networks. In *2018 Joint 10th international conference on soft computing and intelligent systems (SCIS) and 19th international symposium on advanced intelligent systems SCIS-ISIS* (pp. 180–185). https://doi.org/10.1109/SCIS-ISIS.2018.00038
59. Yang, K., Zhong, W., & Li, F. (2020). Leaf segmentation and classification with a complicated background using deep learning. *Agronomy, 10*, 1721. https://doi.org/10.3390/agronomy10111721
60. Zhu, Y., Aoun, M., Krijn, M., & Vanschoren, J. (2018). Data augmentation using conditional generative adversarial networks for leaf counting in arabidopsis plants. *BMVC, 2018*, 121–125.
61. Zhu, W., Xie, L., Han, J., & Guo, X. (2020). The application of deep learning in cancer prognosis prediction. *Cancers (Basel), 12*, 1–19. https://doi.org/10.3390/cancers12030603

Anuj Kumar is a Ph.D. in Computer Applications. Areas of interest include pattern recognition, machine learning. He has more than 10 years of experience in teaching undergraduates and graduates in the field of Computer applications. He is currently working as Assistant Professor in the department of Computer Science and Applications, Panjab University,Chandigarh. He has presented various papers in national/international conferences and has published papers in various quality journals.

Silky Sachar has completed Master's in Computer Applications from Vellore Institute of technology, Vellore, India. She has 2 years of work experience in IT Industry, post which she cleared UGC-National Eligibility Test to work as an Assistant Professor. She has 5 years experience in under-graduates and graduates in the field of Computer applications. She is currently a research Scholar in the department of Computer Science and Applications, pursuing Ph.D in the field of Automatic Medicinal Plant species Identification using Leaf images. She has presented research papers in national/International conferences and have achieved "Best Paper Award" in two of them.