International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Meticulous Leaf Detection from Videos under Leaf Venation and Margin using Multi-Spectral CapsNet and Relational Prototypical LSTM

[1] Vidyashankar, [2] Hemantha Kumar G
[1] *DoS in Computer Science*, [2] *DoS in Computer Science*,
[1] *University of Mysore*, Mysuru, INDIA

---***---

Abstract - Leaf detection is essential for understanding leaf behavior, agricultural monitoring, and ecological studies, which is necessary for species identification and understanding of environmental impacts on leaf growth. Hence a novel approach Multi-spectral CapsNet and Relational Prototypical LSTM, is proposed to overcome challenges in leaf detection from videos. Initially, videos are converted into image frames, which are then pre-processed to remove noise and boost contrast, followed by a watershed approach for segmentation. Existing algorithms struggle to extract essential data from secondary vein patterns due to their vein's hierarchical structure, which extends beyond local spatial patterns. Hence Adaptive Centricity Multi-spectral CapsNet is implemented to extract features from the leaf vein patterns. Which utilizes the Adaptive Centricity Hough Line Detection (ACHL) Algorithm to extract local features like vein spacing and branching angles. Multi-spectral Attentional CapsNet (MA-CapsNet) captures global features like contextual information by focusing on spectral channels containing relevant data for identifying vein patterns. Furthermore, in leaf detection, existing methods analyze input data sequentially without including relative nearby regions information, which limits the network's capacity to distinguish between regions with varied depths and to grasp the global spatial relations between serrations Hence, Bidirectional Relational Prototypical LSTM (Bi-RP LSTM) is introduced to capture spatial relationships between serrations and analyze under serration depth levels thereby improving leaf detection accuracy. Finally, the proposed approach is implemented in Python, making it easier and more accurate than existing models for leaf detection in terms of accuracy, recall, precision, sensitivity, and F1 score.

Keywords - Leaf detection, Vein spacing, Angle of branching, Serration depth, Medial Axis Transform

I. Introduction

In video analysis, leaf detection is the process of identifying and monitoring leaves over a series of frames. This procedure is essential for many applications, including environmental research, disease detection, and plant growth monitoring. Leaf identification methods usually start by segmenting the video frames to identify areas that are likely to contain leaves using computer vision techniques. Then, characteristics like color, texture, and form are used to set leaves apart from the surrounding area. To analyze patterns of development and identify anomalies, tracking algorithms are used to trace the movement of leaves between frames. Deep learning and other machine learning techniques showed promise in enhancing the precision and resilience of leaf detection systems.

Researchers and practitioners can more effectively analyze plant behavior, evaluate environmental conditions, and support agricultural management methods by automating the process of leaf detection in films. In addition to form, texture, and color, additional characteristics including leaf venation, edge, apex, base, and so forth are also important for accurate detection [1-4].

The vein pattern in a leaf that carries carbohydrates, nutrients, and water is recognized as leaf venation. It is divided into two primary categories: reticulate venation, which forms a branching network of veins, and parallel venation, which is prevalent in monocots and involves veins running parallel to one another (typical in dicots). Veins that run parallel to each other through the base to the tip of the leaf are known as parallel venation, and these vessels usually do so without creating an intricate network. In monocotyledonous plants like grasses and lilies, this pattern is common. Reticulate venation, on the other hand, shows a branching network of veins that creates a complex pattern all over the leaf surface. Dicotyledonous plants, such as roses and maple trees, frequently have this kind of venation. Plant physiology research, leaf categorization, and plant species identification all depend on feature extraction from venation patterns. Venation pattern analysis algorithms currently in usage range from straightforward thresholding and edge detection approaches to more intricate strategies using graph theory and machine learning [5-8].

However, there are still issues with venation pattern analysis, such as managing intricate vein patterns, dealing with a range of leaf sizes and shapes, and making the system resilient to noise and image artifacts. For real-time applications, scalability to enormous datasets and processing performance are also essential. To solve these problems, multidisciplinary teams with backgrounds in botany, computer vision, and machine learning are needed. This will eventually result in algorithms for venation pattern analysis that are more precise and effective. Furthermore, the leaf margin, or the edge or border of a leaf, is another characteristic that is necessary for detection. Leaf edges can be whole, serrated, lobed, or toothed, among other variations. Analysis of the leaf edge is essential for leaf detection because it offers distinguishing characteristics that help with leaf identification and categorization. Different plant species are distinguished from one another and abnormalities including illness or damage are easily identified by differences in the form, serration, and curvature of the leaf margin. Furthermore, knowing the different forms of leaf margins helps with more extensive ecological and taxonomic research [9–11].

The term "serration" in leaf margins describes the tiny, pointed teeth that are present along the leaf's edge. The size, shape, and spacing of these teeth differ, giving important information for leaf identification. Analyzing serration patterns aids in differentiating between species and identifying anomalies in leaf detection. Convolutional Neural Networks (CNNs), contour-based techniques such as Active Contour Models (Snakes), and edge detection techniques like canny edge detection are some of the algorithms used to identify leaves based on their margins. However because of differences in leaf morphology, texture, and occlusions in the video frames, it is still difficult to identify serrations with accuracy. Furthermore, these algorithms can act poorly because of the intricacy of serration patterns and the existence of noise in video data. Furthermore, the problem of robustly recognizing leaves based on their margins is further complicated by real-world variables including illumination fluctuations and environmental influences. To overcome these obstacles, developments in multi-modal information integration, algorithmic resilience, and feature extraction techniques are needed for improved leaf recognition accuracy in video analytic applications [12–15]. Therefore, more innovative algorithms need to be developed to identify leaves based on a variety of leaf properties, including venation and margin.

The following is the paper's primary contribution.

- To improve the accuracy of feature extraction from the leaf venation patterns, the Adaptive Centricity MA-CapsNet technique is utilized, in which Adaptive Centricity Hough Line Detection is employed to extract local characteristics from leaf vein patterns, and Multi-spectral Attentional CapsNet is used to improve the model's capacity to acquire contextual information beyond local spatial patterns in leaf recognition tasks.
- To enhance the accuracy of leaf detection under serration depth, Bi-RP LSTM is introduced for leaf detection, which aims to improve overall detection accuracy by analyzing global spatial connections between serrations and separating regions with various depth levels.

The work is divided into five chapters, the first of which is an introduction, and the second of which is a survey of the literature. Section 3 provides a description of the suggested approach.

It also addresses the various research methods used. Section 4 discusses performance and comparative analysis. Section 5 discusses the work's conclusion in depth.

II. Literature survey

Pankaja et al. [16] suggested classifying and identifying plant leaves utilizing the hybridization of the whale optimization algorithm (WOA) and the random forest (RF). This research made use of the Swedish and Flavia leaf datasets. Pre-processing was done initially to enhance the data's quality or remove noise before feature extraction. To address the dimensionality problem, WOA was used. Additionally, the RF classifier was used to identify the leaf. The recommended approach has a high accuracy of 97.58% with a lower execution time when compared to alternative methods. This work ensured improved classification and identification of plant leaves for therapeutic purposes. Because they don't naturally undertake feature engineering to adequately describe the qualities of plant leaves, this method falls short in finding significant features.

Huixian et al. [17] designed to extract leaf features and identify plant species through image analysis. First, a variety of methods were used to segment plant leaf images. The texture and shape information from leaf sample photographs was then extracted using a feature extraction method. Subsequently, the comprehensive characteristic information of plant leaves was generated using the comprehensive characteristic information. Three approaches are examined and contrasted in this study: support vector machines (SVM), Kohonen networks based on self-organizing feature mapping techniques, and K-Nearest Neighbour (KNN)-based classification. Fifty plant leaf databases were used as the test and comparative datasets. Ginkgo leaves were shown to be easier to identify when seven distinct plants' leaves were compared at the same time. For leaf pictures with complex backgrounds, a good recognition impact has been achieved. Since the segmentation background was straightforward, more study is required to identify leaves in their natural environment and against a variety of backgrounds in this work.

Yang et al. [18] provided an innovative method that utilized tactile and morphological aspects to detect plant leaves. First, shape and texture attributes were offered as the basis for the recommended plant leaf identification approach. The suggested multiscale triangle descriptor (MTD) was used to characterize the shape information of a plant leaf, and the local binary pattern histogram Fourier (LBP-HF) was used to capture the texture feature. Next, the texture and form attributes of an image of a leaf were integrated using weighted distance measuring. L1 distance was used for form characteristics while chi-square distance was used for texture features. The suggested method combined the complementing MTD and LBP-HF traits to produce a potent descriptor for the plant leaf recognition task.

The proposed method has been extensively evaluated on three benchmark leaf datasets. Moreover, this approach lacks accuracy for leaf detection since it takes only consideration of its shape and texture attributes, disregarding other important visual aspects.

Bisen et al. [19] proposed a plant identification method that was automated and used a leaf to identify the species of the plant. This challenge was completed using a deep convolutional neural network to get higher accuracy. Pre-processing the picture, feature extraction, and recognition were the three main stages of identification that were taken into evaluation. The proposed CNN classifier learned plant characteristics, such as leaf categorization, using hidden layers that include convolutional, max-pooling, dropout, and fully connected layers. The model was used to identify the correct category of an unknown plant with the lowest losses and 97% accuracy. It learns features using the Swedish leaf dataset, which consists of 15 tree classes. The segmentation and identification of leaves from photos of foliage using different leaf datasets is not improved by the proposed method.

Azadnia et al. [20] developed a modern, reliable automated image-processing method to recognize medicinal plants under regulated lighting conditions with speed and accuracy. The algorithm was applied to extract texture, color, and form features from the acquired pictures. Artificial neural networks were used to classify a number of the medicinal plants that were the subject of the study. The best classifier was selected based on factors such as accuracy, correlation, and error. The optimal classifier model was generated after feeding the model with the effective attributes. To prevent overfitting and ensure that there is enough capacity to capture the variety within the plant species, this study requires.

Bao et al. [21] observed that even for seasoned botanists, the procedure of recognizing species was difficult. As a result, this study proposed two methods for handling the problem of plant species identification using leaf patterns. The characteristics of a traditional recognition shallow architecture were extracted using the Histogram of Oriented Gradients (HOG) vector. The SVM algorithm was then applied to categorize based on these characteristics. Second, a deep CNN was employed in this study to accomplish recognition. To confirm the findings, tests were carried out with two leaf data sets: the Swedish leaf data collection and the Flavia leaf data set. The key disadvantage of this work is that it does not include a comparison review of other detection and identification approaches to assess the benefits and drawbacks of the proposed strategy.

Hajam et al [22] used ensemble learning techniques to improve the accuracy of medicinal plant identification. Transfer learning was used with pre-trained models like VGG16, VGG19, and DenseNet201 to extract useful characteristics from medicinal plant leaf photos. To build ensemble models using CNNs, compare three popular CNN architectures: VGG16, VGG19, and DenseNet201. Transfer learning was used to harness three three-component classifiers without their upper layers.

This adaption enabled them to identify critical properties in medicinal leaf pictures, which were subsequently merged into thick layers trained on a dataset including 30 different classes of medicinal leaves using a softmax classifier. Training deep neural network models, specifically ensemble models, necessitates enormous computing resources and time.

Salve et al [23] developed a multimodal plant classification system using spectral signatures and leaf venation patterns as essential elements. The VISLeaf dataset was used to test the suggested methodologies. A variety of techniques were used to extract features from the dataset, such as vein characteristics, morphological features, spectral reflectance for non-imaging spectral signatures, and HOG descriptors from scanned leaf pictures. Concatenation was used to fuse the retrieved features and produce a multimodal feature set. By combining data from several feature sets, feature fusion algorithms increase classification performance over the use of individual feature sets alone. However, the combination of feature sets and the fusion of several feature types make the classification system's calculation more challenging.

Zhao et al [24] utilized DoubleGAN, a two-stage generative adversarial network (GAN), to produce precise images of diseased plant leaves. The intention was to tackle the problem of imbalanced datasets in plant disease diagnosis, wherein photos of diseased leaves were often less numerous than those of healthy leaves. There were two steps in the process: To obtain a pre-trained model, a Wasserstein generative adversarial network (WGAN) was trained using images of both healthy and diseased leaves. Next, 64x64 pixel pictures of sick leaves were created by applying the pre-trained model to photographs of the leaves. The dataset was expanded by using a super-resolution generative adversarial network (SRGAN) to increase the resolution of the produced 64x64 pixel pictures to 256x256 pixel images. Comparable to other GANs, DoubleGANs are susceptible to mode collapse, training instability, or poor picture quality due to hyper parameters including learning rates, batch sizes, and network designs.

Ashwinkumar et al [25] proposed the use of an optimum mobile network-based convolutional neural network (OMNCNN) in an automated model for the detection and classification of plant leaf diseases. Preprocessing, segmentation, feature extraction, and classification were the distinct steps at which the suggested OMNCNN model functions. Bilateral filtering (BF) based pre-processing and Kapur's thresholding-based image segmentation were utilised to identify the affected regions of the leaf picture.

Furthermore, the emperor penguin optimizer (EPO) method was used to optimize the hyper parameters of the Mobile Net model as a feature extraction strategy to improve the rate of plant disease identification. Lastly, a classifier based on an extreme learning machine (ELM) was used to provide the proper class labels to the submitted plant leaf photos.

However, dependence on manually defined features and adjusted parameters limits scalability and flexibility when dealing with different datasets.

The above statement stated that [16] Lacks inherent feature engineering, delaying the effective representation of plant leaf characteristics. [17] Requires further research to improve leaf recognition in natural habitats and against varied backgrounds. [18] Only considers shape and texture, disregarding other crucial visual features, leading to accuracy limitations. [19] Fails to enhance leaf segmentation and identification from diverse foliage photographs. [20] Struggles to balance network complexity, risking overfitting while capturing plant species variability. [21] Omits comparative analysis with alternative techniques, hindering assessment of suggested approach benefits. [22] Demands significant computational resources and time for training, limiting scalability. [23] Faces challenges in calculation due to combining multiple feature sets, impacting efficiency. [24] Experiences training instability and poor picture quality due to hyper parameter tuning issues. [25] Relies on manual feature definition and parameter adjustment, restricting scalability and flexibility across datasets. Addressing these difficulties through more research and innovation is necessary for establishing more robust, efficient, and scalable techniques for identifying plant leaves and detecting diseases.

III. Leaf Detection with Venation and Margin using Multi-spectral Caps Net and Relational Prototypical LSTM

Leaf detection is an important step in video analysis since it allows us to analyze dynamical leaf behavior and helps in agricultural observation and ecological studies. To address those obstacles and detect leaves in the video, an elusive intelligent method called Multi-spectral Caps Net and Relational Prototypical LSTM is presented. This innovative approach begins with converting the video into image frames, followed by noise removal and contrast enhancement through pre-processing techniques. Subsequently, watershed segmentation is employed for accurate leaf segmentation. Each leaf has distinct vein patterns, which include parallel and reticulate venation. Secondary vein connectivity and branching patterns are unique in leaves with reticulate venation, and these characteristics are essential for recognizing a species. The hierarchical nature of vein networks makes it difficult to extract essential data from secondary vein patterns, such as vein spacing, branching angles, and connection with higher-order veins. Primary veins comprise the basic structure, with secondary, tertiary, and higher-order veins branching off of them. Existing algorithms failed to capture contextual details beyond local spatial patterns, limiting their capacity to analyze the long-range. Hence for the feature extraction phase, the Adaptive Centricity Multi-spectral Caps Net is introduced, which

Utilizes the Adaptive Centricity Hough Line Detection (ACHL) Algorithm to extract local features like vein spacing and branching angles. The Adaptive Hough Transform adjusts parameters based on local image characteristics to detect primary and secondary veins, while the MAT extracts the skeleton of the leaf veins by capturing the essential structure of the vein network, including the secondary veins and their branching patterns. Thus, ACHL identifies the distance between the veins, computing vein spacing, then analyses the branching points, calculating the angle at which secondary veins branch off from primary veins. Multi-spectral Attentional CapsNet (MA-CapsNet) extracts global features from leaf vein patterns, which incorporates Multi-spectral Channel Attention to selectively attend to spectral channels containing useful information for distinguishing vein patterns, enhancing contextual understanding beyond local spatial patterns. The Capsule Network represents vein segments hierarchically, capturing orientation, position, and connectivity information to handle long-range dependencies and complex spatial relationships effectively. Thus, MA-CapsNet provides a robust solution for extracting features from leaf vein patterns and addressing challenges associated 0with hierarchical organization and contextual information capture.

Moreover, the serration depth in leaf margins is a significant feature for identifying plant species. It is quantified as the distance between the base of a serration and its tip perpendicular to the leaf margin. Serrations fluctuate within a leaf due to genetic, developmental, and environmental influences, resulting in a gradient that spans regions. Detecting and characterizing this gradient requires algorithms that analyze the spatial distribution of serrations and differentiate between regions with different levels of depth. In leaf detection, existing algorithms process input data sequentially, layer by layer, without considering relative information from neighboring regions, restricting the network's ability to grasp the global spatial relationships between serrations and distinguish between regions with varying depths. Therefore, Bidirectional Relational Prototypical LSTM (Bi-RP LSTM), is implemented for leaf detection. Relational prototype networks collect relative information between serrations, which improves the clarity of global spatial connections. A prototype represents each class or degree of serration depth, allowing for more effective separation between depth levels. Bi-LSTM analyses serrations sequentially along the leaf margin, processing input data in both directions to take into account information from neighboring areas. This method enhances leaf detection accuracy by providing contextual awareness, relative location comprehension, spatial connection analysis, and depth level distinction within the leaf margin.

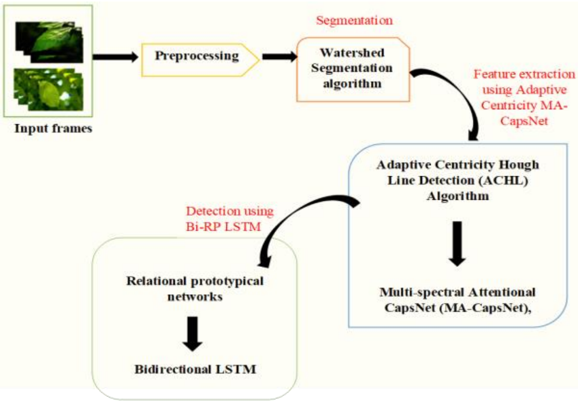


Fig.1.architecture of the Relational Prototypical LSTM and Multi-spectral Caps Net

Figure 1 illustrates the architecture of the Relational Prototypical LSTM and Multi-spectral CapsNet for leaf detection with venation and margin. Leaf videos are first converted into picture frames and pre-processed for contrast improvement and noise reduction. The watershed approach is used for segmentation. The Adaptive Centricity Multi-spectral CapsNet is used for feature extraction, along with the Medial Axis Transform for vein skeletonization and the ACHL Algorithm comprises the Adaptive Hough Transform for vein spacing and branching angle extraction. MA-CapsNet improves contextual knowledge by focusing on certain spectral channels. By capturing hierarchical vein organization, the Capsule Network makes it easier to recordlong-range dependencies. Leaf detection, which follows feature extraction, presents the Bi-RP LSTM and uses relational prototype networks to achieve global spatial knowledge. Sequential serration configurations are analyzed using Bi-LSTM, which considers surrounding areas for context awareness. The methodological aspects of the proposed leaf detection are discussed in the upcoming sections.

Fig. 1. Architecture of the Leaf detection with venation and margin using Multi-spectral CapsNet and Relational Prototypical LSTM.

Data collection

The data collection starts with a 20-second leaf video. The input leaf video is translated into 611 distinct frames, each representing a snapshot of the leaf at a given instant. This step ensures that each frame is treated separately during leaf detection and analysis. By considering each frame as an independent object, the model properly records and analyzes the delicate characteristics of the leaf's structure and venation patterns at various periods, allowing for robust and dependable leaf recognition. This collection contains 488 frames for training the model, which provides sufficient information for the algorithm to recognize the detailed patterns of leaf features such as venation and marginal characteristics. The remaining 123 frames are put aside for performance testing, which provides an entire evaluation set to confirm the model's ability to recognize leaves correctly.

Pre-processing

After data collection, preprocessing takes place to prepare a dataset with valuable information and remove unwanted ones. Random changes in pixel values frequently occur in videos due to factors such as camera sensor noise and compression artifacts. Noise reduction technique, median filtering is used to smooth out these variances and generate a clearer image. This procedure improves frame quality and removes unnecessary distractions in the final analysis.

Leaves in a video vary in brightness and contrast due to lighting conditions, shadows, or camera settings. Increasing contrast enhances overall image quality and helps recognize details. In this proposed pre-processing stage for boosting the contrast of the image a simple method named, linear grayscale transformation is used. By mapping the image's original intensity values to a new range, this approach stretches or compresses the intensity values to raise the overall contrast. Using this approach, the contrast is adjusted by linearly transforming the image's pixel intensities. The processed image, h(x,y), is intended to contain pixel values ranging from [e,f]. In contrast, the original image, I(x,y), has pixel intensities ranging from [c,d]. The pixel intensities are subjected to a linear expansion to achieve this. This compression or stretching of the intensity range increases the image's dynamic range and sharpens the contrast between its features. The h(x,y) functions are expressed in the following equation (1)

$$h(x,y) = \begin{cases} f & I(x,y) > d \\ \frac{f-e}{d-c}[I(x,y) - c] + e & c \leq I(x,y) \leq d \\ e & I(x,y) < c \end{cases} \quad (1)$$

Where c and d stand for the original image's grey transformation range and e and f for the processed image's grey transformation range. Performing this pre-processing process on the frames of images taken from the leaf video increases the quality of the images and makes it simpler to detect and analyze the features of the leaves. After preprocessing, segmentation takes place, which is explained in the following section.

Watershed segmentation method

By determining the borders between elements of an image, the watershed segmentation method is a technique used to separate them. The technique is based on the idea of a topographic map, in which elevation levels are deduced from the intensity values of the images. The grayscale image is treated as a topographic surface, and the heights of the surface are represented by the intensity values of the image. Next, utilizing the surface gradient as a guide, watershed lines are employed to divide various areas. The ability to identify possible borders between different image sections depends on this information. Equation (2) expresses the gradient of the proposed model.

$$\nabla f(x,y) = \sqrt{(\frac{\partial_f}{\partial_x} + \frac{\partial_f}{\partial_y})^2} \quad (2)$$

Where ∇ is the gradient operator and f(x,y) is the intensity function of the image at pixel (x,y). After that, the algorithm uses the markers it created as the first seeds for segmentation. The markers help identify the regions of interest and direct the segmentation process. The gradient image is often modified or masked to highlight ROI and reduce unwanted gradients, which leads to more efficient image segmentation. The masked gradient is the sensible combination of the gradient image and the marker function. Based on the image's gradient or intensity information, the algorithm divides the low-contrast region into distinct sections. After that, visualize the image as a topographic surface with gradients or pixel intensities that match elevations by using a suggested watershed transformation. The watershed transform is applied to the gradient or intensity image treating the intensity values as elevations. This turns the grayscale image into a segmented image with watershed lines separating the various groups. The watershed transform W_T, is often stated mathematically as in equation (3):

$$W_T(x,y) = \nabla f(x,y) \quad (3)$$

The image is divided into regions once the watershed transform has been calculated. The borders between the areas are established by the watershed lines. Equation (4) expresses the resulting segmented regions.

$$Se(W_T) = s_{e1}, s_{e2}, s_{e3}, \ldots., s_{en} \quad (4)$$

The segmentation process assigns each pixel in the image to an expected region based on the watershed lines. The watershed segmentation approach, used after pre-processing, reliably segments individual leaves in the image, even when the borders between leaves are not clearly defined or overlapping features.

Adaptive Centricity Multi-Spectral Capsnet

The existing algorithms are not able to extract features from the hierarchical vein structure of leaves effectively because they are unable to capture global contextual information and long-range connections. Hence the new algorithm named Adaptive Centricity Multi-spectral CapsNet is introduced for feature extraction. ACMC algorithm is utilized for feature extraction from leaf vein patterns, in this an ACHL Algorithm extracts local features like vein spacing and branching angles, and MA-Caps Net is primarily used to capture contextual information.

Adaptive Centricity Hough Line Detection (ACHL) Algorithm:

The ACHL Algorithm is an effective method in image processing and computer vision for extracting detailed local features from images, particularly those seen in natural structures such as leaf veins. ACHL functions by methodically analyzing a leaf's vein network to extract critical information such as vein spacing and the angles at which secondary veins branch out from the main veins. Firstly, the AHT, a part of ACHL is employed, that adapts its parameters in response to local image features. This adaption helps it to accurately distinguish primary and secondary veins in leaf images. The fundamentals of the AHT, which represent characteristics of interest such as veins, are specified with polygons and parametrically defined. The approach converts the shapes of interest to a parameter space that is accurately analyzed. In the

Cartesian system of coordinates (x_l,y_l), a line is defined as in equation (5):

$$y_l = q_l x_l + b_l \quad (5)$$

The intercept with the y-axis is represented by bl, while the slope of the line is represented by q_l. Every point (x_l, y_l) in Cartesian coordinates corresponds to a line in (q_l, r_l). This space displays all conceivable lines based on their slope and intercept. Equation (6) depicts the translation from Cartesian coordinates to parameter space or Hough space.

$$b_l = -q_l x_l + y_l \quad (6)$$

The Hough Transform is stated in polar coordinates (θ, ρ), where θ is the line's angle with the horizontal axis and ρ is the minimum distance from the origin. These parameters are connected to b_l and q_l through Equations (7) and (8):

$$\rho = x_l \cdot \cos(\theta) + y_l \cdot \sin(\theta) \quad (7)$$

$$\theta = \arctan\left(-\frac{y_l}{1}\right) \quad (8)$$

This representation is useful for detecting veins at any angle. The AHT adjusts its settings in response to local visual properties such as intensity gradients and curvature. This adaptation enables the AHT to detect both primary and secondary veins by tailoring its parameters to the vein patterns. The AHT detects primary and secondary veins in the leaf by transforming the image into Hough space. The parameter adaption ensures that the AHT is responsive to vein characteristics, allowing it to extract vein information more effectively. Once the veins are spotted, the Medial Axis Transform (MAT) is utilized to extract the skeleton of the leaf veins. The MAT represents the basic structure of the vein network, including the secondary veins and their branching patterns, as a skeleton.

The first step is to create a binary picture of the leaf veins, with foreground pixels and background pixels. The method often starts to compute the distance transform of the binary image that contains the leaf veins. This procedure determines the distance from each pixel in the picture to the nearest border. The output is a distance map, with each pixel value representing the distance to the nearest border. The distance transform (D_T) for a binary image (I_m) is mathematically defined as follows in equation (9):

$$D_T(x, y) = \min_{(x', y') \epsilon I_m} \|(x, y) - (x', y')\| \quad (9)$$

Where the positions of the pixel being processed are represented as (x, y), (x', y') are the positions of border pixels, and $\|\cdot\|$ is the Euclidean distance. After obtaining the distance transform, locations along the vein structures where the distance function displays local maxima are identified to determine the skeleton of the leaf vein network. To do this, the binary image is thinned, which is an iterative process that involves removing pixels from the boundaries of the object while maintaining its general structure and connectivity. The vein network's topology is maintained during the thinning process, which is usually achieved by morphological operation. The leftover pixels create the skeleton, referred to

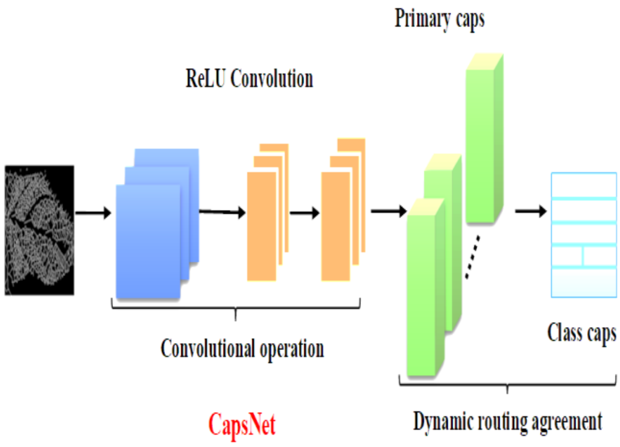


Fig.2. Multi-spectral Caps Net

As the Medial Axis, which accurately depicts the center vein system of the leaf as the reduction in thickness proceeds. The next step is connecting the local maxima points obtained during thinning to construct the skeleton. This is achieved by tracing paths between neighboring local maxima points, ensuring that the lines follow the object's centerline. By following these steps, the MAT accurately extracts the skeleton of the leaf veins, capturing both the secondary veins and the object's branching patterns.

After getting the leaf vein skeleton, ACHL continues to extract local characteristics, such as vein spacing and the angles at which primary veins split off into subsidiary veins. Vein spacing, which gives important details about the general distribution of veins all over the leaf, is calculated by measuring the distance between adjacent veins along the skeleton. In the meanwhile, the angles generated between the primary and secondary veins are calculated by examining the junction locations along the skeleton to identify the angles of branching. Mathematically, the computation of vein spacing S_v is represented as in equation (10):

$$S_v = \frac{1}{M}\sum_{i=1}^{M} D_T \quad (10)$$

Where D_(T)is the distance between adjacent vein segments and M is the total number of vein segments. Also, calculating branching angles requires figuring out the angle that forms at a branching point between two adjacent vein segments. This is stated as follows in equation (11):

$$\theta_j = arccos\left(\frac{v_i.v_{i+1}}{\|v_i\|\|v_{i+1}\|}\right) \quad (11)$$

Where the angles between the two adjacent vein segments are denoted by θ_j, and the unit vectors indicating the directions of the two vein segments are represented by v_i and v_{i+1}. Adaptive Hough Transform and Medial Axis Transform together create an excellent basis for extracting specific information from leaf vein patterns. When combined, these algorithms allow for the calculation of vein spacing, the precise position of branching points, and the extraction of the skeletonized vein network, which improves the understanding of leaf shape and facilitates a range of uses, including the identification of species.

Multi-spectral Attentional CapsNet (MA-CapsNet)

A novel deep learning architecture called MA-CapsNet is used to extract the global features from leaf vein patterns in images. To efficiently capture the hierarchical structure and contextual data present in vein networks, it combines the features of multi-spectral channel attention and CapsNet, each offering unique advantages in enhancing the model's performance.

attention mechanism is expressed as follows in equation (12):

$$C_i = \sigma(W_f * F_i) \cdot F_i \quad (12)$$

The channel-wise attention weights for the i^{th} feature map is denoted by C_i, while $F_i$represents the i^{th} feature map and W_f denotes the weight of the feature map. Using this attention mechanism, Multi-spectral Channel Attention enables the model to flexibly attend to spectral channels containing significant information, hence improving its capacity to capture distinguishing aspects of vein patterns across multiple spectral bands. Following the multi-spectral channel attention, MA-CapsNet employs CapsNet, which excels in capturing the hierarchical organization of vein networks.

Fig. 2. CapsNet structure

Figure 2 depicts the structure of the proposed CapsNet. The CapsNet is made up of three layers: convolutional, primary, and digital. The CapsNet enables a hierarchical depiction of venous networks. Each vein segment is represented as a capsule, containing information about its orientation, location, and connection to higher-order veins. CapsNet uses dynamic routing to discover the hierarchical connections among various venous segments. The dynamic routing method iteratively adjusts the weights to more frequently allocate capsule outputs from the previous layer to capsules in the following layer. This routing makes it easier for the model to represent intricate spatial linkages and long-range interdependence between vein segments. The dynamic routing process involves three main steps, initially, the routing weights are calculated using a softmax function. Equation (13) calculates the dynamic routing weight D_{ij} where i indexes capsules in the lower-level layer and j indexes capsules in the higher-level layer.

$$D_{ij} = \frac{e^{b_{ij}}}{\sum_m e^{b_{im}}} \quad (13)$$

The outputs of the low-level capsules are weighted and summed to obtain the input of the high-level capsule j: Lower-level capsules collect the local properties of vein segments, whereas higher-level capsules capture more abstract representations such as the spatial relationships between vein segments,

$$R_j = \sum_{i=1}^{m} D_{ij} u_{j|i} \quad (14)$$

Equation (14) computes the weighted sum R_j of inputs from lower-level capsules for each higher-level capsule j. Then the weighted sum R_j is passed through a nonlinear activation function to obtain the output of the high-level capsule z_j:

$$z_j = \frac{\|R_j\|^2}{1+\|R_j\|^2} \cdot \frac{R_j}{\|R_j\|^2} \quad (15)$$

Where $\|R_j\|$ is the Euclidean norm of the R_j. This activation function enables the CapsNet to capture complex spatial interactions between vein segments and express them hierarchically. MA-CapsNet, which combines Multi-spectral Channel Attention and CapsNet, provides an effective method for feature extraction from leaf vein patterns. This integration process involves considering not only the spatial arrangement of vein segments but also their relationships across different spectral bands, allowing the model to infer richer contextual information about the leaf vein patterns. It tackles the limitations of vein hierarchical structure and successfully maintains contextual information over local spatial patterns.

Bidirectional Relational Prototypical LSTM

Once after extracting the features, the unique neural network architecture called Bi-RP LSTM is implemented for leaf detection. This novel approach combines two essential components: RPN and Bi-LSTM networks. Together, these transform leaf analysis by improving its understanding of global spatial connections and the sequential arrangement of sharp edges along the leaf margin. Initially, RPNs utilize prototype vectors to represent each class or category. This prototype vector serves as a reference point, encapsulating the class's core characteristics. In the scenario of leaf detection, each class represents a distinct amount of serration depth.

The main aim of RPNs is to capture the relative information between serrations in a leaf by expressing each class (or degree of serration depth) with a prototype. The set of serrations on a leaf is represented as $S = \{s_1, s_2, \dots \dots \dots, s_N\}$, where N is the total number of serrations. Each serration s_i is represented as a feature vector x_i. Also have a collection of prototype vectors representing different levels of serration depth, represented by $P = \{p_1, p_2, \dots \dots, p_K\}$ where K is the number of classes or depth levels. The distance between a serration (x_i) and a prototype (p_k) is computed using a distance metric, which is expressed in the following equation (16).

$$d(x_i, p_k) = \|x_i - p_k\|^2 \tag{16}$$

A softmax function is then employed to characterize the connection between a serration and a prototype, resulting in a probability distribution across the classes, expressed in equation (17).

$$P(y_i = k|x_i) = \frac{e^{-d((x_i, p_k)}}{\sum_{j=1}^{K} e^{-d((x_i, p_k)}} \tag{17}$$

RPNs capture the relationships between different serrations, enabling the network to understand the global spatial relationships within the leaf. Furthermore, the Bi-LSTM is designed to analyze data sequences, such as analyzing the sequential arrangement of serrations along the leaf margin. Each serration is considered a data point in the sequence, and because Bi-LSTM is sequential, it processes these data points in the proper sequence. The input data for leaf margin analysis frequently appears as a series of features, with each feature representing various aspects of the leaf border, such as serration length, angle, or curvature. The input sequence of serration characteristics is denoted as $S = \{s_1, s_2, \dots \dots \dots, s_N\}$, where N is the number of serrations.

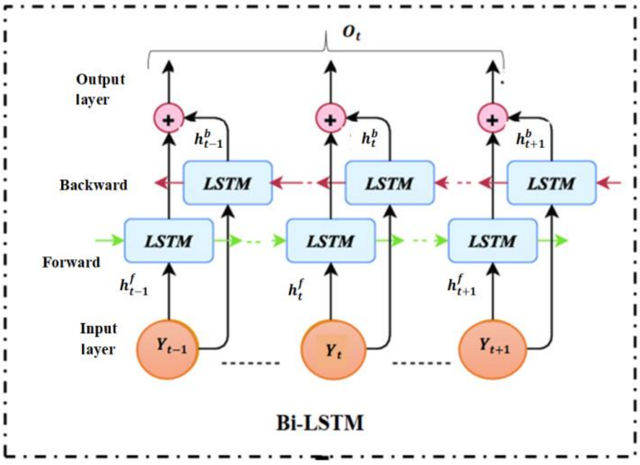


Fig. 3. **Schematic diagram of Bidirectional LSTM**

A schematic diagram of Bidirectional LSTM is shown in above Figure 3. Bi-LSTM consists of two LSTM layers they are forward and backward. The forward LSTM processes the sequence from left to right, whereas the reverse LSTM processes it from right to left. By processing incoming data bi-directionally, the network takes into account information from neighboring areas when evaluating each serration. This enables the network to gather contextual information and identify the relative locations of serrations along the leaf margin.

The bidirectional processing of the input sequence allows the network to grasp the precise position of serrations inside the leaf margin. By considering input from both directions, the network infers spatial correlations between serrations, such as their distances and structure patterns. Recognizing relative placements is essential for correctly detecting and classifying serrations with varied depths. The LSTM unit comprises multiple components, including an input gate, a forget gate, and an output gate. They are determined as follows in equation (18), (19) and (20):

$$I_t = \sigma(Y_t W_{hi} + h_{t-1} W_{hi} + b_i) \tag{18}$$

$$F_t = \sigma(Y_t W_{hf} + h_{t-1} W_{hf} + b_f) \tag{19}$$

$$O_t = \sigma(Y_t W_{ho} + h_{t-1} W_{ho} + b_o) \tag{20}$$

At time step t, the hidden states of the forward and backward LSTMs are designated as h_t^f and h_t^b, respectively.

Forward and reverse processing is expressed as follows in equations (21) and (22):

$$h_t^f = \emptyset\left(Y_t W_{yh}^{(f)} + h_{t-1}^f W_h^{(f)} + b_h^{(f)}\right) \quad (21)$$

$$h_t^b = \emptyset\left(Y_t W_{yh}^{(b)} + h_{t-1}^b W_h^{(b)} + b_h^{(b)}\right) \quad (22)$$

These forward and backward hidden states capture information from the past and future contexts of each serration. The combination of the hidden state is expressed in the following equation (23)

$$h_t = h_t^f \oplus h_t^b \quad (23)$$

Where ⊕ represents the summation by element, which is used to sum the elements of the forward and reverse outputs. Bi-LSTM accurately captures contextual information by integrating the outputs of both forward and backward LSTMs. Each hidden state comprises information from both the present serration and its neighboring serrations in both directions. It allows the network to understand the relative positions of serrations within the leaf margin, thus enhancing its ability to accurately identify and differentiate between different features of the leaf. The final output of the Bi-RP LSTM for each serration is obtained by combining the outputs of the relational prototypical network and the bidirectional LSTM:

$$Y_{final} = \{h_t^f, h_t^b, P(y_i = k|x_i)\} \quad (24)$$

Bi-RP LSTM provides potential improvements in leaf detection tasks by combining the features of RPNs with Bi-LSTM, allowing for the investigation of spatial connections and the distinction between regions of differing depths. The RPN component captures the global spatial correlations between serrations, even while the Bi-LSTM component examines the sequential arrangement of serrations along the leaf margin using both local and contextual information. This extensive approach improves detection accuracy by using both spatial and sequential information obtained from leaf structures.

Overall, this approach uses advanced image processing, deep learning, and sequence modeling approaches to reliably recognize and analyze leaf attributes in video frames. It solves issues such as noise, vein hierarchy, and spatial interactions at the leaf margin, eventually improving leaf identification accuracy.

Result and discussion

This section provides a detailed description of the implementation results and the performance of the proposed system to ensure that the proposed technique performs better and provides an accurate detection of leaf from videos with leaf venation and margin.

System configuration

The proposed system is simulated in Python, and this section provides a detailed description of the implementation findings and performance of the proposed system, as well as a comparison section to ensure that the proposed system works properly.

Software	: Python
OS	: Windows 10 (64-bit)
Processor	: Intel i5
RAM	: 8GB RAM

Simulation output of the proposed model

The simulation modeling strategy of the proposed Multi-spectral CapsNet and Relational Prototypical LSTM involves training the models on a diverse dataset of leaf images with annotated vein patterns and serration depths.

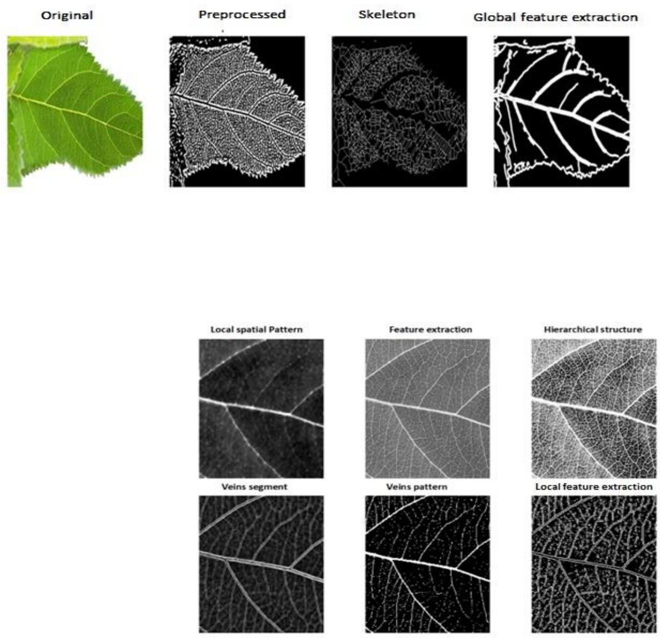


Fig. 4. **Detected leaf.**

Figure 4 shows the detected leaf. The original frame has a close-up shot of a green plant with multiple leaves. The second frame seems to be a modified version of the first. The identical leaves are visible, but one of them has a prominent yellow line painted across it, following the central vein from base to tip. This line appears to highlight or select the leaf, presumably as part of an image-processing assignment for identifying and isolating healthy leaves. The detected leaf is a close-up image of a single leaf from the same plant for further processing.

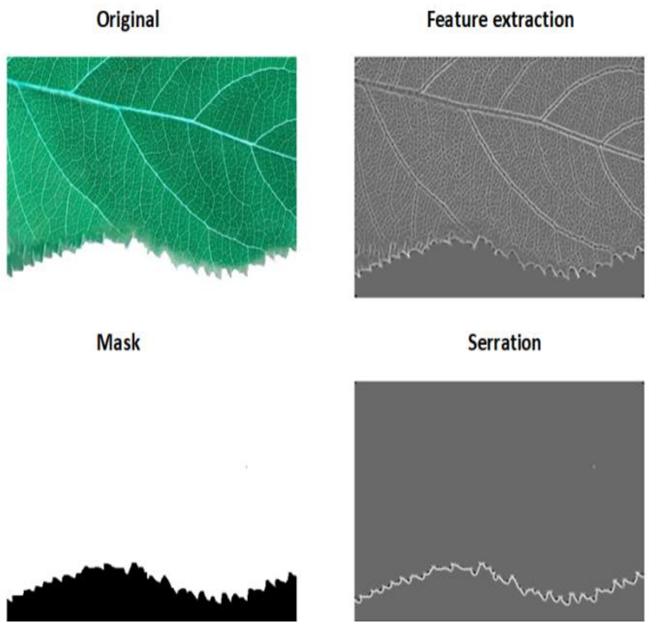


Fig. 5. **Local feature extracted image.**

Figure 5 illustrates the local feature extracted image. The leaf's skeleton image, which was obtained using the Medial Axis Transform, shows the basic layout of the vein network, including the primary and secondary veins. This representation of the leaf's vascular system is clear and concise. Vein spacing and branching angles are emphasized in the local feature extracted image by using the Adaptive Hough Transform to better capture the fine details of the leaf's vein patterns.

depicts the global feature extracted image. Vein networks are arranged hierarchically, with principal veins serving as the main framework and secondary, tertiary, and higher-order veins branching off of them. Multi-spectral channel attention ability to extract vein segments reveals intricate spatial details that are essential for species identification, such as vein spacing and branching angles. MA-CapsNet's vein patterns identify spectral channels that hold pertinent information, improving contextual awareness beyond local spatial patterns. The hierarchical structure of vein networks is represented by global features that the Capsule Network extracts from leaf images. These features successfully capture long-range interdependence by encapsulating orientation, location, and connection information.

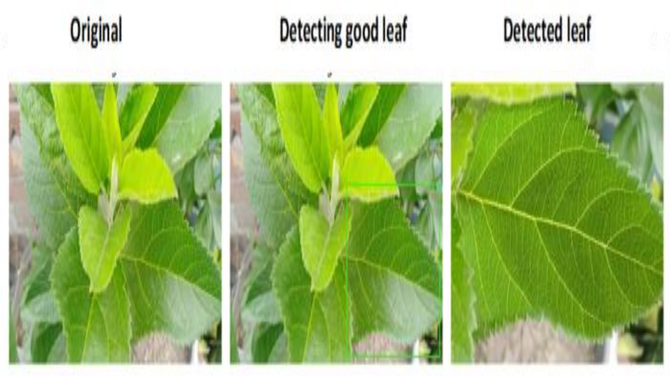
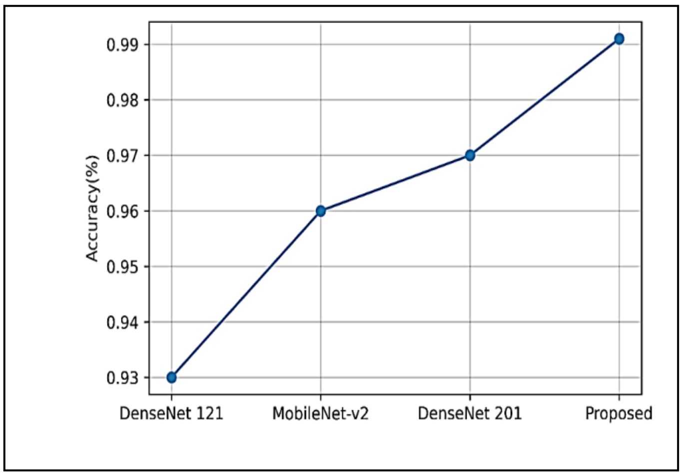


Fig.6.G**lobal feature extracted image**

Figure 7 demonstrates the spatial relationships within the leaf margin. The image of the serration mask displays the unique forms and sizes of the serrations along the leaf border



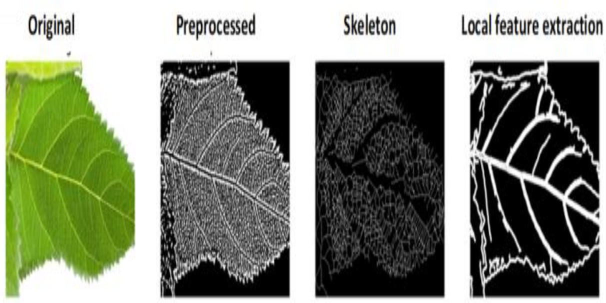
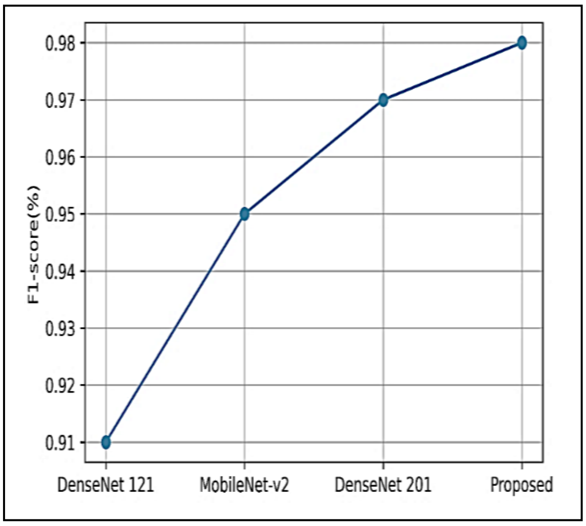


Fig.7.Serration

And also their defined lines. Complex details like vein spacing and branching angles are shown in the image with the extracted serration features, providing data on the structural properties of serrations. Lastly, a complete perspective of the serrations is provided by the serration image, which shows their various depths and locations along the leaf margin. In addition to providing essential inputs for the Bi-RP LSTM to effectively identify and analyze spatial connections within the leaf margin, these pictures collectively demonstrate the intricacies of leaf serrations.

Comparative analysis of the proposed model

This section highlights the proposed Multi-spectral CapsNet and Relational Prototypical LSTM leaf detection model with the traditional models and the achieved outcome was explained in detail in this section by comparing it with DenseNet 121 [12], MobileNet-v2 [12], and DenseNet 201 [12], InceptionV3 [26], InceptionResNetV2 [26], MobileNetV2 [26], efficientNetB0 [26] NN [27], SVM [27], KNN [27], RNN [27],



and BILSTM [27] showing their results based on various metrics such as accuracy, recall, precision, false positive rate, sensitivity, specificity, loss and F1-score.

Figure 18 shows the accuracy comparison of the proposed leaf detection method with the existing methods. The existing methods such as DenseNet 121, MobileNet-v2, and DenseNet 201 achieve an accuracy of 0.93%, 0.96%, and 0.97%. Compared with existing models the proposed model attains the highest accuracy of 0.9912%. This indicates that the proposed method better identifies leaves from images, leading to fewer misclassifications.

Fig.8.Comparison of accuracy

A comparison of recall of the proposed leaf detection model to the existing models is illustrated in the figure 6 Compared with existing techniques the proposed technique achieves the highest recall value. The existing techniques such as DenseNet 121, MobileNet-v2, and DenseNet 201 achieve a recall value of 0.95%, 0.93%, and 0.96% respectively, and the proposed model achieves a recall value of 0.9913%.vThis suggests that the proposed model is more effective at capturing all relevant instances of leaf features,

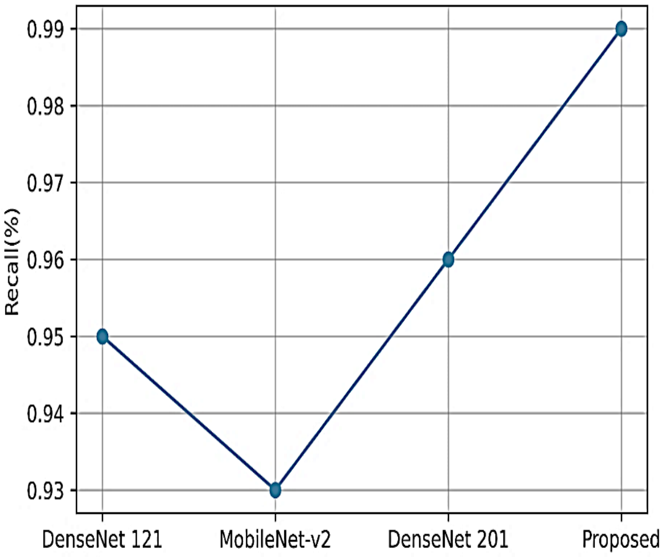


Fig.9.Comparison of recall

A comparison of the F1-score of the proposed leaf detection model to the existing models is illustrated in the figure 20. Compared with existing techniques the proposed technique achieves the highest F1-score value. The existing techniques such as DenseNet 121, MobileNet-v2, and DenseNet 201 achieve an F1-score value of 0.91%, 0.95%, and 0.97% respectively, and the proposed model achieves an F1-score value of 0.9823%. This indicates improved accuracy and recall balance, demonstrating the suggested method's capacity to detect leaves properly while minimizing false positives and negatives

Overall, in the result area, from the proposed methodology a comparison is made with the existing methods, and the techniques were explained using graphs. This shows that the technique that is used in the leaf detection with venation and margin using Multi-spectral CapsNet and Relational Prototypical LSTM has comparatively higher accuracy, precision, recall, and less computation time than the previous techniques that are taken for the comparison.

Conclusion

In conclusion, the "Multi-spectral CapsNet and Relational Prototypical LSTM" technique that has been suggested offers an exhaustive method to address the difficulties associated with automated leaf detection. Using the integration of novel techniques in pre-processing, feature extraction, and detection, this system demonstrated outstanding abilities in precisely recognizing leaves from images. While MA-CapsNet improved contextual awareness through multi-spectral attention, the combination of Adaptive Centricity Multi-spectral CapsNet and ACHL Algorithm ensured robust feature extraction by collecting minute details of leaf vein patterns. Moreover, Bi-RP LSTM presented an innovative approach to leaf detection by using bi-directional sequential analysis and relational prototype networks to identify intricate spatial interactions inside leaf margins. Compared with existing models such as DenseNet 121, MobileNet-v2, DenseNet 201, InceptionV3, InceptionResNetV2, RNN, and BILSTM the proposed model achieves a high specificity value of 0.99%, sensitivity of 0.98%, precision 0.98%, and accuracy of 0.9912%. The proposed model attains a low computation time of 0.015s and a loss value of 0.008 a high F1-score of 0.9823%, recall value of 0.9913%. This proves that the proposed leaf detection approach performed well when compared to other existing techniques. Its great precision, resilience, and computing time make it promising for a variety of uses in agriculture, and environmental monitoring and it make accurate and effective leaf analysis possible in a variety of conditions.

Conflict of Interest Statement:

Regarding the publishing of this work, the authors affirm that they have no conflicts of interest. All the research, analysis, and conclusions presented in this study have been carried out independently, without any financial, personal, or professional interests that could have influenced the work.

Reference

[1]Parekh, R., Bhattacharya, S., and Chaki, J. (2020). A hierarchical technique to classifying plant leaves using numerous attributes. Computer and Information Sciences Journal of King Saud University, 32(10), pp. 1158–1172.

[2] Ahmad, N., Saleem, G., Younus, M.U., Asif, H.M.S., Anwar, S., and Anjum, M.R., 2021. Plant disease detection using leaf images by utilising texture and colour characteristics. 1139–1168 in Wireless Personal Communications, 121(2).

[3] In 2020, Muneer, A. and Fati, S.M. effective and automated method for classifying herbs using deep learning based on texture and shape characteristics. Access IEEE, 8, pp. 196747-196764.

[4] Singh, P.K., and M. Sood, 2020. hybrid system that uses qualitative textural feature analysis to identify and categorise plant diseases. Computer Science Procedia, 167, pp. 1056–1065.

[5].Su, J., Chen, Q., Wu, Z., and Wang, M. (2020). Quick identification of plant leaves with KNN optimisation and enhanced multiscale triangular representation. IEEE Access, 8, 208753–208766 pages.

[6]Triki, A., Mahdi, W., Gaikwad, J., and Bouaziz, B. (2021). Detecting and segmenting leaves from digital photos of herbarium specimens using mask R-CNN is known as "deep leaf." pp. 76–83 in Pattern Recognition Letters, 150.

[7]
Dubey, A.K., Thomas, M.T., and Thanikkal, J.G. (2020). Medicinal plant identification using a unique shape descriptor algorithm (SDAMPI) and a condensed image database. IEEE Journal of Sensors, 20(21), 13103–13109.

[8]In 2021, Zhang, L., Weckler, P., Lan, Y., Lee, J.M., **a, C., **ao, D. a method for detecting leaves that uses joint segmentation and leaf skeleton identification to go from coarse to fine. Engineering of Biosystems, 206, pp. 94–108.

[9]Wang, S., Kong, W., Li, D., Shi, G., and Chen, Y. (2020). A framework for extracting phenotypic features and segmenting leaves from multiview stereo plant point clouds. IEEE Journal of Selected Topics in Remote Sensing and Applied Earth Observations, 13, pp. 2321–2336.

[10] Amean, Z.M., Hancock, N., and Low, T. (2021). Stereovision-based automatic leaf segmentation and **overlap** leaf separation. 12, p. 100099; Array.

[11]Alves, F.M., Gonçalves, E.G., de Alencar Figueiredo, L.F., Abreu, U.G., Arruda, R., Bao, F., Weber, V., Bambil, D., Pistori, H., and Bortolotto, I.M., 2020. Machine learning and artificial neural networks are used to identify plant species based on colour, shape, texture, and learning resources. Systems and Decisions in the Environment, 40(4), pp. 480–484.

[13]Wang, S., Kong, W., Li, D., Shi, G., and Chen, Y. (2020). A framework for extracting phenotypic features and segmenting leaves from multiview stereo plant point clouds. IEEE Journal of Selected Topics in Remote Sensing and Applied Earth Observations, 13, pp. 2321–2336.

[14] Amean, Z.M., Hancock, N., and Low, T. (2021). Stereovision-based automatic leaf segmentation and **overlap** leaf separation. 12, p. 100099; Array.

[15] Alves, F.M., Gonçalves, E.G., de Alencar Figueiredo, L.F., Abreu, U.G., Arruda, R., Bao, F., Weber, V., Bambil, D., Pistori, H., and Bortolotto, I.M., 2020. Machine learning and artificial neural networks are used to identify plant species based on colour, shape, texture, and learning resources. Systems and Decisions in the Environment, 40(4), pp. 480–484.

[16] Deng, Z., Wu, Q., Zhang, K., and Ji, M., 2020. Convolutional neural networks-based multi-label learning for the identification and assessment of crop leaf diseases. 24 Soft Computing, pp. 15340-1527.

[17] Suma, V., and Pankaja, K., 2020. Plant leaf classification and recognition using random forest (RF) and whale optimisation algorithm (WOA). Series B, 101, pp. 597–607, Journal of the Institution of Engineers (India).

[18] J. Huixian, 2020. study of deep learning and artificial neural networks for plant picture recognition. Access, IEEE, 8, pp. 68828–68841.

[19] C. Yang (2021). By combining texture and shape characteristics, plant leaves can be recognised. p. 107809 in Pattern Recognition, 112.

[20] Hajam, M.A., Khanday, A.M.U.D., Arif, T., and Neshat, M. (2023). A Successful Fine-Tuning Ensemble Convolutional Learning Model for Medicinal Plant Leaf Recognition. p. 618 in Information, 14(11).

[21] Salve, P., Sardesai, M., and Yannawar, P. (2022). Multimodal plant identification utilising a hybrid feature fusion method that combines hyperspectral data from imaging and non-imaging sources. Computer and Information Sciences Journal of King Saud University, 34(1), pp. 1361–1369.

[22] Zhang, Z., Hu, J., Song, W., Xiong, Q., Gao, X., Chen, Z., and Zhang, Y. (2021). Detection of plant diseases using DoubleGAN-generated leaves. IEEE/ACM Transactions on Bioinformatics and Computational Biology, 19(3), pp. 1817-1826.

[23]Manimaran, V., Rajagopal, S., Ashwinkumar, S., and Jegajothi, B. (2022). automatic identification and categorisation of plant leaf diseases with the use of convolutional neural networks based on MobileNet. Proceedings of Materials Today, 51, pp. 480–487.

[24]Leonowicz, Z., Maji, A.K., Jasiński, M., Hassan, S.M., and Jasińska, E. (2021). Plant-leaf disease detection with CNN and transfer-learning methodology. p. 1388 in Electronics, 10(12).

[25] Kumar, G.P., Rajesh, V., and Dudu, B., August 2022. Classification of Plant Leaves using Bidirectional Long Short-Term Memory and Deep Feature Fusion. (pg. 68-73) in the 2022 International Conference on Innovations in Science and Technology for Sustainable Development (ICISTSD). IEEE.