Applied Biological Research 25(3): 361-371; (2023)
DOI: 10.5958/0974-4517.2023.00040.X

SIMILAR PLANT SPECIES CLASSIFICATION WITH DUAL INPUT TRANSFER LEARNING MODELS

Parul Sharma* and Pawanesh Abrol

[1]*Department of Computer Science and Information Technology, University of Jammu, Baba Saheb Ambedkar Road, Gujarbasti, Jammu - 180 006, Jammu & Kashmir (India)*
*e-mail: *sharmaparul1620@gmail.com, pawanesh.abrol@gmail.com*

(Received 30 March, 2023; accepted 23 July, 2023)

ABSTRACT

The classification of plant species is crucial for preserving biodiversity, identifying endangered species, preventing diseases, and controlling weeds. However, with the vast number of species and their similar physical appearances, identifying and classifying them correctly can be difficult for human experts. Artificial Intelligence, such as Convolution Neural Networks (CNNs), can aid in automatic plant species recognition, especially for identifying similar species that require the learning of subtle differences. In the present investigation, transfer learning models such as MobileNet, AlexNet, and GoogLeNet were used to classify ten citrus species that look alike. A dataset of images of fruits and leaves of these citrus plants was compiled, and all images were segmented and their backgrounds were subtracted to create a new segmented dataset. In addition, instead of using only one organ as input to the CNN, dual inputs of both fruits and leaves were also used. The classification accuracy of 78.25% was achieved when MobileNet model was used on original dataset of fruits and it extended to 97.25% when GoogLeNet model was used on segmented dataset when both the organs were used as input. Evidently, present study provides innovative methods and techniques to accurately distinguish and classify visually similar plant species.

Keywords: Citrus species classification, dual-input CNN, Grad-CAM, similar species classification, transfer learning.

INTRODUCTION

Plants are essential for human survival, providing oxygen, food, and medicine, while also maintaining ecological balance and reducing air pollution. With over 390,000 plant species, there is a need for supervision and governance of their abundance, conservation, and classification (Dueck, 2023). Limited experts in this domain make it challenging to identify, classify, detect disease, and prevent weeds. The variability in appearance and characteristics of plants necessitates simplification to fit them into a species. This highlights the importance of using automatic techniques using Artificial Intelligence (AI), to accurately identify plant species.

Automatic plant species recognition (Gaston *et al.*, 2004) is a valuable tool for professionals like scientists, farmers, and students, which uses image recognition techniques to reduce the taxonomical gap (Wäldchen *et al.*, 2018). Deep learning, specifically Convolution Neural Network (CNN), is a powerful form of artificial intelligence that has enabled plant recognition systems to perform similarly to human experts (Lee *et al.*, 2015; Goëau *et al.*, 2018, Sarvamangala *et al.*, 2021). CNN can also be implemented by using the transfer learning approach in which the weights learned by the pre-trained model can be applied to a new model, which can then be fine-tuned on a new dataset as implemented by Wang *et al.* (2019). Some of the names of such transfer learning architectures are AlexNet

(Krizhevsky *et al.*, 2012), MobileNet (Andrew *et al.*, 2017), GoogLeNet (Szegedy *et al.*, 2015), etc. The transfer learning approach has also gained popularity for plant species classification purposes. They were used by Kamal *et al.* (2019) for the leaf classification belonging to 28 categories and achieved 99.74% accuracy. Liu *et al.* (2019) recognized the Chrysanthemum flowers belonging to 104 cultivars, with the help of VGG16 and ResNet50 in a very less amount of time. In another study conducted in Pereira *et al.* (2019), the AlexNet model was used on re-processed images to identify Grape variety and the test accuracy came out to be 77.30%. Barbedo *et al.* (2018) used GoogLeNet transfer learning approach on background subtracted images and achieved the accuracy of 87% which increased from 84% when the original dataset was used. Leaves have traditionally been used for plant species classification due to their distinct features (Smith *et al.* (2022)), but according to findings by Lee *et al.* (2018); Rzanny *et al.* (2019), considering other plant organs and combining multiple inputs can enhance classification accuracy.

The present investigation was aimed to classify ten citrus species using high-quality natural images of leaves and fruits. A new balanced dataset was created with the purpose to enable the Convolutional Neural Network (CNN) to distinguish subtle differences between similar-looking plant species, making it more accurate in identifying and classifying them in real-life scenarios. The research used three criteria: two datasets (original and segmented), three fine-tuned transfer learning models (MobileNet, AlexNet, and GoogLeNet), and three inputs (leaves, fruits, and both).

MATERIALS AND METHODS

Proposed method

The key steps involved in the present research process is described in Fig. 1. It started with dataset collection wherein new dataset was prepared with the images of fruits and leaves from ten citrus species. Then a second dataset 'the segmented dataset' was prepared by segmenting and removing their background from all original images. The original and segmented dataset had images of fruits and leaves. These two were considered as input individually and combined. A total of six selections were made which were then given to the CNN model. Three transfer learning approaches namely MobileNet, AlexNet and GoogLeNet were used to obtain classification results of citrus species. Through this systematic approach, the study aimed to achieve accurate citrus species identification via image analysis.

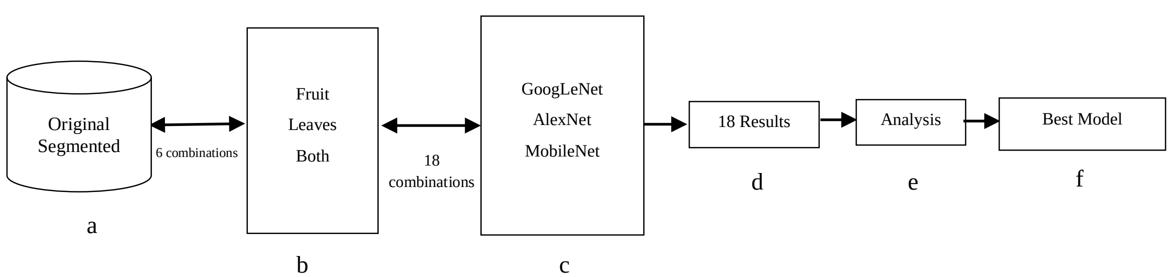

Fig. 1: The flowsheet of the work. The figure displays six stages of workflow, including a) Preparation of two datasets, b) Selection of input to CNN, c) Selection of transfer learning model, d) Recording of experiment results, e) Analysis of results, and f) Analysis of results for best model

Dataset

Previous research in plant species classification has predominantly focused on using images of leaves only, limiting the scope of input data to a single organ of the plant for machine learning models (Wäldchen *et al.,* 2018). Additionally, the majority of studies as can be observed in table 1, have relied on lab-controlled images to train these models which might not fully represent the complexity and variability of real-world conditions.

Table 1: Standard datasets for plant species classifications

Dataset	Reference	Components	Images description
Leafsnap	Kumar *et al.* (2012)	Leaves	Scan and photos in plain background
Flavia	Wu *et al.* (2007)	Leaves	Scan and photos in plain background
Swedish leaf	Söderkvist (2001).	Leaves	Scan in plain background
ICL	Hu *et al.* (2012)	Leaves	Scan and photos in plain background
Oxford Flower 17	Nilsback & Zisserman (2006)	Flowers	Photos in real environment
Oxford Flower 102	Nilsback & Zisserman (2008)	Flowers	Photos in real environment
DRGV	Pereira *et al.* (2018)	Fruits-grapes	Photos in real environment

In present study, two new datasets were created with the images of leaves and fruits of ten citrus plants *viz*., *Citrus X sinensis* (sweet orange), *Citrus nobilis x citrus deliciosa* (kinnow), *Citrus jambhiri* (rough lemon), *Citrus pseudolimon* (hill lemon), *Citrus* × *paradisi* (grapefruit), *Citrus maxima* (pomelo), *Citrus limetta* (sweet lime), Citrus × *aurantiifolia* (lime), × *Citrofortunella microcarpa* (calamondin orange), and *Citrus limon* (lemon). Fig. 2 shows one sample image of both fruit and leaf from original datasets of these species.

The images were captured from the citrus orchard of S.K. University of Agricultural Science and Technology Jammu, J&K and some from local parks and gardens. Datasets were created by adding

Fig. 2: Image samples of fruits and leaves from ten citrus species (Narangi, Lemon, Kinnow, Lime, Grapefruit, Pomelo, Rough Lemon, Hill Lemon, Sweet Lemon, and Sweet Orange). The first row displays Narangi, Lemon, Kinnow, Lime, and Grapefruit, while the second row shows Pomelo, Rough Lemon, Hill Lemon, Sweet Lemon, and Sweet Orange from left to right

200 images for leaves and fruits as well of ten citrus plants. A total of 4000 images in complete dataset were processed for segmentation and background removal to form new segmented datasets. All the images were captured using mobile devices like Samsung galaxy A71, Samsung galaxy M52, i-phone 11 and DSLR Nikon D5300. The variability of capturing devices was kept to improve the generalizability of model. The resolution and size of captured images were variable and very high for CNN processing. GoogLeNet, AlexNet and MobileNet were designed to take the input image of size 224 x 224 pixels, 256 x 256 pixels and any resolution > 32 x 32 pixels, respectively. The optimization and classification of models were done by using their specified input size. Augmentation techniques like horizontal flip, vertical flip, rescaling, cropping, rotation by 30^0, height and width shifting were also employed.

Segmentation and background removal

The images captured in natural environment tend to have a lot of noise and undesired objects. Initially CNN treats all the components of images equally, and then with subsequent training it learns about important features and patterns. To have an insight of what the final convolution layer looks at such images as an important feature, Selvaraju *et al.* (2017) used Gradient-weighted Class Activation Mapping (Grad-CAM) to produce course heat map of image highlighting the important regions. The 2nd and 3rd column of Fig. 3A-D rows highlights the fruit, part of human hand, other plant leaf, and background fruit as important features, respectively. In all images segmentation followed by background subtraction was carried out to remove the undesired parts.

A brief explanation of the procedure of both the processes are given as follows:

Segmentation: The images were first cropped from all the sides to give the plant component maximum space in the image. Segmentation was

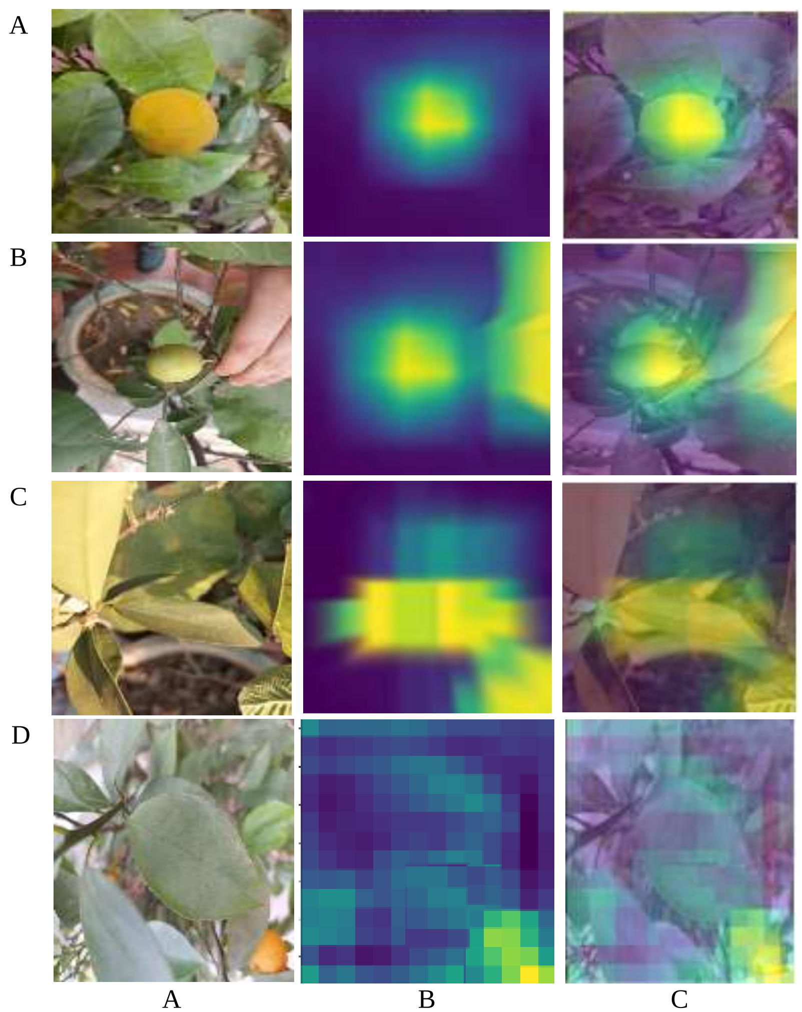

Fig. 3: Input image of leaf and fruit of citrus plant using Grad-CAM. The figure has three columns: a) the original image, b) a heatmap highlighting important regions, and c) the output generated by overlapping the original image and the heatmap. The rows represent different scenarios, including A) when only fruit is present, B) when a part of human hands is present, C) when leaves of other plants are overlapping, and D) when fruits in the background are existing.

carried out with the help of Canny edge detector (Canny, 1986), followed by morphological operations to remove the background. For Canny edge detector algorithm, the coloured RGB image was first converted into a grayscale image. The grayscale image was then smoothed using Gaussian distribution function (Sekehravani *et al.*, 2020).

$$G(p,q) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{p^2+q^2}{2a^2}} \quad \text{(i)}$$

Where p and q represent horizontal and vertical distances from the origin and σ the distribution standard deviation. Less σ value means low level of smoothening where edges are protected and vice versa is for more σ value. At each pixel, horizontal and vertical gradients were computed with the help of image convolution with following derivative filters (Canny, 1986).

$$\text{Horizontal filter} \quad G_p = \frac{dG(p,q)}{dp} \quad \text{(ii)}$$

$$\text{Vertical filter} \quad G_q = \frac{dG(p,q)}{dq} \quad \text{(iii)}$$

Next the gradient magnitude is evaluated using following equations (Sekehravani *et al.*, 2020):

$$\text{Edge gradient amplitude } G = \sqrt{G_p^2 + G_q^2} \quad \text{(iv)}$$

$$\text{Edge gradient direction } \theta = \tan^{-1} = \frac{G_q}{G_p} \quad \text{(v)}$$

Two threshold values selected were: the high threshold H_{th} and a low threshold L_{th}. The H_{th} helps to keep the main edge but may discard small lines which result in discontinuous edges, whereas L_{th} retains more small edges. L_{th} helps in continuing the discontinued edges by H_{th}.

Background removal: At this point the edges of main foreground object were obtained. The background removal is the process of removing unwanted background from image as conducted by Mohamed *et al.* (2010). Two morphological operations namely erosion and dilation were used, wherein the structural element replaced the current pixels to the minimum and maximum values of pixels defined in set, respectively. The output obtained from Canny edge was processed with dilation first and then by erosion. The resultant convex went through the processes of polynomial fill (convex, colour). After that, the resultant mask was dilated and then eroded once again. then Gaussian blur was applied to further reduce the noise. Finally, the masks of three channels were concatenated on original image to get the background subtracted image.

Deep convolution neural network

For classification purposes various transfer learning models were benchmarked. To evaluate their efficacy in complex task of classifying similar plant species three popular transfer learning architectures of CNN namely GoogLeNet (Szegedy *et al.*, 2015), AlexNet (Krizhevsky *et al.*, 2012), and MobileNet (Andrew *et al.*, 2017) were implemented. They were originally designed as a response to worldwide challenge "Large Scale Visual Recognition Challenge" (ILSVRC) [Russakovsky *et al.*, 2015] on the ImageNet dataset (Deng *et al.*, 2009). GoogLeNet architecture is based on 9 inception modules on the principal of "network in network". A module comprises three parallel convolutional blocks of size 1 x 1, 3 x 3 convolutions and 5 x 5 convolutions along with max pooling layer. It is capable of capturing a greater number of features from image. Although it is deeper with 22 layers and wider than AlexNet, it has only 5 million parameters. AlexNet model has 5 convolutions layers followed by 3 fully connected layers. The first two convolution layers consist of both normalization and pooling layer whereas fifth convolution layer has only one pooling layer. The features computed from these convoluted layers are fed to three fully connected layers of which the last layer output classes are set to 10. The softmax function does the exponential normalization (0 to 1) of input from last layer and computes the distribution of 10 classes. The first seven layers comprise Rectified Linear Unit (ReLu) activation function and the last two fully connected layers consist of 50% dropout value. There are about 60 million parameters generated in AlexNet. MobileNet is a model that is designed for lightweight execution of problems for mobile phones and embedded system. It is made up of depth wise convolution that mainly comprises two operations which are depth-wise convolution followed by point-wise convolution. It has 28 layers and relatively very fewer parameters but it is capable of providing high accuracies.

In such models the important features from the earlier layers of the pre-trained network are used to train smaller network with much lesser number of parameters. Having learnt the features from previous training, these networks only need to learn about the relations of the present problem. The layers in these models are either froze to make its parameters non-trainable or set free to enable its parameters to learn new features from current training. The model can be used as feature extractor in which all the layers of network, except the last fully connected layer, are freeze and last layer is altered by placing suitable classifier. The other way is to fine tune the network in which all the layers are retrained which gives better classification results buy may result in overfitting. In this present research work, the first few layers are kept frozen while leaving rest of the higher layers trainable as depicted in Fig. 4. It gives better performance at cost of higher computational resources.

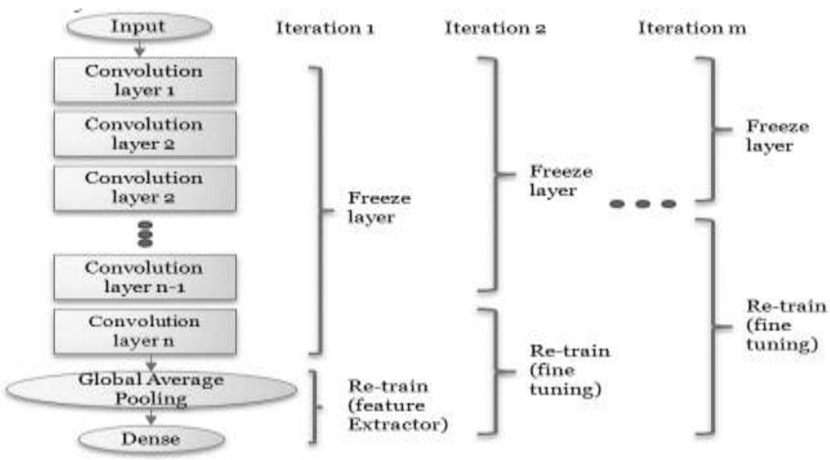

Fig. 4: Fine-tuning of layers in transfer learning approach

The hyper parameters of the three test models were kept uniform for fair comparison of their classifications. The hyper parameter values were: Empirical optimization: Stochastic gradient descent; Momentum: 0.9; Weight decay: 0.001; Activation function: Rectified linear unit; Base learning rate: 5×10^{-3}, Learning rate schedule: decreased by factor of 10 whenever the validation error stagnated. The training and testing data ration was kept as 80 and 20, respectively, across all the experiments which made 160 and 320 images for training per species when single and both organs were used as input, respectively. Likewise, 40 and 80 images were left for testing. A consistent batch size of 40 was employed throughout all experiments, and the epoch value was set at 50 since the results converged within this limit.

Evaluation metrics

Evaluation metrics are essential tools in assessing the performance of machine learning algorithms and classifiers. Performance metrics based on True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP) provide valuable insights into the effectiveness of model. TP occurs when the model correctly predicts the positive class when it is indeed positive. TN happens when the model correctly predicts the negative class when it is indeed negative. FN occurs when the model mistakenly predicts the negative class when it should have been positive. Lastly, FP takes place when the model incorrectly predicts the positive class when it should have been negative. By considering these outcomes, we can calculate various performance metrics like accuracy, specificity, precision, recall, and F1-score.

Accuracy: It is one of the most straightforward evaluation metrics, representing the proportion of correct predictions made by a model over the total number of predictions. (Sokolova *et al.,* 2009).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{(vi)}$$

Specificity: The specificity focuses on the true negative rate, calculating the proportion of correctly predicted negative instances relative to the total number of actual negative instances (Sokolova *et al.*, 2009).

$$\text{Specificity} = \frac{TN}{TN+FP} \quad \text{(vii)}$$

Precision: Precision, also known as positive predictive value, measures the proportion of true positive predictions over the total number of positive predictions made by the model. It indicates the model's ability to avoid false positives, making it essential in scenarios where false positives are costly (Davis *et al.,* 2006).

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{(viii)}$$

Recall: Recall, also known as sensitivity or true positive rate, measures the proportion of correctly predicted positive instances relative to the total number of actual positive instances. Recall becomes crucial when the cost of false negatives is high, as it shows the model's effectiveness in identifying positive cases (Davis *et al.,* 2006).

$$\text{Precision} = \frac{TP}{TP+FN} \quad \text{(ix)}$$

F1 score: It is the harmonic mean of precision and recall, offering a balanced evaluation metric that considers both false positives and false negatives (Davis *et al.*, 2006).

$$\text{Precision} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad \text{(x)}$$

RESULTS AND DISCUSSION

The model-wise classification accuracies of three CNN models demonstrated that when using original and segmented images of fruits, leaves, and both organs as inputs. It was found that GoogLeNet outperformed the fine-tuned models of AlexNet and MobileNet when using segmented datasets as inputs, achieving accuracies of 91.50, 94.50, and 97.25% for fruits, leaves, and both organs respectively. Interestingly, leaves were found to be more accurate in classifying species than fruits, with an increase in accuracy ranging from 0.75 to 1.25% for original datasets and 3 to 4% for segmented datasets.

Table 2: Classification accuracies (%) of three CNN models

Plant organ	Dataset	MobileNet	AlexNet	GoogLeNet
Fruit	Original	78.25	80.50	82.50
	Segmented	88.75	90.00	91.50
Leaves	Original	79.25	81.75	83.25
	Segmented	92.75	93.75	94.50
Both	Original	87.50	89.25	91.25
	Segmented	94.25	95.50	97.25

The species-wise classification accuracies of the ten citrus species has been presented in Table 3. The results are given for segmented datasets when 3 CNN models were used to classify these species using the images of leaves, fruits and both organs-based as inputs. The results indicated that rough lemon had the highest average accuracy of 95.79% among the species, while sweet orange had the lowest of 86.43%. The highest individual accuracy was achieved for Pomelo when both of its organs were used as input for GoogLeNet model, with a 100% accuracy rate. On the other hand, the lowest accuracy of 76.62% was observed for Grapefruit when its fruit images were classified using the MobileNet model.

Impact of using segmented new dataset on the classification accuracy

The difference in accuracies between the original dataset 'A' and the segmented dataset 'B' is displayed in Fig. 5. The range of this difference varied from 6.0 to 13.5%, depending on the dataset used, the model applied and the type of organ (leaves or fruits or both) used as input. The difference in accuracies

Table 3: Species wise classification accuracy (in percentage %) on segmented dataset

Citrus plant name	Botanical name	MobileNet			AlexNet			GoogLeNet			Class Mean
		Fruits	Leaves	Both	Fruits	Leaves	Both	Fruits	Leaves	Both	
Sweet orange	*Citrus X sinensis*	82.50	84.13	85.25	82.89	86.89	87.72	84.51	90.00	94.00	86.43
Lime	*Citrus × aurantiifolia*	95.18	87.50	93.44	96.15	90.32	96.49	95.83	93.33	98.04	94.03
Lemon	*Citrus limon*	88.89	93.94	96.77	89.61	93.65	98.25	90.28	95.00	98.00	93.82
Hill lemon	*Citrus pseudolimon*	89.87	93.94	95.08	90.67	95.24	94.64	91.30	95.00	96.00	93.53
Rough lemon	*Citrus jambhiri*	95.12	93.85	95.08	96.10	95.24	96.49	97.22	95.00	98.00	95.79
Narangi	*Citrofortunella microcarpa*	91.36	92.42	93.33	92.21	92.06	92.73	92.96	91.67	96.00	92.75
Grapefruit	*Citrus × paradise*	76.62	96.97	93.33	77.78	96.83	92.73	79.10	96.67	93.88	89.32
Kinnow	*Citrus nobilis x citrus deliciosa*	88.75	87.69	86.67	89.33	87.10	87.50	91.43	88.14	87.76	88.26
Pomelo	*Citrus maxima*	96.34	92.31	95.00	96.10	91.94	96.43	95.77	91.53	100.00	95.05
Sweet lime	*Citrus limetta*	82.50	90.91	90.16	82.89	90.48	92.86	84.51	90.00	93.88	88.69

for leaves was higher, ranging from 11.25 to 13.5% as compared to the fruits wherein accuracies ranged from 9.0 to 10.5%. This could be attributed to the fact that the images of naturally photographed leaves tend to have more noise than those of fruits, which the segmentation process has helped to mitigate, resulting in higher accuracies for dataset B. Overall, the accuracies of models were generally higher for dataset B, ranging from 88.75 to 97.25%, as compared to the dataset A, which ranged from 78.25 to 91.25%, depending on the model used and the type of organ used as input. The highest accuracies were achieved in GoogLeNet with both organs as input in both datasets.

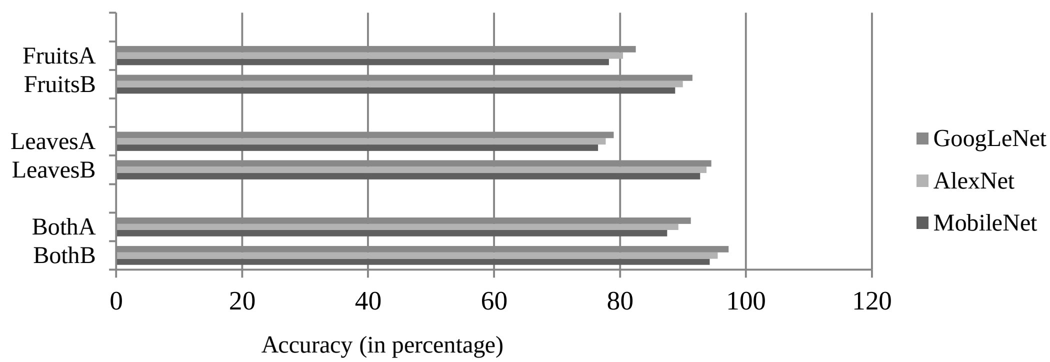

Fig. 5: Accuracy difference between original (A), and segmented dataset (B).

Impact of citrus species organ appearance on evaluation metrics with focus on specificity, precision, recall and F1 score

The use of various metrics, *viz*., specificity, precision, recall, and F1-score was examined to assess the effect of species appearance on classification accuracy. Table 4 shows a breakdown of these metrics in a case where the segmented dataset was analysed using GoogLeNet. The highest specificity was achieved for rough lemon and Pomelo when both organs were used as input, as well as for grapefruit leaves, and rough lemon and Pomelo fruits, which yielded similar results. It revealed that the precision was high for lemon, rough lemon, and Pomelo when either of the organ was used as input. Specifically, these fruits had 100% precision rate showing that the model made no false positive predictions. The same was observed for grapefruit leaves, and rough lemon and Pomelo fruits, showing that the model was highly accurate in identifying positive examples for these inputs. In case both the components of lime and Pomelo were used, there was 100% recall, meaning that the model accurately identified all the instances of these species. The F1 score signified that Pomelo and kinnow plants were most accurately classified and least accurately classified, respectively, when both components were used as input to the GoogLeNet model. The model also performed well in identifying grapefruit leaves, rough lemon and Pomelo fruits, but struggled in classifying kinnow leaves and grapefruit fruits. Rzanny *et al.* (2019) used five perspectives of plants (entire plant, flower frontal- and lateral-view, leaf top- and back-side view) to classify 101 plant species using CNN and achieved highest top-1 accuracy of 97.1% when all the perspectives were combined. Lee *et al.* (2017) pre-trained the two path CNN with the features of ImageNet dataset by Deng *et al.* (2009), and then both the paths were re-purposed to train on organs and others on generic information. In species layer, both the generic and organ layer information were encapsulated to classify species. Lee *et al.* (2018) extended their work and incorporated Recurrent Neural Network instead of CNN to allow classification based on different views of a plant, capturing one or more of its organs and optimizing the relationships between them. He *et al.* (2016) demonstrated that using multiple plant organs as input to a multi-column deep CNN achieved superior results as compared to the single-organ input. Notably, combining leaf with flower produced highest score of 91.7%, while the entire plant and branch combination lowered the score to 28.5%. This highlights the importance of considering multiple organs for accurate classification. The perusal of literature reveals that no existing information addresses the identification of similar-looking plant species. Lack of prior research highlights the unique nature of this study and the novelty of approach. The results of present study are expected to bridge the gap in knowledge and contribute to the field of plant identification.

Table 4: Species-wise classification specificity (S), precision (P), recall (R) and F1-score (F) (data in %)

Citrus plant name	Fruits				Leaves				Both			
	S	P	R	F	S	P	R	F	S	P	R	F
Sweet orange	97.77	82.22	92.50	8705	98.88	90.47	95.00	92.68	99.44	95.12	97.50	96.29
Lime	99.72	97.43	95.00	96.20	99.44	95.00	95.00	95.00	99.72	97.56	100.00	98.76
Lemon	98.61	88.37	95.00	91.56	99.72	97.43	95.00	96.20	100.00	100.00	97.50	98.73
Hill lemon	99.72	97.22	87.50	92.10	99.72	97.43	95.00	96.20	99.72	97.50	97.50	97.50
Rough lemon	100.00	100.00	95.00	97.43	99.72	97.43	95.00	96.20	100.00	100.00	97.50	98.73
Narangi	99.44	94.87	92.50	93.67	99.16	92.68	95.00	93.82	99.72	97.50	97.50	97.50
Grapefruit	97.77	80.48	82.50	81.48	100.00	100.00	95.00	97.43	99.72	97.43	95.00	96.20
Kinnow	99.44	94.73	90.00	92.30	98.88	90.24	92.50	91.35	98.88	90.47	95.00	92.68
Pomelo	100.00	100.00	95.00	97.43	99.44	94.87	92.50	93.67	100.00	100.00	100.00	100.00
Sweet lime	98.05	83.72	90.00	86.74	98.88	90.47	95.00	92.68	99.72	97.43	95.00	96.20

An enhanced deep learning model that incorporates dual-input transfer learning on segmented dataset has been developed for classification of similar looking citrus species. Different state of the art deep learning models such as MobileNet, AlexNet and GoogLeNet were fine-tuned and used for the classification. Citrus dataset was constructed with naturally captured images after the unwanted data which affect the performance of CNN models as demonstrated by Grad-CAM results were removed. All three fine-tuned CNN models performed well in classifying 10 species by using original dataset and it increased significantly by using segmented dataset. For fruits, leaves and both the difference were 10.50, 13.5 and 6.75% by using MobileNet; 9.5, 12.0, 6.25% by AlexNet and 9.0, 11.25, and 6.0% by using GoogLeNet. GoogLeNet, followed by AlexNet performed considerable well in each scenario. It was better at classifying images from original dataset as compared to other two. The results also indicate that when two organs: fruits and leaves were combined as input to CNN models it resulted in enhanced classification with jump in about accuracy about 2% (only leaves as input) - 8% (in case of fruits) than when only single organ was used. Another promising finding was that the leaves were better at discriminating citrus species than fruits. The classification accuracy was 78.25% when MobileNet model was used on original dataset of fruits and it extended to 97.25% when GoogLeNet model was used on segmented dataset when both the organs were used as input.

The research results are constrained by the small number of citrus species studied, limited use of transfer learning, and limited input organs. Therefore, exploring a more comprehensive dataset that includes citrus species from various regions and climates could provide more robust and generalizable outcomes. Secondly, there is possible dataset bias and limited generalizability issues as the dataset used in the study was captured from a specific citrus orchards and local gardens. To mitigate this limitation, future research must focus on data from broad range locations to ensure that the results are not influenced by the characteristics of specific sample. Besides, there is need to explore alternative transfer learning models and input organs to complement and reinforce the study's conclusions.

Conclusion: The present work focused on developing innovative methods and techniques to accurately distinguish and classify visually the similar plant species. By leveraging advanced technologies such as machine learning algorithms and image processing techniques, one can have a reliable and efficient solution to the challenges posed in the identification of visually similar plant species. The research findings not only expand our understanding of plant diversity but also reveal the potential future use of dual input transfer learning model in various domains such as conservation, agriculture, and environmental management.

REFERENCES

Andrew, G. and Menglong, Z. 2017. Efficient convolutional neural networks for mobile vision applications. *Mobilenets*. *arXiv preprint arXiv:1704.04861*.

Barbedo, J.G.A. 2018. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and Electronics in Agriculture,* **153:** 46-53.

Canny, J. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**: 679-698.

Davis, J. and Goadrich, M. 2006. The relationship between Precision-Recall and ROC curves. pp. 233-240. **In:** *Proceedings of the 23rd international conference on Machine learning*. 26-29 June 2006. Pittsburgh, Pennsylvania, USA.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. pp. 248-255. **In:** *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 20, 2009. Miami, USA.

Dueck, K. 2023. How many plants are in the world? AZ Animals. [*https://a-z-animals.com/blog/how-many-plants-are-in-the-world/*].

Gaston, K.J. and O'Neill, M.A. 2004. Automated species identification: Why not? *Philosophical Transactions of the Royal Society of London B Biological Sciences.* **359**(1444): 655-667.

Goëau, H., Joly, A., Bonnet, P., Lasseck, M., Šulc, M. and Hang, S.T. 2018. Deep learning for plant identification: How the web can compete with human experts? *Biodiversity Information Science and Standards.* **2**: e25637. [*https://biss.pensoft.net/article/25637/*].

He, K., Zhang, X., Ren, S. and Sun, J. 2016. Deep residual learning for image recognition. pp. 770-778. **In:** *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 27-30 June, Las Vegas, USA.

Hu, R., Jia, W., Ling, H., and Huang, D. 2012. Multiscale distance matrix for fast plant leaf recognition. *IEEE Transactions on Image Processing,* **21**(11): 4667-4672.

Kamal, K.C., Yin, Z., Li, B., Ma, B. and Wu, M. 2019, September. Transfer learning for fine-grained crop disease classification based on leaf images. pp. 1-5. **In:** *10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing*. IEEE. 24-26 September 2019, Amsterdam, the Netherlands.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 25, (eds. F. Pereira, C. Burges, L. Bottou, and K.Q. Weinberger). Curran Associates Incorporated

Lee, S.H. Chan, C.S., Wilkin, P. and Remagnino, P. 2015. Deep-plant: Plant identification with convolutional neural networks. pp. 452-456. **In**: *IEEE International Conference on Image Processing*, 27-30 September, Quebec City, Canada.

Lee, S.H., Chan, C.S. and Remagnino, P. 2018. Multi-organ plant classification based on convolutional and recurrent neural networks. *IEEE Transactions on Image Processing*, **27**(9): 4287-4301.

Lee, S.H., Chang, Y.L., Chan, C.S. and Remagnino, P. 2017. HGO-CNN: Hybrid generic-organ convolutional neural network for multi-organ plant classification. pp. 4462-4466. **In:** *2017 IEEE International Conference on Image Processing*. IEEE. 17-20 September, 2017, Beijing, China.

Liu, Z., Wang, J., Tian, Y. and Dai, S. 2019. Deep learning for image-based large-flowered chrysanthemum cultivar recognition. *Plant Methods,* **15**(1): 1-11.

Mohamed, S.S., Tahir, N.M. and Adnan, R. 2010. Background modelling and background subtraction performance for object detection. pp. 1-6. **In**: *2010 6th International Colloquium on Signal Processing & its Applications*. IEEE. 21-23 May, 2010, Malacca, Malaysia.

Nilsback, M.E. and Zisserman, A. 2006. A visual vocabulary for flower classification. pp. 1447-1454. **In:** *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06),* **2**: IEEE, 17-22 June, 2006, New York, USA.

Nilsback, M.E. and Zisserman, A. 2008. Automated flower classification over a large number of classes. pp. 722-729. **In:** *Sixth Indian Conference on Computer Vision, Graphics & Image Processing*. IEEE. 16-19 December, 2008, Bhubaneswar, India.

Pereira, C.S., Morais, R. and Reis, M.J. 2018. Pixel-based leaf segmentation from natural vineyard images using colour model and threshold techniques. pp. 96-106. **In:** *International Conference Image Analysis and Recognition*. (eds. A. Campilho, F. Karray, & B. ter Haar Romeny). Image Analysis and Recognition, Lecture Notes in Computer Science. Vol 10882. Cham. Springer.

Pereira, C.S., Morais, R. and Reis, M.J. 2019. Deep learning techniques for grape plant species identification in natural images. *Sensors,* **19**(22): 4850 [*https://www.mdpi.com/1424-8220/19/22/4850*].

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. and Berg, A.C. 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, **115**(3): 211-252.

Rzanny, M., Mäder, P., Deggelmann, A., Chen, M. and Wäldchen, J. 2019. Flowers, leaves or both? How to obtain suitable images for automated plant identification. *Plant Methods,* **15**: 1-11.

Sarvamangala, D.R. and Kulkarni, R.V. 2021. Convolutional neural networks in medical image understanding: A survey. *Evolutionary Intelligence*, 1-22. [*https://doi.org/10.1007/s12065-020-00540-3*].

Sekehravani, E.A., Babulak, E. and Masoodi, M. 2020. Implementing canny edge detection algorithm for noisy image. *Bulletin of Electrical Engineering and Informatics,* **9**(4): 1404-1410.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. pp. 618-626. **In:** *Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy.

Sokolova, M. and Lapalme, G. 2009. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, **45**(4): 427-437.

Smith, J., Johnson, L. and Kim, S. 2022. A Review on classification methods for plants leaves recognition. *Journal of Plant Research*, **135**(4): 437-452.

Söderkvist, O. 2001. *Computer Vision Classification of Leaves from Swedish Trees*. Master's Thesis, Department of Electrical Engineering, Computer Vision, Linkoping University, Sweden.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D. and Rabinovich, A. 2015. Going deeper with convolutions. pp. 1-9. **In***:* *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 7-12 June, 2015, Boston, USA.

Wäldchen, J., Rzanny, M., Seeland, M. and Mäder, P. 2018. Automated plant species identification - Trends and future directions. *PLoS Computational Biology,* **14**(4). [*https://doi.org/10.1371/journal.pcbi.1005993*].

Wang, K., Gao, X., Zhao, Y., Li, X., Dou, D. and Xu, C.Z. 2019. September. Pay attention to features, transfer learn faster CNNs. ***In:*** *International Conference on Learning Representations.* April 30, 2020. Addis Ababa, Ethiopia.

Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F. and Xiang, Q.L. 2007. A leaf recognition algorithm for plant classification using probabilistic neural network pp. 11-17. **In:** *IEEE International Symposium on Signal Processing and Information Technology*, IEEE. Giza, Egypt.