High Performance Ensembled Convolutional Neural Network for Plant Species Recognition

S. Anubha Pearline(✉) and V. Sathiesh Kumar

Madras Institute of Technology, Anna University, Chennai 600044, India
anubhapearl@mitindia.edu

Abstract. Recognition of plant species is challenging due to variation in leaves such as its arrangement, color variation, leaf venation, varying margin, inter- and intra-similar species. Photographed images possess difficulties in recognition due to differing lighting conditions, rotations, viewpoints, and color backgrounds. To minimize these issues, a Bi-channel Convolution Neural Network (CNN) involving two pre-trained CNNs (VGG-16 and SqueezeNet) is adopted. The networks are trained individually and their predictions are fused to obtain the final prediction scores. The Bi-channel CNN is evaluated on three datasets, namely, Flavia, Swedish leaf, and Leaf-12 datasets. The optimal γ values obtained for the three datasets (Flavia, Swedish Leaf and Leaf-12) are 0.5, 0.9 and 0.5, respectively. The performance metrics (accuracy, precision, recall and f1-score) obtained with the models trained on Leaf-12 dataset are 97.70%, 0.98, 0.98 and 0.98, respectively. The above-specified values are obtained for a γ value of 0.5.

Keywords: Plant species recognition · Deep learning · Bi-channel architecture · Late fusion

1 Introduction

Plant species recognition [1] using leaves is challenging, as it has different shapes, simple or compound structure and color variations (due to aging or seasonal conditions). Computer vision challenges such as camera viewpoint variation, illumination changes, scale variation and background clutter also prevail as a hindrance in plant species classification.

Convolutional Neural Networks (CNN) are prominent since 2012, as it performs both feature extraction and classification with the aid of convolution, pooling and fully connected layers (FC). Also, in recent times, the fusion of multiple CNNs are carried out to significantly boost the performance of the system. Hence, in this paper, a dual network is adopted for plant species recognition. In the following subsections, related works, methodology, experimental details and the results are discussed.

S. K. Singh et al. (Eds.): CVIP 2020, CCIS 1378, pp. 526–538, 2021.
https://doi.org/10.1007/978-981-16-1103-2_44

2 Related Works

Figueroa et al. [2] proposed a Convolutional Siamese Network (CSN) for plant species recognition. The method is tested on smaller datasets (No of Images/Class = varied between 5 to 30). CSN method consists of twin CNNs. Each CNN holds 3 sets of convolution with ReLU activation function, Maxpooling followed by a global average pooling and 2 fully connected layers. The similarity metric, Euclidean distance is utilized to find the similarity between the pairs of leaves.

Dual Deep Learning Architecture (DDLA) [3] is a feature extraction method incorporated for plant species recognition. The authors integrated two deep learning architectures, namely, MobileNet and DenseNet-121 for DDLA. The extracted features from DDLA are classified using machine learning classifiers. DDLA is evaluated on four leaf datasets (Flavia, Swedish leaf, Folio and custom created Leaf-12).

Rizk [4] proposed a dual path Convolutional Neural Network (CNN) for plant species recognition. Dual-path CNN incorporates two simple CNNs. Each path consists of four sets of convolution and max-pooling layers. The first path extracts the shape features and the second path extracts the venation features. The feature maps from each path are merged and forwarded to a fully connected layer. The augmented Flavia dataset is used in the evaluation of the method.

Siamese-Inception (S-Inception) network [5] is suggested for plant species recognition. It incorporates a technique called few-shot learning. S-Inception architecture utilizes two Inception CNNs connected in a parallel fashion. Spatial Structure Optimizer (SSO) metric is constructed for enhancing the performance of S-Inception. Then, the plant species are classified using K-Nearest Neighbor. The efficacy of S-Inception is determined using Flavia, Swedish leaf, and LeafSnap datasets.

Lee et al. [6] designed a hybrid fusion methodology, Ensemble CNN for recognition of plant species. It includes two networks, namely, Hybrid-Generic Convolutional Neural Network (HGO-CNN) and Plant-StructNet. HGO-CNN integrates the organ and generic knowledge gained from single plant images. Plant-StructNet modelling comprises of Recurrent Neural Network (RNN) emphasizing on multi-images. PlantCLEF 2015 dataset is used for assessing the methodology.

He et al. [7] proposed a bi-channel network for plant species recognition. VGG-16 and SqueezeNet pre-trained CNNs form the Bi-channel system. A stacking layer is used for fusing the predicted probabilities attained from the two pre-trained CNNs. The collected web-crawled images of Orchidaceae plant family is used in the evaluation of the method. The integrated technique achieved an accuracy of 96.81%.

Hu et al. [8] proposed a Multi-Scale Fusion Convolutional Neural Network (MSF-CNN) for plant species recognition. MSF-CNN utilized multiple CNNs for different image size and concatenated the output feature maps from the two consecutive CNNs. Experimental results are reported by using the MalayaKew leaf and the LeafSnap datasets.

Ghazi et al. [9] integrated two pre-trained fine-tuned CNNs (GoogleNet and VGG) for plant species recognition. Also, the authors explored the pre-trained CNNs such as GoogleNet, VGGNet and AlexNet architectures. LifeCLEF 2015 plant dataset is experimented on the integrated CNNs and attained an accuracy of 80%.

Pre-trained CNN models such as GoogleNet and VGG-16 are used in plant species classification. Similarly, pre-trained CNN models are employed in plant disease prediction. Mukherjee et al. [10] recommended early classification of plant disease through GoogleNet to preserve plant species from damage. Wang et al. [11] suggested VGG-16 fine-tuned model for identifying disease in apple leaf.

From the extensive literature survey, it is observed that there is a large number of research works focusing on either hierarchical, early or late fusion of deep learning networks. In recent years, the bi-channel framework gained attention because of its fusion methodology. Hence, in this paper, a bi-channel framework is adopted for plant species recognition. This approach is tested on three leaf datasets, namely, Flavia, Swedish leaf, and Leaf-12 (custom-developed).

3 Methodology

A Bi-channel Convolutional Neural Network (B-CNN) framework with the late fusion approach is proposed for plant species recognition [7]. The following subsections, details about the image preprocessing, bi-channel method and training algorithm.

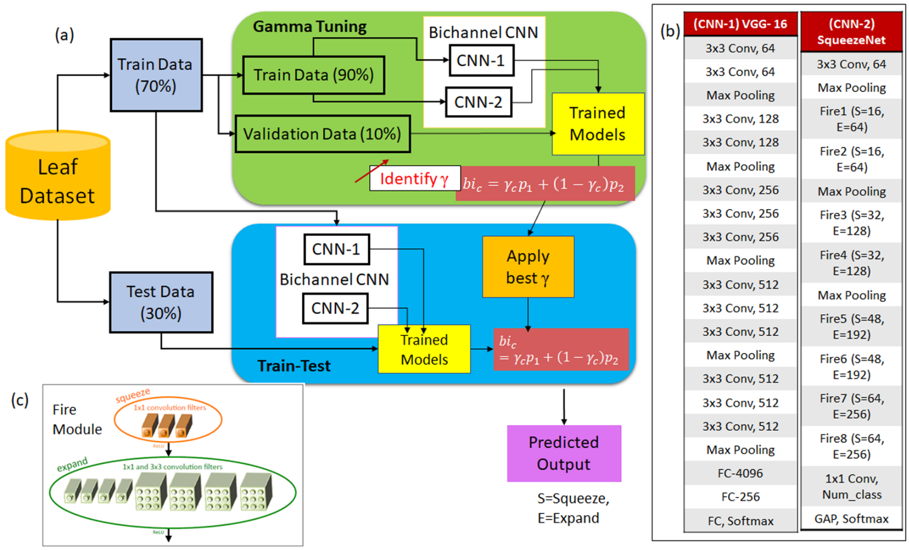

Fig. 1. Workflow for Bi-channel Convolution Neural Network. (a) Workflow (b) CNN-1 and CNN-2 (c) SqueezeNet's Fire Module [11]

Figure 1 illustrates the workflow of Bi-channel Convolution Neural Network. The leaf dataset is split as train (70%) and test (30%) data. The training data are passed to B-CNN architecture. B-CNN architecture consists of two CNN (CNN-1=VGG-16, CNN-2=SqueezeNet) architectures connected in a parallel manner. The two CNN frameworks are trained and validated individually to obtain prediction scores (p1, p2).

Based on the prediction, the weights (γ) are assigned to the individual architecture. It is performed to obtain the overall prediction of the bi-channel architecture. Optimal γ is identified from the Gamma tuning segment. The best Gamma value is used in the retrain process cum test data evaluation mechanism. Finally, performance metrics are estimated and reported.

3.1 Image Preprocessing

Three datasets, namely, Flavia, Swedish leaf and Leaf-12 are used. B-CNN method is also experimented on augmented datasets. The augmentation operations employed in this paper include horizontal flip, vertical flip and Gaussian blur. Also, the augmented dataset comprises the existing images from the dataset.

The images from each of the datasets are reconstructed to 300 × 300 using image resizing function by maintaining the aspect ratio and image quality. Hence, the images are further resized to 100 × 100 × 3 for saving computational time.

3.2 Bi-channel Convolutional Neural Network

Bi-channel CNN [7] is adopted for processing the leaf images. Bi-channel CNN includes two pre-trained CNNs, VGG-16 (CNN-1) and SqueezeNet (CNN-2) architectures. The prediction score from the two architectures is fused, as stated in Eq. (1).

$$bi_c = \gamma_c p_1 + (1 - \gamma_c)p_2, c = 1, 2, \ldots C \tag{1}$$

Where bi_c represents the final prediction score

γ_c is the weight
p_1 is the prediction score of CNN-1
p_2 is the prediction score of CNN-2
c is the leaf classes 1,2….C

CNN-1. VGG-16 [12] architecture is used in one of the branches of bi-channel CNN. It consists of thirteen 3 × 3 convolutions followed by three fully connected (FC) layers. This architecture comprises of five max-pooling layers inserted in between two convolution layers as shown in Fig. 1(b). The number of filters is increased from 64 to 512 as the architecture is deepened.

CNN-2. SqueezeNet [13] architecture is used in the second branch of the bi-channel framework. It consists of a convolution layer, followed by eight fire modules. The last fire module is followed by 1 × 1 convolution and Global Average Pooling (GAP). Fire module consists of operations such as squeeze and expand as shown in Fig. 1(c). Squeeze operation includes of 1 × 1 convolutions. The outputs from the Squeeze operation is fed into the expand operation. The expand operation consists of 1 × 1 and 3 × 3 convolutions. 1 × 1 Convolutions decrease the number of feature maps and functions similar to an FC layer.

3.3 Training of Bi-channel CNN

Bi-channel CNN is trained using SGD (Stochastic Gradient Descent) optimizer [14] with a learning rate of 0.001 and 50 epochs. The results deteriorated when the optimizer (Adam with different learning rates 0.001 and 0.0001) is varied. Algorithm 1 shows the procedure for Bi-channel CNN. The optimal γ is identified by incorporating Manual search mechanism. After training the data, the prediction score (validation data) obtained from CNN-1 and CNN-2 i.e. p1 and p2 are fused for each class by utilizing the equation represented in Step 4 of the algorithm.

The arbitrary weight γ is varied from 0, 0.1, 0.2, 0.3….1 and the final accuracies are predicted. Based on the highest accuracy, optimal weight γ_c is fixed. The best γ obtained is used in the final prediction process.

Algorithm: Late Fusion in Bi-channel CNN

Inputs: Images, I; Models, *m1* and *m2*; prediction score, p_1 and p_2; arbitrary weight, γ; Number of epochs, ep=50, Number of classes, $c \epsilon 1,2..C$

Output: Predicted label, l

Step 1: Resize the Input Images, I to 100x100 pixels

Step 2: Train the images using the Model, *m1* for *ep* number of epochs and obtain the prediction score, p_1 using the validation data.

Step 3: Train the images using the Model, *m2* for *ep* number of epochs and obtain the prediction score, p_2 by using the validation data.

Step 4: Using p_1 and p_2, find the optimal arbitrary weight, γ by employing equation (1).

$$bi_c = \gamma_c p_1 + (1 - \gamma_c) p_2, c = 1, 2, \ldots C$$

Step 5: Retrain the bi-channel CNN architecture with best γ value acquired from validation data and evaluate the method by using the test data.

Step 6: The performance metrics are estimated and reported.

4 Experimental Results and Discussion

The experiments on plant species recognition system are implemented using Windows 10, 64-bit OS, CPU with Intel i7 processor 20 GB RAM and NVIDIA Titan X GPU with 3584 CUDA cores. The code implementation is carried out using Python 3.5, supporting python packages such as Keras with Tensorflow backend, OpenCV, scikit-learn, and os. Three leaf datasets (Flavia [15], Swedish leaf [16], and Leaf-12) are used in the experimental investigation.

4.1 Flavia Dataset

Flavia [15] is a standard leaf dataset consisting of 32 plant species. For each class, 50 images are utilized. The performance metric (Accuracy) of the individual CNN models is obtained by varying the size of input images in the dataset. The dimensions of the input images considered in the studies are $100 \times 100 \times 3$ and $224 \times 224 \times 3$. The estimated results are listed in Table 4. It is observed that the accuracy in the prediction of plant species is improved for larger input images (224 × 224 × 3) compared to smaller images (100 × 100 × 3). It also results in increased computation time for larger images. Also, the model utilizes parameters on a high scale for larger images. Hence, to save computational time and easier model reuse, image size of 100 × 100 × 3 is further employed for experimental evaluation (Table 1).

Table 1. Results obtained by varying input image size

Image Size	Number of parameters	Individual model training time (s)	B-CNN accuracy (%)
100 × 100 × 3	**CNN-1: 34, 650, 208**	**CNN-1: 74.91**	**94.38%**
	CNN-2: 722, 496	**CNN-2: 26.24**	
224 × 224 × 3	CNN-1: 118, 536, 288	CNN-1: 285.76	96.88%
	CNN-2: 738 ,912	CNN-2: 81.34	

The augmented dataset consists of 6400 images with 200 images per each class. Using train-validation set, the optimal γ value is identified as γ = 0.5 (validation accuracy = 97.99%) and is depicted in Fig. 2. The non-augmented dataset (Number of images = 1600) resulted in an optimum γ value of 0.7.

The accuracies achieved by using the augmented dataset (97.71%) is higher compared to the dataset without augmentation (94.38%). Other metrics such as precision, recall and F1-score are 0.98, 0.98 and 0.98, respectively. Also, Fig. 3 shows the confusion matrix curve plotted for the augmented Flavia dataset.

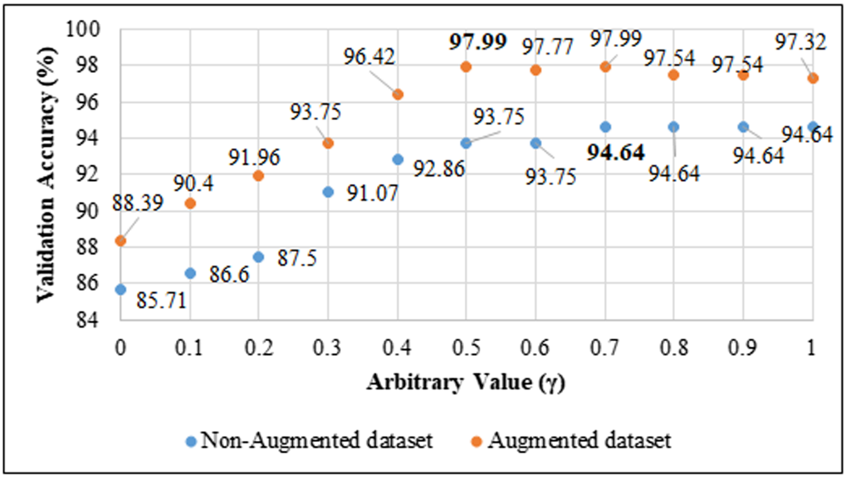

Fig. 2. Accuracies obtained by varying γ for Flavia dataset

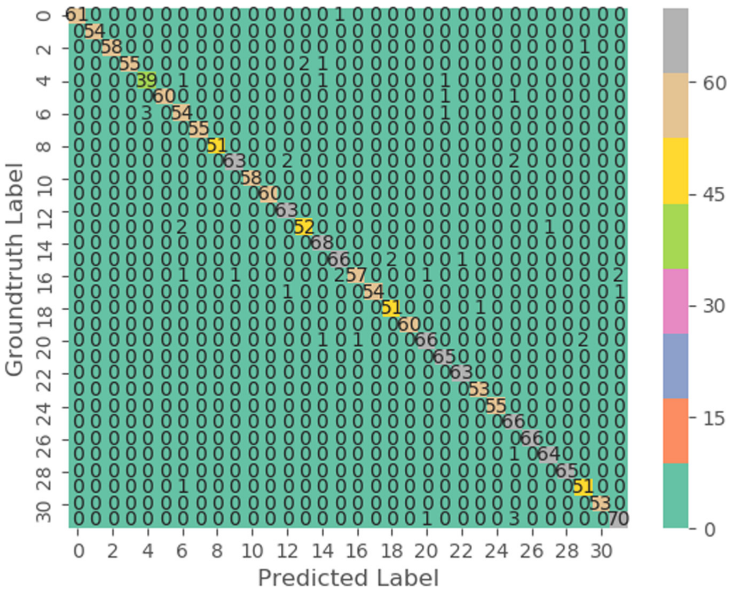

Fig. 3. Confusion matrix for Augmented Flavia dataset

4.2 Swedish Leaf Dataset

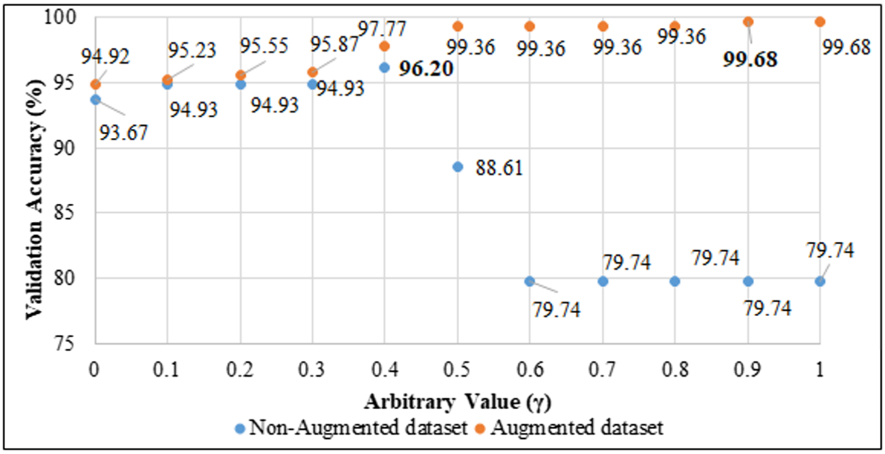

Fig. 4. Accuracies obtained by varying γ value for Swedish leaf dataset

Swedish leaf [16] is a standard dataset consisting of 15 plant species. Each of the plant species contains 75 images. The augmented dataset consists of 4,500 images with 300 images for each class. The optimal γ is estimated to be 0.9 (validation accuracy = 99.68%) from validation data and it is shown in Fig. 4. Similar to Flavia dataset, the dataset with and without augmentation resulted in different optimum γ values.

The augmented dataset resulted in an accuracy, precision, recall and F1-score values of 98.67%, 0.99, 0.99 and 0.99, respectively. The accuracy of Swedish leaf dataset without augmentation is 94.97%. Confusion matrix obtained by using the augmented Swedish leaf dataset is represented in Fig. 5.

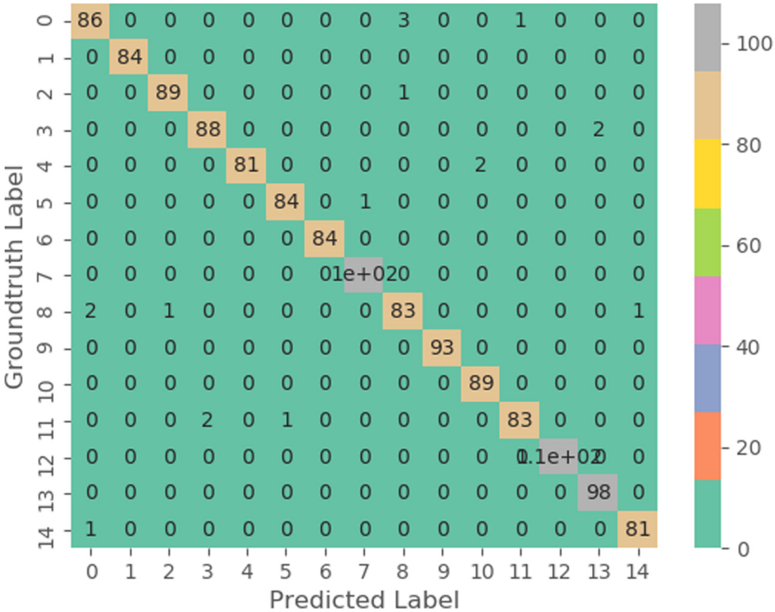

Fig. 5. Confusion matrix for Augmented Swedish leaf dataset

4.3 Leaf-12 Dataset

Leaf-12 is a self-collected real-time dataset. It consists of 12 leaf classes and 320 images per class. The total number of images in the augmented real time dataset is of 15,360. From the train-validation data, the optimal γ is estimated to be 0.5 (validation accuracy = 97.76%). Figure 6 shows the accuracies obtained on varying γ (train-validation). For Leaf-12 dataset, the γ value remained the same for non-augmented as well as the augmented dataset.

For the train-test data, the estimated accuracy, precision, recall and F1-score are 97.70%, 0.98, 0.98 and 0.98. The non-augmented dataset resulted in an accuracy of 95.49%. Confusion matrix obtained for the augmented Leaf-12 dataset is shown in Fig. 7.

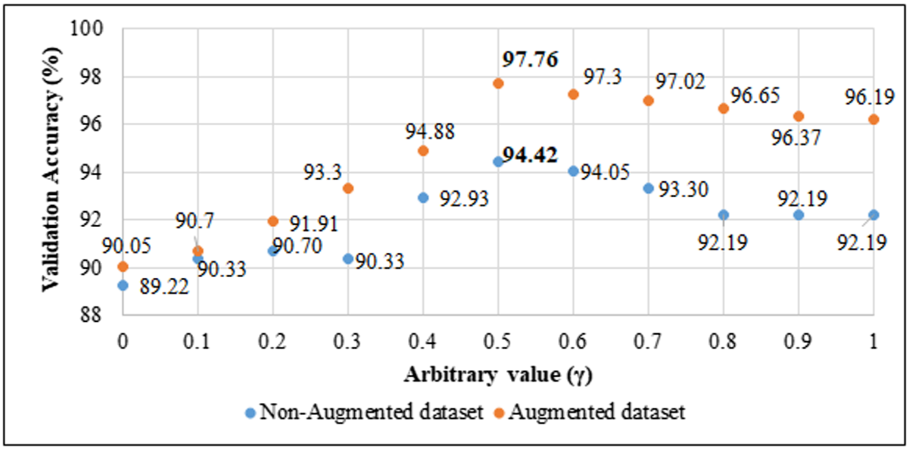

Fig. 6. Accuracies obtained by varying γ for Leaf-12 dataset

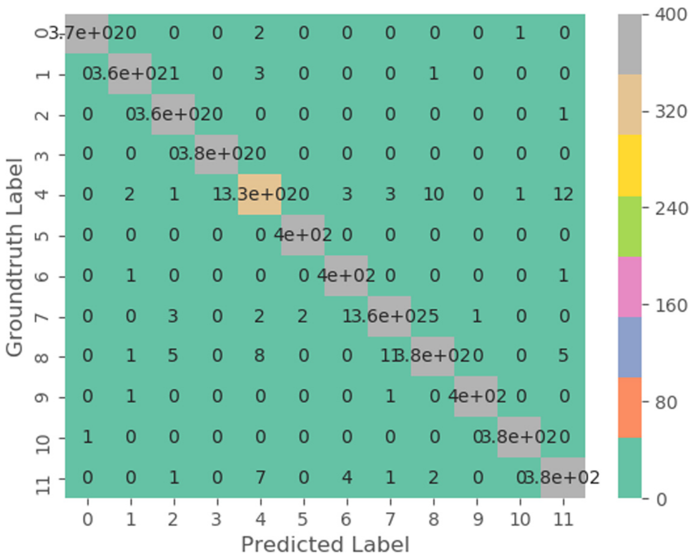

Fig. 7. Confusion matrix for Augmented Leaf-12 dataset

4.4 Discussion

Table 2 shows the estimated performance metrics for plant species recognition using a Bi-channel CNN. Accuracies of plant species recognition obtained from augmented Flavia, Swedish leaf and Leaf-12 datasets are 97.71%, 98.67% and 97.7%, respectively. The augmented datasets showed improved accuracies for standard and real-time datasets. Even though the accuracies proved to be better for the augmented datasets, it is proportionately increasing the computational time. The computational time increased four folds as the dataset size is increased by 4 times. Table 3 shows the computation time for CNN-1 (train and test time), CNN-2 (train and test time), gamma identification time and B-CNN prediction time.

Precision, Recall and F1-score for Flavia and Leaf-12 datasets obtain a similar value of 0.98. Whereas, for the Swedish leaf dataset, the obtained precision, recall and F1-score values are 0.99, 0.99 and 0.99, respectively. Using the Scikit-learn package, the weighted metrics (Precision, Recall and F1-score) are considered. In the experimental studies, the number of images per class is equal, eradicating the class imbalance problem. In addition to it, the weighted metrics manages the class imbalance problem for randomly distributed dataset between the train and test data groups.

Table 2. Results of Bi-channel CNN for standard and real-time dataset

Metrics	Flavia		Swedish leaf		Leaf-12	
	Without augmentation	With augmentation	Without augmentation	With augmentation	Without augmentation	With augmentation
Number of images	1,600	6,400	1,125	4,500	3,840	15,360
Accuracy (%)	94.38	**97.71**	94.97	**98.67**	95.49	**97.7**
Precision	0.95	**0.98**	0.95	**0.99**	0.95	**0.98**
Recall	0.94	**0.98**	0.95	**0.99**	0.95	**0.98**
F1-Score	0.94	**0.98**	0.95	**0.99**	0.95	**0.98**
Gamma	0.7	**0.5**	0.4	**0.9**	0.5	**0.5**

Table 3. Computation time for Bi-channel CNN

Computation Time (in seconds)	Flavia		Swedish leaf		Leaf-12	
	Without augmentation	With augmentation	Without augmentation	With augmentation	Without augmentation	With augmentation
i) CNN-1 Train time	77.65	289.42	53.77	209.83	181.32	699.34
ii) CNN-1 Test time	0.48	1.66	0.38	1.22	1.07	3.79
iii) CNN-2 Train time	26.44	106.36	19.41	74.52	63.07	250.88
iv) CNN-2 Test time	0.22	0.59	0.19	0.46	0.43	1.28
v) Gamma Identification time	0.04	0.07	0.04	0.06	0.04	0.15
vi) B-CNN Prediction time	0.01	0.02	0.01	0.01	0.01	0.04

From Table 4, it is observed that the accuracies obtained by using the Bi-channel CNN are higher compared to other methods such as conventional [17, 18], single deep learning architectures [17, 19] and double deep learning techniques [5].

Table 4. Comparison of accuracies between Bi-channel CNN framework and reported literatures

Methods	Flavia	Swedish leaf	Leaf-12
Conventional Method [17]	89.17%	88.46%	82.38%
Multiscale Distance Matrix [18]	90.33%	93.60%	–
S-Inception [5]	92.34%	91.67%	–
Inception-V3 [17]	92.50%	94.67%	90.28%
End-to-End CNN [19]	91.08%	96.06%	–
B-CNN	**97.71%**	**98.67%**	**97.70%**

The real-time images are collected and tested with the bi-channel CNN method (Trained model with the best gamma value of 0.5 on Leaf-12 dataset). The prediction results are shown in Fig. 8.

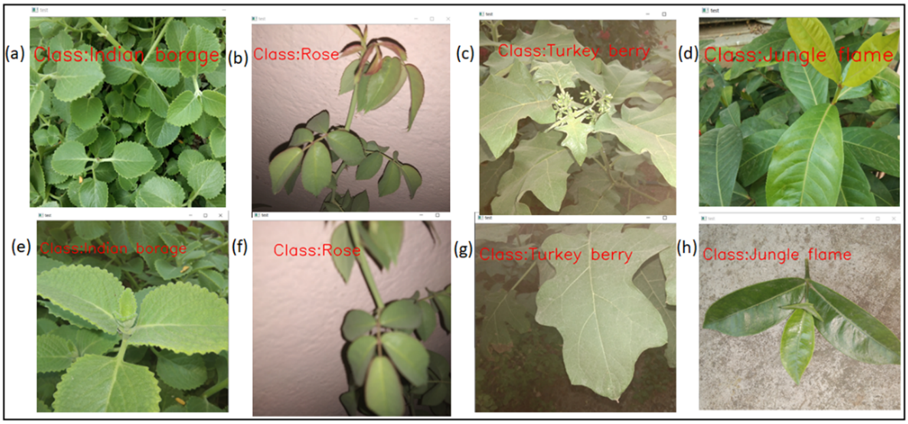

Fig. 8. Prediction of real-time images from Leaf-12 dataset using Bi-channel CNN. (Predicted class is mentioned over the image in Red font.) (a) Groundtruth: Indian Borage (b) Groundtruth: Rose (c) Groundtruth: Turkey Berry (d) Groundtruth: Jungle Flame (e) Groundtruth: Indian Borage (f) Groundtruth: Rose (g) Groundtruth: Turkey Berry (h) Groundtruth: Jungle Flame

5 Conclusion

Plant species recognition is performed using Bi-channel CNN. The proposed architecture consists of two CNN parallel branches (CNN-1 = VGG-16, CNN-2 = SqueezeNet). Both the CNNs are trained individually and prediction scores are obtained. The weighted prediction from each of the branches is fused (late fusion approach) to obtain the final predicted score. The weight γ parameter is optimized to achieve the highest performance metrics. The Bi-channel CNN method is assessed on three datasets namely, Flavia (94.38%), Swedish leaf (94.97%), and Leaf-12 (95.49%). Augmented datasets provided better performance for Bi-channel CNN. The accuracies of Flavia, Swedish leaf, and Leaf-12 datasets are 97.71%, 98.67%, and 97.70%, respectively.

Acknowledgment. The authors would like to thank NVIDIA for providing NVIDIA Titan X GPU under the University Research Grant Programme.

References

1. Wäldchen, J., Rzanny, M., Seeland, M., Mäder, P.: Automated plant species identification—trends and future directions. PLoS Comput. Biol. **14**(4), e1005993 (2018)
2. Figueroa-Mata, G., Mata-Montero, E.: Using a convolutional siamese network for image-based plant species identification with small datasets. Biomimetics **5**(1), 8 (2020)
3. Raj, A.P.S.S., Vajravelu, S.K.: DDLA: dual deep learning architecture for classification of plant species. IET Image Process. **13**(12), 2176–2182 (2019)
4. Rizk, S.: Plant leaf classification using dual path convolutional neural networks. Doctoral dissertation, Notre Dame University-Louaize (2019)
5. Wang, B., Wang, D.: Plant leaves classification: a few-shot learning method based on siamese network. IEEE Access **7**, 151754–151763 (2019)
6. Lee, S.H., Chan, C.S., Remagnino, P.: Multi-organ plant classification based on convolutional and recurrent neural networks. IEEE Trans. Image Process. **27**(9), 4287–4301 (2018)
7. He, G., Xia, Z., Zhang, Q., Zhang, H., Fan, J.: Plant species identification by bi-channel deep convolutional networks. In: Journal of Physics: Conference Series, vol. 1004, no. 1, pp. 1–6 (2018)
8. Hu, J., Chen, Z., Yang, M., Zhang, R., Cui, Y.: A multiscale fusion convolutional neural network for plant leaf recognition. IEEE Signal Process. Lett. **25**(6), 853–857 (2018)
9. Ghazi, M.M., Yanikoglu, B., Aptoula, E.: Plant identification using deep neural networks via optimization of transfer learning parameters. Neurocomputing **235**, 228–235 (2017)
10. Mukherjee, S., Kumar, P., Saini, R., Roy, P.P., Dogra, D.P., Kim, B.G.: Plant disease identification using deep neural networks. J. Multimedia Inf. Syst. **4**(4), 233–238 (2017)
11. Wang, G., Sun, Y., Wang, J.: Automatic image-based plant disease severity estimation using deep learning. Comput. Intell. Neurosci. (2017)
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
13. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv preprint arXiv:1602.07360 (2016)
14. Ruder, S.: An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747 (2016)
15. Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F., Xiang, Q.L.: A leaf recognition algorithm for plant classification using probabilistic neural network. In: IEEE International Symposium on Signal Processing and Information Technology, pp. 11–16. IEEE (2007)
16. Söderkvist, O.J.O.: Computer vision classification of leaves from Swedish trees. Master's thesis, Linkoping University (2001)

17. Anubha Pearline, S., Sathiesh Kumar, V., Harini, S.: A study on plant recognition using conventional image processing and deep learning approaches. J. Intell. Fuzzy Syst. **36**(3), 1997–2004 (2019)
18. Wang, X., Du, W., Guo, F., Hu, S.: Leaf recognition based on elliptical half gabor and maximum gap local line direction pattern. IEEE Access **8**, 39175–39183 (2020)
19. Kaya, A., Keceli, A.S., Catal, C., Yalic, H.Y., Temucin, H., Tekinerdogan, B.: Analysis of transfer learning for deep neural network based plant classification models. Comput. Electron. Agric. **158**, 20–29 (2019)