Tanzania Journal of Forestry and Nature Conservation, Vol 90, No. 3 (2021) Special Issue*: Embracing Science and Technology in Nature Conservation. pp 93-103*

Plant Species Identification from Leaf Images Using Deep Learning Models (CNN-LSTM Architecture)

[*]J. Banzi, T. Abayo

Department of Tourism and Recreation
Sokoine University of Agriculture, Tanzania

*Correspondence: *jamalbanzi@sua.ac.tz*

ABSTRACT

Species knowledge is important for biodiversity conservation. Identification of plants by conventional approach is complex, time consuming, and frustrating tor non-experts due to the use of botanical terms. This is a challenge for learners interested in acquiring species knowledge. Recently, an interest has surfaced in automating the process of species identification. The combined availability and ubiquity of relevant technologies, such as digital cameras and mobile devices, advanced techniques in image processing and pattern recognition makes the idea of automated species identification become real. This paper elucidates development of convolutional neural network models to perform plant species identification using simple leaves images of plants, through deep learning methodologies. Training of the models was performed by using an open database of 100 plant species images, containing 64 different element vectors of plants in a set of 100 distinct classes of plant species. Several state-of the- art model architectures were trained, with the proposed model attaining performance of 95.06% success rate in identifying the corresponding plant species. The significant success rate makes the model very useful identifier or/and advisory tool. The approach could be further expanded to support an integrated plant species identification system to operate in real ecosystem services.

Keywords: Biodiversity - Computer vision - Convolutional neural networks - Plant species - Deep learning.

INTRODUCTION

The diversity of plant species plays a very important role in various areas such as foodstuff, medical science, industrial growth, and environment protection. Many productive activities of all human beings depends on plants as it provides a lot of food and some necessaries. It also helps to maintain the balance of carbon dioxide and oxygen in the atmosphere. It is estimated that more than half of the world's medicines come from natural plant synthesis, and 1/4 of them are extracted directly from plants or plants are the sole raw materials. However, the increase of human production activities, over-reclamation, rapid urban development, over-logging, global warming, lack of awareness of plant species knowledge, and pollution destroyed the ecological environment of living organisms, which have resulted in the extinction of hundreds of plant species while many plant species constantly die every year. Extinction of a large numbers of plant species will cause a serious consequence on human and ecosystem, such as flash floods, flooding, desertification, and regional climate change and so on (Canuti *et al.* 2009). Actually, the extinction of one species is likely to result in the disappearance of another. Evidence indicates that with the disappearance of biodiversity, the functions of natural and artificial ecosystems are also altering (Chang *et al.* 1990). Therefore, it is imperative to protect plant species. The first step of protecting plant species is to identify them and understand what they are and where they come from. Nevertheless, there are more than 270,000 plant species that have been named on the earth (Hervé *et al.* 2013), and many of them are still unknown yet. It is therefore difficult, time-consuming and almost impossible for human being to identify such a large number of existing plant species, especially for non-expert individuals, hence there is a high need for an automatic plant species identification tool.

Automatic plant species identification is very important for phytotaxonomy (Wiens 2016). It leverages image based methods which are considered a promising approach for plant species identification (Zhang *et al.* 2019). With automatic plant species identification tool, a user can take a picture of a plant in the field with the build-in camera of a mobile device and analyse it with an installed recognition application to identify the species or at least to receive a list of possible species if a single match is impossible. By using a computer-aided plant identification system, even a non-professional can take part in this process. Therefore, it is not surprising that large number of research bodies are devoted to automate the plant species identification process. For instance, ImageCLEF, one of the foremost visual image retrieval company, is hosting a plant identification challenge since 2011 (Zhang *et al.* 2019. It is predicted that the interest will further grow in the foreseeable future due to the constant availability of portable devices incorporating myriad precise sensors. These devices provide the basis for more sophisticated ways of guiding and assisting people in species identification. Furthermore, emerging trends and technologies such as augmented reality, data glasses, and 3D-scans give this research topic a long-term perspective.

This research places an attempt towards application of deep learning architectures for identification of plant species. Application of deep learning models represented by convolutional neural network (CNN) have presented great success in many image-based fields of computer vision, for example, traffic detection, medical image recognition, Scenario text detection, expression recognition, and face recognition (Purohit *et al.* 2016).

Leaf geometric morphology

For a leaf to be identified, the system should first understand the geometric structure of a leaf. A leaf is described as a usually green, flattened, lateral structure attached to a stem and functioning as a principal organ of photosynthesis and transpiration in common plants (Cerutti *et al.* 2014). Leaf shape is considered more heritable and often favoured over leaf geometry (Wu *et al.* 2007). Different species have different leaves in case of leaf tip, apex, vein, teeth blade, and so on as shown in Figure.1. All leaves can be coarsely divided into two categories: simple with a single leaf blade Figure.1 (left) and compound with several leaflets, as shown in Figure.1 (right).

.

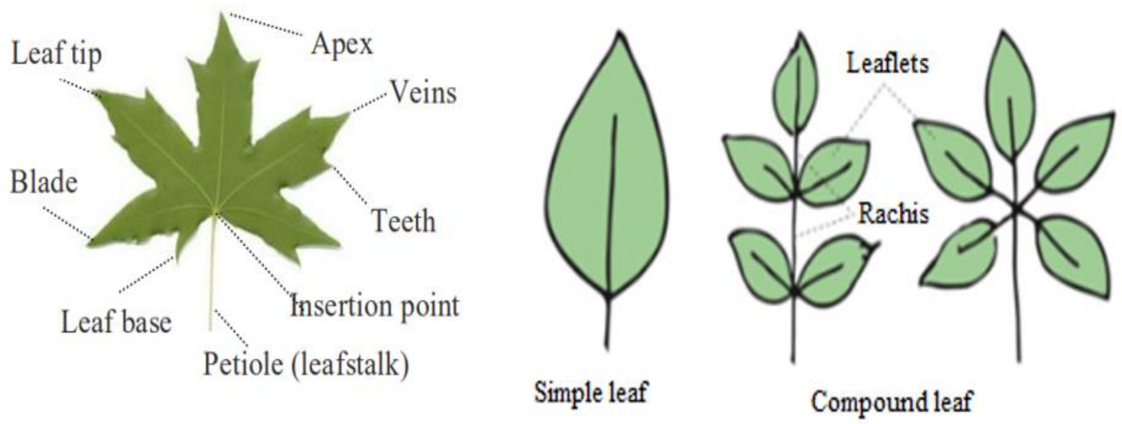

Figure 1: Leaf geometric structure.

Compound leaves are described as pinnately, palmately, and doubly compound (Armstrong *et al.* 2015). Most studies use simple leaves for species classification, while several studies considered compound leaves in their experiments, because compound leaves can be regarded as a collection of simple leaf-like structures called leaflets. A lot of characteristics of leaf image are very useful for botanists to classify plant species. A leaf contains a lot of features that can be extracted for automated classification and identification including tooth spacing, number per centimetre, and qualitative descriptions of their convex and concave flanks, which are often employed by botanists for manual classification of species and could also be utilized for an automated identification system.

Image recognition

Compared with convectional image recognition systems, the image recognition technology based on deep learning does not need to extract specific features. With deep learning, plant species features are only extracted through iterative learning which is capable of capturing appropriate image features, acquire global and contextual features, and has strong robustness and higher recognition accuracy (Konstantinos *et al.* 2018). Features are extracted from the structure of the leaf image and then pass-through pre-processing stages like conversion to grayscale, cropping, and edge detection see Figure 2(c) before being processed into higher stages from recognition, classification, and ultimately identification.

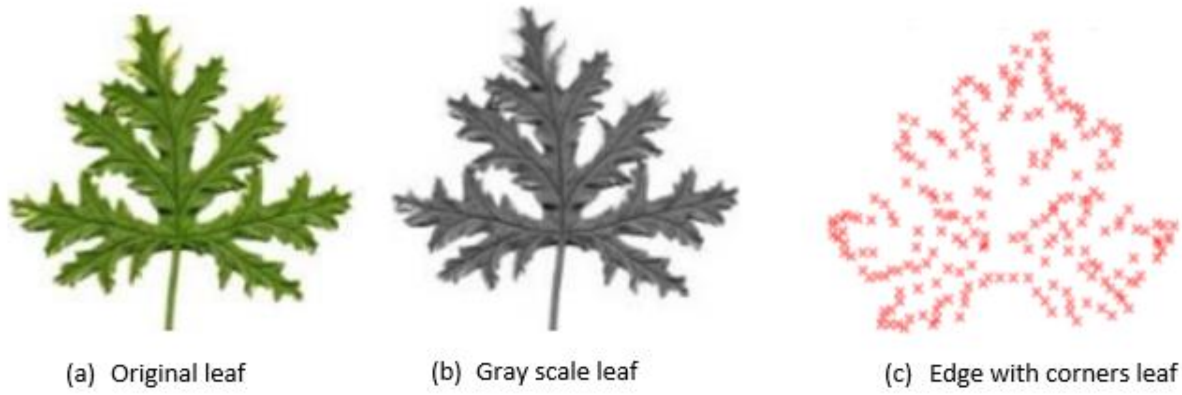

Figure 2: Leaf image features.

Convolutional neural network models

Artificial neural networks (ANNs) are mathematical models that mimic the general principles of brain function with their neurons and synapses that interconnect them (Gu *et al.* 2018). The main characteristic of ANNs is the ability to be trained through the process of supervised learning. During that process, neural networks are "trained" to model the system with the use of existing data that contain specific matchings of inputs and outputs of the system to be modelled. CNNs are an evolution of traditional artificial neural networks, focused mainly on applications with repeating patterns in different areas of the modeling space, especially image recognition (LeCun *et al.* 1998). In CNNs the layers convolute by drastically reducing the required number of structural elements i.e., number of artificial neurons which is far better compared to traditional feedforward neural networks. Figure 3 shows the structure of CNN that can be used in many applications. For image recognition applications, several baseline architectures of CNNs have been developed, which have been successfully applied to complicated tasks of visual imagery. The common CNN architectures that are largely used in image data recognition includes: (i) AlexNet (Krizhevsky *et al.* 2012), (ii) PredNet (Banzi *et al.* 2019), (iii) GoogLeNet (Szegedy *et al.* 2015), (iv) Overfeat (Sermanet *et al.* 2014), and (v) VGG (Simonyan *et al.* 2014). These models and their training and testing processes, are freely available for transfer learning purpose.

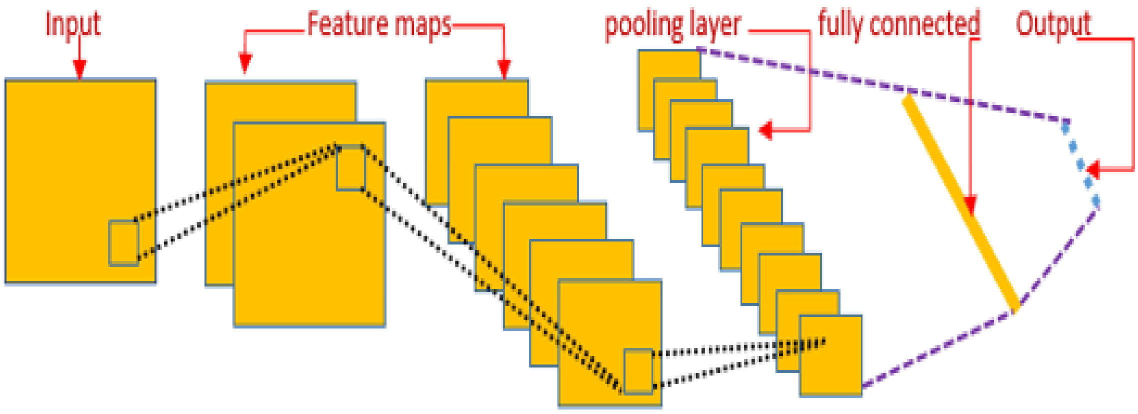

Figure 3: The Structure Convolutional Neural Network.

MATERIALS AND METHODS

Dataset and pre-processing

There are three features used to describe each leaf; the shape, texture and margin. For Each feature, a 64 element vector is given per sample of leaf. These vectors are taken as a contiguous descriptor (for shape) or histograms (for texture and margin) shown in Figure 4. One file for each 64-element feature vectors describing a single leaf image. Each row begins with the class label and the remaining 64 elements is the feature vector for a single instance. The images were taken with mobile devices (mostly iPhones) in outdoor environments. These images appear in controlled backlit and front-lit versions, with several samples per species. They vary considerably in sharpness, noise, illumination patterns, and shadows. Since the dataset is not large enough for deep learning to overcome the problem of overfitting, data augmentation was used to enlarge the dataset for network training and overcoming overfitting. The dataset used in this work can be downloaded at UCI Machine Learning Repository: One-hundred plant species leaves data set Data Set database. Figure 5 shows sample of image data of 24 species in grey scale from UCI Machine learning repository.

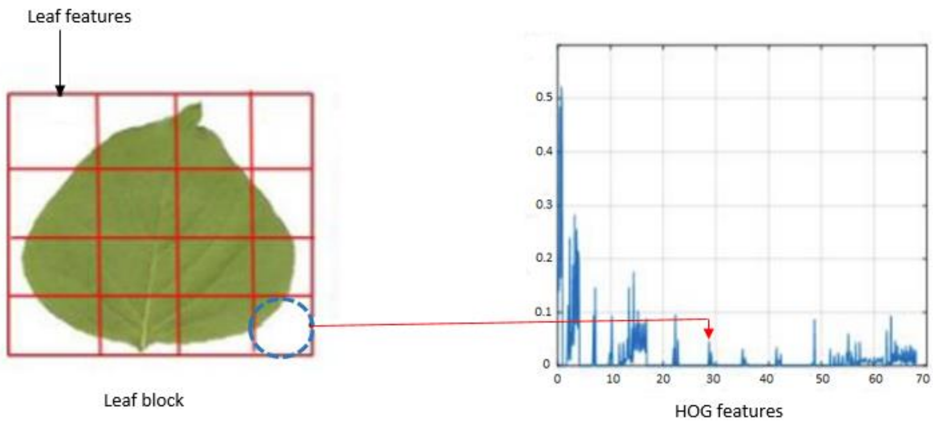

Figure 4: Feature extraction from a plant leaf.

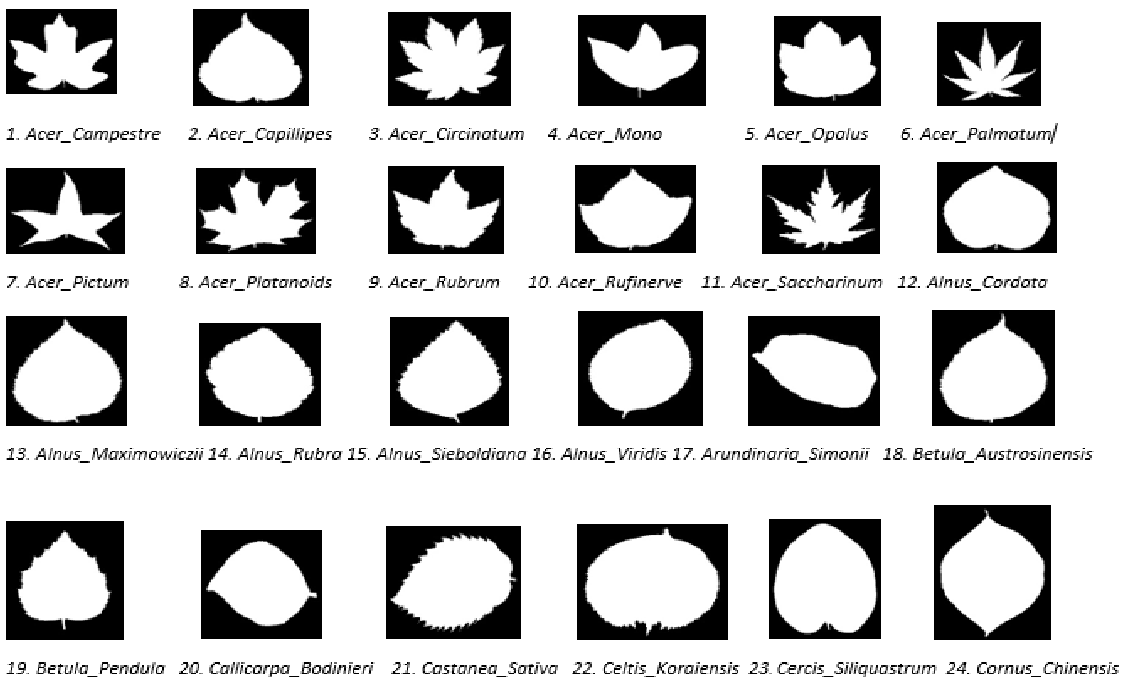

Figure 5: A sample plant species image from UCI repository.

Identification of leaf pattern

Leaf pattern is known as an important clue for human beings to identify plant species. It can indicate several systematically informative features and is extremely useful for species classification. Most leaf based plant species recognition methods make use of leaf pattern characteristic to recognize plant species (Singh *et al*. 2015). The reason is that different plant species have obvious leaf shape differences for distinguishing species, particularly for non-botanists who have limited knowledge of plant characters. The leaf texture may be unclear and its color may change with the seasons and illumination, while its shape always remains unchanged in almost the whole growth period. Shape is regarded as the most discriminative feature. It can provide ironic information for describing plant species. In almost all automatic leaf based plant species recognition, leaf shape is the most common feature used for plant species identification (LeCun 1995).

Leaf shape has five fundamental geometry features, such as the longest distance between any two points on a leaf border (L), length of main vein (lengthwise-Lv), widest distance of a leaf (Crosswise-W), leaf area (A) and the leaf perimeter (P). Then, several features can be derived from these 5 basic features by some mathematical operations, such as, smoothness of a leaf image, form factor, aspect ratio (Lv/W), narrow factor (L/Lv), difference between a leaf and a circle (4πA/P2), rectangularity (LvW/A), ratio of perimeter to longest distance (P/L), ratio of perimeter to the sum of main vein length and widest distance (P/(Lv+W)), and 5 structural features obtained by applying morphological opening on grayscale image. These features are extracted and used in the subsequent process of plant species recognition.

CNN-LSTM Architecture – The combined model

The learning pattern of a convolutional neural network(CNN) is generally based on aggregation of feature maps derived at multiple levels. As a consequence of this aggregation occurring in the deep layers of the CNN, it tends to lose the significance of the fine granular details learnt by the initial layers. The traditional CNN based methods for plant species recognition system focusses on learning the features of the leaf pattern in an orderly fashion starting from basic image level features like edges and move towards complex texture based differences (Zhang *et al.* 2019). Consequently, some significant details are not passed to the deeper layers of the network.

In this work, a combined architecture is proposed where by the CNN layer is embedded with the long short-term (LSTM) memory layer to pass those significant features extracted in the initial layers to the deeper layers of the network. This supports effective aggregation of feature maps for precise classification and ultimately accurate recognition.

The proposed CNN-LSTM architecture is presented in Figure 6. Structurally, CNN-LSTM architecture consists of a sequence of CNN blocks to extract complex internal features from leaf images connected to LSTM which is used as a classifier. The number of channels increases from 32 to 128 along the depth of the network. The first block has a convolution receptive area of size 7x7, trailed by a 3x3 kernel for the LSTM block and finally a 3x3 filter for the third fully connection block. Then, it applies the average pooling over the feature map. This enables the classifier to model a reduced set of features without much loss of context and also avoids the risk of overfitting. The entire model is followed by a sequence of 1x1 convolutional layers after the last Fully connection block.

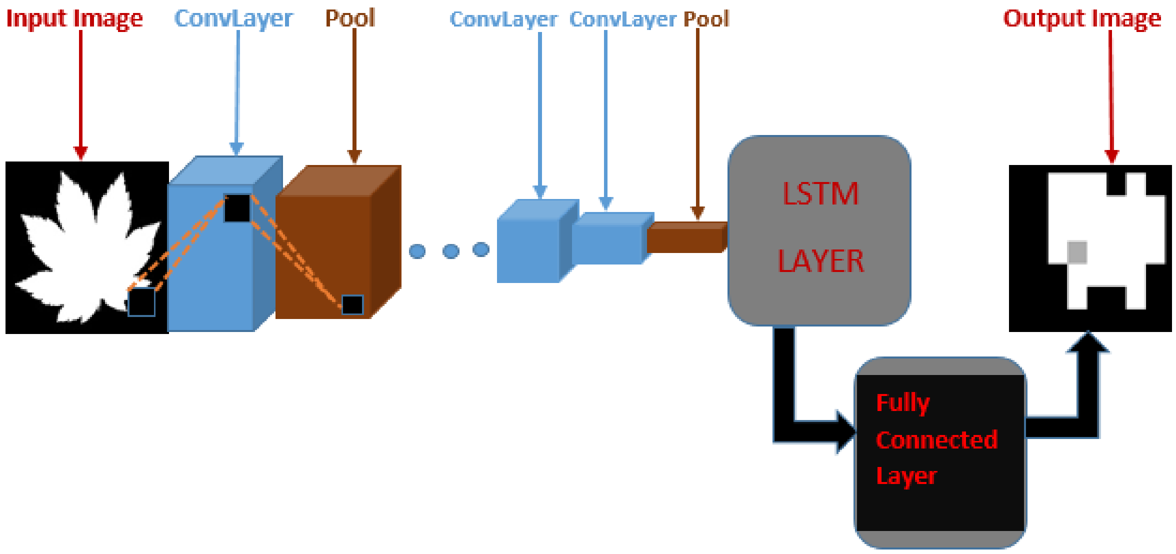

Figure 6: The proposed fused CNN network for plant species identification.

Generally, the proposed network system has 20 layers: 12 convolutional layers, five pooling layers, one Fully Connected layer, one LSTM layer, and one output layer with the softmax function for classification. The convolutional layer with a size of 3 × 3 kernels is used for extracting feature and it is activated by the ReLU function. The max-pooling layer with a size of 2 × 2 kernels is applied to reduce the dimensions of an input image. In the last part of the architecture, the function map is shifted to the LSTM layer to extract time information required for classification purpose. Using the reshape method, the input size of the LSTM layer becomes (196,512). After analyzing the time features, the model sorts the plant species images through a fully connected layer to predict whether they belong under any of the 100 categories of the plant species.

Experimental set up

We perform a complete experiment to demonstrate the performance of the proposed approach. The CNN-LSTM Model was implemented with the python library, using Theano and Keras. The model parameters are optimized using gradient descent Adam algorithm with all parameters set to default values. The system specification details are presented in table 1, while the detail of the tensor in each layer of the proposed architecture is shown in the summary presented in Table 2.

Table 1: System requirement for plant specifies identification.

System requirement	Specifications
Central processing unit (CPU)	Intel(R)Core i7-4790@ 3.6GHz
Operating system (OS)	Microsoft window
Random access memory (RAM)	16GB
System architecture	64bit

Table 2: Summary of the tensor in each layer of the proposed species identification system.

Layer	*Type*	*Kernel Size*	*Stride*	*Kernel*	*Input Size*
1	Convolution2D	3 × 3	1	32	224 × 224 × 3
2	Pool 2	2 × 2	2	-	224 × 224 × 3
3	Convolution2D	3 × 3	1	64	112 × 112 × 32
4	Convolution2D	3 × 3	1	64	112 × 112 × 64
5	Convolution2D	3 × 3	1	64	112 × 112 × 64
6	Pool 2	2 × 2	2	-	112 × 112 × 64
7	Convolution2D	3 × 3	1	128	56 × 56 × 64
8	Convolution2D	3 × 3	1	128	56 × 56 × 128
9	Pool 2	2 × 2	2	-	56 × 56 × 128
10	Convolution2D	3 × 3	1	256	28 × 28 × 128
11	Convolution2D	3 × 3	1	256	28 × 28 × 256
12	Convolution2D	3 × 3	1	256	28 × 28 × 256
13	Pool 2	2 × 2	2	-	28 × 28 × 256
14	Convolution2D	3 × 3	1	512	14 × 14 × 256
15	Convolution2D	3 × 3	1	512	14 × 14 × 512
16	Convolution2D	3 × 3	1	512	14 × 14 × 512
17	Pool 2	2 × 2	2	-	14 × 14 × 512
18	LSTM	-	-	-	196 × 512
19	FC	-	-	32	100,352
20	OUTPUT	-	-	100	32

Data preprocessing

Before the training stage, we first need to pre-process the raw data from the dataset. The input of the pose estimator is the cropped image, but the original ground truth of the image is the absolute position in the entire raw image. Therefore, there is a need to first transform the ground truth into a relative position with respect to the center of the hand. Finally, the images are sized to 96x96 and in grey-scale with values normalized between 0 and 1.

Training

We train the proposed deep CNN-LSTM regressively with the learning time gradually decreased. The model was based on the pre-trained model [38] and was trained to classify plant species and therefore identify the new incoming unknown species. The loss was taken as the sum of the firing rates of the error neurons in the zeroth pixel layer. A random hyperparameter search was performed over fourth- and fifth-layer models of the leaf features positions. Our model consists of 20 layers with 3 by 3 filter sizes of all convolutions and stack size per layer of (1,32,64,64,128,256). The initial training rate is set to 0.001 dropped by learning ratio of 10 after every 60 epochs.

Softmax

The proposed system uses k-way softmax classifier to classify image to one among k plant species. The loss due to this architecture is given by

$$CE = -\sum_i^k t_i log(f(s)_i)$$

Where *f*(s) is the output condition probability $P(y = \hat{y}_i | s_i)$ for some training example *Si*

This probability function for softmax activation is given in equation 2

$$f(s)_i = \frac{e^{s_i}}{\sum_j^k e^{s_i}}$$

RESULTS

The training process took around 10 hours on the CPU. The accuracy obtained reaches 95% on 140 epochs. We first measure the performance of the proposed model fig 6. and then compared it with the state-of-the-art approaches to justify its performance.

Experiment show that the proposed CNN-LSTM performs better in classifying plant species than the convectional CNN as it attains the accuracy of 95.06% while the literatures (Zhang *et al.* 2019) have reported only 91% accuracy.

Performance Evaluation

We evaluate the performance of our approach on three publicly available datasets for plant species identification: The PlantCLEF dataset (Villegas *et al.* 2016), the ICL leaf dataset (Bonnet *et al.* 2016), and the LeafSnap dataset (Kumar *et al.* 2012) and the resultant observations are presented in Figure 7. It is observed that the proposed CNN-LSTM outperformed the all the competent state-of-the-art approaches applied in three datasets. This performance is attributed by the use of LSTM layer which captures significant features lost by traditional CNN during training.

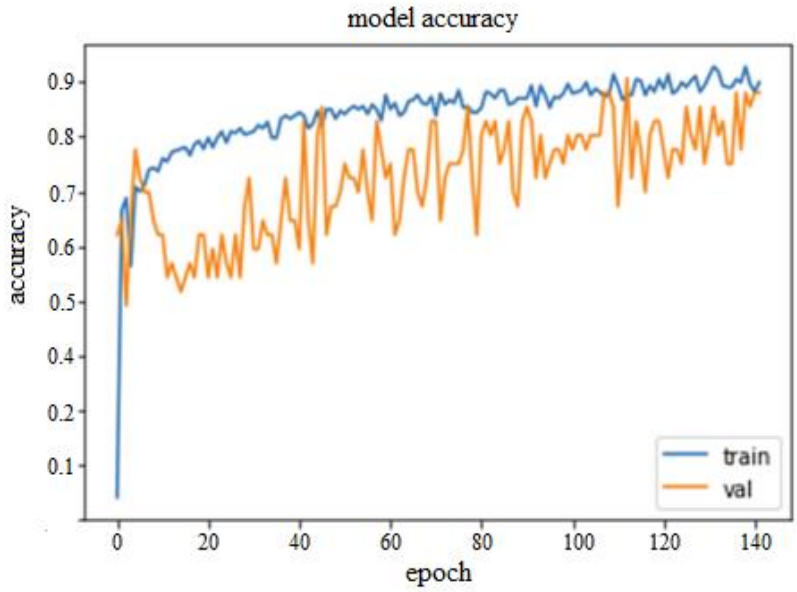

Figure 7: The performance of the CNN-LSTM architecture in classifying plant species.

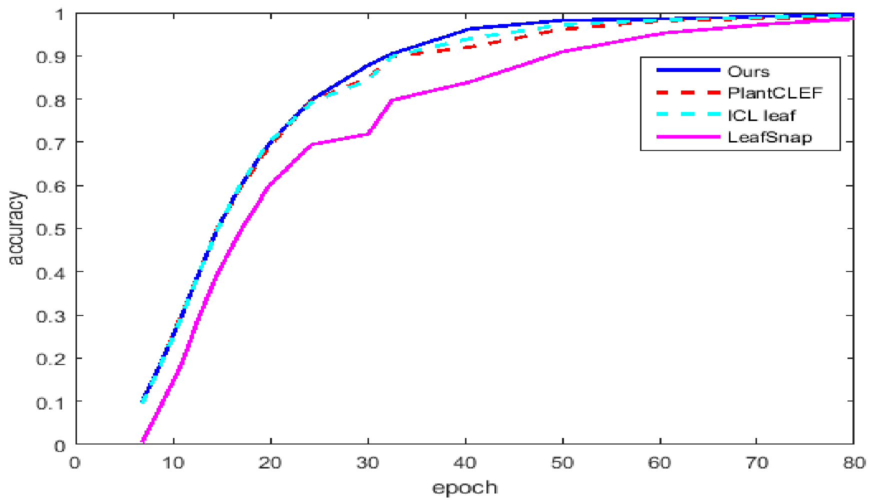

Figure 8: Comparison of CNN-LSTM with the state-of-the-art approaches.

CONCLUSION

This research presents an efficient way to classify and identify plant species from the leaf images using an automated tool. To the best of our knowledge, this is the first attempt to employ the LSTM layer in CNN to classify plant species. The main contribution of this work is the integration of LSTM on top of the CNN network for effective feature learning. It helps to selectively weigh the features at different layers at the inception of a single layer. Hence, the receptive field at a layer is extended to look at feature maps from different levels of the processing hierarchy. The current layer can now process its input with more contextual information and hence capture all the significant features. Learning at the layers preceding the current layer is now aided by the perception of the features at the current layer. This is due to back propagation of the tensors along the skip connections. The proposed network learnt around 100K parameters to detect the type of leaf image and if it belongs to a certain class, which is comparatively less than the existing deep learning approaches reported in the literature. Because of this least number of parameters, we performed data augmentation to allow for in deep features to be extracted and teach the model to accurately identify the classes. We also performed benchmark experiment with the three state-of-the-art approaches of plant species recognition. The empirical results indicate that the proposed CNN-LSTM was able to classify the particular class of each image with an accuracy of 95% far better than all the three combatants. This indicates that the proposed system could be expanded to support an integrated plant species identification system to operate in real ecosystem services.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

REFERENCES

Babatunde, O., Armstrong L. & Diepeveen, D. 2015. A survey of computer-based vision systems for automatic identification of plant species. Journal of Agricultural Informatics, 6(1).

Banzi, J., Bulugu, I., & Ye, Z. 2020. Learning a deep predictive coding network for a semi-supervised 3D-hand pose estimation. IEEE/CAA Journal of Automatica Sinica, 7(5), 1371-1379.

Bonnet, P., Joly, A. & Hervé Goëau. 2016. Plant identification: Man vs. Machine. LifeCLEF 2014 plant identification challenge. Multimedia Tools & Applications, 75(3):1647-1665.

Canuti P. 2009. Landslides - disaster risk reduction. Springer Ebooks. 72(72):31-61.

Cerutti G., Tougne L., Coquin D. & Vacavant A. 2014. Leaf margins as sequences: a structural approach to leaf identification. Pattern Recognit Lett 49:177–184. doi: 10.1016/j.patrec.

Chang SC. 1990. Landslides and their environmental impacts in northern Taiwan (1968-1986). Journal of Geographical Science pg11-27.

Ferentinos, K.P. 2018. Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*, 311-318.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B. & Chen, T. 2018. Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354-377.

Hervé Goëau, Joly A., & Bonnet, P. 2013. The imageCLEF plant identification task. ACM International Workshop on Multimedia Analysis for Ecological Data.

Krizhevsky, A., Sutskever, I. & Hinton, G.E. 2012. ImageNet classification with deep convolutional neural networks. In: Advances in neural information Processing systems, pp. 1097-1105.

Kumar, N., Belhumeur, P.N. & Arijit, B. 2012. Leafsnap: A Computer Vision System for Automatic Plant Species Identification. European Conference on Computer Vision, Springer Berlin Heidelberg, 2012. DOI:10. 1007/978-3-642-33709-3_36

LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. 1998. Gradient-based learning applied to document recognition. Proc. IEEE 6 (11), 227-2324.

Loffe, S. & Szegedy, C, 2015. Batch Nomalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint: 1502.03167.

Purohit S., Viroja R. & Gandhi S. 2016. Automatic plant species recognition technique using machine learning approaches. IEEE International Conference on Computing & Network Communications.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R. & LeCun, Y., 2014. Overfeat: Intergrated recognition, Localization and Detection using convolution networks. In: International conference on learning Representations.

Simonyan, K. & Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Singh, V., Varsha & Misra, A.K. 2015. Detection of unhealthy region of plant leaves using image processing and genetic algorithm. International Conference on Advances in Computer Engineering and Applications, pp. 1028-1032, 2015.

Villegas, G., Liao, W., Criollo, R., Phillip, W. & Ochoa, D. 2017. Detection of leaf structures in close-range hyperspectral images using morphological fusion, Geo-spacial Information Science, 20:4, 325-332.

Wiens, J.J. 2016. Climate-Related Local Extinctions Are Already Widespread among Plant and Animal Species. PLoS Biology, 14(12):e2001104.

Wu, S. Bao, F. & Xu, E. 2007. A leaf recognition algorithm for plant classification using probabilistic neural network. IEEE International Symposium on Signal Processing and Information Technology.

Zhang, X., Zhou, F., Lin, Y. & S. Zhang. 2016. Embedding lable structures for fine-grained features representation. In 2016 IEEE Conference on Computer Vision and Pattern Recognition