¹Mufeeda V K

²Dr. R. Suganya

Enhanced Recurrent Neural Network With C-Efo-Based Feature Selection for Plant Leaf Classification



Abstract: - Plants are an important factor in human life and other living things around the world. Plants are recognized as important influencers of changes in natural cycles. It is an important producer that sustains human life, as it is known to be the only organism capable of converting light energy obtained from the sun into food energy for humans and other organisms. Animals cannot produce food because they depend directly and indirectly on plants for food energy. Automated plant recognition seeks more attention in computer vision and machine learning. A lot of research has been done to solve the problems related to plant classification. The knowledge and ability to distinguish various medicinal plants was locked in by early people before the development of computer systems and digital cameras. The new plant leaf classification was developed in the first stage using improved segmentation techniques and optimal feature selection. Experimental images were collected from the Swedish leaf dataset and subjected to a preprocessing step. The preprocessed image is obtained through grayscale conversion, median filtering, and histogram equalization. Therefore, an optimized UNet model is used to obtain key regions of leaves to improve accuracy. Features of shape, texture and color were obtained. Since they contain the longest length of the resulting features, the best features are chosen to reduce training time and dimensionality reduction. These optimal characteristics are achieved through a modified hybrid algorithm called C-EFO (Crow Search Electric Fish Optimization), where the traditional EFO (Electric Fish Optimization) is combined with the CSO (Crow Search Optimization) algorithm. Once the best features were obtained, the newly developed E-RNN deep learning model was used for classification, where the hyperparameters were best fitted using the C-EFO algorithm. Finally, the experimental results are validated and the proposed model achieves better performance metrics. Experiments show that the proposed C-EFO method outperforms traditional methods in terms of accuracy.

Keywords: Feature Extraction, Plant Species Identification, Segmentation, SVM, Neural Network, CFO

1. INTRODUCTION

Sorting of plant leaves is the most important process in agriculture. It helps botanists detect and identify species unknown to many researchers. In agriculture-based research, much existing work focuses on improving the use of plant datasets and implementing new feature extraction methods. The traditional method uses the characteristic traits of plant leaves for disease identification, which leads to the complexity of the training stage. The plant recognition model involves deep learning because it can handle object detection environments, even with large amounts of training information. Deep learners have improved performance in various fields such as speech analysis, object detection, image detection and recognition. However, because "complex background environments" must be handled, plant classification approaches in conventional models are different from object detection techniques. The occurrence of phase overlap and interference will reduce the identification accuracy. Therefore, addressing these challenges requires the development of deep learning-based models to identify plant diseases. In this chapter, a new classification model for plant leaves is proposed by improving the segmentation and optimal feature selection process. The first stage of the proposed model is a preprocessing stage, in which specific techniques such as "conversion from RGB to grayscale, histogram equalization, and median filtering" are intervened.

The proposed model is also equipped with an optimized U-Net model. After the segmentation process, the features of the image are obtained using various techniques such as shape, color, and texture. However, the duration of these features is not ideal for network training. Therefore, the selection of the best features is included in the proposed model to minimize the dimensionality of the data and build a highly robust classification model. In optimal feature selection, a newly developed hybrid metaheuristic algorithm (C-EFO) is integrated by mixing EFO and CSA. In the final stage, with the help of C-EFO, the plant diseases are classified

Copyright©JES2024on-line: journal.esrgroups.org

¹ Ph.D Research Scholar, Department of Computer Science, Maruthu Pandiyar College, affiliated to Bharathidasan University, Vallam, Thanjavur, Tamilnadu, **mufeedavk85@gmail.com**

²Research Advisor, Department of Computer Science, Maruthu Pandiyar College, affiliated to Bharathidasan University, Vallam, Thanjavur, Tamilnadu, mailtosuha@gmail.com

using E-RNN deep learning technology. Simulation results show that the developed model based on two plant leaf databases is more efficient and feasible in performance compared to the existing models.

2. LITERATURE REVIEW

Sue Han Lee et. al., [2022] states that although species in the same genus and family may have comparable feature traits, it is still beneficial to make a distinction between the genus and family group in order to lessen the possibility of misclassifying species. Additionally, research can be done to enhance the overall HFTL forecasts through the utilisation of data on plant attributes.

Hellmann et. al., [2022] identifies potential distributional shifts in North America of allelopathic invasive plant species under climate change models. The trends reported here appear to be influenced by both time s, with rates of carbon emissions being correlating with changes in temperature and precipitation patterns. Distribution contractions and dispersions generally correspond to temperature and/or hydrological constraints resulting from warming and changing precipitation patterns.

Xiaodong Tang et. al., [2022] developed a method for extracting leaf veins from photos with complex backgrounds in order to retrieve a sample leaf. The three channels of the HSI colour space were divided and produced distinct gradient images using the marker-controlled watershed segmentation approach. The segmented pictures used to extract the sample leaf were estimated using the solidity (integrity) measure, which also served to validate the final leaf extraction outcomes.

Egerton-Warburton et. al., [2022] pointed out that the AMF community composition was constructed by nitrogen addition, as changes in the cation-anion balance could have a direct effect on soil pH and lead to changes in the AMF community composition.

Tian et. al., [2022] The segmented production of ROS and the limited passive diffusion capacity outside the production sites not only limit the toxicity of ROS by effective removal of ROS by appropriate antioxidant combinations and concentrations, but also allow the controlled transport of ROS from organelle to organelle, either through, for example, aquaporins or direct association, e.g., chloroplasts and nuclei via the stroma

Khalil et. al., [2022] the article summarizes the research on speech-based emotion recognition using deep learning technology and expands the deep learning technology for speech-based emotion recognition. The author simulates multimodal emotion recognition and the experimental results.

Ren Ye et. al., [2022] the leaf area of the seedlings was extracted and located using the single connected domain analysis algorithm, and then the leaf area was used to determine whether the seedlings were suitable for transplanting. The recognition accuracy was achieved. Sometimes, due to seed quality and mechanical damage, empty holes appeared on the seedling tray.

3. PROPOSED METHODOLOGY

3.1 Developed architectural representation of plant leaf classification

Mainly plant leaf classification models seek to focus more on automatically learning characteristics of plant species. Such research ensures excellence in environmental and climate change research. Recently, "automatic plant classification models" have been studied based on new feature learning methods, but they have certain defects. The "model" is researched based on new feature learning methods, but has certain shortcomings. It is worth noting that the image background given in plant leaf classification can pose serious challenges for plant identification. Due to the complexity of plant leaf images, training takes more time. To minimize the impact of data dimensionality, traditional approaches focus on improving improved extraction techniques. Features extracted from leaf images lead to higher processing complexity and require better classifiers for feature training.

To increase system efficiency and reduce complexity, optimization techniques are also incorporated. The use of heuristic algorithms aims to increase the accuracy of the classification process. It also helps minimize the impact of data dimensionality. A new hybrid optimization algorithm was developed to improve the performance of the

plant leaf classification process. Figure 3.1 shows the proposed model based on Opti-U-Net framework and improved RNN.



Fig 3.1: Proposed deep learning model to classify plant leaves

3.2 The Pre-Processing using Median Filtering

Suppose the number of leaf images in the database are Di and Li. An image in a database is represented as $Di = \{Imj\}\ 1 \le j \le Li$, where the term *I*m denotes the *j*th image database. Preprocessing is first performed on the leaf recognition model, where the noise present in the *I*mj image is removed using a median filtering technique. One of the non-linear filters is the median filter, which is often used to remove unwanted noise in images. This is done by taking noisy pixels and replacing them with neighboring values. Equation (3.1). Show Median Filter Equation

$$Im^{q}(x, y) = Me\{Im_{j}(x - x_{1}, y - y_{1}); x_{1}, y_{1} \in Fm\}$$
(3.1)

Here, the term Fm refers to the 2D median filter mask with elements (x1, y1) for ignoring noise, and Imq(x, y) refers to the preprocessed image. Finally, the preprocessed leaf image is obtained and denoted as Im^q Dataset Description

Classification models for plant leaves are based on input leaf images collected from two different sources: the Swedish Leaf Dataset [1] and the D Leaf Database [2].

The Swedish dataset: The Swedish Leaf Database embeds a large number of leaf images numbering 75 and an estimated 15 various plant species. Images taken from the database are part of a "manual alignment" process that provides better spatial information. Figure 3.2 shows a few sample images from this database.

		*		
Acer	Fagus Silvatica	Quercus	Salix sinerea	Tilia

Alnus incana	Populus tremula	Salix alba	Sorbus aucuparia	carpinifolia
Betula	Pubescens	Populus	Salix aurita	Sorbus

Fig. 3.2: Standard images from the Swedish database

D-leaf image dataset: The leaf image dataset D consists of image samples collected from various leaves of tropical plants. Contains a collection of 43 plants, each with approximately 30 images. The standard image of the D-Leaf dataset is shown in Figure 3.3.

Dipterocarpus grandiflorus	Cynometra malaccensis	Cinnomomum iners	Cassia fistula

Bauhinia blakaena	Barringtonia racemosa	Acacia auriculiformis	Alstonia

Fig. 3.3: Example images from the D-leaf image dataset



3.3 Optimal Feature Selection

The proposed model uses C-EFO developed for feature selection with maximizing accuracy as its main goal. The segmented image Im^{segmt} is used to extract features Ftg of size $1 \times nf$. Using primitive functions during the training phase leads to high dimensional complexity of the functions.

The goal of this process is to ensure that the training of the classifier is done accurately. This process involves selecting the necessary features to achieve the best results with the help of the suggested C-EFO shown in Figure 3.6.

The figure below depicts the process of selecting the best features F tkopt of size 1 * nf using C-EFO developed to improve classification accuracy. The results show that the size of the selected features is much smaller compared to the extracted features, which improves the training strategy.



Fig. 3.5: Developed C-EFO-aided optimal feature selection

4. Results and Discussions

The proposed model is compared with several existing algorithms and classifiers to evaluate the performance of the model. and existing classifiers. Optimization algorithms such as "Particle Swarm Optimization (PSO) [3], Gray Wolf Optimizer (GWO) [4], EFO, and CSA" and various classifiers such as "k-NN [5], VGG16 [6], LSTM [7]], and RNN [8]" were included in the comparative analysis.

4.1 Validation Measures

Validation measures are used to measure the performance of supervised computer models on various datasets. They are important in the selection process because they help reveal test results. The confusion matrix, also known as the error matrix, shows the number of mistakes the supervised model made. The rows in this table represent the actual number of cases in each category, while the columns represent the predicted number of categories. The term true negative (n) refers to situations where a model makes a negative prediction when trying to predict a given outcome. On the other hand, true positives (p) are considered positive predictors. A false negative (nf) is a negative prediction of a positive case, and a false positive (fp) is a negative prediction of a positive case.

This work focuses on classification performance using ten validation measures, as described below.

Sensitivity
$$Sntv = \frac{n}{n+fp}$$
......(3.2)
 FDR $Fd = \frac{fp}{n+fp}$(3.3)
Specificity FDR $Fd = \frac{fp}{n+fp}$(3.4)
 MCC $Mc = \frac{p*n-f*nf}{\sqrt{(p+f)(p+nf)(n+f)(n+nf)}}$(3.5)
 $Snv = \frac{p}{p+nf}$(3.6)

F1 - Score
$$F1 s = \frac{Sny * Pcn}{Sny + Pcn}$$
.....(3.9)

4.2 Validation Results

Segmentation results obtained from the developed O-U-Net, with the help of implemented C-EFO and other related algorithms such as PSO, GWO, CSA, and EFO, are shown in Figure 4.1 (Swedish leaf dataset) and Figure 4.2 (D-leaf dataset).



EFO	
Proposed C- EFO	

Fig. 4.1: U-NET-based Segmentation outcomes using the Swedish leaf database

Original images		•		
Ground Truth			•	
Processed image)	•	٠	٠
PSO			•	
GWO)			
CSA)		•	•
EFO)	•		
CEFO)	•		•

Fig. 4.2: U-NET-based Segmentation outcomes using the D-leaf database.

5. Performance Evaluation

The classifier introduced in the proposed framework achieves high performance in all evaluation metrics. The evaluation is performed by analyzing plots of two different datasets in Figure 5.1 and Figure 5.2 to demonstrate accurate classification performance.

The proposed model is tested against various machine learning algorithms such as K-nn, LsTM, RNN and Vgg-16 in D-leaf and Swedish databases. Figure 5.3 shows the analysis results of the proposed model for various machine learning algorithms on the Swedish tree leaves dataset. The proposed model performed well in 85% of the learning evaluations. It improves accuracy by 4% over Vgg-16, 4.5% over LsTM, 3.6% over RNN, and 3% over K-nn.

Evaluation of the proposed model by the F1-score method shows high performance, which is 1.64%, 2.3% and 2.4% better than k-NN, VGG16 and LsTM, respectively. The proposed model performs better in the evaluation of negative measures.





Fig. 5.1: Assessment of proposed model with diverse ML Algorithms on Swedishdataset.

Figure 5.2 shows the classification evaluation on the D-leaf dataset using different learning percentages. The results are more accurate and precise compared to the proposed model.



(a) Accuracy

(b)F1-Score



















(i)Sensitivity

(j)Specificity

Fig. 5.2: Proposed Framework Evaluation with the D-leaf database with differentML. algorithms

When the generated model was put to the test, it outperformed current algorithms and classifiers based on two distinct leaf image databases, the D-leaf and the Swedish leaf database, in terms of classification performance for plant leaves. [15] Tables 5.1 and 5.2 present the model's analysis as it was applied to the Swedish leaf database.

Table 5.1 provides model evaluation metrics for various heuristics. According to the proposed model, it is 1% more accurate than PSO, 2% more accurate than EFO, 0.45% more accurate than WOA, and 0.7% more accurate than CSA Table 5.2 shows the presentation of different M.L. Methods in the Swedish leaf database. The proposed method shows significant improvement over various M.L. algorithm Instead, the proposed classification model was tested under the D-sheet database of Table 5.3 and Table 5.4.

Metrics/Tec	PSO	WOA	EFO	CSA	Proposed
hniques	(%)	(%)	(%)	(%)	method (%)
"Accuracy"	96.3	96.8	95.3	96.5	97.3
"Sensitivity"	98.3	98.8	97.2	98.5	99.2
"Specificity"	48.4	50.6	49.2	50.5	49
"Precision"	97.9	97.9	97.9	97.9	98
"FPR"	51.6	49.4	50.8	49.5	50
"FNR"	1.7	1.2	2.8	1.5	0.80
"NPV"	48.4	50.6	49.2	50.5	49.9
"FDR"	2.1	2.1	2.1	2.1	2.0
"F1-Score"	98.1	98.3	97.5	98.2	98.6
"MCC"	49.3	54.9	42.8	52.6	58.7

Table 5.1: Heuristics-based evaluation on Suggested C-EFO Model using Swedish Leaf Database

Metrics/Tech	k-NN	VGG16	LSTM	RNN	Proposed
niques	(%)	(%)	(%)	(%)	method (%)
Accuracy"	94.4	93.2	93.1	93.8	97.3
"Sensitivity"	98.0	96.8	96.6	97.5	99.2
"Specificity"	50.0	49.6	50.0	49.5	49.9
"Precision"	96.0	95.9	96.0	95.9	98.0
"FPR"	50.0	50.4	50.0	50.5	50.1
"FNR"	2.0	3.2	03.4	2.5	0.8

"NPV"	50.0	49.6	50.0	49.5	49.9
"FDR"	4.0	4.1	4.0	4.1	2.0
"F1-Score"	97.0	96.4	96.3	96.7	98.6

Table 5.3: Heuristics-based evaluation on Suggested C-EFO Model using D-Leaf Database

Metrics and	KNN (%)	VGG16 (%)	LSTM(%)	RNN (%)	Proposed Method
Techniques					
"Accuracy"	96.5	96.1	95.2	95.6	97.1
"Sensitivity"	98.4	98.0	97.0	97.5	99.0
"Specificity"	50.3	49.4	50.6	48.6	51.0
"Precision"	97.9	97.9	98.0	97.8	98.0
"FPR"	49.7	50.6	49.4	51.4	49.0
"FNR"	1.6	2.0	3.0	2.5	1.0
"NPV"	50.3	49.4	50.6	48.6	51.0
"FDR"	2.1	2.1	2.0	2.2	2.0
"F1-Score"	98.2	97.9	97.5	97.7	98.5

Table 5.4: Classification-based evaluation on Suggested C-EFO Model using D-LeafDatabase

Metrics and	KNN (%)	VGG16 (%)	LSTM(%)	RNN (%)	Proposed Method
Techniques					
"Accuracy"	92.7	93.8	93.8	94.5	97.1
"Sensitivity"	96.1	97.3	97.4	98.1	99.0
"Specificity"	50.4	49.9	49.9	50.6	51.0
"Precision"	96.0	96.0	95.9	96.1	98.0
"FPR"	49.6	50.1	50.1	49.4	49.0
"FNR"	3.9	2.7	2.6	1.9	1.0
"NPV"	50.4	49.9	49.9	50.6	51.0
"FDR"	4.0	4.0	4.1	3.9	2.0
"F1-Score"	96.1	96.7	96.7	97.1	98.5
"MCC"	46.9	51.6	52.1	55.9	57.2

6. CONCLUSION

In this study, a novel model for classifying plant leaves is proposed. It includes a suite of techniques designed to train and classify plant leaves. [19]These include segmentation, feature selection, and preprocessing. The plant leaf model optimization strategy is called C-EFO, which stands for Continuous Improvement in Feature Selection. The plant leaf model is designed to use optimized U-NET for segmentation and introduce E-RNN to improve the classification process. The results of the study showed that the model was able to improve its accuracy and performance compared to other methods.

REFERENCES

- [1] Baez, Luis Alonso, Tereza Tichá, and Thorsten Hamann. "Cell wall integrity regulation across plant species." *Plant Molecular Biology* 109.4-5 (2022): 483-504.
- [2] Wang, Anson, et al. "Potential distributional shifts in North America of allelopathic invasive plant species under climate change models." *Plant Diversity* 44.1 (2022): 11-19.
- [3] Salve, Pradip, Pravin Yannawar, and Milind Sardesai. "Multimodal plant recognition through hybrid feature fusion technique using imaging and non-imaging hyper-spectral data." *Journal of King Saud University-Computer and Information Sciences* 34.1 (2022): 1361-1369.
- [4] Hasan, ASM Mahmudul, et al. "Weed recognition using deep learning techniques on class-imbalanced imagery." *Crop and Pasture Science* (2022).

- [5] Smith, Lucy T., et al. "Revised species delimitation in the giant water lily genus Victoria (Nymphaeaceae) confirms a new species and has implications for its conservation." *Frontiers in Plant Science* 13 (2022): 883151.
- [6] Chase, Mark W., et al. "Down, then up: non-parallel genome size changes and a descending chromosome series in a recent radiation of the Australian allotetraploid plant species, Nicotiana section Suaveolentes (Solanaceae)." *Annals of Botany* 131.1 (2023): 123-142.
- [7] Wang, Congyan, et al. "Which factor contributes most to the invasion resistance of native plant communities under the co-invasion of two invasive plant species?." *Science of the Total Environment* 813 (2022): 152628.
- [8] Li, Dawei, et al. "PlantNet: A dual-function point cloud segmentation network for multiple plant species." *ISPRS Journal of Photogrammetry and Remote Sensing* 184 (2022): 243-263.
- [9] Bradley, Ellie L., et al. "Secreted glycoside hydrolase proteins as effectors and invasion patterns of plantassociated fungi and oomycetes." *Frontiers in Plant Science* 13 (2022): 853106.
- [10] Ngou, Bruno Pok Man, et al. "Concerted expansion and contraction of immune receptor gene repertoires in plant genomes." *Nature Plants* 8.10 (2022): 1146-1152.
- [11] Muneer, Muhammad Atif, et al. "Differential response of mycorrhizal fungi linked with two dominant plant species of temperate grassland under varying levels of N-addition." *Applied Soil Ecology* 170 (2022): 104272.
- [12] Huimin, L. I. A. O., et al. "Study on dust retention rank and pattern recognition of typical garden plant leaves in Changsha." *Ecology and Environment* 31.1 (2022): 110.
- [13] Lu, Xiaofeng. "Deep learning based emotion recognition and visualization of figural representation." *Frontiers in psychology* 12 (2022): 818833.
- [14] Sun, Yanqing, et al. "Twenty years of plant genome sequencing: achievements and challenges." *Trends in Plant Science* (2022).
- [15] Zewdie, Beyene, et al. "Plant biodiversity declines with increasing coffee yield in Ethiopia's coffee agroforests." *Journal of Applied Ecology* 59.5 (2022): 1198-1208.
- [16] Tian, Zhiwei, et al. "Application status and challenges of machine vision in plant factory—A review." *Information Processing in Agriculture* 9.2 (2022): 195-211.
- [17] Haq, Shiekh Marifatul, et al. "A cross-cultural analysis of plant resources among five ethnic groups in the Western Himalayan region of Jammu and Kashmir." *Biology* 11.4 (2022): 491.
- [18] Fedick, Scott L., and Louis S. Santiago. "Large variation in availability of Maya food plant sources during ancient droughts." *Proceedings of the National Academy of Sciences* 119.1 (2022): e2115657118.
- [19] Lu, Yuzhen, et al. "Generative adversarial networks (GANs) for image augmentation in agriculture: A systematic review." *Computers and Electronics in Agriculture* 200 (2022): 107208.
- [20] Guedes, Raul Narciso C., Rachel R. Rix, and G. Christopher Cutler. "Pesticide-induced hormesis in arthropods: Towards biological systems." *Current Opinion in Toxicology* 29 (2022): 43-50.
- [21] Kubera, Elżbieta, et al. "Detection and recognition of pollen grains in multilabel microscopic images." *Sensors* 22.7 (2022): 2690.
- [22] Gao, Ming, et al. "A Resilience Evaluation Framework on Ultrasonic Microphone Jammers." *IEEE Transactions* on Mobile Computing (2023).
- [23] Ferreira, Savio S., Charles E. Anderson, and Mauricio S. Antunes. "A logical way to reprogram plants." *Biochemical and Biophysical Research Communications* (2023).
- [24] Trieflinger, Stefan, et al. "The discovery effort worthiness index: How much product discovery should you do and how can this be integrated into delivery?." *Information and Software Technology* 157 (2023): 107167.
- [25] Sabo, Autumn E., et al. "Herbaceous plant height is an early indicator of groundlayer response to an experimental manipulation of forest structure and deer pressure." *Forest Ecology and Management* 536 (2023): 120902.