Study on Identification and Classification of Medicinal Plants Using Machine Learning and Deep Learning

Nilesh S. Bhelkar[1], Dr. Avinash Sharma[2]
[1]*Research Scholar at Madhyanchal Professional University , Ratibad , Bhopal -.462044, MP, INDIA.*
[2]*Professor Dept of Computer Science and Engineering, Madhyanchal Professional University , Ratibad , Bhopal -.462044, MP, INDIA*

ABSTRACT
Now a day's pharmaceutical industry facing a challenges to create a medicine for new illnesses. The major resource to find new medicines are medicinal plants which are available in abundant but most of the Professional like medical practitioners, botanists, medical representatives and pharmacist are not in position to recognize the right therapeutic plant. We benefit greatly from recent advances in computer vision technology such as deep learning, machine learning, and image processing. The researcher and academician are developing many automatic identification and recognition system which are used to recognize and categorize the medicinal plants so as to pharmacist prepare a medicine within less time which overcome the limitation of traditional manual identification system. The performance of identification and classification system is greatly depends on the feature extraction part of the system. One of the most recent technology due its deep features extraction capability and classification accuracy used for identification and classification is deep learning. This study reviews effectiveness of several deep learning and machine learning methods.
Keywords: Classification, Deep learning, Feature extraction, Identification, Medicinal plants, Machine learning.

I. INTRODUCTION

The nature is the greatest resource through which human being fulfil their need like food, shelter, medicine etc. Due to the environmental changes many life danger viruses and illnesses are created. The pharmaceutical and medical field has many challenges how to deal with such new viruses and illnesses. How they will come up with immediate solution. We know all of us just face the challenge of COVID-19. To deal with such life danger viruses and illnesses immediate solution are required. The pharmacist and medical practitioners need new medicinal properties to defend from such life danger viruses and illnesses. The pharmacist and medical practitioners will get new medicinal properties through one of the great treasure that is medicinal plants.

The ancient treasure has around 400000 medicinal plants but only 270000 plants were identified by botanists and ayurdedic practitioners. The rest of the Medicinal plants are unidentified and it is difficult to identify manually due to similar morphological structure of plants. The manual identification is prone to error. Plants are vital to all life on Earth, including human life. Plants have a major impact on how the natural cycle works. Plants are incredibly useful since they are the foundation of the food chain and are the origin of numerous treatments. Despite the numerous imaginative advancements made in the field of botany, there are still a huge number of unexplored, unidentified, and unutilized plants. The biology and agribusiness of plants both depend on accurate plant identification. Environmental preservation, assessing plant resources, and teaching are just a few applications for automated plant classification systems. Two important components of plant taxonomy that are crucial to these activities are plant identification and plant categorization.

The many researcher and academician developed an automatic plants identification and categorization system. A system for identifying and recognising plants can be created thanks to technology advancements in the fields of computer vision. The many researchers are used image processing techniques to developed face recognition and plants disease detection system. The detection is done using plants morphological and geometrical features in that parts of plant like leaves,

DOI:10.46647/ijetms.2023.v07i03.109 ISSN: 2581-4621

flowers, fruits, stem and roots are used. Most of the researchers are used leaf features due to its availability in all season.
Computer-assisted plant identification has developed into a fascinating area of study in order to obtain the data. The desire for automated technologies that would enable non-botanical people to carry out important field work of identifying and characterising plants has increased due to the global shortage of specialist taxonomists. These instruments are necessary in many fields, such as forestry, agronomy, and pharmaceutical science. Leaf recognition comes first throughout the design and development of such instruments. Plant identification based on leaf image is the most successful and reliable method when compared to alternative methods like cell and molecule biology. Because inexpensive digital cameras are readily available, taking samples of leaves and taking pictures of them is practical and viable.
The application of deep learning, that is currently very well-liked and in use, has been made in a number of industries, including biology, medicine, computer vision, speech recognition, and others. When compared to traditional machine learning techniques, deep learning can extract more precise information. This study reviews effectiveness of several deep learning and machine learning methods.

II. RELATED WORK

Begue et.al. [1] utilized computer vision and machine learning method for automatic recognition of medicinal plants. Author collected data from tropical island of Mauritius. It consist of 720 images of distinct plant species The shape and colour features are extracted for classification and applied 10-fold cross validation procedure using random forest classifier and achieved 90.1% accuracy.
The author [2] collected the 32 different arurvedic medicinal plants species data. The colour and shape features were extracted. The machine learning support vector machine method applied on optimised feature sets and 99.66% accuracy was achieved.
In this [3] research six type of medicinal plant species data consider for classification. The 65 texture and multispectral features were extracted and chi-square feature selection optimization strategy is applied. Only optimised 14 features were used for classification. MLP, RF, logit-boost, basic logistic and bagging these five machine learning classification techniques were applied and promising 99.01% accuracy was achieved through multilayer perceptron classifier.
A 200-image collection of medicinal plants from India's western ghat was generated by the author [4]. There are 20 separate classes, each with 10 leaves from various plants. The HOG (vein features) and SURF (feature descriptor) experiments using the kNN classifier demonstrated accuracy of about 99.6%.
Researcher [5] was created Indian medicinal plants leaf dataset namely Ayur Bharat of having 10 distinct classes of around 10000 images. InceptionV3, VGG16, and ResNet101, 3 pre-trained CNN-based architectures, were chosen and trained on the ImageNet database. To actualize the transfer learning technique these three model were applied on Ayur Bharat dataset and the performance was improved by canny edge detection method. The greatest validation accuracy and f1 score 0.9732 and 0.9653 respectively achieved by InceptionV3 architecture with pre-processing.
Bahri et.al.[6] developed a Dynamic CNN (DCNN) classification model for Morocco aromatic and medicinal plants. The author created dataset of 16 plant species with 320 images. In this study, entropy impurity was used to dynamically integrate CNN classification results. Inception V3, ResNet50, and VGG16 are employed. The dynamic Resnet50 method used by CNN was 97.4% accurate.
In this study [7], face94 dataset is utilised. It include 5 distinct plant species and 100 images of each species. 80:20 ratio of plant species is utilised for training and testing. The classification of leaf photos using the CNN deep learning model yielded an accuracy of 86%.
A CNN-based system with an accuracy of 95.58 percent was utilised in this study [8] leaf feature is used to recognize the species. The author collected images of leaves from six groupings, which are referred to by their regional names: Tulshi, Sojne, Pathorkuchi Darchini, Tejpata, and Neem, to develop this taxonomy.

Website: ijetms.in Issue: 3 Volume No.7 May - June – 2023
DOI:10.46647/ijetms.2023.v07i03.109 ISSN: 2581-4621

Sachar and Kumar [9] created a deep ensemble learning model for automatically identifying medicinal leaves, and utilizing pre-trained models like ResNet50 InceptionV3 and MobileNetV2, their accuracy on the test set was 99.66%, and their average accuracy was 99.9%. 30 classes make up the dataset for medicinal leaves.

This study [10] describes a transfer learning strategy. Out of six supervised learning algorithms, DenseNet201 performed the best on average (87%), while GoogLeNet performed the worst (79%). Log Gabor filters and DL algorithms were used to construct a computer vision model for the recognition of therapeutic plants. The model's accuracy on a dataset of 49 therapeutic plant species was 98%.

In this [11], fifty medical images from Google images were used for edge detection. When CNN's texture patch was applied for classification, the findings revealed that it was 98% accurate.

The author [12], employed CNN to identify and categorise deseases of soybean plants. The Network's classification accuracy was 99.32% after training on photos obtained in the natural world. This demonstrates CNN's promising capacity to extract significant information from the environment that are needed for the classification of plant diseases.

In [13], Dhingra et al. created a computer vision technique for recognising and categorising leaf diseases. In diagnosing six different kinds of leaf diseases, their method achieved an accuracy of 97.22 percent.

This [14] work proposes AyurLeaf, a deep learning-based CNN model to classify medicinal plants using shape and colour characteristics of the leaves. 40 therapeutic plants species leaf samples dataset is used. Alexnet based DNN is used for effective feature extraction. SVM and Softmax classification methods were applied and after five cross validations, had a classification accuracy of 96.76%.

Based on the traditional CNN, the authors of [15] integrated a pooling layer and an inception structure to diagnose leaf disorders. The number of parameters was decreased in this model, and the identification accuracy increased by up to 91.7%.

A CNN-based automated system for classifying medicinal plants is shown in the approach suggested in [16]. High-level characteristics are extracted for classification using a 3-layer CNN. A data augmentation technique is used to support the procedure in order to increase efficiency. The testing findings indicate that the method's recognition accuracy rate is roughly 71.3%.

A mixed neural network system is being developed by this research [17]. "AousethNet" is an altered version of AlexNet. Accuracy of AousethNet is 99% with the Mendeley dataset (MD2020).

The AlexNet and GoogLeNet architectures were fine-tuned by Pawara et al. [18], He achieved 99% accuracy on the Swedish Leaf dataset, 94% accuracy on the Flavia dataset, and 98% accuracy on the Folio dataset.

On the Folio, Flavia, and Swedish leaf datasets, VGG19 architecture was combined with a logistic regression classifier by Pearline et al. [19] and achieved accuracy of 96%, 96%, and 99%, respectively.

In this study[20] 185 plants image dataset is utilized. AlexNet, GoogLeNet, VGG- 19, ResNet50, and MobileNetV2 these 5 CNN based model are pretrained. 92.3% greatest accuracy achieved by MobileNetV2 with ADAM optimizer.

A plant identification system was created by Karahan and Nabiyev [21] using the pretrained network MobileNetV2. 5,345 flower and plant photos from 76 different species were utilised in the collection. The suggested model had accuracy on the training set of 99.71% and accuracy on the test set of 98.97% after 15 iterations.

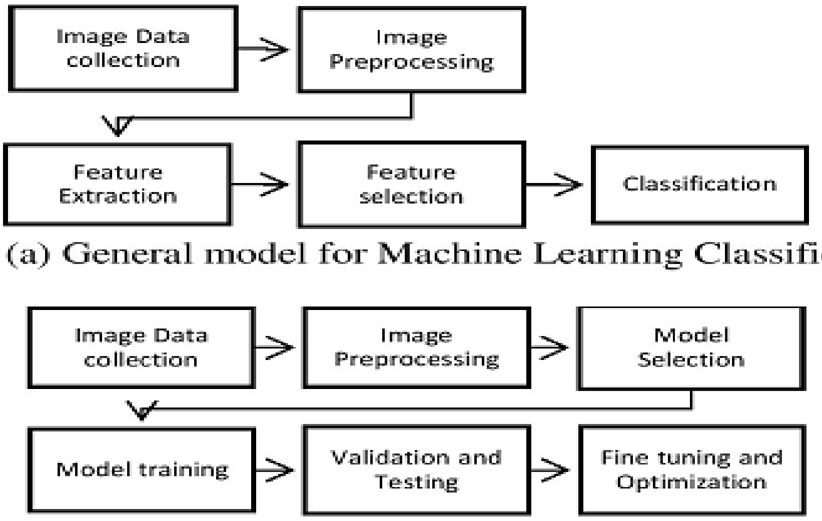
In their CNN-LSTM network design by Banzi and Abayo [22], 20 layers were included with the softmax function for classification. An open collection of 100 plant species photos was used to train the models. According to experiments, the suggested CNN-LSTM works better than the convectional CNN in classifying plant species because it achieves an accuracy of 95.06%.

Table I summarised the classification methods of ML and DL of this study.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.7 May - June – 2023
DOI:10.46647/ijetms.2023.v07i03.109 ISSN: 2581-4621

III. GENERALIZED MODEL

In this study we have found that there are different techniques are invented by researchers and academician to recognise and group the therapeutic plants. The general model for identification and classification of medicinal plants is depicted in fig.1. The fig.1 (a) shows the general model for identification and classification of medicinal plants using image processing and machine learning technique. It includes several steps, including gathering image data, processing it, extracting features, choosing features, and classifying them. The image data collection can be done through digital camera, scanner, smart phone and dataset of medicinal plant species. Image preprocessing operation will perform for noise removal, resizing image, segmentation and contrast to improve overall quality of image. The leaf image shape, texture, colour and vein features are extracted using different image processing and computer vision techniques. Feature selection step play very important role in identification and classification process. Improper feature selection may lead to incorrect classification. The supervised and unsupervised ML classifiers are utilized to classify therapeutic plants based on selected features of leaf. Some of the classification techniques are SVM, RF, DT, KNN, Artificial Neural Network, LDA, Naïve Bayes, K-means and Fuzzy C-means.



(a) General model for Machine Learning Classification

(b) General model for Deep Learning Classification

Fig.1 Generalized model for Medicinal plant Identification and Classification

Deep learning has been used in many fields, including automatic voice recognition, computer vision, audio and video recognition, and natural language processing, with outstanding results [23, 24]. Deep learning method also used to identify and categorise medicinal plants due to its detail feature extraction capability. Huge amounts of data are used by deep learning, which was inspired by how the human brain processes information, in order to map a given input to particular labels. Deep learning architecture includes recurrent neural networks, convolutional neural networks, deep belief networks, and deep neural networks. The generic model for classifying and identifying medicinal plants using deep learning is displayed in Fig. 1(b). The first two step process are same as fig.1 (a). The third step is model selection mostly Convolution Neural Network based architecture is used which is automatically extract the features from raw data and the model will train for same features to get the required results. Some of the CNN based models are AlexNet, GoogLeNet, VGG- 19, ResNet50, and MobileNetV2, DCNN. The trained model then validate and test on different dataset for performance check. To improve the performance of model fine tuning and optimization process will apply by adjusting hyperparameters. Adaptive Moment Estimation (Adam), Root Mean Square Propagation (RMSProp), Stochastic Gradient Descent (SGD), AdaGrad (Adaptive Gradient Descent), Momentum, and Adadelta are a few well-known optimizers [22]. The performance of both approach will measure by f1 score, precision, recall and accuracy metrics.

DOI:10.46647/ijetms.2023.v07i03.109 ISSN: 2581-4621

TABLE I. ML AND DL LEAF CLASSIFICATION STUDY SUMMARY

Author and Reference	**Year**	**Dataset**	**Method**	**Features**	**Accuracy**
Begue et. al.[1]	2017	720 images of Medicinal plants	RF classifier	Shape, colour	90.1%
Dahigaonkar et.al.[2]	2018	32 different ayurvedic medicinal plants species	SVM	Colour, texture and Shape	96.66%
S. Naeem et al [3]	2021	6 distinct plant species	MLP	14 optimised features	99.01%
Sabu et.al. [4]	2017	200 medicinal plants images from India's western ghat	kNN	Vein, feature descriptor	99.6%
Pandian et al. [5]	2021	Ayur Bharat of having 10 distinct classes of around 10000 images.	Pre-trained InceptionV3 architecture		97.37%
Bahri et al. [6]	2022	dataset of 16 plant species with 320 images	Dynamic CNN	Vein	97.4%
Diqi and Mulyani [7]	2020	500 images from 5 different plant species	CNN	Geometric, colour	86%
Musa et. al. [8]	2021	5 plant species images	CNN		95.58%
Sachar and Kumar [9]	2022	30 classes plant species	Deep ensemble learning model		99.66%
Oppong et al. [10]	2022	49 plant species.	Log Gabor filters and CNN DL (Deep learning) systems	Texture	97.3%
Sabarinathan et al. [11]	2018	fifty medical photos sourced from Google images	CNN	Shape	98 %.
Wallelign et. al.[12]	2018	12,673 soybean leaf samples	CNN	Colour	99.32%
Dhingra et al.[13]	2019		Computer Vision method		97.22%
Dileep et. al. [14]	2019	Dataset of 40 medicinal plants from south western coast of India	CNN, Softmax and SVM classifiers	Shape, Colour	96.76%
Anami et.al [15]	2020	6000 *images*	CNN	Shape	92.89%
Akter et.al[16]	2020	3570 *images*	CNN	Shape Vein	71.30%
Elizabeth et.al [17]	2022	Mendeley dataset (MD2020)	AousethNet		99%,
Pawara et. al. [18]	2017	Swedish leaf dataset Folio	AlexNet and GoogLeNet		99% 98%
Pearline et.al.[19]	2019	Folio Flavia	VGG19, LR classifier		96% 96%
Adetiba el. Al [20]	2021	Leafsnap image dataset	MobileNetV2		92.3%

Karahan et. al. [21]	2021	5345 images of 75 plant species	MobileNetV2		98.97%
Banzi et.al. [22]	2021	100 plant species images	CNN-LSTM network with 20 layers		95.06%.

IV. CONCLUSION

The correctness of manual identification process is greatly depends on human expertise. Manual identification of medicinal plants is labour-intensive and subject to human error. These problems might be resolved by automatic plant identification, the recent computer vision's ML and DL methods are utilised to achieve atomisation in identification and classification of medicinal plants but creating such a system would need a substantial investment of time and money together with a comprehensive knowledge of plant morphology. The majority of recent studies on autonomous plant identification systems test their efficacy by analysing pre-existing datasets that were produced in a controlled environment. So, further research must be done on images taken in various lighting situations and against complex backgrounds.

Several aspects that might have an impact on the classification process' accuracy have been discussed and studied in our particular analysis. Features, classifiers, and testing datasets are some of these variables. We discovered that adding additional features has a beneficial effect on the categorization procedure. It is also advised to use a large dataset to enhance training. The created identification system's accuracy would increase as a result. The investigation also revealed that a variety of variables, including as shape, vein, colour, and texture, would significantly affect how well an object might be categorised. An autonomous ML and DL method for detecting plants would be highly helpful for environmental preservation, and with greater accuracy, the usage of therapeutic plants in the medical field may become more prevalent.

REFERENCES

[1] Begue, Adams, et al.(2017) "Automatic Recognition of Medicinal Plants Using Machine Learning Techniques." International Journal of Advanced Computer Science and Applications, vol. 8, no. 4, The Science and Information Organization, 2017. Crossref, https://doi.org/10.14569/ijacsa.2017.080424.

[2] Dahigaonkar, T.D.; Kalyane, R. Identification of ayurvedic medicinal plants by image processing of leaf samples. Int. Res. J. Eng. Technol. (Irjet) 2018, 5, 351–355.

[3] S. Naeem et al., "The classification of medicinal plant leaves based on multispectral and texture feature using machine learning approach," Agronomy, vol. 11, no. 2, 2021, doi: 10.3390/agronomy11020263.

[4] A.Sabu, K. Sreekumar, and R. R. Nair. Recognition of ayurvedic medicinal plants from leaves: A computer vision approach. In 2017 Fourth International Conference on Image Information Processing (ICIIP), pages 1–5, Dec 2017

[5] K. Arunaggiri Pandian, T. S. Sai Kumar, S. Thabasum Aara, and A. Prabalakshmi, "Identification of Indian Medicinal Plants from Leaves using Transfer Learning Approach," in Proceedings of the 5th International Conference on Trends in Electronics and Informatics, ICOEI 2021, 2021, doi: 10.1109/ICOEI51242.2021.9452917.

[6] Bahri, Y. Bourass, I. Badi, H. Zouaki, K. El moutaouakil, and K. Satori, "Dynamic CNN Combination for Morocco Aromatic and Medicinal Plant Classification," Int. J. Comput. Digit. Syst., vol. 11, no. 1, 2022, doi: 10.12785/ijcds/110120.

[7] M. Diqi and S. H. Mulyani, "Implementation of CNN for Plant Leaf Classification," Int. J. Informatics Comput., vol. 2, no. 2, 2021, doi: 10.35842/ijicom.v2i2.28.

[8] M. Musa, M. S. Arman, M. E. Hossain, A. H. Thusar, N. K. Nisat, and A. Islam, "Classification of Immunity Booster Medicinal Plants Using CNN: A Deep Learning Approach," in Communications in Computer and Information Science, 2021, vol. 1440 CCIS, doi: 10.1007/978-3-030-81462-5_23.

DOI:10.46647/ijetms.2023.v07i03.109 ISSN: 2581-4621

[9] Sachar S, Kumar A (2022). Deep ensemble learning for automatic medicinal leaf identification. Int J Inf Technol. (2022);14(6):3089-3097. doi: 10.1007/s41870-022-01055-z. Epub 2022 Aug 12. PMID: 35975198; PMCID: PMC9373896

[10] Oppong SO, Twum F, Hayfron-Acquah JB, Missah YM (2022). A Novel Computer Vision Model for Medicinal Plant Identification Using Log-Gabor Filters and Deep Learning Algorithms. Comput Intell Neurosci. 2022 Sep 27;2022:1189509. doi: 10.1155/2022/1189509. PMID: 36203732; PMCID: PMC9532088.

[11] Sabarinathan, C.; Hota, A.; Raj, A.; Dubey, V.K.; Ethirajulu, V. Medicinal plant leaf recognition and show medicinal uses using convolutional neural network. Int. J. Glob. Eng. 2018, 1, 120–127.

[12] Wallelign, S.; Polceanu, M.; Buche, C. Soybean plant disease identification using convolutional neural network. In Proceedings of the Thirty-First International Flairs Conference, Melbourne, FL, USA, 10 May 2018.

[13] Dhingra, G.; Kumar, V.; Joshi, H.D. A novel computer vision based neutrosophic approach for leaf disease identification and classification. Measurement 2019, 135, 782–794.

[14] M. R. Dileep and P. N. Pournami, "AyurLeaf: A Deep Learning Approach for Classification of Medicinal Plants," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 321-325, doi: 10.1109/TENCON.2019.8929394.

[15] B. Anami, N. Malvade and S. Palaiah, "Deep learning approach for recognition and classification of yield affecting paddy crop stresses using field images," Artificial intelligence in agriculture, vol. 4, pp. 12– 20, 2020.

[16] R. Akter and M. Hosen, "CNN-based leaf image classification for Bangladeshi medicinal plant recognition", Emerging technology in computing, communication and electronics, pp. 1-6, doi: 10.1109/ETCCE51779.2020.9350900.

[17] C. P. Blesslin Elizabeth and S. Baulkani, "Novel Network for Medicinal Leaves Identification," IETE Journal of Research, pp. 1–11, 2022.

[18] P. Pawara, E. Okafor, L. Schomaker, and M. Wiering, *Data Augmentation for Plant Classification," Advanced Concepts For Intelligent Vision Systems*, Springer International Pub- lishing, Midtown Manhattan, New York City, pp. 615–626, 2017.

[19] S. Anubha Pearline, V. Sathiesh Kumar, and S. Harini, "A study on plant recognition using conventional image processing and deep learning approaches," *Journal of In- telligent & Fuzzy Systems*, vol. 36, no. 3, p. 1997, 2019.

[20] T. Karahan and V. Nabiyev, "Plant identification with convolutional neural networks and transfer learning," *Pamukkale University Journal of Engineering Sciences*, vol. 27, no. 5, pp. 638–645, 2021.

[21] J. Banzi and T. Abayo, "Plant species identification from leaf images using deep learning models (CNN-LSTM architec- ture)," *Tanzania Journal of Forestry and Nature Conserva- tion*, vol. 90, no. 3, pp. 93–103, 2021.

[22] S. Postalcıoğlu, "Performance analysis of different optimizers for deep learning- based image recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 02, Article ID 2051003, 2019

[23] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, 2020.

[24] T. R. Gadekallu, DS Rajput, MPK Reddy et al., "A novel PCA–whale optimization-based deep neural network model for classification of tomato plant diseases using GPU," *Journal of Real-Time Image Processing*, vol. 18, no. 4, pp. 1383–1396, 2021.