Journal of Advanced Zoology

ISSN: 0253-7214

*Volume **44** Issue **S-5 Year 2023** Page **1080:1090***

Applying Deep Learning and Machine Learning Algorithms for The Identification of Medicinal Plant Leaves Based on Their Spectral Characteristics

Nidhi Tiwari[1], Bineet Kumar Gupta[2*], Abhijityaditya Prakash[3], Mohammad Husain[4], Devendra Singh[5]

[1]Research Scholar, Department of Computer Science Engineering, Shri Ramswaroop Memorial University, Lucknow, India
[2,3]Faculty of Computer Science and Information Systems, Shri Ramswaroop Memorial University, Lucknow, India*
[4]Faculty of Computer Science and Information Systems, Islamic University of Madinah, Al Madinah Al Munawarah, Saudi Arabia
[5]Faculty of Biotechnology, Institute of Biosciences and Technology, Shri Ramswaroop Memorial University, Lucknow, India

***Corresponding Author**: Bineet Kumar Gupta[2]*
Email: hod.csis@srmu.ac.in[2]

Article History

Received: 05 June 2023
Revised: 14 Sept 2023
Accepted: 17 October 2023

Abstract

The study and consideration of medicinal plants have been ongoing throughout history due to their significant role in maintaining the well-being of mammals. Although identifying medicinal plants can be a valuable skill, it is often time-consuming, tedious, and requires the expertise of a specialist. The project works on the technique of image processing, which identifies the various medicinal plants. There has been a strong emphasis on improving efficiency through the application of technology, with a focus on incorporating digital image processing and pattern recognition techniques. To ensure accurate plant identification, proposals involving the application of computer vision neural network techniques have been advanced. This approach involves neural network models such as CNN, SVM, KNN, and Navie Bay for identifying the medical plants based on their respective features. After the validation step, the project provides a classification of 92.3 precision and 90.56 F1 score.

Keywords: Medicinal Plant detection, CNN, ML, AI, DL

1. Introduction

Plants play a crucial role in promoting biodiversity and sustaining life on Earth.[1] They provide essential resources such as oxygen and water, and many species of plants also have medicinal properties that can be used to treat various ailments. The knowledge of medicinal plants has been transmitted from one generation to another, underscoring its importance. Maintaining knowledge of medicinal plants is crucial because finding a plant that possesses medicinal properties and all the necessary therapeutic properties can be incredibly challenging. Great promise is shown by various techniques such as pattern recognition, image processing, and computer recognition, in aiding the identification and classification of medicinal plants. It is also known that herbal treatments have no adverse effects, therefore a patient can only die if any incorrect plant is given to him or her for healing. This necessitates the subsequent stage to adopt a fully automatic approach for precise identification of medicinal plants. The implementation of Ayurvedic treatments is enabled by the classification and identification of these plants. Forest department officials, ayurvedic practitioners' agronomists, botanist and all those who are engaged in the manufacture of ayurvedic pharmaceuticals get a good benefit from appropriate classification and categorization of medicinal plants. Expert taxonomists are very less available to this area. In contrast, ayurvedic remedies are preferred by people over all others [2]. Additionally, plants are classified according to their characteristics by taxonomists, including their leaves, flowers, stems, and branches. Species are categorized derived on their leaves, which are present year-round. Throughout Mother Earth, organisms rely on photosynthesis for their food. In order to survive and breathe, humans and animal race rely entirely on plants.

There have been different ways of knowing the properties of plants throughout the history of all cultures. Learners with limited knowledge of plants can be provided with a combination of botany, the study of plants, and information technology. This is done through automated identification systems that provide information on plant properties [3]. With the help of an automated system, the user can identify species of plants by taking images of them.

In recent years, ML algorithms have emerged as a powerful tool for the recognition of plant species on the basis of their spectral characteristics. Spectral analysis provides a non-destructive method for identifying plant species by measuring their unique spectral signatures. ML algorithms can then be given to analyze these spectral data and classify the plant species accurately.

Within this context, the utilization of machine learning algorithms for the recognition of medicinal plant leaves [4] based on their spectral features has garnered considerable interest in the realm of medicinal plant research. Through the deployment of ML algorithms, the identification of medicinal plants can be refined, leading to enhanced efficiency and precision. This, in turn, facilitates the production of high-quality herbal medicines. The objective of this paper is to conduct a review of the latest advancements in the utilization of ML algorithms for the recognition of medicinal plant leaves, with a focus on their spectral characteristics.

2. Materials and Methods

To effectively train and assess object recognition algorithms [5], it is essential to have access to datasets at various developmental stages. The Medicinal Plant Leaf dataset [6], which encompasses 30 distinct plant species, each represented by images sized at 1600X1200, serves as a crucial resource. However, to streamline and optimize its utility in advancing Artificial Intelligence models, particularly within the realms of machine learning and deep learning applications, a deliberate curation process was undertaken. This process involved the selection of 10 plant species, and their associated images were resized to 256X256, yielding a dataset that has played a pivotal role in furthering the capabilities of AI models. This dataset acts as a valuable asset for researchers and computer scientists, enabling the recognition of plant species, disease detection, and exploration of the herb's presence and medicinal characteristics. By making this dataset available to the wider community, it aims to inspire and accelerate research in the field of medicinal plants, addressing the critical shortage of publicly available datasets, which has long hindered progress in this vital domain.

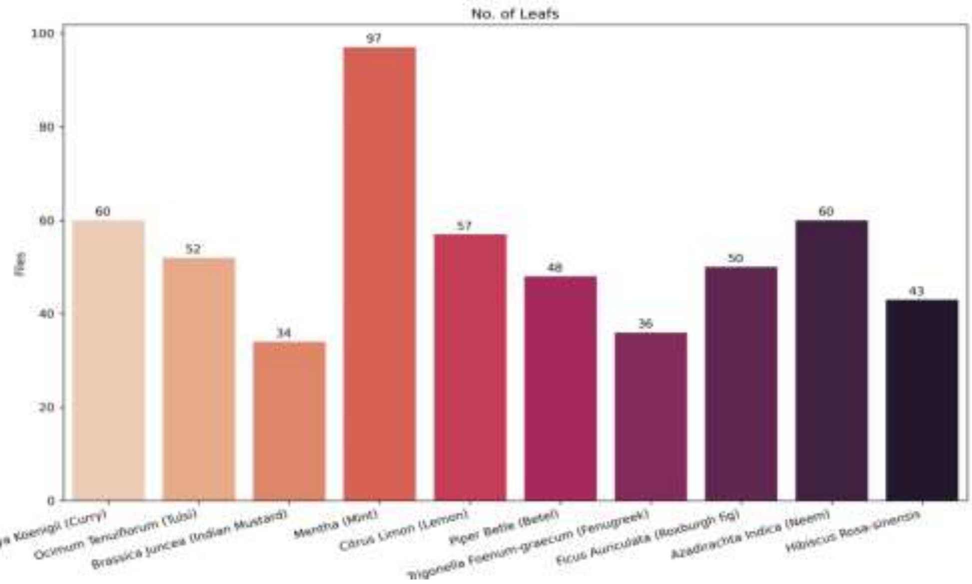


Fig 1. Plant Species for training and testing data

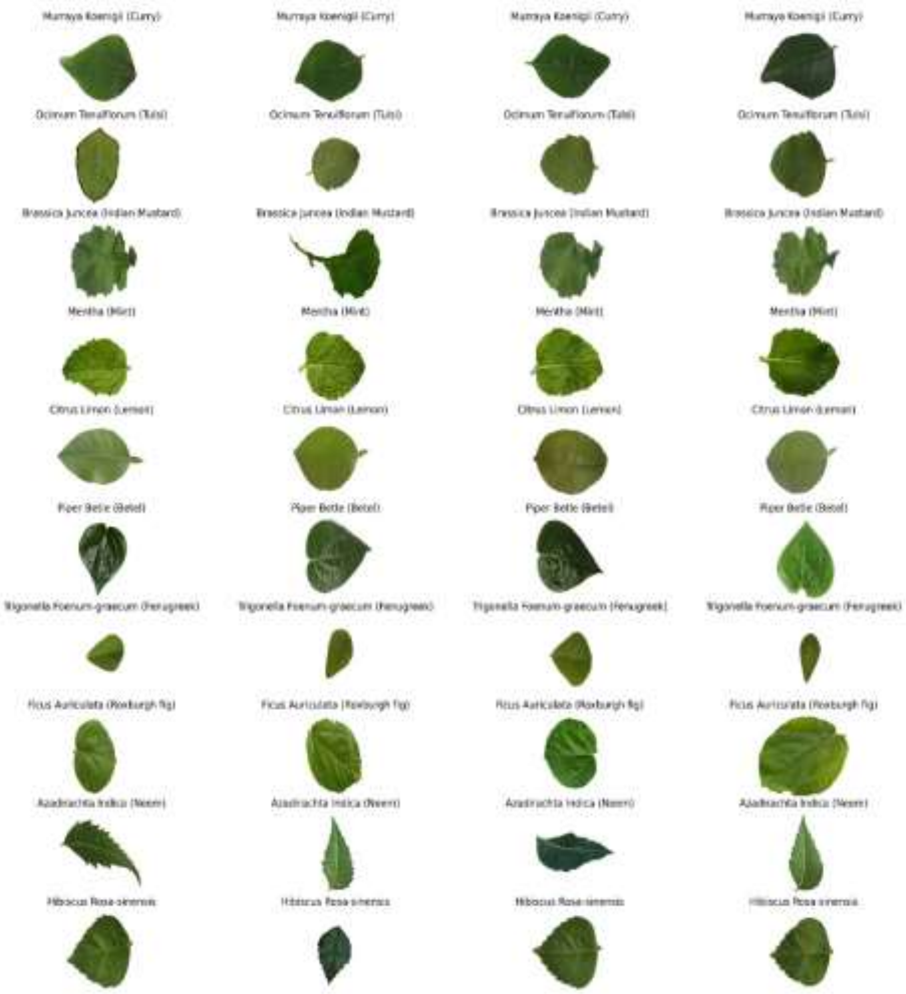


Fig 2: Dataset Description

As depicted in the figure, a sample dataset is presented, comprising four samples for each plants for the purpose of analysis.
These medicinal plants that are selected for the analysis:
0 : Azadirachta Indica (Neem)
1 : Brassica Juncea (Indian Mustard)
2 : Citrus Limon (Lemon)
3 : Ficus Auriculata (Roxburgh fig)
4 : Hibiscus Rosa-sinensis
5 : Mentha (Mint)
6 : Murraya Koenigii (Curry)
7 : Ocimum Tenuiflorum (Tulsi)
8 : Piper Betle (Betel)
9 : Trigonella Foenum-graecum (Fenugreek)

Machine Learning based classification

Plant recognition is a pivotal task accomplished through the categorization of different plant species [7]. Diverse ML algorithms are at the disposal of researchers and practitioners to execute this classification process

effectively. These algorithms span a range of methodologies, including Support Vector Machines (SVM), K-Nearest Neighbour (KNN), and Naive Bayes (NB).
In the realm of computer vision, natural language processing (NLP) and bioinformatics, machine learning algorithms play a substantial role and find widespread application. Among these, Convolutional Neural Networks (CNNs), renowned for their deep learning capabilities, have revolutionized image recognition tasks. SVMs, on the other hand, are adept at handling both linear and nonlinear classification problems, making them versatile tools for various domains. Additionally, K-Nearest Neighbors (K-NNs) have been extensively employed for their simplicity and effectiveness in pattern recognition tasks.
These ML algorithms collectively emphasize to the advancement of plant recognition systems, with each having its unique benefits & suitability for specific scenarios, ensuring the continued progress of this critical field.

Convolutional Neural Networks (CNNs)
Within the domain of deep learning, Convolutional Neural Networks (CNNs) distinguish themselves as specialized algorithms meticulously crafted for tasks centered on the comprehension of images and videos [11]. These CNNs feature a multi-layered architecture finely tuned to decipher the intricacies of visual data.
At the core of CNNs[12] are convolutional layers, where filters and convolution operations are employed to systematically detect intricate features within the input image. These layers essentially act as the eyes of the network, discerning essential patterns and structures. Complementing the convolutional layers are pooling layers, which play a vital role in down sampling the feature maps. By reducing their spatial dimensions, pooling layers not only help manage computational complexity but also prevent the network from overfitting, ensuring it generalizes well to new data. The ultimate responsibility for image classification lies with the fully connected layers, which leverage the knowledge of features gleaned from the previous layers to make precise classifications. One of the remarkable traits of CNNs is their innate ability to autonomously learn pertinent features directly from the data, obviating the need for extensive manual feature engineering. This remarkable capability makes CNNs particularly effective and indispensable for tasks that hinge on visual data, reaffirming their significance in the ever-evolving landscape of deep learning and image-related applications.

Artificial Neural Network (ANN)
In the area of ANN, the neurons in a brain form a network, which then processes information and produces results. This is analogous to what happens in the brain when you connect all of the neurons together to form a network. ANN make computers and machines so smart they can operate just like humans, understanding, experiencing and making decisions, just like they do in the human brain. Neural networks are unique because they can understand things by themselves, experience mistakes, and learn from them.
ANNs [9] currently dominate the field of machine learning. The concept of Artificial Neural Networks (ANNs) bears a resemblance to the neurons in the human brain. ANNs' performance is influenced by factors like the quantity of neurons and the inclusion of hidden layers. In this particular implementation, feedforward neural networks with a single hidden layer were utilized, and the training phase employed the Scaled Conjugate Gradient method. The stopping criteria for training were based on reaching a minimum gradient.

K-Nearest Neighbour (KNN)
Plant species can also be identified using KNN methods. Feature space distance metrics are the basis of the KNN algorithm [9]. Based on neighbor weights, KNN classifiers determine the label for the sample based on k nearest neighbors. An instance-based learning algorithm, K-Nearest Neighbors (k-NNs), is powerful and simple for both classification and regression. On the basis of the majority class among its k closest neighbors, it classifies a new data point according to its k closest neighbors. Due to their ability to classify new data using similarity to the training data, k-NNs are especially good for tasks involving small datasets and simple decision boundaries.

Naive Bayes (NB)
Statistical methods draw upon Bayesian classifiers [13][14], employing probability as their foundation for classification tasks. The Bayes algorithm, grounded in Bayes' Theorem, is a probabilistic approach for classification. In the case of the Naive Bayes algorithm, it leverages input features to compute class

probabilities and selects the class with the highest probability, assuming feature independence as a simplification. This "naive" assumption streamlines calculations. Bayesian probability analysis is notably effective in handling high-dimensional data and large datasets due to its computational efficiency and straightforward implementation.

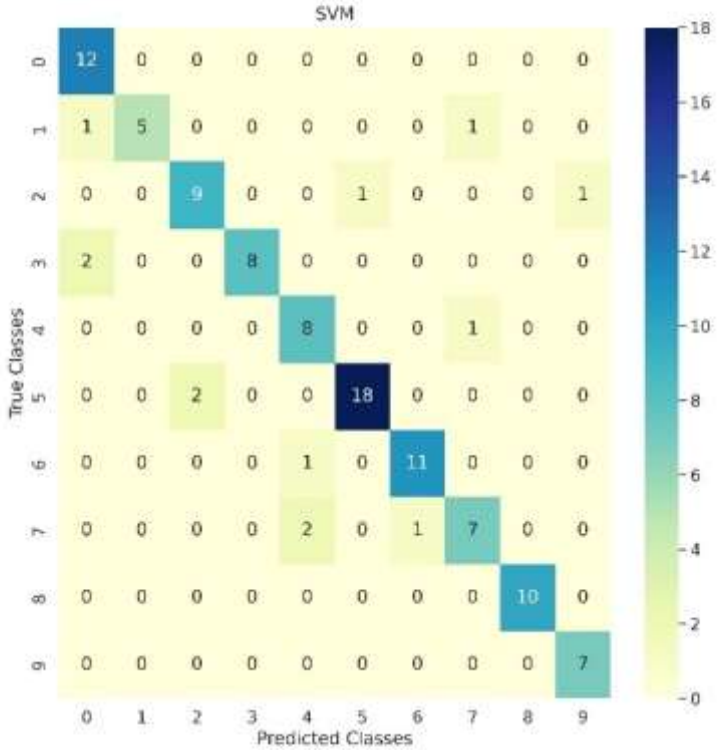
Support Vector Machines (SVMs)
A SVM, as based on supporting vectors (SVMs)[10], serves as an algorithm for conducting classifications and regression analyses within supervised learning models, utilizing associated algorithms. The efficacy of SVMs has been demonstrated and validated, particularly in high-dimensional spaces and scenarios with more dimensions than samples. Notably, SVMs are memory-efficient and quick in decision-making, as they rely on a subset of the training data known as support vectors. In the present study, an SVM classifier was deployed to distinguish between healthy and diseased plant varieties, showcasing its utility in supervised learning for both classification and regression tasks.

Performance Matrix
Confusion matrix [15] is a mathematical formula that is used to calculate how many classes are confused with each other in the system. It is named thus because it makes it easy to discover if the system is mislabeling one class as the other.
The basic layout of a confusion matrix is as follows:

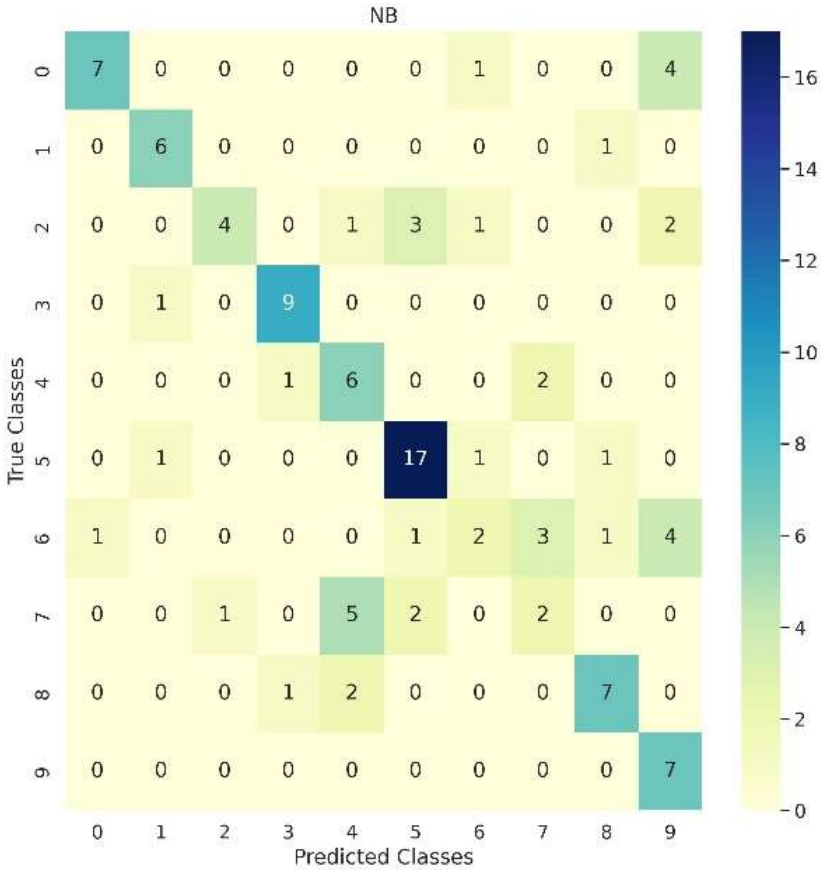
	Actual Positive	Actual Negative
Predicted Positive	True Positive	False Positive
Predicted Negative	False Negative	True Negative

3. Results and Discussion
This experiment is conducted on 10 different species of medicinal plants. The following are the performance matrix that is used for determining the efficiency of models.
- Precision
- Recall
- F1-Score
- Confusion Matrix

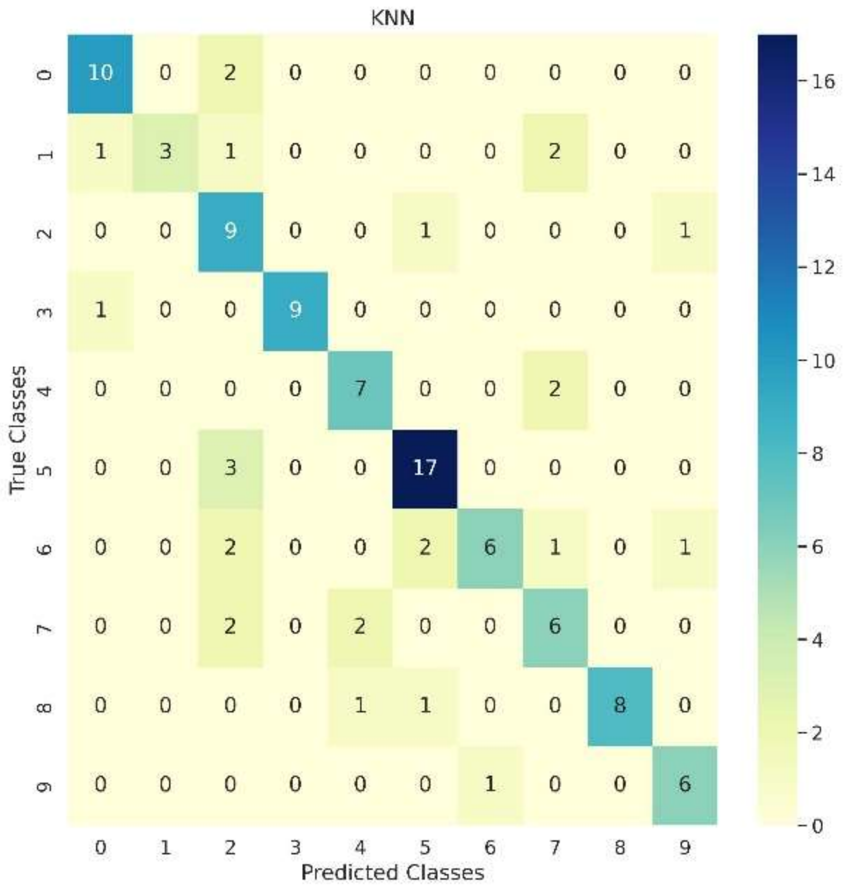
Various configurations were tested for each model, and the results were recorded to evaluate performance.


SVM
True Classes
Predicted Classes

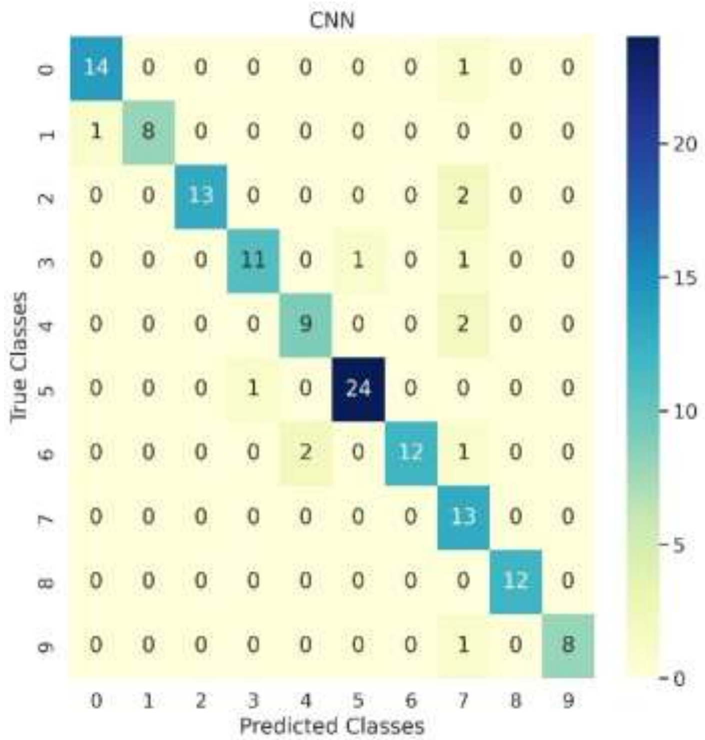
(i)


NB
True Classes
Predicted Classes

(ii)


KNN
True Classes
Predicted Classes

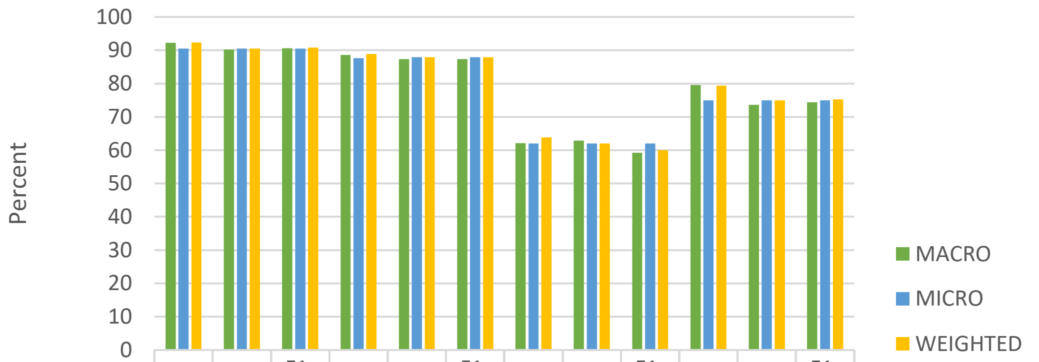
(iii)


CNN
True Classes
Predicted Classes

(iv)

Fig 3: Confusion matrix for 10 class of Medicinal plants (i) SVM (ii) Naïve's Bay (iii) KNN (iv) CNN

Analyzing the figure depicted in figure 2 confusion matrix for SVM, NB, KNN, and CNN, it is evident that CNN outperforms the other models. In the case of CNN, the true positive values for every class are consistently the highest, indicating a higher level of accuracy. Furthermore, the recall, precision and F1 score metrics find support in the insights gleaned from the confusion matrix.



	CNN			SVM			Naïve's Bay			KNN		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
MACRO	92,24	90,27	90,6	88,62	87,38	87,39	62,08	62,87	59,23	79,6	73,65	74,38
MICRO	90,51	90,51	90,51	87,62	87,96	87,96	62,03	62,03	62,03	75	75	75
WEIGHTED	92,36	90,51	90,76	88,89	87,96	87,92	63,8	62,03	60,03	79,34	75	75,25

Fig 5: Bar Chart for all the models.

The analysis of the performance metrics reveals that the CNN (Convolutional Neural Network) model outperformed the other models in terms of precision, recall, and F1-score, achieving scores ranging from an impressive 90.51% (micro) to 92.36% (weighted). These high scores indicate that the CNN model excels in accurately classifying medicinal plant leaves while maintaining a fine balance between precision and recall. Similarly, the SVM model also demonstrated strong performance with good precision, recall, and F1-score, ranging from 87.62% (micro) to 88.89% (weighted). Although slightly lower than the CNN model, these scores affirm the SVM model's capability for accurate leaf detection. Also the Naïve Bayes model achieved comparatively lower precision, recall, and F1-score, ranging from 62.03% (micro) to 63.80% (weighted), suggesting that it may not be as accurate in identifying medicinal plant leaves as the CNN and SVM models. The KNN (K-Nearest Neighbors) model, while performing better than Naïve Bayes, still fell short of the CNN and SVM models, with scores ranging from 75.00% (micro) to 79.34% (weighted)

Overall, both the CNN and SVM models demonstrate superior accuracy in medicinal plant leaf detection, making them the top performers among the models considered. The Naïve Bayes and KNN models, while still viable, exhibit lower accuracy scores in this task. Therefore, for the classification of the dataset, macro techniques could be considered, but it's evident that the CNN and SVM models stand out as the most reliable choices for accurate leaf detection.

ANALYSIS

Among the many powerful machine learning algorithms available today, CNNs, SVMs, K-NNs, and Naive Bayes all have a wide range of applications across a wide range of industries. SVMs are efficient for high-dimensional data and small datasets and CNNs are well suited for image and video recognition tasks, but CNNs are well-suited for image recognition tasks as well. SVM excel in handling high-dimensional data and small datasets. They are well-suited for scenarios with limited data. Conversely, KNNs are suitable for tasks with small datasets and straightforward decision boundaries, while Naive Bayes is effective in managing high-dimensional data and large datasets. As a result of the variety of tasks that leaf detection covers in the fields of botany, agriculture, and forestry, leaf detection plays a very substantial role in each of these fields. Several techniques have been used over the past few decades to detect leaves accurately, such as CNNs and Random Forests, with the goal of improving detection accuracy and reducing the amount of human error. Recent years have witnessed the successful implementation of a number of leaf detection systems using these techniques, and these systems have shown promising results when used to identify tree species based on their leaf images.

The evaluation of medicinal plant leaf detection was conducted using four machine learning models: CNN, SVM, Naïve Bayes, and KNN. The precision, recall, and F1-score metrics were reported for macro, micro, and weighted averaging. Additionally, the accuracy and loss function [16] were plotted against the number of epochs, revealing that the testing and training accuracy improved as the number of epochs increased, and the loss function decreased gradually with increasing epochs.

Upon examining the confusion matrix, a clear pattern emerges: the Convolutional Neural Network (CNN) consistently demonstrates superior leaf classification accuracy when compared to its counterparts, namely, the Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naive Bayes. This observation underscores the CNN's remarkable proficiency in distinguishing between different types of leaves, showcasing its effectiveness as a classification tool.
The CNN's ability to discern intricate leaf features and patterns, combined with its capacity to automatically learn relevant representations from the data, contributes significantly to its high accuracy. In contrast, while SVM, KNN, and Naive Bayes offer respectable performance, they appear to lag behind the CNN in terms of precision and accuracy in leaf classification tasks.

This outcome emphasizes the CNN's pivotal role in achieving precise and reliable leaf classification results, a testament to its robustness and suitability for complex image-based recognition tasks. It underscores the significance of choosing the appropriate algorithm for specific applications, where the CNN proves to be the optimal choice for accurate leaf classification.

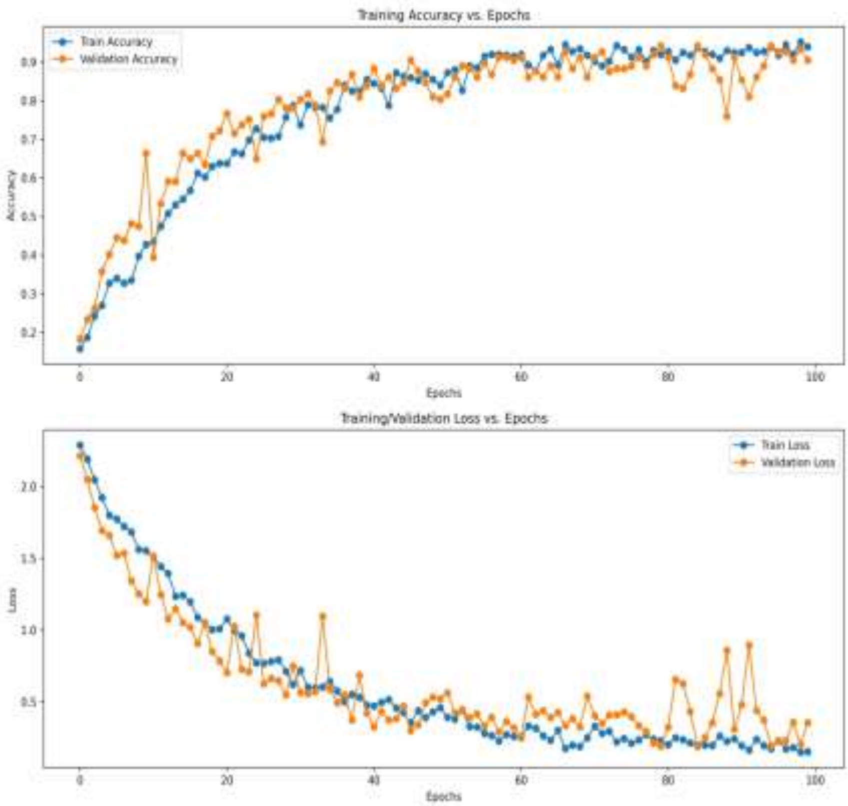


Fig 6: CNN 5-layer model training/validation accuracy vs epochs and training/validation loss vs epochs

Analyzing the graphical representation (Fig 5) depicting the connection between training accuracy and epochs, it becomes evident that with an increase in the number of epochs, the models' efficiency likewise improves. However, at a certain point, it plateaus and remains relatively constant. Furthermore, when examining the graph illustrating the relationship between training and validation loss with respect to epochs, it is evident that as the number of epochs increases, the loss decreases.

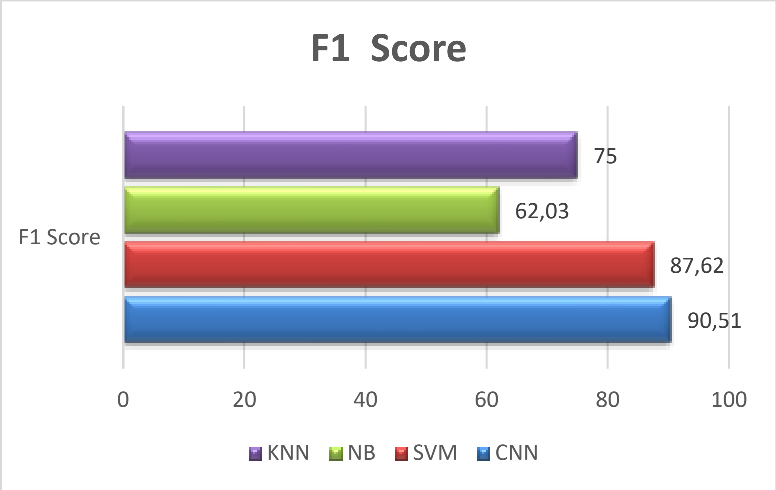


Fig 7: Micro F1 Score of all models.

Overall, from the figure 6, the highlights that the Convolutional Neural Network (CNN) outperforms other machine learning algorithms in leaf detection using the Medicinal Plant Leaf dataset, achieving the highest F1 score of 90.51%. SVM also performed well with an F1 score of 87.62%, showcasing its effectiveness in this application. Naive Bayes and KNN showed respectable but comparatively lower F1 scores, suggesting that they may not be the best choices for this specific task. Researchers and practitioners may consider these results when selecting an algorithm for leaf detection in medicinal plant species.

4. Conclusion

This paper revolves around the utilization of machine learning-based classification techniques for the purpose of feature extraction and stratification of medicinal plant species. In the area of ML and DL, several algorithms come into play, including SVMs, KNNs, CNNs and Naive Bayes. The focal point of this study is the application of these algorithms on a meticulously curated dataset of segmented leaves, which serves as the foundation for both training and testing.

Remarkably, the CNN algorithm shines as it attains an impressive F1 score of 90.5, demonstrating its prowess in medicinal plant species detection. A hierarchical layer structure further improves the effectiveness for the detection process. The research project encompasses a comprehensive delineation of phases, commencing with the utilization of a dataset comprising ten distinct classes for the preparation and testing of the prediction model. This is seamlessly followed by a crucial preprocessing phase, integral to refining the performance of the CNN algorithm.

The important outcomes of this endeavor are the capacity to identify and classify leaf diseases through advanced image processing strategies. The meticulous construction of the model, coupled with the precise implementation steps, ensures its computational accuracy, surpassing existing systems. This proposed system is poised to deliver swift and accurate results, contingent upon the input image, thereby empowering farmers to avert potential losses. In essence, this research presents an innovative and practical solution that promises to revolutionize the field of medicinal plant species detection and disease prevention.

References

1. Rana, K. L., Kour, D., Kaur, T., Devi, R., Yadav, A. N., Yadav, N., ... & Saxena, A. K. (2020). Endophytic microbes: biodiversity, plant growth-promoting mechanisms and potential applications for agricultural sustainability. *Antonie Van Leeuwenhoek*, *113*, 1075-1107.
2. Manojkumar P., Surya C. M., and Varun P. Gopi, “Identification of Ayurvedic Medicinal Plants by Image Processing of Leaf Samples”, 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pp 978- 1- 5386-1931-5.
3. Lobet, G., Draye, X., & Périlleux, C. (2013). An online database for plant image analysis software tools. *Plant methods*, *9*, 1-8.
4. Pushpanathan, K., Hanafi, M., Mashohor, S., & Fazlil Ilahi, W. F. (2021). Machine learning in medicinal plants recognition: a review. *Artificial Intelligence Review*, *54*(1), 305-327.
5. Blanz, V., Schölkopf, B., Bülthoff, H. C. B. V. V., Burges, C., Vapnik, V., & Vetter, T. (1996). Comparison of view-based object recognition algorithms using realistic 3D models. In *Artificial Neural Networks—ICANN 96: 1996 International Conference Bochum, Germany, July 16–19, 1996 Proceedings 6* (pp. 251-256). Springer Berlin Heidelberg.
6. S, Roopashree; J, Anitha (2020), “Medicinal Leaf Dataset”, Mendeley Data, V1, doi: 10.17632/nnytj2v3n5.1
7. Taha, B., & Shoufan, A. (2019). Machine learning-based drone detection and classification: State-of-the-art in research. *IEEE access*, *7*, 138669-138682.
8. Specht, D. F. (1990). Probabilistic neural networks. *Neural networks*, *3*(1), 109-118.
9. Hossain, E., Hossain, M. F., & Rahaman, M. A. (2019, February). A color and texture based approach for the detection and classification of plant leaf disease using KNN classifier. In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp. 1-6). IEEE.
10. Yang, X., & Guo, T. (2017). Machine learning in plant disease research. *March*, *31*, 1.
11. Anubha Pearline, S., Sathiesh Kumar, V., & Harini, S. (2019). A study on plant recognition using conventional image processing and deep learning approaches. *Journal of Intelligent & Fuzzy Systems*, *36*(3), 1997-2004.
12. Du, C., & Gao, S. (2017). Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network. IEEE access, 5, 15750-15761.
13. McCann, S., & Lowe, D. G. (2012, June). Local naive bayes nearest neighbor for image classification. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3650-3656). IEEE.
14. Webb, G. I., Keogh, E., & Miikkulainen, R. (2010). Naïve Bayes. *Encyclopedia of machine learning*, *15*(1), 713-714.
15. Hung, Y. H., Huang, M. L., & Chen, K. S. (2003). Service quality evaluation by service quality performance matrix. *Total Quality Management & Business Excellence*, *14*(1), 79-89.
16. Ma, C., Wu, L., & Ying, L. (2022). The multiscale structure of neural network loss functions: The effect on optimization and origin. *arXiv preprint arXiv:2204.11326*.
17. Bineet Kumar Gupta, Neeraj Kumar and Abhishek Bajpai, Integrating User Centred Design Process and Ergonomics of Human Computer Interaction to Assist Data Quality Issues “, International Journal of Advanced Studies in Computers, Science and Engineering (IJASCSE), ISSN 2278 7917, Volume 6 (10), pp 14-19, October 2017. http://www.ijascse.org/ijascse-volume-6-issue-10 (UGC Journal No: 45435)
18. N. Kumar, B.K. Gupta, V. Sharma, V. Dixit, and S.K. Singh, “E-Health: Stipulation of mobile phone technology in adolescent Diabetic Patient Care” Paediatric Diabetes, Jon Wily & Sons A/A, Volume 14(18), October 2013, ISSN P: 1399-543X, O: 1399-5448, p-90, 90. (DOI: 10.1111/pedi.12075). http://onlinelibrary.wiley.com/doi/10.1111/pedi.12075/abstract. (IF 3.17 WoS/ UGC Journal Number: 48012).
19. Rajeev Kumar, Bineet Kumar Gupta et. al., “Evaluating the Impact of Malware Analysis Techniques for Securing Web Applications through a Decision-Making Framework under Fuzzy Environment” in International Journal of Intelligent Engineering and Systems, Vol.13, No.6, July 2020, pp. 94-109, DOI: 10.22266/ijies2020.1231.09, http://www.inass.org/2020/2020123109.pdf. (IF 1.29)