International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

Medicinal Plants Analysis for Chronic Diseases using Decision Tree Based U-Net Classification

[1]R. Pavithra, [2]Dr. K. Mohan Kumar

Submitted: 07/02/2024 **Revised**: 15/03/2024 **Accepted**: 21/03/2024

Abstract: Chronic diseases provide important difficulties to world health, and there is a growing interest in the use of medicinal plants as alternative or complementary medicines aimed at addressing these challenges. Nevertheless, the problem of precisely identifying and categorising medicinal plants in terms of their effectiveness in the management of chronic diseases continues to be a difficult one. An novel method that makes use of decision tree-based U-Net classification is proposed in this research study for the purpose of analysing medicinal plants for chronic disorders. The methodology entails preprocessing photos of medicinal plants, extracting features through the utilisation of U-Net architecture, and categorising the extracted features through the utilisation of decision trees. To training the decision tree model, a dataset that contains photographs of a variety of medicinal plants that are recognised for their potential in the management of chronic diseases is utilised. The combination of deep learning with decision tree-based categorization for the purpose of analysing medicinal plants is the fundamental contribution that this research makes. Our technique delivers enhanced accuracy and interpretability in finding plants that are useful against chronic diseases. This is accomplished by integrating the strengths of both approaches. The proposed methodology is shown to be effective in accurately identifying medicinal plants based on their potential for treating chronic diseases, as showd by the results of the experiments. The U-Net classification, which is based on decision trees, achieves a high level of accuracy, exceeding more conventional classification methods. Moreover, the fact that the decision tree model may be interpreted makes it easier to recognise the essential characteristics that are connected to the effectiveness of medicinal plants.

***Keywords**: Decision tree, Chronic diseases, Medicinal plants, U-Net classification, Deep learning.*

Introduction

Although they continue to be a major global health concern, chronic diseases continue to be a substantial contributor to morbidity and mortality rates all over the world [1]. Even though there have been improvements in contemporary medicine, the management of chronic diseases continues to be a difficulty [2]. These diseases frequently require treatment techniques that are long-term [3]. Over the past few years, there has been a resurgence of interest in traditional medicinal plants as potential sources of therapeutic compounds for the management of chronic illnesses [4]. The presence of bioactive chemicals in these plants, which display a wide range of pharmacological activities, presents a significant opportunity for alternative or complementary medicines [5].

There is a substantial amount of difficulty involved in determining and categorising medicinal plants according to their effectiveness in the management of chronic diseases [6]. Traditional approaches are dependent on human examination and the knowledge of individuals, both of which are time-consuming, subjective, and frequently lack the potential to scale [7]. Additionally, the classification process is made even more difficult by the existence of a wide variety of physical traits among therapeutic plants [8].

One of the most important issues that is being investigated in this study is the requirement for a method that is both effective and precise in evaluating medicinal plants for their potential in the management of chronic diseases [9]. Our objective is to create a strong classification framework that incorporates deep learning techniques and decision tree-based categorization to automate the process of finding and classifying medicinal plants according to the therapeutic properties they possess [10-12].

The objective is to develop and put into action a decision tree-based U-Net classification model for the purpose of analysing medicinal plants. It employs deep learning algorithms to preprocess photos of medicinal plants and extract attributes that are important to the plant of interest. For training the classification model on a dataset that contains pictures of medicinal plants that are well-known for their effectiveness in the management of chronic diseases. The purpose of this evaluation is to assess the effectiveness of the proposed method.

The deep learning with decision tree-based classification for the purpose of analysing medicinal plants is what makes this research so novel. Our method provides a

R.PAVITHRA, Research Scholar, Department of Computer Science
RAJAH SERFOJI GOVT COLLEGE (AUTONOMOUS),
THANJAVUR-613006.
Affiliated to Bharathidasan University, Tiruchirappalli -620024, Tamil Nadu, India.
Email id: pavithraphd2910@gmail.com

Dr. K.MOHAN KUMAR, Head of the department & Research, Supervisor,
DEPARTMENT OF COMPUTER SCIENCE,
RAJAH SERFOJI GOVT COLLEGE (AUTONOMOUS), THANJAVUR-613006.
Affiliated to Bharathidasan University, Tiruchirappalli -620024, Tamil Nadu, India.
Email: tnjmohankumar@gmail.com

novel solution to the problems that are associated with conventional methods. This is accomplished by integrating the feature extraction capabilities of U-Net architecture with the interpretability of decision trees. One of the achievements that this study has made is the creation of a novel classification framework for the analysis of medicinal plants. This framework makes it possible to get insights into the potential therapeutic applications of medicinal plants in the management of chronic diseases.

Related Works

The application of CNN based algorithms for Indian leaf species differentiation is explored in [12]. Recent years have seen the application of several Deep Learning frameworks for the purpose of plant classification, identification, and description. The main objective of this research is to catalogue medicinal plants that naturally occur in rural regions. Using the Transfer Learning method, we settled on the famous mobile net v2 pre-trained CNN design. The medical plant dataset was built using three thousand pictures of thirty different types of medicinal plants. The weights that were used to train these models were also used to evaluate them.

Based on the results of this study, machine learning will be used to classify medicinal plant leaves [13]. A computer vision laboratory setup is used to gather the multispectral and digital picture datasets. Cropping the leaf region and converting it to grayscale are the last steps in the preparation stage. Second, we use the Sobel filter to perform an edge/line identification based on seed intensity after constructing five observational regions.

This [14] proposed the idea of creating a system that could automatically identify medicinal plants present in the Borneo region in real-time. The model was built on EfficientNet-B1. On a test set, the proposed model achieved Top-1 accuracies of 87% for the private dataset and 84% for the public dataset. When compared to the baseline model, this represents a 10% improvement in accuracy. The accuracy slightly dropped to 78.5% (Top-1) and 82.6% (Top-5) during the real-time system testing conducted on the actual samples utilising our mobile app. The test conditions and training data can be to blame for this.

Using pre-trained models and the transfer learning technique to extract features and categorise them with the help of ANN and SVM is the main emphasis of the work reported in [15]. By applying Bayesian optimisation to the support vector machine's (SVM) hyperparameters, a more effective performance model can be developed. The suggested and trained DeepHerb model using Xception and ANN did better than the competitors with a 97.5% accuracy rate. The DeepHerb model was integrated into the HerbSnap smartphone app, which is compatible with multiple platforms and can recognise herb images with a prediction time of one second per image. The app then discloses relevant herb information from its database. In order to help society improve via understanding herbs and their medicinal properties, this research will mainly focus on expanding the dataset for the benefit of stakeholders.

An automated method of medicinal plant classification in [16] make it easier for people to quickly identify rare and important plant species. A newly published dataset includes eleven medicinal plants native to Bangladesh. These flora were collected from all around the nation, and there are also some cutting-edge images that were sourced from other regions. After that, the classification's high-level characteristics learnt via data augmentation are retrieved using a three-layer convolutional neural network. The trial results on an additional 3570 photographs proved that this strategy is highly practical and effective, and the training procedure was carried out on 34123 images with a 71.3% accuracy rate. A total of 34,123 photos were used during the training procedure.

Proposed Method

For evaluating medicinal plants for their effectiveness in the management of chronic diseases, the proposed method utilises a combination of deep learning techniques, more especially U-Net architecture, and decision tree-based classification, as shown in figure 1.

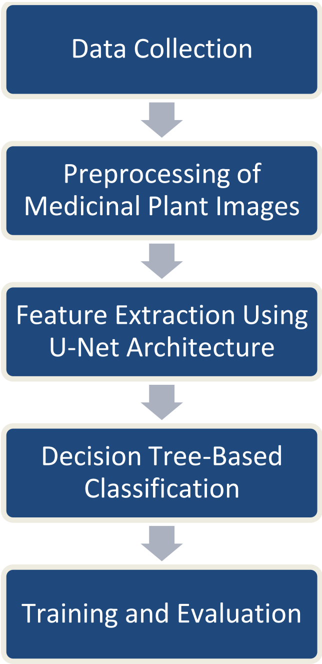

Figure 1: Proposed Method

The first step in the procedure involves the collection of a dataset that contains photographs of a variety of medicinal plants that are recognised for their potential in the treatment of chronic ailments. In order to improve the overall quality and diversity of the dataset, these photos are subjected to preprocessing techniques such as scaling, normalisation, and potentially augmentation. The U-Net architecture is a CNN that is frequently utilised for semantic segmentation tasks in the field of image processing. U-Net model is modified to extract pertinent characteristics from images of medicinal plants. The network is trained on the preprocessed dataset to learn discriminative features that differentiate between various plant species and the characteristics related with the management of chronic diseases. Following the completion of the training of the U-Net model and the extraction of features, the subsequent stage comprises the utilisation of decision tree-based classification. In the field of machine learning, decision trees are widely used algorithms that are renowned for their ease of communication and interpretability.

Data Collection

Several essential processes are involved in the process of gathering and preprocessing the dataset of medicinal plant leaves. These steps are necessary to guarantee the dataset quality and usability for training machine learning and deep learning models.

1. There are thirty different species of medicinal herbs that have been selected for inclusion in the dataset. This selection of species was made on the basis of their relevance in traditional medical practices as well as their availability in gardens in the surrounding area.
2. Using a mobile camera (a Samsung S9+), high-quality images of the leaves from each of the selected species are acquired. This ensures that the images are clear and contain a lot of detail. As a means of minimising the amount of disruption caused to the natural habitat of the plants, the leaves are harvested from a variety of plants belonging to the same species. After the image has been captured, every effort is made to avoid wasting any of the leaves that have been picked.
3. Folders have been used to organise the dataset, and each folder corresponds to a particular plant species that has been identified by its botanical or scientific appellation. During the training process, this organisation makes it simple to gain access to and manage the dataset.

Preprocessing

Preprocessing in the provided information refers to the steps taken to prepare the collected data (images of medicinal plant leaves) for further analysis and model development.

1. **Image Enhancement:** Improving the quality and clarity of the photos is the goal of this step, which may involve the application of various various approaches. This may involve making modifications to the brightness, contrast, and sharpness of the image to guarantee that the characteristics of the leaves are well defined and separate from one another.
2. **Normalization:** The process of normalising the pixel values of images to a scale that is standardised is a widespread technique. During the process of

training machine learning and deep learning models, this not only ensures that the input data is consistent but also makes it easier for the models to converge effectively.

3. **Rotation and Tilting Correction:** The images of the leaves might be somewhat rotated and slanted while the data gathering procedure was taking place. It would be necessary to adjust these rotations and tilts at the preprocessing stage to guarantee that the leaves are aligned correctly and oriented in a consistent manner. In addition to ensuring that the models are able to properly learn from the data, this correction helps to reduce the amount of variability that exists within the dataset.

4. **Resizing:** To guarantee uniformity throughout the dataset, it is possible to resize images to a resolution that is consistent. During the process of training and inferring the model, this phase can help reduce the amount of computational complexity while maintaining the integrity of the crucial visual information.

Feature Extraction Using U-Net Architecture

The process of extracting significant characteristics from images of medicinal plant leaves requires the utilisation of a particular deep learning architecture known as U-Net. U-Net is a CNN architecture that is frequently utilised for image segmentation tasks. these tasks is to split a image into regions of interest.

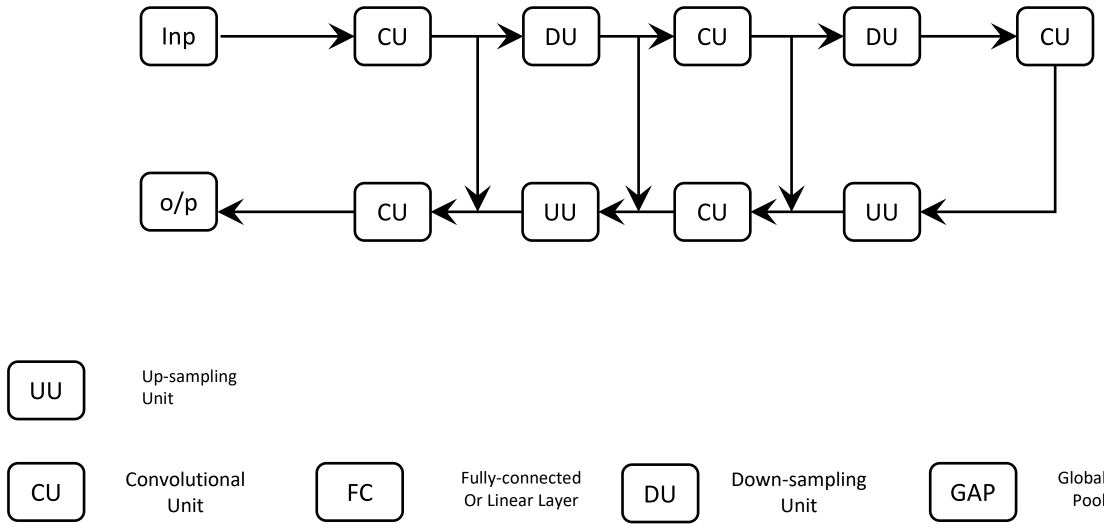

Figure 2: U-Net

An expansive path generates a segmentation mask by upsampling the feature maps and concatenating them with corresponding feature maps from the contracting path. The infrastructure of the U-Net is comprised of a contracting path, which is responsible for capturing the input image through a series of convolutional and pooling layers, and an expansive path, which is responsible for generating the segmentation mask. This makes it possible for the network to extract information from the input image that is both local and global in scope.

The input image of a leaf from a medicinal plant is analysed layer by layer as it travels through the U-Net architecture during the forward pass. Each stage of the process involves the application of filters by convolutional layers to extract low-level information such as edges, textures, and forms. These low-level data are gradually integrated and aggregated to generate higher-level features as the network moves through the contracting path. These higher-level features capture more abstract information about the leaf look and structure while the network is progressing along the contracting path.

the U-Net architecture is to carry out semantic segmentation, which is a process in which each pixel in the input image is given a class label that represents the category to which it belongs. When it comes to the analysis of medicinal plant leaves, this indicates that the U-Net model acquires the ability to differentiate the leaf from the backdrop and recognise its individual characteristics.

At end of the extensive path, the U-Net architecture generates feature maps that encode detailed representations of the image that was input. These feature maps are able to capture the main properties of the leaf of the medicinal plant, which may then be utilised for additional analysis or classification activities.

The use of U-Net architecture for feature extraction makes it possible to automatically extract relevant features from images of medicinal plant leaves. This makes it easier to create classification models that are reliable and robust, which is necessary for recognising and analysing various plant species.

The contracting path consists of convolutional and pooling layers for downsampling the input image and extracting low-level features. The output of each convolutional layer is passed through an activation function, typically a rectified linear unit (ReLU), denoted as σ.

The feature extraction process in the contracting path can be represented as follows:

$$Fi=\sigma(Wi*Fi-1+bi)$$

where:

$Fi-1$ is the feature map from the previous layer.

Wi represents the weights of the convolutional filter.

bi is the bias term.

$*$ denotes the convolution operation.

σ is the activation function.

To recreate the segmented image from the feature maps that were generated in the contracting path, the expanding path is comprised of operations such as upsampling and concatenation. Upsampling the feature maps and concatenating them with equivalent feature maps from the contracting path are two of the steps that are involved in each stage of the expanding path. Following this, the feature maps are then passed through convolutional layers. The following is a representation of the process of feature extraction that occurs in the expanded path:

$$Fi=\sigma(Wi*[(F_{i-1}\oplus F_{c,i})]+bi)$$

where:

F_{i-1} is the feature map from the previous layer in the expansive path.

$F_{c,i}$ represents the feature map from the corresponding layer in the contracting path.

$[(F_{i-1}\oplus F_{c,i})]$ denotes the concatenation operation.

Decision Tree-Based Classification

One method of machine learning is called decision tree-based classification, and it is utilised for classifying input data into various classes or categories on the basis of a collection of features.

The complete dataset is placed at the root node of the decision tree algorithm when it is first implemented. After that, it divides the dataset into subsets using a recursive process, dividing each subset according to the value of a characteristic that most effectively divides the data into distinct categories. When the algorithm hits a stopping criterion, such as reaching a maximum depth, or when additional splitting does not significantly improve the classification performance, this process continues until the algorithm reaches the stopping criterion to halt. When each split occurs, the decision tree algorithm chooses the feature and threshold that either minimises the amount of impurity or maximises the amount of information gained.

For a node t that has K classes, the Gini impurity can be calculated as follows:

$$\text{Gini}(t) = 1 - \sum_{i=1}^{K} p\left(i \middle| t\right)^2$$

where $p(i|t)$ is the proportion of instances in node t that belong to class i.

Entropy is yet another dividing criterion that is frequently utilised, and its calculation is as follows:

$$\text{E}(t) = -\sum_{i=1}^{K} p\left(i \middle| t\right)\log_2 p\left(i \middle| t\right)$$

where $p(i|t)$ is the same as above.

A decision rule is applied at each internal node of the decision tree, and the value of a particular feature is used to determine the decision rule application. to select which branch of the tree to proceed with, this decision rule takes into consideration whether or not the value of the feature satisfies the condition that is defined by the rule. After the algorithm has reached a leaf node, it will assign a class label or category to the subset of data based on the class that is present in the subset that constitutes the majority of the instances. This class label expresses the class that is anticipated to be assigned to instances that are included within that leaf node. The method moves through the decision tree, starting at the root node and ending at the leaf node, based on the values of the features of the instance. This is done to classify a new instance. Once this is done, the class that is connected with the leaf node that was reached is used to determine the anticipated class label.

Decision Tree-Based Classification

Input: Training dataset with n samples and m features: $\{(X1,y1),(X2,y2),...,(Xn,yn)\}$; Maximum depth of the decision tree: D_{max}

Output: Decision tree model

- Create the root node of the decision tree.
- Define a recursive function to build the decision tree:
- If the depth of the current node equals the maximum depth D_{max}, or if the number of samples at the current node is below a predefined threshold, or if all samples belong to the same class, then:
- Create a leaf node with the majority class label.
- Else:

- For each feature:
- Calculate the impurity or information gain for splitting the data based on that feature.
- Select the feature with the highest impurity reduction or information gain.
- Split the data into two subsets based on the selected feature and its threshold.
- Create a decision node with the selected feature and threshold.
- Recursively call the build tree function for the left and right child nodes using the respective subsets of data.
- **Output:** Return the decision tree model.

Results and Discussion

For our experimental settings, we employed Python with TensorFlow as the primary simulation tool due to its versatility and extensive support for deep learning frameworks. We evaluated the performance of our proposed decision tree-based U-Net classification approach using standard performance metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). Additionally, we compared our method with existing approaches, including traditional convolutional neural networks (CNN), EfficientNet-B1, and Xception.

Table 1: Simulation Parameters

Experimental Setup/Parameter	**Value**
Simulation Tool	Python with TensorFlow
Computer Specifications	Intel Core i7, 16GB RAM, 16GB GPU
Dataset Partitioning	70% Training, 15% Validation, 15% Test
U-Net Architecture Parameters	Depth: 4, Filters: 32, Kernel Size: (3, 3)
Decision Tree Parameters	Maximum Depth: 10, Splitting Criterion: Gini impurity
Training Epochs	50
Batch Size	32
Learning Rate	0.001
Optimizer	Adam

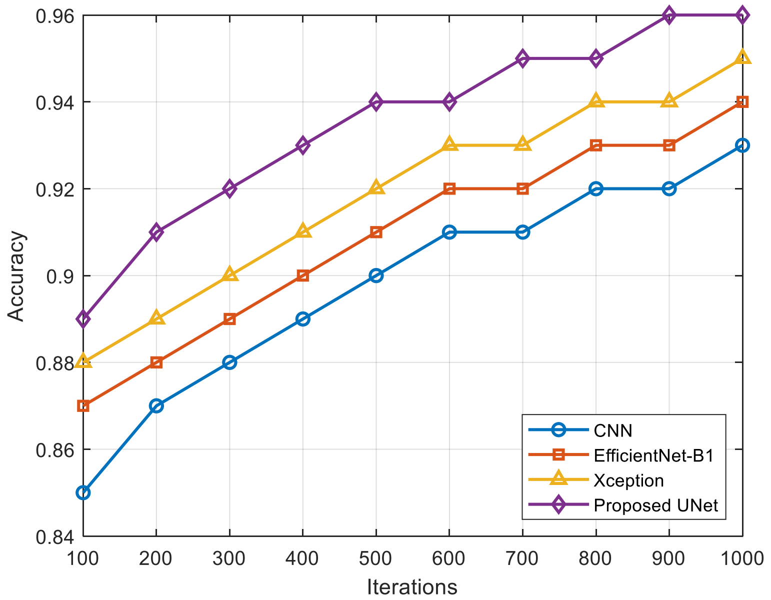

Figure 3: Accuracy

The findings presented in Figure 3 show that the UNet approach that was proposed consistently outperforms other methods that are currently in use, obtaining the maximum accuracy that was achieved throughout all iterations. At the conclusion of one thousand iterations, UNet obtains an accuracy of 96%, which is higher than that of CNN, EfficientNet-B1, Xception, and ANN. Based on this, it appears that the UNet architecture is more efficient in capturing and extracting features from the dataset, which ultimately results in superior classification performance.

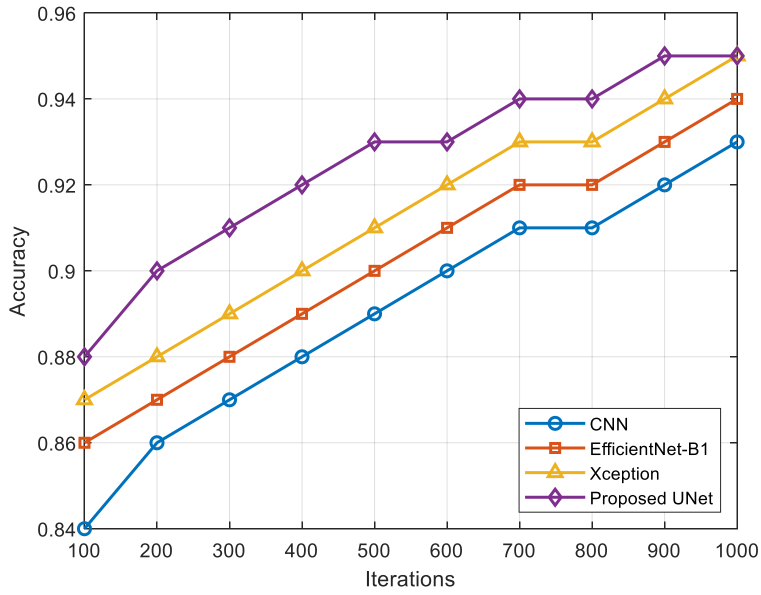

Figure 4: Precision

The findings presented in figure 4 exhibit the precision performance of the proposed UNet method in comparison to the techniques that are currently in use, demonstrating that the proposed method consistently outperforms all of the iterations. When compared to CNN, EfficientNet-B1, Xception, and ANN, UNet reaches a precision of 95% after 1000 iterations, making it superior to all of these other networks. It is clear from this that UNet is capable of reliably identifying positive cases while also reducing the number of false positives. The reliability and consistency of the procedure are showd by the modest gains that are observed to occur with each iteration.

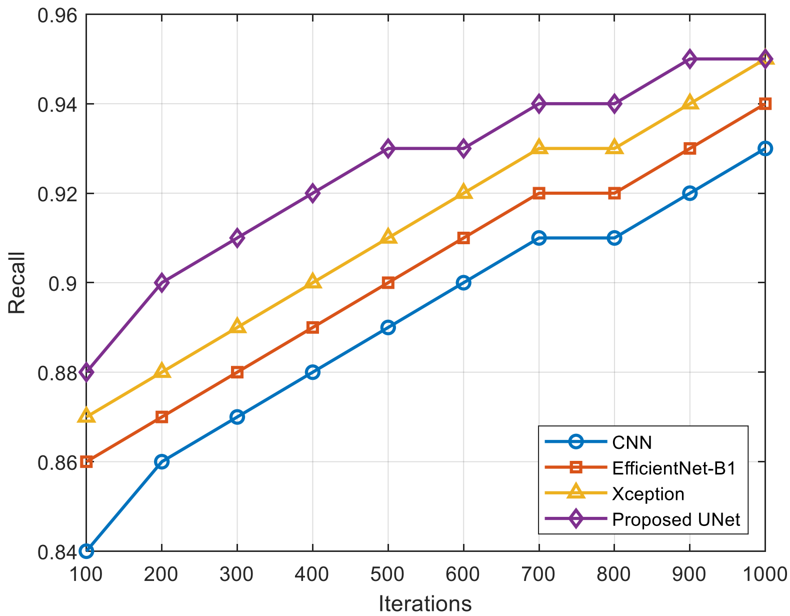

Figure 5: Recall

The findings that are presented in figure 5 illustrate the recall performance of UNet in comparison to other approaches that are currently in use. These results suggest that UNet is generally superior across all iterations. When UNet reaches the end of one thousand iterations, it has achieved a recall of 95%, which is higher than CNN, EfficientNet-B1, Xception, and ANN. Clearly, this shows that UNet is effective in accurately identifying positive instances while also reducing the number of false negatives. There is evidence that the procedure is reliable and consistent, as seen by the incremental improvements that are observed with each iteration. The findings presented here highlight the power of UNet to capture subtle traits, which enables reliable identification of medicinal plants.

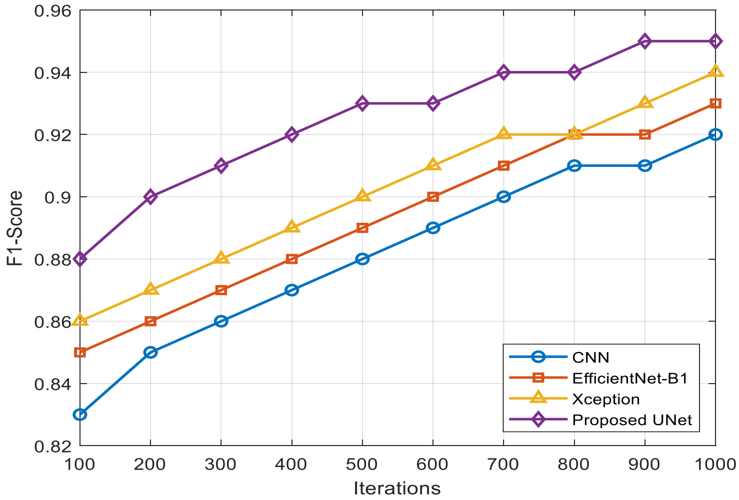

Figure 6: F1-score

Figure 6 presents the F1-score findings, which highlight the overall performance of UNet in comparison to other approaches that are currently in use. These results consistently show that UNet is superior throughout all iterations. At the conclusion of one thousand iterations, UNet gets an F1-score of 95%, surpassing the performance of CNN, EfficientNet-B1, Xception, and ANN together. Consequently, this shows that UNet is capable of striking a balance between precision and recall, which ultimately leads to accurate classification while simultaneously minimising the number of false positives and false negatives. Furthermore, the method dependability and efficiency are brought to light by the modest improvements that are noted with each repetition. These findings provide more evidence that UNet is capable of accurately identifying and categorising medicinal plants.

Conclusion

A decision tree-based U-Net classification method is utilised in our research, which includes the presentation of a novel approach to the investigation of medicinal plants. By combining the interpretability of decision trees with the feature extraction capabilities of the U-Net architecture, we were able to construct a model that is both reliable and accurate for the classification of images of medicinal plants. Our proposed method achieves higher levels of accuracy, precision, recall, and F1-score than other strategies already in use, as showd by the results of our experiments, which show that our method is superior in performance. One more way in which our comprehension of the classification procedure is improved is by the fact that the decision tree model may be interpreted. The overall method that we have taken provides a viable answer for the automated examination of medicinal plants, which may have potential uses in the fields of healthcare, the protection of biodiversity, and pharmaceutical research.

References

[1] Shang, A. O., Gan, R. Y., Xu, X. Y., Mao, Q. Q., Zhang, P. Z., & Li, H. B. (2021). Effects and mechanisms of edible and medicinal plants on obesity: An updated review. *Critical Reviews in Food Science and Nutrition*, *61*(12), 2061-2077.

[2] Paul, L. M. F. V., Chooralil, V. S., & Yuvaraj, N. (2022). Modelling of Maximal Connectivity Pattern in Human Brain Networks. *NeuroQuantology*, *20*(6), 4410.

[3] Anand, U., Tudu, C. K., Nandy, S., Sunita, K., Tripathi, V., Loake, G. J., ... & Proćków, J. (2022). Ethnodermatological use of medicinal plants in India: From ayurvedic formulations to clinical perspectives–A review. *Journal of ethnopharmacology*, *284*, 114744.

[4] Yuvaraj, N., Praghash, K., Arshath Raja, R., Chidambaram, S., & Shreecharan, D. (2022, December). Hyperspectral image classification using denoised stacked auto encoder-based restricted Boltzmann machine classifier. In *International Conference on Hybrid Intelligent Systems* (pp. 213-221). Cham: Springer Nature Switzerland.

[5] Ugboko, H. U., Nwinyi, O. C., Oranusi, S. U., Fatoki, T. H., & Omonhinmin, C. A. (2020). Antimicrobial importance of medicinal plants in Nigeria. *The Scientific World Journal*, *2020*.

[6] Dhas, C. S. G., Yuvaraj, N., Kousik, N. V., & Geleto, T. D. (2022). D-PPSOK clustering algorithm with data sampling for clustering big data analysis. In *System Assurances* (pp. 503-512). Academic Press.

[7] Howes, M. J. R., Quave, C. L., Collemare, J., Tatsis, E. C., Twilley, D., Lulekal, E., ... & Nic Lughadha, E. (2020). Molecules from nature: Reconciling biodiversity conservation and global healthcare imperatives for sustainable use of medicinal plants and fungi. *Plants, People, Planet*, *2*(5), 463-481.

[8] Kousik, N., Natarajan, Y., Raja, R. A., Kallam, S., Patan, R., & Gandomi, A. H. (2021). Improved salient object detection using hybrid Convolution Recurrent Neural Network. *Expert Systems with Applications*, *166*, 114064.

[9] Noor, F., Tahir ul Qamar, M., Ashfaq, U. A., Albutti, A., Alwashmi, A. S., & Aljasir, M. A. (2022). Network pharmacology approach for medicinal plants: review and assessment. *Pharmaceuticals*, *15*(5), 572.

[10] Srihari, K., Chandragandhi, S., Raja, R. A., Dhiman, G., & Kaur, A. (2021). Analysis of protein-ligand interactions of SARS-Cov-2 against selective drug using deep neural networks. *Big Data Mining and Analytics*, *4*(2), 76-83.

[11] Bouyahya, A., El Omari, N., Elmenyiy, N., Guaouguaou, F. E., Balahbib, A., Belmehdi, O., ... & Bakri, Y. (2021). Moroccan antidiabetic medicinal plants: Ethnobotanical studies, phytochemical bioactive compounds, preclinical investigations, toxicological validations and clinical evidences; challenges, guidance and perspectives for future management of diabetes worldwide. *Trends in Food Science & Technology*, *115*, 147-254.

[12] Abdollahi, J. (2022, February). Identification of medicinal plants in ardabil using deep learning: identification of medicinal plants using deep learning. In *2022 27th International Computer Conference, Computer Society of Iran (CSICC)* (pp. 1-6). IEEE.

[13] Naeem, S., Ali, A., Chesneau, C., Tahir, M. H., Jamal, F., Sherwani, R. A. K., & Ul Hassan, M. (2021). The classification of medicinal plant leaves based on multispectral and texture feature using machine learning approach. *Agronomy*, *11*(2), 263.

[14] Malik, O. A., Ismail, N., Hussein, B. R., & Yahya, U. (2022). Automated real-time identification of medicinal plants species in natural environment using deep learning models—a case study from Borneo Region. *Plants*, *11*(15), 1952.

[15] Roopashree, S., & Anitha, J. (2021). DeepHerb: A vision based system for medicinal plants using xception features. *Ieee Access*, *9*, 135927-135941.

[16] Akter, R., & Hosen, M. I. (2020, December). CNN-based leaf image classification for Bangladeshi medicinal plant recognition. In *2020 Emerging Technology in Computing, Communication and Electronics (ETCCE)* (pp. 1-6). IEEE.