Explainable Approach for Species Identification using LIME

1st Mihir Nikam
Department of Information Technology
Sardar Patel Institute of Technology
Mumbai, India
mihir.nikam@spit.ac.in

2nd Ameya Ranade
Department of Information Technology
Sardar Patel Institute of Technology
Mumbai, India
ameya.ranade@spit.ac.in

3rd Rushil Patel
Department of Information Technology
Sardar Patel Institute of Technology
Mumbai, India
rushil.patel@spit.ac.in

4th Prachi Dalvi
Department of Computer Engineering
Sardar Patel Institute of Technology
Mumbai, India
prachi_dalvi@spit.ac.in

5th Dr. Aarti Karande
Department of Master in Computer Application
Sardar Patel Institute of Technology
Mumbai, India
aartimkarande@spit.ac.in

***Abstract*—Plant identification has a wide array of applications in the fields of agronomy and the discovery of natural and medicinal products. This research aims to explore various deep learning techniques like InceptionV3, Xpection, and ResNet to identify plants. Highly accurate machine learning models generally lack explainability and interpretability. Neural networks are usually opaque systems and thus a direct understanding of the interpretations becomes necessary. We aim to remove this ambiguity of how the model reaches its conclusion by introducing Explainable AI (XAI) techniques. Explainability aims to break such barriers by diminishing the lack of transparency in Artificial Intelligence and Machine Learning models, thus taking a step toward making AI reliable. In this paper, Convolutional Neural Network has been used to identify Vietnamese medicinal plant images based on the characteristics of the leaves, stems and other parts of the plant. Upon identification, our paper also elaborates on how each model predicts which part of the image helps the CNN model to make a prediction by integrating Explainable AI (XAI) using the Lime package. Through this research, we generated images using LIME package which highlight pixels that determine the result of our plant identification process.**

***Index Terms*—Medicinal Plants Classification, Explainable Artificial Intelligence (XAI), Convolutional Neural Network, Deep Neural Network, InceptionV3, Xpection, ResNet**

I. Introduction

Plants are a significant source of both food and medicine. Most people in the world used plant-based medicines to take care of their medical requirements. To distinguish and identify various plant species, a person must have a basic knowledge of plants. This promotes the pharmaceutical sector while safeguarding the environment. Consequently, sustainable development also makes it possible to boost farm yields.

Using the distinctive uniqueness of leaves, the majority of researchers tried to derive structural features from leaves, flowers, and bark for classifying plants. The model used in this work to identify the plants is Convolutional Neural Networks (CNN), which has revolutionized the way we interpret various features of an image. To read the images, we use three image classification models, namely, InceptionV3, Xpecption, and ResNet50 on the VNPlant-200 dataset [7].

To understand the interpretations of the various models implemented, we integrate LIME (Local Interpretable Model-agnostic Explanations) with our image classification models. LIME identifies the area of an image with the highest correlation with an inference label for image classification tasks. Most machine learning models are black boxes, that is to say, that we are unable to comprehend the internal working of the model. Instead, we only rely on the model's accuracy to gauge the model's prediction [13]. This paper primarily aims to resolve the ambiguity of machine learning models by using Explainable AI (XAI) to identify medicinal plant specifies. The research also helps in highlighting which pixels of the plant image contribute to the model's prediction. Also, we do a comparative analysis of how different state-of-the-art models perform on a similar dataset.

This paper has been structured in the following format. Section 2 presents the literature survey which highlights previously done work. Then, sections 3 and 4 describe the methodology used and the results obtained, respectively. Section 5 summarizes the objectives completed followed by possible modifications and improvements that can be done in the future.

II. Literature Survey

A. Literature survey on plant identification

Detecting medicinal plants has been considered an important task, as they provided cures for various ailments and diseases. As machine learning revolutionized, new techniques were invented to analyze various kinds of plants, and identify the part of the plant that can provide medicinal benefits. Leaves, roots, flowers, and stems were some of the factors that identified the plant to be medicinal. The authors of [9] reviewed various machine learning algorithms to detect if the

978-1-6654-9291-1/22/$31.00 ©2022 IEEE

plant is of medicinal purpose or not, using the leaf of the plant. Different image processing techniques were used to remove the noise and extract essential features of the leaf. However, after realizing that the leaf is not the only medicinal feature, the dataset needed expansion. Hence, VNPlant-200 [7] was curated in order to take more features such as bark, roots, and flowers into consideration. This dataset was used successfully for [8], as they were successfully able to implement different algorithms such as InceptionV3, Resnet50, Densenet121, and Xception. The CNN model used was able to predict what class the plant belonged to among the 200 given classes. [1] employs MobileNet_v2 on a medicinal plant leaf database consisting of 3000 images with 30 images present in each class. Using MobileNet_v2, a validation accuracy of 83% was achieved. However, it is critical to explain how any machine learning model uses a particular set of features to identify plants. In the prior research work conducted in plant identification, models have predicted results accurately. However, humans are unable to understand the predictions and thus explainability is needed. We aim to bridge this knowledge gap by integrating Explainable Artificial Intelligence (XAI) which assists in highlighting the pixels that determine the model's prediction.

B. Literature survey on Explainable AI - Lime

Explainability has become critical in the era of growing data dependence and computational power. Explainable Artificial Intelligence (XAI) is a thriving area of research that has applications in healthcare, knowledge-based systems, science, finance, and many other such disciplines as shown in [5]. Insights from XAI become vital when the working of deep learning models becomes too complex for human understanding. XAI highlights the features that determine the result making it easier for the human to understand how each feature influences the decision made by our machine learning model. XAI employs methods and procedures that provide explanatory information to simplify the results of the artificial intelligence systems as depicted in [3].

LIME is an upcoming library that helps to explain why the model produced a particular result and also elaborates on the parameters that were taken into consideration. Explainability is largely divided into two approaches, local explanations and global explanations. As demonstrated in [3], local explanations try to interpret individual predictions at the level of the single statistical unit, whereas global explanations characterize the model as a whole in terms of which explanatory variables most strongly influence its predictions for all statistical units. Previously, LIME has been used in [6] for the explanation of Service Supply Chain Forecasting. It's essential to understand the cause behind a prediction so as to increase confidence in the model. LIME is a very powerful tool that explains the reason for a particular prediction. It has been observed that a model is able to predict something accurately, but the reason why it does so was a mystery before the concept of Explainable AI (XAI) came into play. This creates transparency in the model prediction which allows you to improve an unscrupulous model and identify why an existing model should be relied on.

III. Methodology

A. Organization and image classification of dataset

Numerous datasets have been made available, including the Swedish leaves dataset, which has 75 leaves per species, and the Flavia dataset, which contains 1,907 photos of leaves from 32 different species. Our work is based on the VNPlant-200 dataset [7]. This dataset comprises 20,000 images totaling 200 different annotated medicinal plants from Vietnam (VNPlant-200). This dataset is available in two sizes, 256 256 and 512 512 pixels. 12,000 of the photos are used for the training set, while the remaining 12,000 are used for the testing set.

The dataset is comparatively noisier than other available datasets since it contains elements like dirt, tree bark, flowers, and many leaves appearing in one image. We have split the data into two factions for training and testing purposes. The size of the training data is 80 percent of the whole dataset and the remaining 20 percent is the test dataset.

B. Convolutional Neural Networks (CNN)

It is a type of feed-forward ANN that has convolutional layers(one or many) and has several use cases such as image processing, image classification, image segmentation and so much more. CNN uses the image and provides us with resources to extract different features from the image to give meaningful results. It takes raw image pixels, trains the model, and then takes out different features on its own for better understanding and improved classification. The architecture of CNN comprises one input layer, one output layer, many hidden layers, and more than a hundred thousand parameters, which allows them to learn complicated objects and patterns. It makes use of convolution and pooling processes to make smaller samples of a given image, after which, the activation function is applied. The resultant image is commensurate with the size of the input image. Pooling can be classified into two types. The first one is, Max-pooling and the latter is, Average pooling. The activation function usually used in the case of CNN is the ReLU function, which outputs the same value as the input, if the input is more than zero, else it results in zero.

In the image dataset used here, there are about 200 classes of plants, which are classified using different features. These features are automatically decided by the neural network. It's unknown how these features are decided and we aim to shed light on that.

C. InceptionV3

InceptionV3 is a convolutional deep neural network that is extensively used for image classification purposes. InceptionV1 [11] is the initial version and InceptionV3 [12] is the newer improved version of it. In 2014, it was introduced by the name of GoogLeNet. Inception models are based on the concept of having various filters of diverse sizes to avoid overfitting the data and improve performance. Inception models are best known for being wider rather than

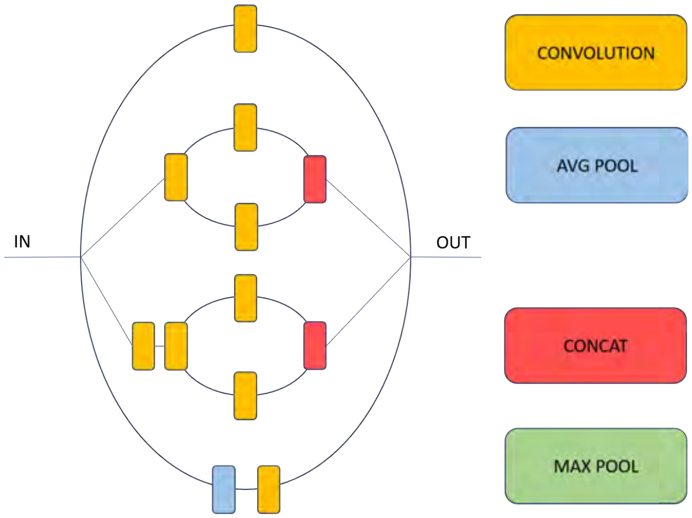

Fig. 1: InceptionV3 Intermediate Block

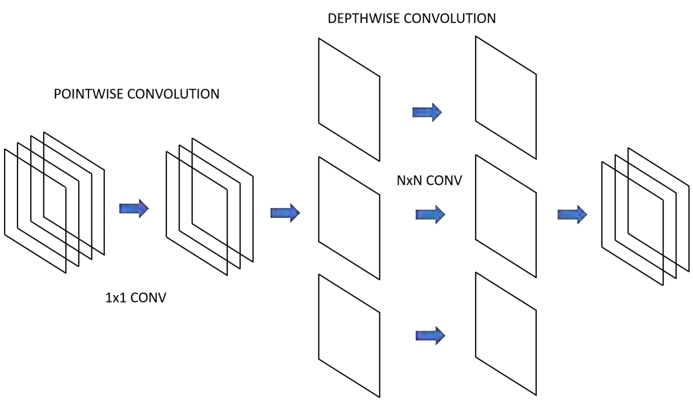

Fig. 2: Xception Convolution

deeper as it has many parallel layers within their architecture. The parallel modules consist of a 1x1 convolution block(computationally light), a 3x3 convolution block, a 5x5 convolution block(computationally heavy), and 3x3 max pooling blocks. We in our research have used the InceptionV3 model specifically as it is deeper than its predecessors V1 and V2 prototypes but the speed remains the same. V3 has much higher efficiency and at the same time, it is computationally less heavy. Major improvizations in V3 are the use of auxiliary classifiers, the use of multiple parallel Convolutions, coherent grid size reduction, and asymmetric Convolutions. Finally, after integration of all these into the Inception architecture a 42-layered model InceptionV3 was built in 2015.

D. Xception

Xception is a convolutional deep neural network presented by Francois Chollet [2], a Google software developer who built Keras. Xception uses the ideology of Inception and is built to be an extreme implementation of it. Xception models take its principle architecture from the Inception model and make certain key architectural changes. Inception uses 1x1, 3x3, and 5x5 convolutional blocks to compress the input image

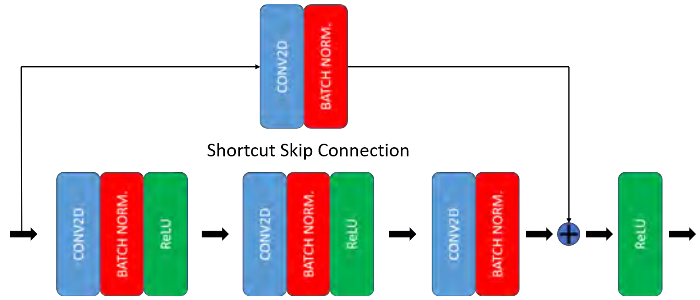

Fig. 3: ResNet Intermediate Block

and then apply the filters to it. On the contrary, Xception inverts this by applying the features first and then compressing the imputed image. This is the new improvisation made to its structure and redesigning depthwise separable convolution. This Xception model also utilizes the residual skip connection concept used in that of ResNet and has resulted in performing great when trained on the ImageNet dataset consisting of over 15 million labeled input images. Xcpetion model has 71 layer deep architecture and comprises millions of trainable parameters making it very efficient for image classification.

E. ResNet50

ResNet is another state of art model architecture built for very high-performance image classification tasks [4]. We have used the ResNet50 for our research purposes available in the Keras library. It has around 23 million parameters that are trainable. Original ResNet had 34 layered architecture but the newer improved versions of it are ResNet50, ResNet101, and ResNet152. It comprises residual networks with skip connections. The shortcut connection made directly between the two adjacent layers was first introduced by the ResNet Architecture. This improvisation was introduced in order to tackle the vanishing gradient issue which is commonly faced in most neural nets which are deep. These shortcut connections regulate the information flow from upper layers to deeper layers without forgetting the vital feature along its forward propagation. ResNet layers have sub-modules of convolutional, ReLu, max pooling, and Batch Normalization. ResNet has excellent benchmarks on some of the vast datasets like ImageNet when compared to its contemporaries like AlexNet, VGG models, and others.

F. Lime Package

Machine learning models are usually treated as “black boxes”, that is to say, the internal functioning of these models is complex and thus cannot be apprehended by us. The more complex the model, the higher the accuracy, and lower the interpretability. It is becoming increasingly crucial to understand these models to gain human trust and acceptance.

LIME is short for **L**ocal **I**nterpretable **M**odel-agnostic **E**xplanations. It currently works on black-box classifier models that categorize images, tabular data, and texts. As the name suggests, LIME consists of three fundamental concepts [14].

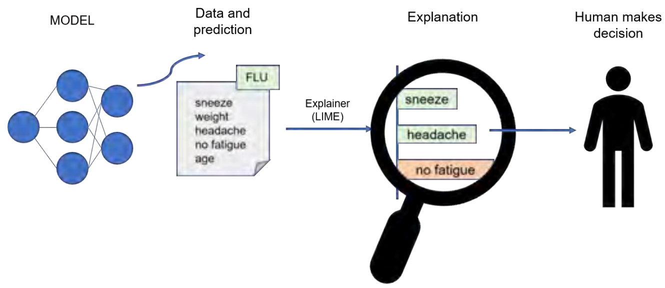

Fig. 4: Internal working of LIME package

- **Model-agnostic** means that LIME is model-independent and not restricted by the type of model used.
- When a model is said to be **Interpretable**, it means that LIME offers a solution to help you comprehend why it performs the way it does.
- **Local** refers to the fact that LIME attempts to approximate the local linear operation of your black-box model in order to find an interpretation for it.

LIME presents textual and visual evidence that provides an understanding of the association between different features of an image.

Fig.5 depicts the methodology used in detail. The Vietnam (VNPlant-200) dataset is first split into training and testing data. Using deep learning techniques like InceptionV3, Xpection, and ResNet, the model is trained on the image dataset. To understand the influence of features on the predictions, we integrate the LIME package. This package explains why our black box models predict one of the 200 classes from the Vietnam (VNPlant-200) dataset. The LIME package then discerns the result given by the model and gives output in the form of the effect a particular feature has namely positive, negative, or neutral. The pixels are accordingly highlighted with the colors green (positive), red (negative), and black (neutral). XAI thus helps us understand the features that either positively or negatively influence the model's decision.

Obviously, if a doctor is given understandable explanations of the machine learning model, the doctor will be better equipped to help the patients. As seen in Fig. 4 [10], the explanation is given by a list of symptoms with relative weights associated with them. These components either support (in green) the prediction or provide evidence against (in red) it. Since a doctor has expertise in the application domain, they can either confirm (trust) or dismiss the prediction.

IV. Results and Discussions

In this research paper, we have compared the three different CNN models on the VNPlant-200 dataset [7]. InceptionV3, Xception and ResNet are all State-of-the-art (SOTA) algorithms when it comes to tasks like image classification. These DNNs vary from each other in terms of their model architecture and hence have varied performance in training. All the models were trained for 20 epochs for the same dataset. In the

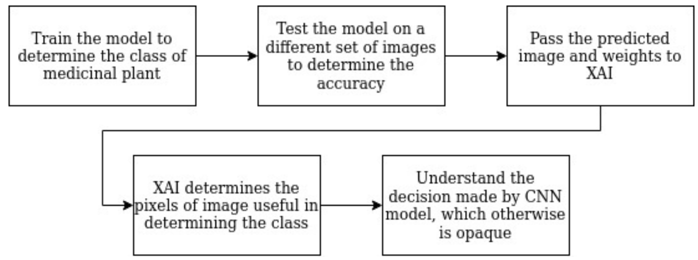

Fig. 5: Methodology

TABLE I: Comparison of model accuracy

Model	Training Accuracy	Testing Accuracy
ResNet	80	62
Xception	86	66
InceptionV3	82	71

subsequent section we have compared the model performances and their explainability with LIME.

A. *InceptionV3*

InceptionV3 is the best fit model out of all the models tried and tested. It gives a training accuracy of 82 percent on the training dataset and a validation accuracy of 71 percent on the test dataset. In the metrics graphs images, we can see the accuracy of the model steadily rises from epoch one to twenty. From the graph we can conclude that the training accuracy is almost always more than validation accuracy which suggests that the model is well trained and is not overfitting. The error in accuracy achieved on the validation data was caused somewhat by the noise in the dataset. The steady increasing slope with no apparent sharp ups or downs and the relative difference between training loss and validation loss is also less further suggestive of a good model fit.

B. *Xception*

Xception is the next best fit model after InceptionV3. Xception gives 86 percent training accuracy on the training dataset and 66 percent validation accuracy on the test dataset. Although Xception's training accuracy is higher than that of InceptionV3, the validation accuracy is lesser than that of InceptionV3. From the accuracy graph, we can infer that the Xception model performs very well while training but isn't able to keep up the performance on the test dataset. The relative difference between the training loss and validation loss is also quite noticeable but is steadily decreasing with the subsequent epochs. This model is suspected to be a little overfitting due to the above-mentioned pointers. On trying out different types of optimizations like changing the learning rate, more epochs, increasing the training dataset size, and more similar performances were observed. In the table below, we have mentioned the accuracy of the various deep learning techniques used in the identification of plants.

C. *Explainability of models using LIME*

After classification, we have used LIME (Local Interpretable Model-agnostic Explanations) to try and explain how

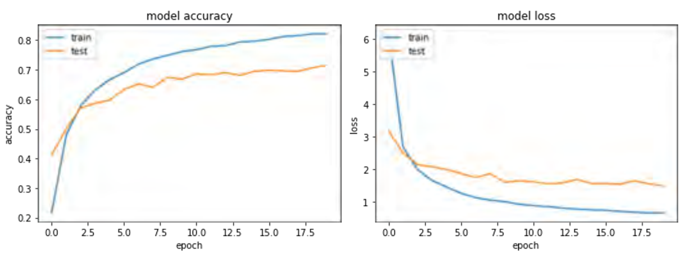

Fig. 6: Accuracy and Loss graphs Vs Epochs for InceptionV3

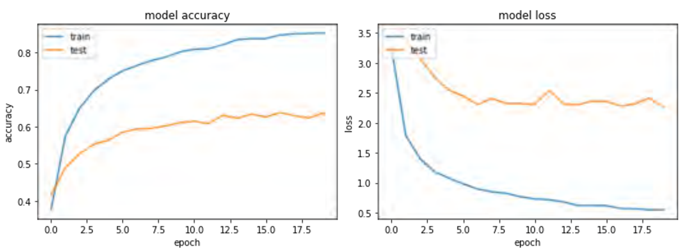

Fig. 7: Accuracy and Loss graphs Vs Epochs for Xception

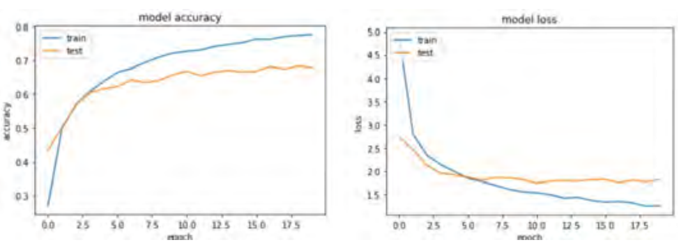

Fig. 8: Accuracy and Loss graphs Vs Epochs for ResNet50

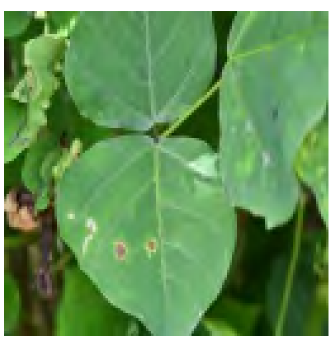

Fig. 9: Original image of 'Erythrina variegat'

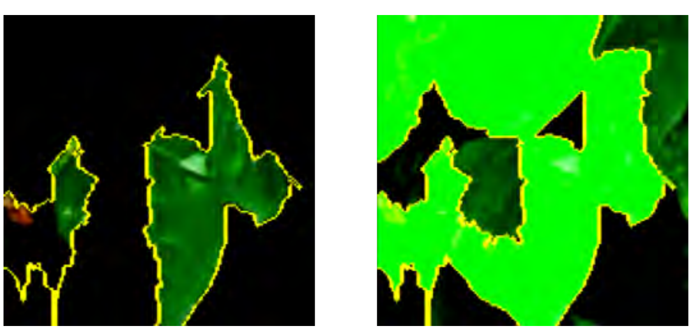

Fig. 10: Lime generated image for InceptionV3

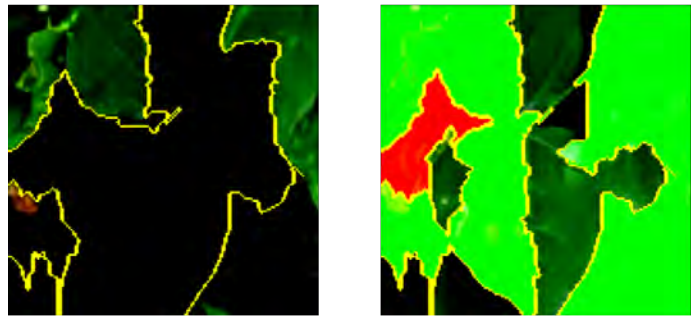

Fig. 11: Lime generated image for Xception

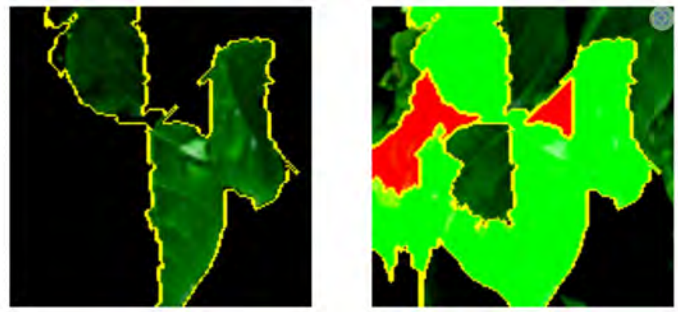

Fig. 12: Lime generated image for ResNet50

the above-used DNN models work to identify the images almost correctly and explain which part of the image triggers the algorithm to make one prediction out of various other classes. Using LIME we can identify the pixels of the image which play a more active part in classification. As in our case, different plants can be identified on the basis of its difference in leaves, stems, flowers, and other parts of the plant. So the pixels that make up the plant in the image are super-pixels which are the basis of the prediction rather than the other non-plant surrounding pixels.

We have generated these images using LIME which highlights these superpixels for InceptionV3, Xception and ResNet50 models on Fig. 9. The left images blacken out all the pixels which are less important and only the superpixels identified by the model are colored. The right side images, the part which is masked green, increase the probability of it being a particular class and the red masked pixels (if any) decrease the probability of it being that particular class. Fig. 10, Fig. 11 and Fig. 12 are LIME-generated images for the class 'Erythrina variegata' from the medicinal plant dataset. Here, it is quite suggestive that the leaf pixels act as superpixels while predicting the images, and hence suggests that the models are behaving as expected. The models are thus identifying the correct pixels which are really relevant rather than other background noise pixels.

With this, we can get insights into the working of different convolutional neural networks used for identification. All three different CNN models identify different parts of plants to predict the plant. The model which performs better is more likely to account for important parts of plants like leaves, flowers stems, or other parts of a particular plant unique to it. The models that are poor in accuracy usually tend to focus on other peripheral features of the plant.

V. Conclusion

Several ML models have been developed which predict results with good accuracy, but transparency still remains a question. There are various fields like healthcare, agriculture, and auto-driving cars where people's trust is as important. To increase the trust, and understand the credibility of the model, explainability is essential which is what has been achieved for the classification of medicinal plants.

The primary objective of this research work is to essentially apply and integrate Explainability (XAI) with our machine learning models to understand the outcomes given by different deep learning techniques such as InceptionV3, Xpecption, and ResNet50 on the VNPlant-200 dataset [7]. We introduce Explainability by using the LIME package which demonstrates the significance of each of the parameters in contributing to the final decision.

VI. Future Scope

In the future, our research work can be extended to an application that recommends plant-based remedies to users. For instance, a user can simply scan the leaves/plants in real-time and the application will identify the leaves/plants with a thorough explanation of the application's prediction, thus building user trust. Another thing that can be added is that, at the moment, the model only predicts which pixels determine the output. It should also be able to determine what of the leaf i.e. the color, the texture of the leaf, the size, and other physical attributes determine the class the plant belongs to.

References

[1] Jafar Abdollahi. Identification of medicinal plants in ardabil using deep learning. 05 2022.

[2] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1807, 2017.

[3] Paolo Giudici and Emanuela Raffinetti. Shapley-lorenz explainable artificial intelligence. *Expert Systems with Applications*, 167:114104, 2021.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. arXiv, 2015.

[5] Markus Langer, Daniel Oster, Timo Speith, Holger Hermanns, Lena Kästner, Eva Schmidt, Andreas Sesing, and Kevin Baum. What do we want from explainable artificial intelligence (xai)? - a stakeholder perspective on xai and a conceptual model guiding interdisciplinary xai research. *Artif. Intell.*, 296:103473, 2021.

[6] Haisheng Li, Wei Fan, Sheng Shi, and Qiang Chou. *A Modified LIME and Its Application to Explain Service Supply Chain Forecasting*, pages 637–644. 09 2019.

[7] Trung Nguyen and Vinh Truong. *VNPlant-200 – A Public and Large-Scale of Vietnamese Medicinal Plant Images Dataset*, pages 406–411. 01 2021.

[8] Trung Nguyen Quoc and Vinh Truong Hoang. Medicinal plant identification in the wild by using cnn. In *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 25–29, 2020.

[9] Kalananthni Pushpanathan, Marsyita Hanafi, Syamsiah Mashohor, and Wan Ilahi. Machine learning in medicinal plants recognition: a review. *Artificial Intelligence Review*, 54, 01 2021.

[10] Marco Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. pages 97–101, 02 2016.

[11] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. arXiv, 2014.

[12] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. 2015.

[13] Giulia Vilone and Luca Longo. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, 76, 05 2021.

[14] Ruben Winastwan. Interpreting image classification model with lime, Jan 2021.