JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A TRANSFER LEARNING-BASED APPROACH TO DETECT MEDICINAL PLANT IN RURAL AREA

[1]Md Sefatullah , [2]Md Shariar Kabir, [3]Mossaddak Hossain, [4]Md Parvez Mosaraf

[1,2,3,4]Student
[1,2,3,4]Department Of CSE
[1,2,3,4]Daffodil International University, Dhaka, Bangladesh

Abstract : According to FAO (Food And agriculture Organization), there are 11.1% of trees in Bangladesh. And we know that trees benefit us in many ways. After oxygen, food and shelter, the only thing we get is medicine. But there is no proper system to find medicinal plants automatically. This is a big challenge for us. If we can find medicinal plants from 11.1% of the trees in Bangladesh, it will be a great success for us. We have selected a model that is ResNet50 model of deep learning. And we have made a dataset with samples from a total of 6 medicinal plants from a garden in Ashulia, Dhaka, Bangladesh. After that our accuracy came to 99% through this algorithm of deep learning. Also, we have used some algorithm that are VGG16 with 96% accuracy and VGG19 with 97% accuracy. Here we have taken 200 pictures of each leaf of the medicinal plant from different angles. And with the help of these algorithm, we have determined different parameters including height, weight, size, and color of leaves. Then we trained in the deep learning algorithms. After training with some algorithms, our best accuracy is 99% with ResNet50 algorithm. If we can apply this model through our mobile app or web, then everyone will be able to understand by scanning the leaves of the tree through it, which medicinal plant it is.

Index Terms: Plants. Trees. Algorithm. Predictions

I. INTRODUCTION

Most of the people of Bangladesh depend on plants in different ways [15]. Medicines are one of them. Ayurvedic medicines made from trees are cheap but good in quality. At present, medicinal plants are very important in Bangladesh. Different types of medicines can be made from trees. Not just medicine, from the beauty of the skin to the hair and nails, a variety of ayurvedic products are very common in the market. There are no side effects like the effectiveness of these ayurvedic [16]. As a result of various side effects of medicines made with chemicals, many people are now leaning on Ayurveda. But due to a lack of proper identifications process, we cannot identify many medicinal plants. That is why despite having many medicinal plants around us, we are not able to find the right way. The present age is the age of science and technology. Now almost every technology is familiar with automation. Advance science has made our life much easier.

In this study, our main focus is how to find the easiest way to find medicinal plants. We used ResNet50 to classify the images. The performance of our deep learning model will help people to find the medicinal plants very easily with less time.

II. RELATED WORK

There are several pieces of research done by different authors based on medicinal plant detection. One of the foremost authoritative works within the field of plant classification has been done by [1] (Begue, Kowlessur, Singh, Mahomoodally, & Pudaruth, 2017). They got 90.1% accuracy with Random Forest Classifier. They worked with 24 medicinal plant leaves.

(M.R. & P.N., 2019) [2] Developed a leaves detection system using Convolutional Neural Network (CNN), They worked with 32 types of medicinal plant leaves and got 98.46% accuracy. There were 2400 total images. 80% were considered for training and 20% were considered for testing.

(Sivaranjani.C, Kalinathan, Amutha.R, Kathavarayan, & Kumar.K.J, 2019)[3] the used different plant features such as color and texture. There was 93.3 accuracy with 5 types of medicinal plants. Each class was 20 images. They consider 70% of data for training and the rest of the conder for the testing. They applied Logistic Regression Classifier to get high-quality accuracy.

(Wu1, et al., 2007) [4] From 5 primary geometric functions, twelve morphological functions are derived after which Principal Component Analysis (PCA)) is used for size discount in order that fewer inputs will be dispatched to a probabilistic neural network (PNN). They accomplished a mean accuracy of 90.3% with the Flavia dataset, that's they created their own creation. They trained 1800 leaves.

(Xue1,4,5, et al., 2019) [5] In the case of the version primarily based totally on spectral facts from leaves (Model B), the ANN version acquired using the averaged VIS/NIR spectra consistent with leaf as inputs confirmed 92.5%curacies for the category of all medicinal flowers used.

(SATTI, SATYA, & SHARMA, 2015) [6] They applied two classifiers. ANN classifier and KNN classifier with 1907 image data taken from the Flavia dataset. ANN performed better than the KNN classifier. Got 85.9% accuracy using the KNN classifier and 93.3% using the ANN classifier.

(Lee, Chan, Pauln, & Remagnino, 2015) [7] processed 43472 images among them 8800 were for testing and 34672 were for training. There were 44 different plant leaves and collected the dataset from the Royal Botanic Gardens, Kew, England. They got 99.5 accuracies with the CNN classifier.

(Grinblat, Uzal, Larese, & Granitto, 2016) [8] proposed a model where they processed 866 leaves samples. The dataset was collected from (Larse et al.(2014a) and it was formed by INTA. They considered classifying three different legume species: white bean, red bean, and soybean. image for soybean, 272 for red bean leaves, and 172 for white bean. They got 99.5% accuracy with the CNN classifier.

(D.V. Nazarenkoa, P.V.Kharyukb, I.V.Oseledetsc, I.A.Rodina, & O.A.Shpiguna, 2016)[9] Logistic regression, SVM, and random forest had been carried out to categorize 36 species of medicinal flowers with enough 95 % accuracy.
(Sabu, K, & Nair, 2017) [10] tried to recognize Ayurvedic Medicinal plants from leaves. They processed their image dataset using computer vision technology. The proposed system they used in their methodology is the combination of SURF and HOG featured extracted from leaf image and they used the K-NN classifier algorithm. Their total dataset is 200 images spread across 20 classes. Their accuracy is nearly 100% when they experimented with the K-NN classifier.

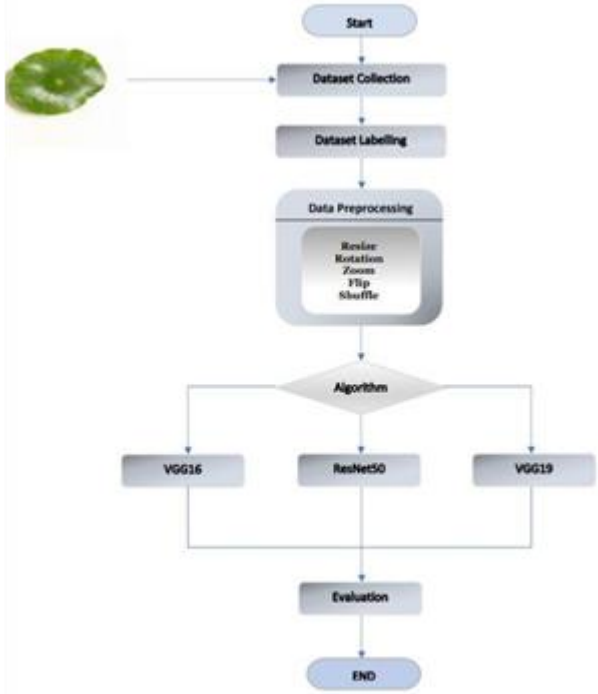
(D. Dahigaonkar1 & T. Kalyane2, 2018) [11] They wrote this article to identify medicinal plants where the main features they identified for identification were the use of leaf color, shape, and texture. They got 96.66% accuracy when they test the sample. This sample is tested using support vectors.

(Chakia, Parekha, & Bhattacharyaa, 2015) [12] proposed an ideal methodology for recognizing plant leaves using a combination of texture and shape features from leaf images. They use Gabor Filter (GF) and GLCM techniques to feature extract. To complete this work, they approach to do with Neuro-Fuzzy Controller (NFC) and Multilayer Perception (MLP) classifier. This comparison is also done by Neural Network and k-Nearest Neighbor classifier. Their total dataset of 32 classes contains 930 images of leaves. All the images of datasets they used for their thesis were captured using a smartphone in their laboratory. The overall accuracy they achieved for all 32 classes is 81.6% by NFC and 87.1% by using the MLP classifier.

Every researcher has used many techniques to identify medicinal plants. Most of the methods are mixed forms of different parameters of plants. They have used leaf color, shape, texture properties, etc. for leaf classification. (BR, C, & P, 2016)[13] they used digital image processing and machine vision technology for classified the ayurvedic or medicinal plant species from leaf images. They recognize plant species by calculating leaf factors from the imputed leaf. They got 93.75% accuracy for their proposal.

III. RESEARCH METHODOLOGY

Our main plan was to create a model through which the medicinal tree could be easily identified. During this study we followed a working method. The following Fig 1 shows the overall working method of this research.



IV. DATASET PRE-PROCESSING

We labeling the dataset manually. Then we pre – processed it by sharing, zooming and shuffling. The value of share range = 10. The value of zoom range = 0.2

V. DATASET DESCRIPTION

In our research, we included six different medicinal plants and 1583 images data collected from a garden in Ashulia, Dhaka, Bangladesh which is totally our own made dataset. Our attributes are, Curry Leaf (179 samples), Arjun Leaf (240), Mint Leaf (240 samples), Neem Leaf (320 samples), Rubble Leaf (290 samples), Marsh Pennywort Leaf (300 samples). For our proposed model we considered 80% of the whole data for training and respectively 10% and 10% for testing and validation.

Fig.2. Dataset sample of six different leaves

Fig.2 shows the sample of six different medicinal plants respectively Curry Tree, Arjun Leaf, Mint Tree, Neem Tree, Rubble Tree, Marsh Pennywort Leaf. We've captured all these photos with a smartphone, and for keeping the background white used white paper.

Tree Name	Total Sample	Training Samples	Testing Samples	Validation Samples
Curry Leaf	179	144(80%)	18(10%)	18(10%)
Arjun Leaf	240	192(80%)	24(10%)	24(10%)
Mint Leaf	244	192(80%)	24(10%)	24(10%)
Neem Leaf	320	256(80%)	32(10%)	32(10%)
Rubble Leaf	290	232(80%)	29(10%)	29(10%)
Marsh Pennywort Leaf	300	240(80%)	30(10%)	30(10%)

VI. Model Description

In this research, we applied 3 models (VGG19, VGG16, ResNet50) for checking out which model provides the best accuracy. ResNet50 gives the best accuracy and kept it as our final model to speed up the detecting process though there is no better option than it for getting the best accuracy for our proposed model. Day by day the community getting wide for the ResNet50.

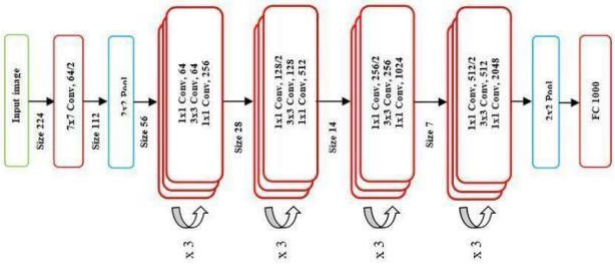


Fig.2 Architecture of ResNet50

ResNet50 is our best accurate proposed model that is 50 layers deep Neural Net. After managing data, the machine takes data as input. While the image acquired just then starts preprocessing. In this stage, occurs image zoom, flip, shuffle, resize, and rotary. After the photos have been preprocessed, they are run through the ResNet50 model. The picture is then categorized, with the primary goal being to extract the feature. After that, it uses the best model to forecast the outcome. Fig.2 shows the architecture of ResNet50 [17].

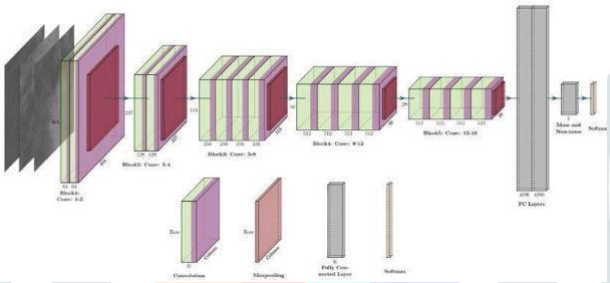


Fig.3 architecture of VGG19

VGG-19 is a 19-layer deep CNN model. The handiest preprocessing turned into eliminating every pixel's mean RGB cost that turned into calculated over the complete education set. Apply the feature to all education, validation, and check datasets to extract all the capabilities and labels. Fig 3 shows the figure of the VGG19 model consisting of a sequence convolutional layer [19].

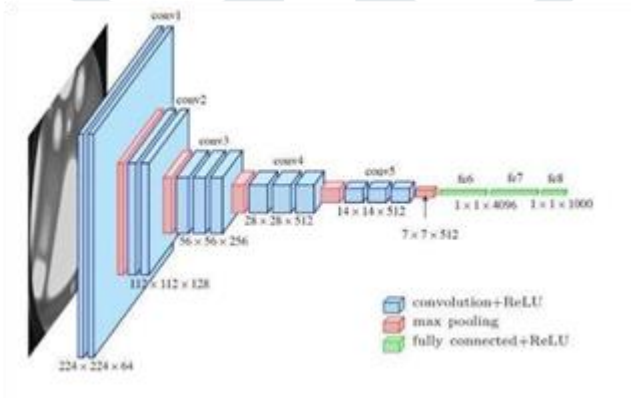


Fig.4 Architecture of VGG16

VGG16 is a convolutional neural network (CNN) architecture that won the 2014 ILSVR(ImageNet) competition. It is regarded as one of the best vision model architectures ever created. The most distinctive feature of VGG16 is that, rather than having a huge number of hyper-parameters, they concentrated on having 3x3 filter convolution layers with stride 1 and always utilized the same padding and max pool layer of 2x2 filter stride 2. Fig.4 shows the architecture of VGG16 [20].

VII. Feature Extraction

The ResNet50 model was utilized to extract the important characteristics from the dataset for this study. The Residual Network, often known as ResNet50, contains 50 layers. The ResNet50 model is a Neural Network as well. The ResNet50 model is a pre-

trained model that was trained on one million ImageNet pictures. Batches with a size of 32 are used for batch normalization. Convolutional layers (conv1, conv2 x, conv3 x, conv4 x, and conv5 x) make up ResNet50.Convolutional block sequences with average pooling make up the ResNet-50 architecture. SoftMax is the model's last layer for categorization. When an input picture is loaded, it is passed through a convolutional layer and a kernel size of (7x7), then maximum pooling is used. The ReLU function is used to introduce all of the hidden layers in this model. This model's desired size is (224,224). Class mode Categorical. Epochs number is 4, Channel is RGB. The ResNet50 model exhibits losses are 0.020, compared to 0.61 VGG19 and 0.51 for VGG16.

VIII. PERFORMANCE

In this paper, we worked with three different models to recognize medicinal plants. In all three models, we got a higher accuracy from the ResNet50 classifier. Our accuracy is 99% when we implemented the ResNet50 classifier. Among these three models, the lowest accuracy we got from VGG19, it gives us only 86% accuracy. Another one is VGG16, which also gives us a good accuracy of 96%. Our suggested classifier is ResNet50.

Model Name	Accuracy	Loss Function
VGG19	87%	0.61
VGG16	96%	0.51
ResNet50 (Suggested Approach)	99%	0.020

Table No. II shows the performance or difference those three model.

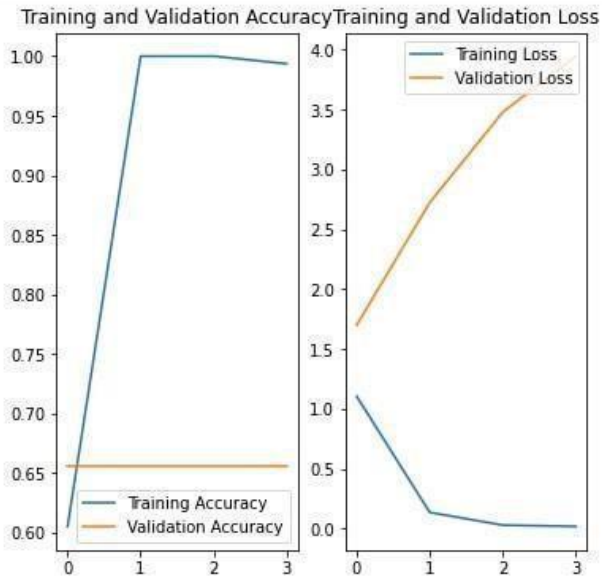
The final result was evaluated using analyzing accuracy score, Precision, Recall, and F1 score for the ResNet50 model.

TABLE III. RESNET50 MODEL ANALYZING SCORE

Classes	Arjun Leaf	Curry Leaf	March Pennyworth Leaf	Mint Leaf	Neem Leaf	Rubble Leaf
Precision	1.00	1.00	1.00	1.00	0.99	1.00
Recall	1.00	1.00	1.00	0.99	1.00	1.00
F1 score	1.00	1.00	1.00	0.99	1.00	1.00

Fig.5 Training and validation accuracy & Training and accuracy lost

For fig-5 the ResNet50 model's training and validation accuracy are depicted as a graph. Here we see two lines with two different colors such as blue and orange. The blue line stands for validation accuracy and the orange lines for training accuracy. In this graph, the blue line is moving from the range of 0.96 to 0.98. And the orange line achieved the range of 1.0 for finalizing the accuracy of training.



Here fig-5 shows the accuracy and losses function of training and validating the data.

From another point of view, the ResNet50 models show the training and validation loss using the graph. On this graph, we indicated the blue line is for training loss and the orange one for validation loss. If we noticed the figure, shows that the blue line is moving in the value from about 0.5 to 0.6, and the orange line also moves ups and down between 0.3 to 0.4. And that's the reason for evaluating the loss for the training and validating set.

IX. ERROR ANALYSIS

Occurs error If somehow the algorithm can't classify or can't detect or does mistake. We've chosen the best model to get high accuracy but have an error for the system limitation. It's occurring while classifying different types of classes, the machine gets confused for recognition.

Fig-3 shows two data conflicts. The machine is confused between Mint leaf and Neem leaf.

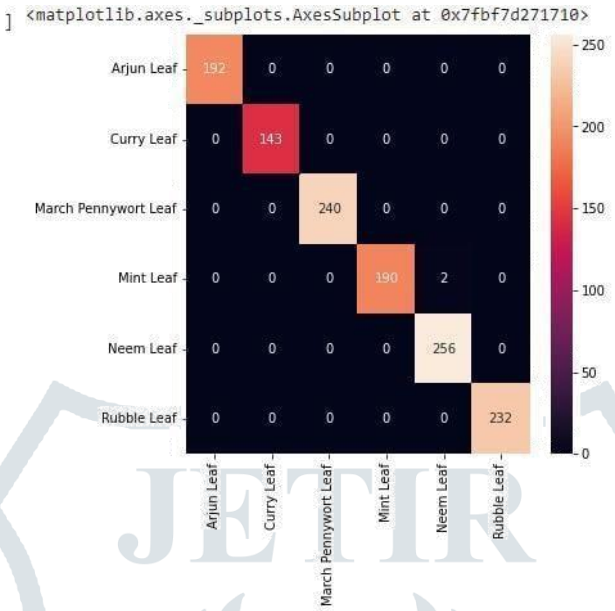


Fig.3: Confusion matrix of ResNet50 model

X. COMPARATIVE ANALYSIS

It has been noticed that several researchers have previously researched different kinds of medicinal plants recognition using various models. They use various types of algorithms and models to achieve their best accuracy. In our paper, we used only 6 types of classes and each class contains various images that were taken using a smartphone. Table No-IV shows the comparison among some paper that was related to medicinal plants detection topic.

TABLE IV. COMPARISON

Research Paper	Name Of Classes	Size Of Data	Experiment on Different Model	Best Model	Accuracy
Adams Begue, Fawzi Mahomoodally, Upasana Singh & Sameerchand Pudaruth	24	720	Random Forest, Multilayer Perceptron Neural Network, Support Vector Machine (SVM), Naïve Bayes, k-Nearest Neighbor	Random Forest	90.1%
Jyotismita Chakia , Ranjan Parekha, and Samar Bhattacharyaa	31	930	NFC multilayer perceptron (MLP)	NFC	97.6%

Tejas D. Dahigaonkar1, Rasika T. Kalyane2	33	2048	Support Vector Machine (SVM), K-nearest neighbor	Support Vector Machine (SVM), K-nearest neighbor	99.66%
Amala Sabu, Sreekumar K, Rahul R Nair	20	200	K-Nearest Neighbor, Support Vector Machine (SVM), ANN	K-Nearest Neighbor	100%

XI. CONCLUSION

After ending the research, we adopted resnet50 (a deep learning-based algorithm) as the main algorithm. We had a total of 6 classes. We trained all classes through deep learning and got 99% accuracy by using the ResNet50 model. There were 1563 pictures of the leaves of our total medicinal trees. Our selected model has given us the best accuracy. Also, our model gives outstanding averages in various assessment measures like precision, recall, and F1. So, we will implement our model through a mobile app. With the help of this app, everyone will be able to find medicinal plants very easily from now on and in a very short time. In order to find it without any system, many people used to use this medicine without knowing the wrong tree. But our app is to save time Besides, it will play a very important role in finding the right tree.

Image	
Scientific name	*Terminalia arjuna*
English name	Arjun Tree
Medicinal uses	It can reduce various heart diseases and also helps reduce high blood pressure.
Image	
Scientific name	*Murraya koenigii*
English name	Curry Leaves
Medicinal uses	Curry leaves helps heart from infections and enliven hair and skin.
Image	

Image	
Scientific name	*Mentha*
English name	Mint Leaf
Medicinal uses	It is one kind of famous herbal remedy, calming stress and anxiety, easing queasy stomachs and helps to restful sleep.

Image	
Scientific name	*Azadirachta indica*
English name	Neem Leaf
Medicinal uses	Its helps in several aspects of health such as controlling blood sugar management, as well as its also help for our hair, skin, teeth, kidney and liver.

Image	
Scientific name	*Agave americana*
English name	American Life Plant
Medicinal uses	Its mainly use for remedy any type of respiratory condition, and arthritis.

XII. REFERENCES

[1] Begue, A., Kowlessur, V., Singh, U., Mahomoodally, F., & Pudaruth, S. (2017). Automatic Recognition of Medicinal Plants using Machine Learning Techniques. International Journal of Advanced Computer Science and Applications, 166-175.

[2] M.R., D., & P.N., P. (2019). AyurLeaf: A Deep Learning Approach for. IEEE, 321-325.

[3] Sivaranjani.C, Kalinathan, L., Amutha.R, Kathavarayan, R. S., & Kumar.K.J, J. (2019). RealTime Identification of. IEEE.

[4] Wu1, S. G., Bao2, F. S., Xu3, E. Y., Wang4, Y.-X., Chang5, Y.-F., & Xiang4, Q.-L. (2007). A Leaf Recognition Algorithm for Plant. IEEE, 11-16.

[5] Xue1,4,5, J., Fuentes2, S., Poblete-Echeverria3, C., Viejo2, C. G., Tongson2, E., Du6, H., & Su1,4,5*, B. (2019). Automated Chinese medicinal plants classification based on machine. 123-131.

[6] SATTI, 9V., SATYA, A., & SHARMA, S. (2015). AN AUTOMATIC LEAF RECOGNITION SYSTEM FOR PLANT IDENTIFICATION USING MACHINE VISION TECHNOLOGY. International Journal of Engineering Science and Technology (IJEST), 874-879.

[7] Lee, S. H., Chan, C. S., Pauln, P., & Remagnino, P. (2015). DEEP-

PLANT: PLANT IDENTIFICATION WITH CONVOLUTIONAL NEURAL NETWORKS. ICIP, 452-456.

[8] Grinblat, G. L., Uzal, L. C., Larese,, M. G., & Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. Elsevier, 418-424.

[9] D.V. Nazarenkoa, P.V.Kharyukb, I.V.Oseledetsc, I.A.Rodina, & O.A.Shpiguna. (2016). MachinelearningforLC–MSmedicinalplantsidentification. Elsevier, 174-180.

[10] Chakia, J., Parekha, R., & Bhattacharyaa, S. (2015). Plant Leaf Recognition using Texture and Shape features with Neural Classifiers. ScienceDirect, 1-13.

[11] Dahigaonkar1, T., & T. Kalyane2, R. (2018). Identification of Ayurvedic Medicinal Plants by Image Processing of. International Research Journal of Engineering and Technology (IRJET), 351-355.

[12] Sabu, A., K, S., & Nair, R. R. (2017). Recognition of Ayurvedic Medicinal Plants from Leaves: A. IEEE, 574-578.

[13] BR, P., C, A., & P, M. N. (2016). Ayurvedic Plant Species Recognition using Statistical Parameters on Leaf. Research India Publications., 5142-5147.

[14]https://www.researchgate.net/figure/Details-of-the-19-layers-of-VGG19-network21used-for-feature-extraction_fig3_334388209

[15] EDQM. (2017). European Regulations for Medicines Place and Role of the European Pharmacopoeia in Europe – Ph. Eur. Concept. EDQM Symposium on Microbiology 10-11 October 2017. [online] Strasbourg: 8. Available at: https://www.edqm.eu/sites/default/files/european_regulations_for_medic inescathie_ vielle-october2017.pdf (Accessed August 15, 2018).

[16] Dybing, F., Dybing, O., Jensen, K. B. (1954). Detection of scilliroside in organic material. Acta Pharmacol. Toxicol. 10 (2), 93–100. doi: 10.1111/j.1600-0773. 1954.tb01326.x

[17] Duckstein, S. M., Lorenz, P., Conrad, J., Stintzing, F. C. (2014). Tandem mass spectrometric characterization of acetylated polyhydroxy hellebosaponins, the principal steroid saponins in Helleborus niger L. roots. Rapid Commun. In Mass Spectrometry 28 (16), 1801–1812. doi: 10.1002/rcm.6959

[18] Desfontaine, C., Guillarme, D., Francotte, E., Nováková, L. (2015). Supercritical fluid chromatography in pharmaceutical analysis. J. pharmaceut. Biomed. Anal. 113, 56–71.doi: 10.1016/j.jpba.2015.03.007

[19]https://www.google.com/search?q=resnet50+architecture&client=msandroidopporev1&prmd=ivn&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjE59Xe4JD3AhVESmwGHfRwDtkQ_AUoAXoECAIQAQ&biw=360&bih=632&dpr=2#imgrc=VxxtkGC2QXgaQM&imgdii=dmadnjpRTUCd6M

[20]https://www.google.com/search?q=vgg16+algorithm+research+gate&tbm=isch&ved2ahUKEwisiaeY9JD3AhVkgGMGHVGZBtQQ2cCegQIABAC&oq=vgg16+algorithm+research+gate&gs_lcp=ChJtb2JpbGUtZ3dzLXdpei1pbWcQAzIECCEQCjoECAAQGDoECB4QClDUL1ibc2CHeWgAcAB4AYAB1QSIAb8akgELMC4zLjcuMi4wLjGYAQCgAQHAAQE&sclient=mobile-gws-wiz-img&ei=QrBWYqzgOuSAjuMP0bKaoA0&bih=632&biw=360&client=msandroidopporev1&prmd=ivn#imgrc=uNMWf O0Df_U9WM