Volume 9, Issue 5, May – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1928

AI-Enhanced Medicinal Plant Identification System with Multilingual Social Media Integration

Sanduni Jayamali Gamage K.G.[1]; Athapaththu P.N.P.[2]; Nandu Gamitha Manawadu[3]; Hansi De Silva[4]
Student[1,2,3]; Lecturer[4]
Faculty of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology Malabe, Sri Lanka

Abstract:- Sri Lanka is a country with a Ayurvedic culture which cannot be experienced anywhere in the world. This cultural system is based on a series of knowledge passed on from generations over 3000 years that could treat a variety of diseases. This traditional ayurvedic system consist of a vast herbal plant collection. Most of the information about this ayurvedic system is written in manuscripts for thousands of years. Sri Lanka lacks a proper system which is specific to ayurvedic sector is a major concern that should be addressed at present. Absence of a system has lead to problems and difficulties in identification and classification of herbal plants, to transfer knowledge about herbal plants and to conserve these ayurvedic plants for the future generation. Another concern is that Ayurvedic undergraduate students face many difficulties when gathering knowledge of these herbal plants and medicinal practices. Sri Lanka does not comprise with a full ayurvedic plant inventory system is another major concern that identified in the country. By considering all the problems an intelligent system has been recognized as a solution. The system will be based on Deep Learning, CNN, GIS, Artificial Intelligence and Machine Learning based principals to cater all the identified problems. The system will be able to identify ayurvedic plant with an image of a leave, flower, or fruit as input. And also, system will be able to classify and provide a detailed description about the identified plant including medicinal value and the distribution of the plant in the island. System will provide a crowdsourcing social media facility with both English and Sinhala languages to share information with fellow herbalist in the country.

Keywords:- *Ayurveda, Deep Learning, CNN, Machine Learning, Artificial Intelligence, NLP, Crowdsourcing, AutoML, GIS.*

I. INTRODUCTION

In Sri Lanka, Ayurveda medicine and herbal plants are playing a significant part in society. According to estimates, traditional medicine is used by 75% of the population [1]. In general, the nation has access to four different conventional. medical systems. They are Deshiya Chikitsa, Siddha, Unani, and Ayurveda. The Ayurvedic and Deshiya Chikitsa methods are seen to be completely ineffective without the use of plants and plant-based medicines [2]. 1430 plant species are considered to have medicinal use. Sri Lanka is home to 174 of these medicinal plants[3]. There have been several challenges in determining whether Ayurveda medicine has played a significant role in Sri Lankan society. For example, most people do not have enough understanding of medicinal plants. As a result, identifying herbal plants and the diseases which can be cured by using them as treatments is difficult. And there are problems when safeguarding herbal plants from numerous risks as people are lacking knowledge in identifying herbal plants and their importance such as home remedies, diseases, and plants outbreak in Sri Lanka. There are many built-in applications in foreign countries like Pl@ntNet [4], PlantSnap [5], iNaturalist [6] and FloralIncognita [7] for plant species identification applications that use mobile devices. There are some applications used in Sri Lanka to detect herbal plants. But there is not any digital inventory system to gather information and there are a handful of applications to detect herbal plants in Sri Lanka [8]. Most of the existing systems use images of leaves for identification of the herbal plants. Plant identification is challenging with only a leaf and its characteristics as there are other unique parts which give an accurate result than the leaf of a plant [9]. These applications only use the English language as instructions or any information regarding plants. Also, they only covered a small number of herbal plants which can be easily found and identified [10]. So, it's better if there is a Sinhala Language supported system because most of the information about the herbal plants is mentioned in Sinhala Language and it is easy to break the language barrier among the local community in Sri Lanka. Further, there are not any specific social media platforms for herbal plants-related stakeholders to communicate their expertise on ayurvedic medicinal plants in built environments, especially among Sri Lankan local community. If there is any platform, it would be beneficial to discover solutions to specific difficulties and share experiences and information with each other. Therefore, it is required to have a system to accurately identify herbal plants as well as recommend home remedies related to identified medicinal plants for cure ailments.

II. RELATED WORKS

Recent works on automatic herbal plant detections have been carried out in various foreign countries as well as in Sri Lanka. A group of three researchers has developed a mobile based system in automatic Identification of medicinal plants in Indonesia using plant leaves [11]. Leaf size, color and texture features were used in the classification process

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1928

where 30 varieties of plant were used with an image collection of 1440. This system mainly provided two features recognition of a medicinal plant and the search for medicinal plants.

Furthermore, [12] suggested an approach for classifying plants into suitable species based on photographs of their leaves. The proposed system includes three key segments. Preprocessing, feature extraction, and classification are all performed. Research was conducted in which 15 leaves were picked at random from 30 different plant species, as well as still photos and videos of herbaceous plants. Then, using video editing software, they are turned into photos. Another study [4] used 17 different varieties of ayurvedic plants to discover a deep learning strategy for the categorization and detection of ayurvedic plant leaves in Sri Lanka. 294 scanned photos and 281 recorded images were obtained. Dileep M.R. and Pournami P.N. used leaf photos to construct a deep learning strategy for classifying 17 different species of medicinal plants in India [5].

In study of [2] have implemented an Android mobile application which uses Probabilistic Neural Network (PNN) for the classification process. Local Binary Patterns (LBP) suggested for leaf texture extraction which was operated in a Linux environment. The resolution of the image is 270 pixels by 240. Captured RGB images from the sensor and then transformed RGB images to grayscale and re-dimensioned to 240x270 pixels. Finally, it provides an accurate rate of 56.33 % when identifying images which was very low. The system only developed in the Android platform and used only leaves for the classification process could be taken as flaws in the system.

Using machine learning technology, a group of researchers created a model for illness prediction. Researchers develop machine learning methods for quick categorization of huge data and accurate illness prediction because of their study. So, forecasting the correct condition is crucial, but processing enormous volumes of data is also necessary, so data mining plays a major part, and categorizing massive datasets using machine learning is straightforward. To accurately anticipate the appropriate illnesses. This study investigates how the KNN machine learning algorithm may be used to quickly classify large amounts of data and accurately forecast illness. The data set was categorized using the KNN algorithm to forecast illness accurately. Additional research based on this study seeks to predict illness using an AutoML technique [6].

Moreover, another study [7] was carried out another approach to disease and remedy suggesting system with artificial intelligence and machine learning. In this research ususes N and machine learning approach to develop the data set of diseases and home remedies. Also, with K-algorithm train the data set to get the high accuracy furthermore using the random frost algorithm and decision tree method also used. Using these machine learning approaches train data with the highest accuracy to suggest the identified diseases and relevant ayurvedic home remedies with a successful rate. Also, with usage of Google API the system identified the best home remedies for suggested diseases. Moreover, in this research brings out how to train data model with classification algorithm with classification rules. Since most of suggesting, predicting system using machine learning with AutoML approaches, it has become a research area which has gained a lot of focus in the recent past [8].

Chatbots are programs that imitate humans' conversation using Artificial Intelligence. It enables to communicate using Natural languages. To identify the indigenous medicinal plants using leaf and leaf and flower [8]. They have used deep learning-based CNN approaches and Machine Learning They have used the OpenCV and TensorFlow technologies. They have built the Chatbot system with AIML and latent Semantic to give knowledge about rare plants. A health care chatbot has been built with the RASA framework and NLP [16]. Chatbot gets the symptoms of the user and display the treatments. But There can be a lot of diseases predicted for the same symptoms but chatbot will give the closest one. [17] For Diabetic education, they have built a chatbot opensource AIML based chatbot. They have used external based knowledge and knowledge based. They have used MYSQL to store the local resources. Using Media Wiki API, they have accessed Wikipedia. Rule – based AIML chatbot for a response for quicker solutions for students' queries [18]. Use Neural Network algorithms are used to obtain accurate efficient responses. TensorFlow and python has use to recognize the user inputs in this ALICE/AIML chatbot [19].

Provide the main idea of the post that has been published will be helpful to users to understand the post. To identify the noise words in social media text, have used ML classifications on Sinhala Unicode characters. The use Sinhala-English mixed text in the social media. They have used SVM, Random Forest, Naïve bayes, Logistic regression. Highest accuracy is the random forest [20]. Keyword extraction for social media short text [21]. These studies used the word2vec and Textrank to solve the unique problem of short text.

We concentrated on the ALICE/AIML and Rasa framework technologies when developing the chatbot. [22] Alice is an open source chatbot that uses natural language processing. It is a program that converses with a person by applying heuristic pattern matching rules to the human's input. AIML is an XML Schema that is used by the software. AIML is an XML dialect used to create natural language software agents. RASA is a development tool for creating unique AI chatbots with Python and Natural Language Understanding (NLU). It provides a platform for creating NLU-powered AI chatbots. Users may train the model and create custom actions. Rasa There are two primary components. RASA NLU and RASA core are the acronyms [23]. RASA NLU is used to infer the purpose of user input and extract the relevant elements. RASA Core generates output by developing a probability model using a Recurrent Neural Network.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1928

Communication using social media create high interactive platforms for individuals, communities, and organization [24]. International plants identification mobile applications like PlantSnap [5] and Pl@ntNet [4] have social platform feature as mentioned in earlier. But they neither support Sinhala Language and nor specified for the herbal plants field. PlantSnap [5] mobile app has a separate community feature. So, er can share posts, photos, their plant collections, and thoughts with other users around the world. Also, they can response (like and comment) to others' posts. Pl@ntNet[4] application has a social media feature to share users' identified plant collections with others. And other users can validate(respond) them.

In [25], Zhang presented a simple and efficient framework, to handle the insight of the crowd on mobile social networks. That framework motivates users to share their knowledge, opinions, or questions through an encouraging online process. Later, the research studies have conducted for allows a user to assign real-time tasks to relevant user in the crowd for execution [26-28] with the purpose of expanding individual or collective human intelligence in the crowdsourcing.

According to the research conducted about personalized social media newsfeed recommendations and other recommendation methods, there is no research conducted on the Sinhala language-supported newsfeed recommendation. In[29], Content-based (CB) Filtering approaches are used to create many contents either by user's comments or feedback. CB filtering model recommends the posts based on the past behavior of the users. Yang *et al.* [30] proposed a Collaborative Filtering (CF) based recommendation system according to a various range of users' data including users' ratings and history data and it compares other normal CF algorithms. Majority of the researchers didn't discuss how social media platforms-based recommendation models deal with data sparsity and cold start problems. Huang *et al.* [31] and Wang*et al.* [32] proposed a hybrid CF model in which a combination of a user-based CF and item-based CF to solve data sparsity and improve the accuracy of the results. Research Paper [33] discussed recommendation algorithms based on users' dynamic information. And also, proposes a new method to get the similarity based on social media users' dynamic information by using behavior information, collections, positive responses, and timestamps. But that research paper never mentioned that it is capable of getting native language data for similarity calculation.

PlantSnap [5] mobile application has an Augmented Reality(AR) feature. But it only shows a 3D demonstration of animated sun, bees, and some other features when the camera moves to the plant. It doesn't demonstrate a 3D view of a plant. The process of 3D model reconstruction using photogrammetry from low-cost RGB cameras is called stereo vision. In [34], the stereo vision method has been applied for plant reconstruction to recognize plant information like plant height and leaf position. In paper [35], the Structure from Motion (SfM) and Multi-View Stereo (MVS) techniques was applied to create 3D models for reconstructing plants of three weed species with contrasting shape and plant structures by fusing images. Also, it is estimated the measuring accuracy of a 3D structure of a plant which is created using the photogrammetric approach based on SfM-MVS and a sequence of photographs taken by a smartphone [36]. Even though, there are some issues and features should be improved in these existing systems.

The proposed mobile application approach in this study is a novel solution for the previously mentioned and recognized drawbacks of existing research works.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1928

III. METHODOLOGY

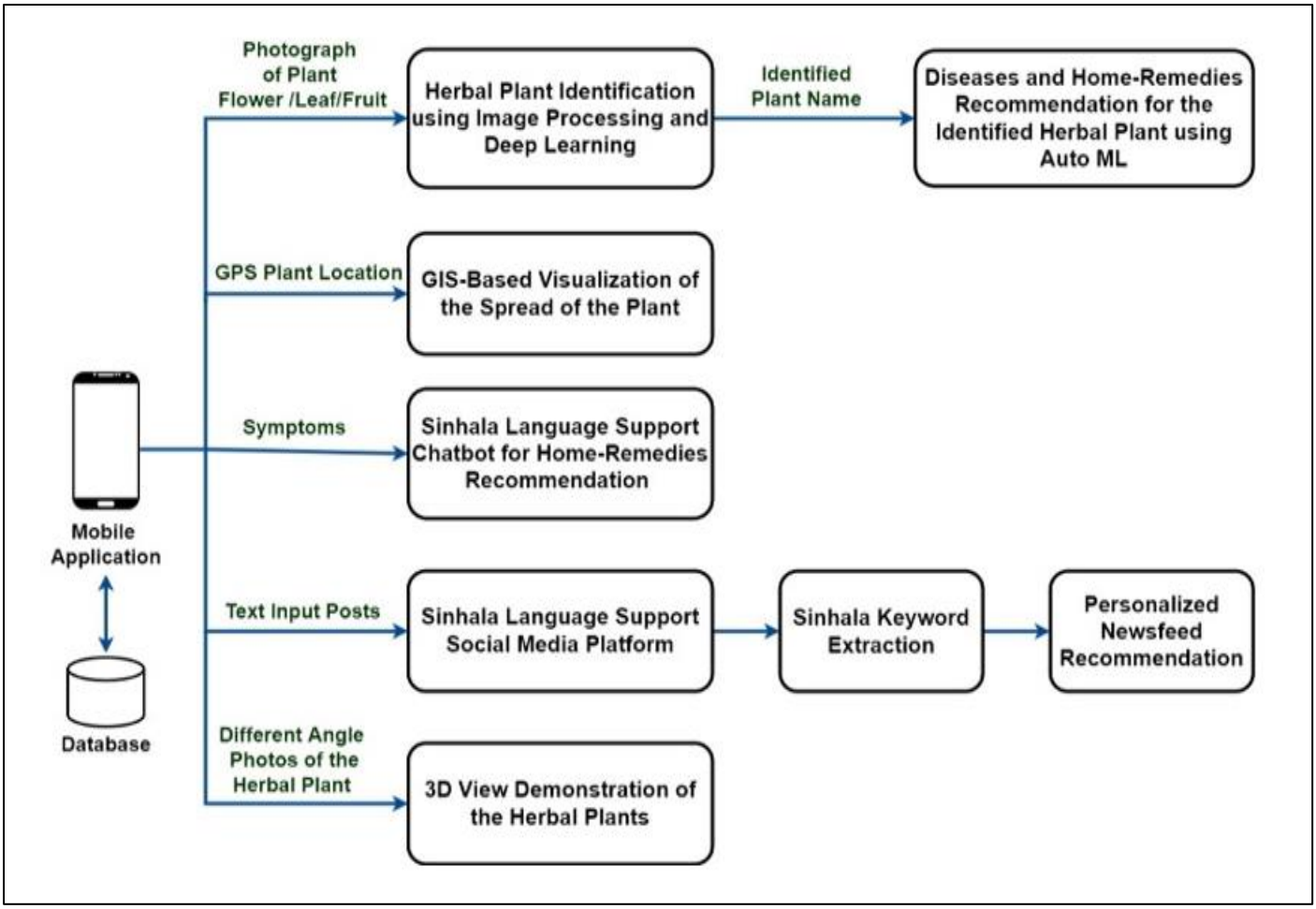

Fig 1: System Diagram

According to the Figure 1, proposed approach consists of 8 main components.

A. Identification of Herbal Plant Type using Image Processing and Deep Learning

The application-scanned picture of the leaf, blossom, or fruit in this study serves as the input for the classification step.

Ayurvedic plant data from Sri Lanka that was recently collected, organized, and labeled was used to train a convolutional neural network (CNN). A noisy dataset was created by gathering photos of the leaves, fruits, and flowers from websites and blogs devoted to Sri Lankan herbal plants, alternative medicine websites, and social media. Of the numerous herbal plants, 15 were selected to be examined further in depth. The dataset was divided into training and test data, each of which has 30% and 70% of the total data.

The Target was to assess and identify the model with the best accuracy among various deep CNN models using newly obtained training and testing data, determine each model's final testing accuracy, and then utilize that model as the final model. The dataset was then enhanced utilizing augmentation approaches to improve accuracy, and the optimal technique with the highest accuracy was chosen. The dataset was then retrained using pre-trained weights on the available convolutional neural network architectures. The most accurate re-trained model was then improved using bottleneck characteristics and data augmentation strategies.

B. Recommendation of Diseases and Home-Treatments for the Identified Herbal Plant using Auto ML and GIS-Based Visualization of the Spread of the Plant

This research brings out how to use AutoML concepts in both the western and ayurvedic medicine fields. Through this, it shows how to use AutoML architecture to get relevant outcomes for the system. Here using the Generalized AutoML category to automate the entire process with high accuracy. When it came to data model training with the AutoML technique, there were two primary steps: data processing and feature selection. Two types of parameters were utilized to select the hyperparameter. Model parameters and hyperparameters are what they are. Optimizing the hyperparameter is a function that minimizes the algorithm's loss and cost, which helps maintain the model's bias and variance balance. Furthermore, this study proposed a strategy for automating the hyperparameter that can provide maximum accuracy. [37].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1928

After the identification of the plant, the proposed system will take the identified plant as the input and display the diseases that can cure using the identified plant. The system will mention the plant part that is used to make home remedies and the system will facilitate users by mentioning home remedies to each disease. This process will be automated using Auto ML (Machine Learning) data models which will use machine learning-based operations. Diseases that can be cured using the identified plant and home treatments that can be made using the plant will be displayed.

The data set was created by using the source and knowledge-based data with help of the Ayurveda Institute at Rajagiriya. So, the dataset contains the plant name in English and Sinhala Languages, diseases that can cure from the relevant plant, and relevant treatments.

PyCaret tool was used for the implementation of the above process. PyCaret is a Python-based fully accessible machine learning framework that automates machine learning procedures. It is a complete machine learning and model management solution that reduces trial cycle time and boosts productivity. When working with datasets, machine learning algorithms usually function in two steps after finishing the data pretreatment step. As a result, we divided the data set into two parts: 70% for training and 30% for testing. The PyCaret framework automates the processes of feature selection and Hypermeter adjustment. Then, compare and analyze the performance of all models before selecting the best model. Then Develop a model with PyCaret's "create_ model()" method. Because we trained the model using the "create_model()" function, it uses the default hyperparameters. To tune hyperparameters, use the "tune_model()" function. This function automatically optimizes a model's hyperparameters on a pre-defined search space using Random Grid Search. Then, evaluate the prediction model's accuracy, choose the best model as the finished model, and prepare for deployment. Finally, save the model before proceeding with the implementation.

And the system maps the plant's detected location using the GIS (Geographic Information System) data model to make aware users of the distribution. A GIS data model will be trained to visualize the data correctly throughout a defined area. After the area of the selected plant is mapped, it will be visualized on a geographical map with GeoPandas.

C. Sinhala Language Support Chatbot and Sinhala Keyword Extraction

The creation of a keyword extraction system that could extract Sinhala phrases was the main objective of this model. This model will first determine the language and then extract the keyword. This will be applied to post suggestions on the social media network. SVM, Naive Bayes, RF, Decision Tree, and Logistic Regression were used to compare the classification of TF-IDF tokenized n-grams and Doc2Vec embedded texts. Yet Another Keyword Extractor (YAKE) was compared against KeyBERT, a technique that uses BERT embedding and cosine similarity to extract keywords independent of text size, language, or topic matter.

The Sinhala language is supported by the chatbot in this suggested component. Chatbots will be built using the best architecture available from the ALICE/AIML and RASA frameworks. The five main components of the chatbot architecture are user input analysis, dialog management, information retrieval, data source, and response generation. The process is started by the user's text entry. The user will enter a disease. The user input analysis component receives the user request and analyzes it to determine the ailment the user has provided and any pertinent data. The chatbot will ask a series of clarifying questions once it has chosen the best interpretation. After the disease has been detected, the home remedies will next be obtained from the data source using the information retrieval component from the data source component. Data source will be preferable to web-based or knowledge-based alternatives. After retrieving data, a response generator will generate a human-like, natural language answer for the user. The dialogue management component monitors and modifies the conversational context.

D. Social Media Platform Framework with Personalized Newsfeed Recommendations and Demonstration of the 3D View of the Herbal Plants

Crowd computing has five main characteristics; They are user crowds, interaction between human and machine, crowd activity with purpose, tasks, and collective intelligence. These elements are used as the base for the proposed crowdsourcing social media platform and describe the functions and architecture of different roles within the framework. First, many users with certain purposes in the herbal plants-related domain gather to form a user crowd as the input. Interactive mobile devices with the application serve as media for user crowds and the system. Then, users post content or questions that should be responded/solved by other users with different backgrounds. This way, the relationship is established among the users. During the processing step, the crowdsourcing social media framework analyzes the information of post collection and user collection and engages the post recommendation model to select the suitable user crowd for the post content. After the processing of crowdsourcing social media platform, the relevant results or strategies will be gained, and data will be stored as the output.

Based on keywords created by the keyword extraction model, the hybrid posts recommendation model of the proposed approach classified the information posted on the social media platform. Google Translation API was used to translate Sinhala keywords to the English language. The proposed hybrid personalized newsfeed recommendation model is a combination of CB and CF filtering models.

- *Content Based (CB) Filtering*

Recommendations are made using features and preferences in the CB filtering. Content similarity is used rather than that of users [38]. Based on the keywords, the similarity of the posts' content was estimated. The recommendation model identified and evaluated the user's interest by checking the user's view post history. Then it delivered personalized posts to the user as the final output.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1928

The most precise and suitable solution for the post was selected after analysis and testing using categorization ranking and metadata-based filtering.

- *Categorization Ranking*

The category mapping function examined the dataset using TfidfVectorizer to detect the post's keywords. The TF-IDF is applied for feature extraction of the text in Natural Language Processing (NLP). So, the model identified the unique words(not consider the repeated values) and calculated the number of times each word appears in a text, weight of each word and the score for the text. Then the posts vectors calculated by the cosine similarity was taken as input to the category similarity ranking.

- *Metadata Based Filtering*

The proposed model combined the raw data in the translated keywords to make a new variable as an input to the TfidfVectorizer. For the cosine similarity calculation, the feature vectors built utilizing the TfidfVectorizer were used as input.

➢ *Collaborative Filtering (CF)*

The dataset contained the data of the posts, users' post view history, and users' profile information. Using that dataset, the required was created and it's used to analyze the matrix preferences and recommend posts for the particular user. Then, the CF-based recommendation approach predicted initial ratings and post preference ranking list based on realistic nearest neighbor, similarity between users and user behavior. After this process, the correlation between users and posts was calculated using spearman's correlation coefficient. As the final result, top-ranked post recommendations were displayed, and the user's newsfeed was restructured according to those ratings.

The proposed model provided a personalized newsfeed recommendations model considering different elements including,

- The existing users of the system are received the top-ranked post recommendations that are viewed by similar users.
- New users are recommended the most popular and trendy posts published by the current users.
- Posts should be recommended based on the involvement of related users and the similarity of the metadata of their posts. which enhanced the quality of the recommendation process in turn.

A structure from motion (SfM) photogrammetric model was proposed in this study to demonstrate the 3D structure of an herbal plant using a collection of 2D photos of the same plant captured from different angles. The photographs were taken all around an individual herbal plant without any preferred position and in order of intervals between them. For this case, 30-50 photos were collected for each plant. The process of implementing this model includes a series of steps. First, the key points and feature descriptors from images of herbal plants were extracted. Then the features between images were matched. After that, the initial two views were found as an appropriate baseline to kickstart the reconstruction and the pose of the baseline was recovered. SfM is a methodology to recover the 3D structure of a scene by using a sequence of 2D images. In the process, this method also recovers the relative pose of a particular view (an image taken by a camera) with the first view in the sequence. Initially, the camera postures of the image inputs were evaluated using the SfM algorithm. So that, the multi-view stereo (MVS) technique is used to create the sparse point cloud and then further rebuild it into a dense point cloud. The MVS algorithm is applied to count the depth and common information for every pixel in each photo and combine them. The Poisson surface reconstruction algorithm was applied for the dense point cloud in order to recover the structure of the 3D surface. MeshLab was used to reconstruct the 3D mesh and dense point cloud and to alter the model. The unwanted objects in the background of the plant were manually deleted and cleared to reduce the noise and enhance the model's performance, accuracy, and visual rendering.

IV. RESULTS AND DISCUSSION

The training and testing of each CNN model are done using our own dataset of herbal plant parts. Table I below shows the completed testing accuracy for each pre-trained deep CNN architecture.

Table 1: Comparison of Accuracies for the Selected CNN Models in Herbal Plant Classification.

CNN Architecture	Final Testing Accuracy
InceptionV3	76.43%
MobileNetV2	60.51%
ResNet50	96.80%
VGG16	96.88%

By comparison, the model with the highest accuracy was examined, and it was the pre-trained model VGG16, which had a testing accuracy of 96.88% and was improved utilizing data augmentation methods and bottleneck characteristics. The best-accurate CNN model was built using the VGG-16 model, which had been pre-trained on the ImageNet dataset. The convolutional basis was then frozen in order to retrieve the features. The following layer was added without changing the batch size. There was also a 6-class classification layer with a "SoftMax" activation method. Before training, the model was built with "categorical cross-entropy" as the loss function, "RMSprop" as the optimizer, "categorical accuracy" as the metrics, and (1e-4) as the learning rate.

After that, the training and testing datasets were loaded, with the training dataset improved using different enhancement techniques and the testing dataset not enhanced for legitimate accuracy purposes. The dataset was then trained across 100 epochs using 100 steps per epoch for training and 50 steps per epoch for testing.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1928

The completed classification model summary of the selected VGG-16 pre-trained model, which had the greatest accuracy, is depicted in Fig.2.

```
 Layer (type)                 Output Shape              Param #
=================================================================
 vgg16 (Functional)           (None, None, None, 512)   14714688

 sequential_7 (Sequential)    (None, 6)                 3078

=================================================================
Total params: 14,717,766
Trainable params: 7,082,502
Non-trainable params: 7,635,264
_________________________________________________________________
```

Fig 2: Finalized Model Summary

The final VGG-16 model is shown in Fig3, along with accuracy and loss charts on the training and validation datasets over training epochs.

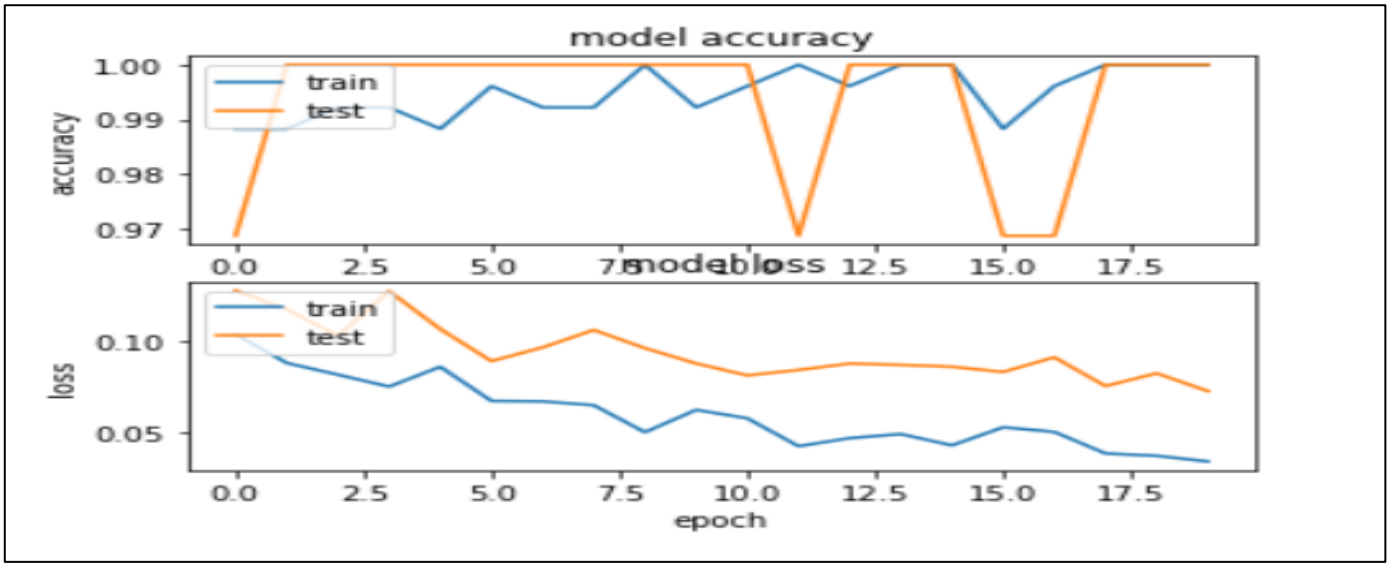

Fig 3: Finalized Model Performance Accuracy and Loss Graphs

In the diseases and treatment data set before selecting the best model Linear Regression, Random Forest, Decision Tree, and Gaussian Processes, SVM, and KNN algorithms are tested for selecting the best performing model. As KNN model provided 93.33% highest accuracy.

```
predict_model(tuned_knn);
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	K Neighbors Classifier	0.9333	0.9929	0.8000	0.8733	0.9018	0.8790	0.8886

Fig. 4. Finalized Model Summary

As the KNN model provides the highest accuracy as shown in figure 4 it will be used for the API creation in the mobile application which will recommend illness and home remedies for the identified plant.

The creation of a keyword extraction system that could extract Sinhala phrases was the main objective of this model.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1928

- Transliteration of the Language of Identification
- Extract The two main elements of the overall scheme were keywords.

The language identification models were evaluated using the accuracy, precision, F1 score, and confusion meter. Table II shows that the accuracy of 0.97 was highest for the 1-gram based Nave Bayes model, SVM, and RF model using Doc2Vec document embedding. The Nave Bayes model, with 0.97 accuracy, outperformed the other two models for both Singlish and English, according to the confusion matrix. In general, models with Doc2Vec document embedding gave more accurate results, as opposed to models that used 3-gram and 5-gram, as can be observed from the data. After taking the results of the language identification phase into account, the final model for Sinhala language detection was selected. It made use of a word-unigram-based Nave Bayes model.

Table 2: Summary of the Accuracies of Language Identification Models

Feature Extract technique		**NB**	**SVM**	**RF**	**DT**
TF – IDF	1 – gram	0.97	0.96	0.96	0.95
	3 – gram	0.78	0.75	0.71	0.68
	5 – gram	0.68	0.68	0.67	0.67
DOC2Vec		**0.96**	**0.97**	**0.97**	**0.96**

In chatbot implementation using the RASA. Also, we have installed the Rasa X, which is a tool for Conversation Driven Development, which provides insight into the chatbot process with users. And also use the retrofit to fetch the APIs from the network service and convert it into JSON.

Table 3: Summary of the Final Results of the Recommendation Model Evaluation

	Mean	**Std**
RMSE	0.7970	0.0034
MAE	0.5908	0.0022

The hybrid newsfeed recommendation model displayed top-ranked posts to the user according to the order. Based user's behavior and history, the recommendation model suggested more relevant posts. After evaluating the final result provided by the proposed approach, it was identified that the hybrid posts recommendation model is more accurate and appropriate than an individual CB or CF model. Metadata based Filtering delivered better output than categorization and ranking. CF recommendation models have issues like cold start problem, data sparsity and inflexibility in adding side queries as mentioned in the earlier. But the suggested method eliminated the above-mentioned downsides and offered an accurate newsfeed recommendation approach. Using RMSE and MAE the proposed model was evaluated, and the validation was done by using 5-fold cross-validation. The proposed hybrid posts recommendation model's mean of RMSE and MAE values are 0.7970 and 0.5908 according to the Table III. Singular vectors in the unstructured and dispersed post-view history data were factorized and ranked using Singular value decomposition(SVD).

Usage of a mobile device can facilitate capturing a sequence of 2D photos for the following 3D reconstruction and the feature extraction , as well as maintain the final outcome's quality. The final outputs of the proposed 3D model indicates that reconstruction of 3D point cloud using the SfM methodology is a suitable approach to demonstrate 3D view of an herbal plant.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an intelligent herbal plant detecting system with the Sinhala language-supported social media platform. In the future, the web version of the application will be released, and the plant scope will be increased which will cover most of the herbal plants in Sri Lanka. Auto ML technology will be utilized which will update the diseases and treatments on a real-time basis with a proper justification process. In this study, the proposed hybrid posts recommendation model solves the drawbacks of the traditional recommendation methods. Also, the insufficiency of the dataset caused the less accuracy of the recommendation model. Further, the accuracy of the overall proposed system can be improved by using scaled datasets.

REFERENCES

[1]. T. N. M. Kankanamalage, R. M. Dharmadasa, D. C. Abeysinghe, and R. G. S. Wijesekara, "A survey on medicinal materials used in traditional systems of medicine in Sri Lanka," *J. Ethnopharmacol.*, vol. 155, no. 1, pp. 679–691, 2014.

[2]. P. B. Weragoda, "The traditional system of medicine in Sri Lanka," *J. Ethnopharmacol.*, vol. 2, no. 1, pp. 71–73, 1980.

[3]. K. Priya, S. Ranjana, and R. Manimegala, "Medicinal Plant Identification using Android Application based on leaf image," *Eur. J. Mol. Clin. Med.*, vol. 7, no. 09, pp. 1496–1506, 2020, doi: 10.31838/ejmcm.07.09.161.

[4]. *Plantnet-project.org*. [Online]. Available: https://identify.plantnet-project.org/. [Accessed: 23-Jan-2022].

[5]. "PlantSnap - plant identifier app, #1 mobile app for plant identification," *Plantsnap - Identify Plants, Trees, Mushrooms With An App*, 24-Aug-2018. [Online]. Available: https://www.plantsnap.com/. [Accessed: 23-Jan-2022].

[6]. "There's an app for us ·," *iNaturalist*. [Online]. Available: https://www.inaturalist.org/pages/mobile_apps_nz. [Accessed: 23-Jan-2022].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1928

[7]. "The flora incognita app," *Floraincognita.com*. [Online]. Available: https://floraincognita.com/flora-incognita-app/. [Accessed: 23-Jan-2022].
[8]. P.V.D Nadeeshan, A.D.A.D.S Jayalath, T.G.A.G.D Amarawansha and H.P Jayasuriya, "Deeghayu - Ayurveda Information Centralized Mobile Application" (2019).
[9]. M. R. Dileep and P. N. Pournami, "AyurLeaf: A Deep Learning Approach for Classification of Medicinal Plants," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2019-October, pp. 321–325, 2019, doi: 10.1109/TENCON.2019.8929394.
[10]. N. Pathiranage, N. Nilfa, M. Nithmali, N. Kumari, L. Weerasinghe, and I. Weerathunga, "Arogya -An Intelligent Ayurvedic Herb Management Platform," in 2020 61st International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), 2020.
[11]. K. Anuradha, D. P. M. Lakshan, and R. P. S. Kathriarachchi, "A Study on the Ayurveda Plant Recognition for Remedial Medications Using Image Processing Techniques," pp. 515–522
[12]. M. Jayanka and T. G. I. Fernando, "Recognising Ayurvedic Herbal Plants in Sri Lanka using Convolutional Neural Networks," *Vidyodaya J. Sci.*, vol. 23, pp. 48–60, 2020
[13]. Dahiwade, G. Patle, and E. Meshram, "Designing disease prediction model using machine learning approach," *Proc. 3rd Int. Conf. Comput. Methodol. Commun. ICCMC 2019*, no. Iccmc, pp. 1211–1215, 2019, doi: 10.1109/ICCMC.2019.8819782
[14]. R. I. S. Bandara, S. Prabagaran, S. A. K. G. Perera, M. N. R. Banu, and K. A. D. C. P. Kahandawaarachchi, "Wedaduru -An Intelligent Ayurvedic Disease Screening and Remedy Analysis Solution," *2019 Int. Conf. Adv. Comput. ICAC 2019*, pp. 151–155, 2019, doi: 10.1109/ICAC49085.2019.9103410
[15]. Nagarajah and G. Poravi, "An extensive checklist for building autoML systems," *CEUR Workshop Proc.*, vol. 2360, no. April,2019.
[16]. J. Gupta, V. Singh and I. Kumar, "Florence- A Health Care Chatbot," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 504-508, doi:K
[17]. S. Hussain and G. Athula, "Extending a Conventional Chatbot Knowledge Base to External Knowledge Source and Introducing User Based Sessions for Diabetes Education," 2018 32nd International Conference on
[18]. Rule-based Chabot for student enquiries to cite this article: Jagdish Singh et al 2019 J. Phys.: Conf. Ser. 1228 012060.
[19]. D. A. D. S. Jayalath, P. V. D. Nadeeshan, T. G. A. G. D. Amarawansh, H. P. Jayasuriya and D. P. Nawinna, "Ayurvedic Knowledge Sharing Platform with Sinhala Virtual Assistant," 2019 International Conference on Advancements in Computing (ICAC), 2019, pp. 220-225, doi: 10.1109/ICAC49085.2019.9103413.
[20]. K. Shanmugalingam and S. Sumathipala, "Language identification at word level in Sinhala-English code-mixed social media text," 2019 International Research Conference on Smart Computing and Systems Engineering (SCSE), 2019, pp. 113-118, doi: 10.23919/SCSE.2019.8842795.
[21]. D. Zhao, N. Du, Z. Chang and Y. Li, "Keyword Extraction for Social Media Short Text," 2017 14th Web Information Systems and Applications Conference (WISA), 2017, pp. 251-256, doi: 10.1109/WISA.2017.12.
[22]. Wikipedia contributors. (2021, December 3). Artificial Linguistic Internet Computer Entity. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Artificial_Linguistic_Internet_Computer_Entity&oldid=1058370204
[23]. S. Meshram, N. Naik, M. VR, T. More and S. Kharche, "College Enquiry Chatbot using Rasa Framework," 2021 Asian Conference on Innovation in Technology (ASIANCON), 2021, pp. 1-8, doi: 10.1109/ASIANCON51346.2021.9544650.
[24]. Anandhan, L. Shuib, M. A. Ismail, and G. Mujtaba, "Social Media Recommender Systems: Review and Open Research Issues," IEEE Access, vol. 6, no. Figure 1, pp. 15608–15628, 2018, doi: 10.1109/ACCESS.2018.2810062.
[25]. Zhang X, Shang L, Yuan Y (2017) A crowd wisdom management framework for crowdsourcing systems. IEEE Access 4:9764–9774
[26]. Xing Y, Wang L, Li Z (2019) Multi-Attribute crowdsourcing task assignment with stability and satisfactory. IEEE Access 7:133351–133361
[27]. Tran L, To H, Fan L (2018) A real-time framework for task assignment in hyperlocal spatial crowdsourcing. ACM Trans Intell Syst Technol 9(3):37:1–37:26
[28]. Alireza S, Shafigheh H, Masoud RA (2018) Personality classification based on profiles of social networks' users and the five-factor model of personality. Hum-Cent Comput Inform Sci 8(1):24–38
[29]. Narducci et al., (2013, Aug). Exploiting big data for enhanced representations in content-based recommender systems. Presented at International Conference on EC Web 2013. [Online] Available :
[30]. Z. Yang, B.Wu, K. Zheng, X.Wang, and L. Lei, ``A survey of collaborative filtering-based recommender systems for mobile Internet applications," *IEEE Access*, vol. 4, pp. 3273_3287, 2016.
[31]. B. Huang, X. Yan, and J. Lin, ``Collaborative filtering recommendation algorithm based on joint nonnegative matrix factorization," *Pattern Recognit. Artif. Intell.*, vol. 29, no. 6, pp. 725_734, Aug. 2016.
[32]. Wang, T. Li, X. Wang, S. Zhu, and C. Ding, ``Community discovery using nonnegative matrix factorization," *Data Mining Knowl. Discovery*, vol. 22, no. 3, pp. 493_521, 2011.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1928

[33]. J. Cheng, Y. Liu, H. Zhang, X. Wu, and F. Chen, "A new recommendation algorithm based on user's dynamic information in complex social network," *Math. Probl. Eng.*, vol. 2015, 2015, doi: 10.1155/2015/281629.

[34]. Takizawa *et al.*, "Plant recognition by integrating color and range data obtained through stereo vision," *J. adv. comput. intell. intell. inform.*, vol. 9, no. 6, pp. 630–636, 2005.

[35]. D. Andújar, M. Calle, C. Fernández-Quintanilla, Á. Ribeiro, and J. Dorado, "Three-dimensional modeling of weed plants using low-cost photogrammetry," *Sensors (Basel)*, vol. 18, no. 4, p. 1077, 2018.

[36]. E. Chong and H. H. Harith, "Performance of structure-from-motion approach on plant phenotyping using images from smartphone," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1038, no. 1, p. 012031, 2022.

[37]. T. Nagarajah and G. Poravi, "A review on automated machine learning (AutoML) systems," in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, 2019

[38]. U. Shardanand, and P. Maes, (1995, May) Social information filtering: Algorithms for automating 'word of mouth. Presented at CHI '95 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. [Online]