E-Krushi: Agricultural Assistance using Deep Learning and Natural Language Processing.

Isha Hemant Desai
Vivekanand Education Society's Institute of Technology
Chembur, Maharashtra, India.
2020.isha.desai@ves.ac.in

Jessica Lalchandani
Vivekanand Education Society's Institute of Technology
Chembur, Maharashtra, India.
2020.jessica.lalchandani@ves.ac.in

Trishala Jeswani
Vivekanand Education Society's Institute of Technology
Chembur, Maharashtra, India.
2020.trishala.jeswani@ves.ac.in

Dr. Mrs. Sujata Khedkar
Vivekanand Education Society's Institute of Technology
Chembur, Maharashtra, India.
sujata.khedkar@ves.ac.in

***Abstract*—Agriculture is one of the main contributing sectors in the Indian economy. There are various issues hampering the growth of agricultural production in India. Plant disease is one major problem among others. The system proposes to aid the farmers in combating and taking quick actions to prevent losses due to disease proliferation. Deep learning tools are employed to generate analysis for the same. The farmers can acquire a quick analysis of the plant's health by providing an image of the plant's leaf. The application provides facilities to interact in various regional languages in order to cater the needs of its vernacular audience. Natural language processing techniques are used to translate the queries of the farmers. A website and an android application are a part of the proposed solution in order to provide a user-friendly interface. Chatbot embedded in the application and the website will assist the farmers with their queries by providing multilingual support.**

Key Words: ***Farmers, Plant Disease detection, Deep Learning, Language translation, NLP.***

I. Introduction

Agriculture, with its allied sectors, is the largest source of livelihood in India. 70 percent of India's rural households still depend primarily on agriculture for their livelihood, with 82 percent of farmers being small and marginal. India is the largest producer of milk, jute and pulses according to the 2012 census.

It is the second-largest producer of rice, wheat, sugarcane, cotton and groundnuts, as well as the second-largest fruit and vegetable producer, accounting for 10.9% and 8.6% of the world fruit and vegetable production, respectively. Farming is central to India's socio-economic development. Therefore, enhancements in technology to aid farmers, will help to accelerate India's growth trajectory.

All species of plants, wild and cultivated alike, are subject to disease. The occurrence and prevalence of plant diseases vary from season to season, depending on the presence of the pathogen, environmental conditions, and the crops and varieties grown. Some plant varieties are particularly subject to outbreaks of diseases while others are more resistant to them.

Plant diseases are posing a new challenge in the growth of India's agriculture. According to the survey conducted by The Economic Times "Crops worth Rs 50,000 crore are lost owing to pest and disease attacks every year".

There are many applications currently available in the market catering to the needs of the farmers. But they do not provide the facility of interacting in different state languages. There are many non-Hindi speaking states in India like West Bengal, Telangana, Tamil Nadu, Karnataka and Kerala. Even farmers in states like Maharashtra would prefer to get the information in Marathi. Therefore, it is the need of the hour to assist the farmers regarding the plant diseases, its causes and solutions in vernacular languages.

The objective of this project is to provide better agricultural assistance to the farmers in their regional language. To begin with the focus will primarily be Marathi and Hindi and later on the scope will be broadened to different state languages.

II. OVERVIEW

A. *Motivation*

The motivation for this project stems from the critical role that agriculture plays in the Indian economy and the numerous challenges faced by farmers, including plant diseases that can lead to significant crop losses. The project aims to empower farmers with advanced technology solutions to combat these challenges effectively. By leveraging deep learning tools, the system can provide quick and accurate analysis of plant health based on images of leaves. This technology-driven approach will enable farmers to take timely actions to prevent disease proliferation and minimise losses.

Furthermore, the project is motivated by the need to cater to the diverse linguistic and regional preferences of Indian farmers. By offering multilingual support through natural language processing techniques, the system ensures that farmers from various regions can easily interact with and benefit from the solution. This inclusivity is essential for reaching a wider audience and making the technology accessible to all.

The development of a user-friendly website and Android application, along with the integration of a chatbot, is another crucial aspect of the project's motivation. These interfaces provide a convenient and accessible means for farmers to access plant health analysis and seek assistance with their queries. Overall, the project's motivation lies in leveraging cutting-edge technology to enhance agricultural productivity, empower farmers, and address the unique challenges faced in the Indian agricultural sector.

B. *Problem Statement*

Plant disease is one of the biggest challenges faced by the farmers of India. The loss incurred each year due to the disease outbreak is huge and is increasing every year. The information regarding the causes, symptoms and treatment of diseases will help the farmers to reduce the losses incurred. But this alone is not enough. The identification of the disease type can be tricky and difficult for the farmers. To accelerate the treatment of the infested crops, an accurate and quick analysis of their state is required. There are various applications in the market producing insights regarding the crop disease, but they are mostly taking inputs and giving outputs in English language. The farmers from within the interiors of Indian villages need multilingual assistance in order to profit from the analysis.

C. *Relevance of the project*

The relevance of this project lies in its potential to address pressing issues within the Indian agricultural sector. Agriculture is a cornerstone of the Indian economy, but it faces numerous challenges, with plant diseases being a major concern. This project is highly relevant because it offers a technology-driven solution that can significantly improve agricultural productivity and reduce crop losses.

By harnessing deep learning tools for plant health analysis, the project empowers farmers to quickly detect and mitigate diseases, thus preserving crop yields and income. The inclusion of multilingual support through natural language processing ensures that the solution is accessible to a diverse range of farmers, making it highly relevant for the Indian context where linguistic diversity is substantial.

Moreover, the development of a user-friendly website and Android application, coupled with a chatbot for real-time assistance, ensures that the technology is not only effective but also practical and user-centric. This relevance is underscored by the potential to positively impact the livelihoods of millions of farmers across India, contributing to food security, economic growth, and sustainable agriculture in the country. In summary, the project's relevance lies in its ability to address critical agricultural challenges in India and empower farmers with innovative technology solutions.

III. LITERATURE REVIEW

A. Survey of Existing Systems

P. Y. Niranjan, V. S. Rajpurohit and R. Malgi et al. [1] conducted a survey on chat-bot systems for the agriculture domain. The paper focused on two chatbots, "ADANS: An Agriculture Domain Question Answering System using Ontologies" and "AGRI-QAS Question-Answering System for Agriculture Domain," which utilize natural language processing and semantic web technologies.

D. Sawant, A. Jaiswal, J. Singh and P. Shah et al. [2] developed AgriBot, an intelligent interactive interface to assist farmers in agricultural activities. The chatbot, designed using DialogFlow API, covers queries related to seeds, fertilizers, market prices, storage facilities, and government schemes, with prediction algorithms like K-nearest neighbor, Random forest, and Decision Tree.

U. Kiruthika, S. K. S. Raja, V. Balaji and C. J. Raman et al. [3] created a chatbot for direct marketing of food crops using e-agriculture. The system connects farmers and consumers directly, ensuring fair pricing. Developed with Python, R language, JavaScript, and PostgreSQL, it facilitates secure communication without third-party mediation.

S. V. Militante, B. D. Gerardo and N. V. Dionisio et al. [4] used deep learning, particularly Convolutional Neural Network (CNN), for plant leaf detection and disease recognition. The study achieved a 96.5% accuracy in detecting diseases in apple, corn, grapes, potato, sugarcane, and tomato leaves.

H. Andrianto, Suhardi, A. Faizal and F. Armandika et al. [5] developed a smartphone application for deep learning-based rice plant disease detection. Using VGG16 architecture, the system captures rice plant leaf images, classifying diseases like Brown Spot, Leaf Blast, and Hispa with a train accuracy of 100% and test accuracy of 60%.

N. Saranya, L. Pavithra, N. Kanthimathi, B. Ragavi and P. Sandhiyadevi et al. [6] employed an Artificial Neural Network (ANN) algorithm to detect banana leaf and fruit diseases. The system distinguishes between healthy and diseased banana plants, addressing leaf diseases (Black Sigatoka, Freckle Leaf) and fruit diseases (Anthracnose, Freckle Fruit).

R. D. Devi, S. A. Nandhini, R. Hemalatha and S. Radha et al. [7] developed an IoT-enabled disease detection system for agricultural applications. Using Random Forest and KNN, the system detects and classifies bunchy top of banana and sigatoka diseases based on parameters like temperature, soil moisture, and pH score.

Kirti and N. Rajpal et al. [8] focused on black rot disease detection in grape plants using color-based segmentation and machine learning. Utilizing the PlantVillage Dataset, the study achieved accuracies of 93.3% (Linear Kernel), 94.1% (RBF Kernel), and 93.9% (Polynomial Kernel) with Support Vector Machine Classifier.

J. S. H. Al-bayati and B. B. Üstündağ et al. [9] utilized an evolutionary optimization approach for apple leaf disease identification. The study incorporated Deep Neural Network (DNN), Robust Speed Up Feature (SURF), and Modified Grasshopper Optimization Algorithm (MGOA), achieving an accuracy of 98.49%.

M. A. Rahman, M. M. Islam, G. M. Shahir Mahdee and M. W. Ul Kabir et al. [10] proposed an improved segmentation approach for tomato leaf disease detection. Using a combination of thresholding and morphological operations, coupled with a deep neural network for classification, the system achieved an accuracy of 99.25%.

B. S. Kusumo, A. Heryana, O. Mahendra and H. F. Pardede et al. [11] developed a machine learning-based system for automatic detection of corn-plant diseases using image processing. Utilizing various image processing features and machine learning algorithms like SVM, DT, RF, and NB, the study achieved optimal performance with SVM using RGB features.

S. Y. Yadhav, T. Senthilkumar, S. Jayanthy and J. J. A. Kovilpillai et al. [12] implemented a CNN model with an optimized activation function for plant disease detection. The study focused on detecting various plant diseases, achieving a new accuracy of 95% through the developed activation function and K-means clustering algorithm for area calculation.

B. Limitations and Research Gap

There are many applications currently available in the market catering to the needs of the farmers.

But they do not provide the facility of interacting in different state languages.

There are many non-Hindi speaking states in India like West Bengal, Telangana, Tamil Nadu, Karnataka and Kerala.

Even farmers in states like Maharashtra would prefer to get the information in marathi.

Therefore, it is the need of the hour to assist the farmers regarding the plant diseases, its causes and solutions in vernacular languages.

Comparison with the existing systems -

1. The existing chatbots don't have multilingual support.
2. Very few provide bilingual support in English and Hindi. Chatbots for farmers in Marathi language are very rare.
3. Even if some chatbots provide multilingual support, they don't have the plant disease detection module available.
4. If the chatbot is not able to solve some query then the helpline numbers and the further course of action are not provided.
5. The existing plant disease detection systems focus on only one particular plant such as tomato, corn, rice, etc.
6. There is no such system available which can detect leaf disease for many plants.

7. The accuracy of the chatbots providing regional language support is not very high due to the lack of proper datasets.
8. Most of the plant disease detection applications are for the house plants and the ones which are available for the crops are very less accurate.

C. *Project Contribution*

The proposed system aims to give a comprehensive solution to plant disease proliferation faced by farmers in India. The system will take in user input in the form of an image, and provide a consolidated analysis on the crop health. This analysis will help the farmers take quick action to save the crop and thus reduce the losses. Deep learning tools will be employed to produce the analysis. The project aims to deliver the information to the farmers in varied languages other than English. In order to reach out to the farmers from remote places in India who are not comfortable with English, the system will be providing multilingual support. An NLP based chatbot will be embedded in the website and android application to provide assistance in vernacular languages.

D. *Dataset Description:*

Total 9 plants are considered in the plant disease detection model. Including the different disease categories for each plant we have total 33 categories of images. The total number of training and testing images are 76180.

Disease Type	Training Images	Testing Images
Strawberry: healthy	1528	752
Strawberry: Leaf scorch	1486	731
Peach: Bacterial spot	1545	761
Peach: healthy	1447	712
Potato: Early blight	1624	799
Potato: healthy	1527	752
Potato: Late blight	1624	799
Corn (maize): Cercospora leaf spot Gray leaf spot	1374	677
Corn (maize): Common rust	1597	786
Corn (maize): healthy	1557	766
Corn (maize): Northern Leaf Blight	1597	787
Pepper, bell: Bacterial spot	1601	789
Pepper, bell: healthy	1664	820
Grape: Black rot	1581	778
Grape: Esca (Black Measles)	1614	795
Grape: healthy	1417	697
Grape: Leaf blight (Isariopsis Leaf Spot)	1441	710
Apple: Apple scab	1688	831
Apple: Black rot	1670	822
Apple: Cedar apple rust	1474	726
Apple: healthy	1681	828
Cherry (including sour): healthy	1528	753
Cherry (including sour): Powdery mildew	1409	694
Tomato: Tomato Yellow Leaf Curl Virus	1642	808
Tomato: Tomato mosaic virus	1499	738
Tomato: Target Spot	1530	753
Tomato: Spider mites Two-spotted spider mite	1457	718
Tomato: Septoria leaf spot	1474	726
Tomato: Leaf Mold	1575	776
Tomato: Late blight	1550	763
Tomato: healthy	1612	794
Tomato: Early blight	1608	792
Tomato: Bacterial spot	1425	701

Table. 1: Plant Disease Images Dataset Description

The dataset is customized using information from KCC, internet articles, FAQs, etc. This dataset provides structured information in a JSON format covering various topics related to agriculture. It includes different "intents(tags)" representing categories or themes, each with associated "patterns" that users might input to inquire about that particular topic. For each tag, there are corresponding "responses," which are concise answers or information related to the user's input.

The dataset covers a range of agricultural subjects, including plant diseases, farming practices, equipment, policies, and more. The tags associated with each intent serve as identifiers, allowing for easy reference and organization of the dataset.


```
683 patterns
227 tags: ['apple_additional_tips', 'apple_advantages', 'apple_black_rot_causes', 'apple_black_rot_cure', 'apple_black_rot_prevention', 'apple_black_rot_symptoms', 'apple
_common_challenges', 'apple_fertilization', 'apple_harvesting', 'apple_scab_causes', 'apple_scab_cure', 'apple_scab_prevention', 'apple_scab_symptoms', 'apple_seeds_condi
tions', 'apple_soil_conditions', 'apple_storage', 'apple_uses', 'apple_water_supply', 'askinghelp', 'bacterial_spot_causes', 'bacterial_spot_control', 'bacterial_spot_pre
vention', 'bacterial_spot_symptoms', 'bell_pepper_advantages', 'bell_pepper_bacterial_spot_causes', 'bell_pepper_bacterial_spot_cure', 'bell_pepper_bacterial_spot_prevent
ion', 'bell_pepper_bacterial_spot_symptoms', 'bell_pepper_challenges', 'bell_pepper_fertilization', 'bell_pepper_harvesting', 'bell_pepper_post_harvest_handling', 'bell_p
epper_seeds_conditions', 'bell_pepper_soil_conditions', 'bell_pepper_storage', 'bell_pepper_uses', 'bell_pepper_water_supply', 'bell_pepper_yield_tips', 'buy', 'buy_agric
ultural_equipment', 'cedar_apple_rust_causes', 'cedar_apple_rust_cure', 'cedar_apple_rust_prevention', 'cedar_apple_rust_symptoms', 'cercospora_leaf_spot_causes', 'cercos
pora_leaf_spot_cure', 'cercospora_leaf_spot_prevention', 'cercospora_leaf_spot_symptoms', 'cherry_plant_advantages', 'cherry_plant_common_challenges', 'cherry_plant_ferti
lization', 'cherry_plant_pollination', 'cherry_plant_pruning', 'cherry_plant_soil_conditions', 'cherry_plant_storage', 'cherry_plant_uses', 'cherry_plant_water_supply', '
cherry_plant_yield_max', 'cherry_powdery_mildew_causes', 'cherry_powdery_mildew_cure', 'cherry_powdery_mildew_prevention', 'cherry_powdery_mildew_symptoms', 'corn_additio
nal_tips', 'corn_advantages', 'corn_common_challenges', 'corn_common_rust_causes', 'corn_common_rust_cure', 'corn_common_rust_prevention', 'corn_common_rust_symptoms', 'c
orn_fertilization', 'corn_harvesting', 'corn_northern_leaf_blight_causes', 'corn_northern_leaf_blight_cure', 'corn_northern_leaf_blight_prevention', 'corn_northern_leaf_b
light_symptoms', 'corn_seeds_conditions', 'corn_soil_conditions', 'corn_storage', 'corn_uses', 'corn_water_supply', 'corn_yield_maximization', 'early_blight_causes', 'ear
ly_blight_control', 'early_blight_prevention', 'early_blight_symptoms', 'goodbye', 'grape_black_rot_causes', 'grape_black_rot_cure', 'grape_black_rot_prevention', 'grape_
black_rot_symptoms', 'grape_esca_causes', 'grape_esca_cure', 'grape_esca_prevention', 'grape_esca_symptoms', 'grape_leaf_blight_causes', 'grape_leaf_blight_cure', 'grape_
leaf_blight_prevention', 'grape_leaf_blight_symptoms', 'grapefruit_additional_tips', 'grapefruit_advantages', 'grapefruit_common_challenges', 'grapefruit_fertilization',
'grapefruit_harvesting', 'grapefruit_seeds_conditions', 'grapefruit_soil_conditions', 'grapefruit_storage', 'grapefruit_uses', 'grapefruit_water_supply', 'grapefruit_yiel
d_maximization', 'greeting', 'indian_farmer_policies', 'indian_farmer_rules', 'late_blight_causes', 'late_blight_control', 'late_blight_prevention', 'late_blight_symptoms
', 'leaf_mold_causes', 'leaf_mold_control', 'leaf_mold_prevention', 'leaf_mold_symptoms', 'peach_additional_tips', 'peach_advantages', 'peach_bacterial_spot_causes', 'pea
ch_bacterial_spot_cure', 'peach_bacterial_spot_prevention', 'peach_bacterial_spot_symptoms', 'peach_common_challenges', 'peach_fertilization', 'peach_harvesting', 'peach_
seeds_conditions', 'peach_soil_conditions', 'peach_storage', 'peach_uses', 'peach_water_supply', 'peach_yield_maximization', 'potato_additional_tips', 'potato_advantages'
, 'potato_common_challenges', 'potato_early_blight_causes', 'potato_early_blight_cure', 'potato_early_blight_prevention', 'potato_early_blight_symptoms', 'potato_fertiliz
ation', 'potato_harvesting', 'potato_late_blight_causes', 'potato_late_blight_cure', 'potato_late_blight_prevention', 'potato_late_blight_symptoms', 'potato_seeds_conditi
ons', 'potato_soil_conditions', 'potato_storage', 'potato_uses', 'potato_water_supply', 'potato_yield_maximization', 'sell', 'sell_agricultural_equipment', 'septoria_caus
es', 'septoria_control', 'septoria_prevention', 'septoria_symptoms', 'spider_mite_causes', 'spider_mite_control', 'spider_mite_prevention', 'spider_mite_symptoms', 'straw
berry_additional_tips', 'strawberry_advantages', 'strawberry_common_challenges', 'strawberry_fertilization', 'strawberry_harvesting', 'strawberry_leaf_scorch_causes', 'st
rawberry_leaf_scorch_cure', 'strawberry_leaf_scorch_prevention', 'strawberry_leaf_scorch_symptoms', 'strawberry_seeds_conditions', 'strawberry_soil_conditions', 'strawber
ry_storage', 'strawberry_uses', 'strawberry_water_supply', 'strawberry_yield_maximization', 'target_spot_causes', 'target_spot_control', 'target_spot_prevention', 'target
_spot_symptoms', 'thanks', 'tomatoblackrot', 'tomatoblights', 'tomatoblossomendrot', 'tomatocanning', 'tomatodiseaseprevention', 'tomatodiseases', 'tomatofertilization',
'tomatofertilizationdeficiencies', 'tomatofertilizer', 'tomatoflowering', 'tomatofruitsetting', 'tomatogrowingseason', 'tomatoharvesting', 'tomatohumidity', 'tomatoleafcu
rl', 'tomatolightrequirements', 'tomatonutrientdeficiency', 'tomatopaste', 'tomatopests', 'tomatopestscontrol', 'tomatoplanting', 'tomatopowderymildew', 'tomatopreservati
on', 'tomatopruning', 'tomatosauce', 'tomatoseeds', 'tomatosoil', 'tomatostorage', 'tomatosupport', 'tomatotemperature', 'tomatotransplanting', 'tomatotrellis', 'tomatotr
ellising', 'tomatovarieties', 'tomatowatering', 'tomv_causes', 'tomv_cure', 'tomv_prevention', 'tomv_symptoms', 'tylcv_causes', 'tylcv_cure', 'tylcv_prevention', 'tylcv_s
ymptoms']
```


Fig 1: Dataset showing 683 Patterns and 227 Tags

IV. PROPOSED DESIGN

A. *Architecture*

The section explains the architecture and gives a granular view of the working modules. The work flow followed throughout the project is also elaborated.

4.1 Block Diagram :

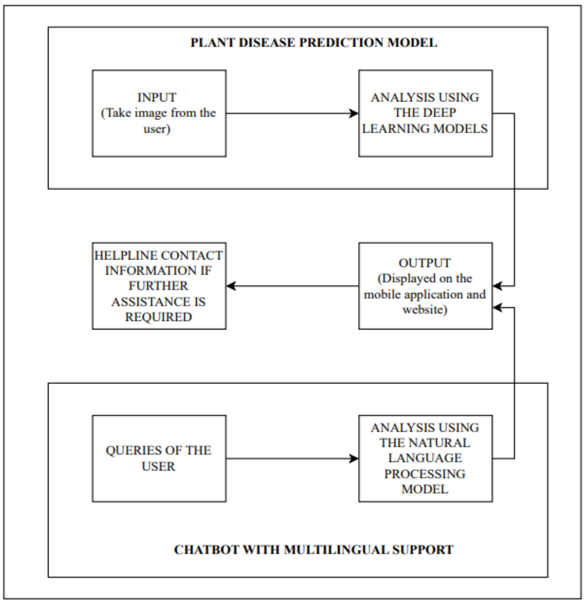

Figure 2 : Block Diagram

The entire system is divided into 3 major parts – Plant Disease Prediction Model, Chatbot with multilingual support and Mobile and Web Application.

Plant Disease Prediction Model :

1. The input image is taken from the user. The image is of the leaf of the plant which is to be tested.
2. Analysis of the image is done using the Deep Learning models. The models are trained by using the Plant Village dataset which has approximately 14 different varieties of plant leaf images. The different stages for building the model are – Data Preprocessing and Cleaning, Quantization, Deployment, etc. The output of the final analysis is displayed on the mobile and web application.

Chatbot with Multilingual Support :

1. The user can type the query in the chatbot. The language options given are English, Hindi, Marathi, etc.
2. Analysis of the query is done using the Natural Language Processing model. The model is trained using the data from Kisan Call Centre (KCC), Open Government Data (OGD) Platform India, etc. The different stages for building the model are – Language Understanding, Information Retrieval, Response Generation, etc. The response is displayed on the mobile and web application.

Mobile and Web Application :

1. The mobile and web applications have a very easy to navigate user interface. The output of both the models is displayed.
2. If further assistance is required then the helpline contact information is also available.

B. *Modular Diagram*

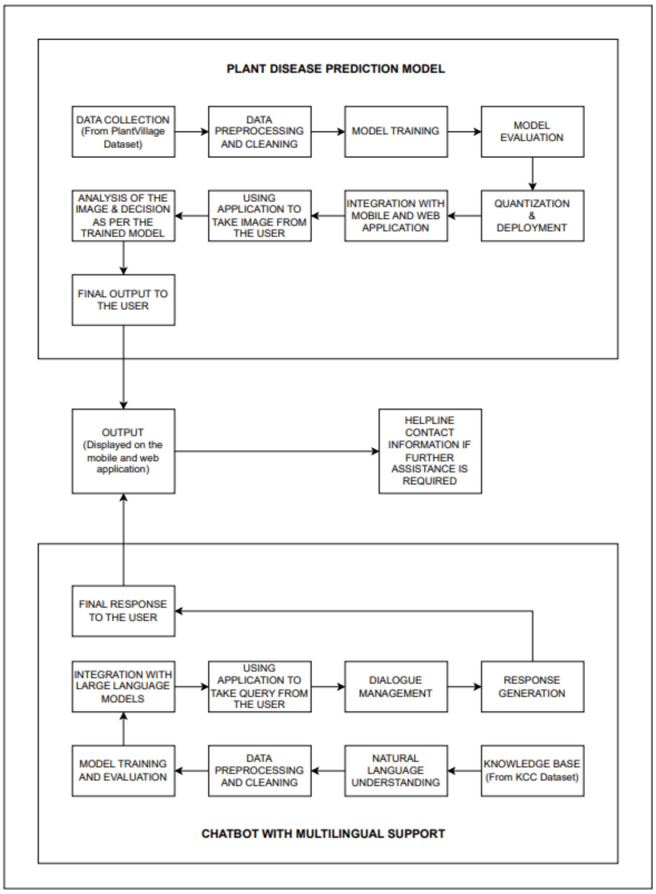

Figure 3 : Modular Diagram

Plant Disease Prediction Model

1. Data Collection :

Gathering a comprehensive dataset of plant images, including healthy and diseased samples, is essential. PlantVillage dataset is used which contains 9 types of plants along with several diseases pertaining to each plant.
Collect data related to different plant varieties and diseases.

2. Data Preprocessing and Cleaning :

This block involves tasks like image resizing, noise reduction, and data augmentation to prepare the dataset for training.
Data cleaning ensures high-quality input for the model.

3. Model Training :

Train deep learning models (e.g., Convolutional Neural Networks - CNNs) using the preprocessed dataset.
The training process involves feature extraction and model optimization.

4. Model Evaluation :

Assess the model's performance using validation datasets and metrics such as accuracy, precision, recall, and F1-score.
Fine-tune the model based on evaluation results.

5. Quantization & Deployment :

Quantization is a technique to optimise the model for deployment on resource-constrained devices (e.g., mobile devices).
Deploying the model to production environments using tflite.

6. Integration with mobile and web application :

Ensuring seamless integration of the Plant Disease Prediction Model with the Mobile and Web Application.
Developing APIs or endpoints for communication between the application and the model.

7. Input image from user :

Implement functionality in the application to allow users to upload plant images for analysis.
Process user input to prepare images for prediction.

8. Analysis of the image :

The inference engine runs predictions on user-submitted images using the trained model.
It interprets the model's output, determining the likelihood and type of disease.

9. Final output to the user :

Present the analysis results to users in an understandable format.
Include information about the detected disease, severity, and recommended actions.
Present the analysis results to users in an understandable format.
Include information about the detected disease, severity, and recommended actions.

Chatbot with Multilingual Support

1. Knowledge Base :

This contains the information and data that the chatbot uses to respond to user queries.
It can include structured databases, FAQs, documents, or other data sources.

2. Natural Language Understanding :

NLU is responsible for processing user queries and understanding their intent, entities, and context.
Building blocks within NLU may include:
Intent Recognition: Identifying the user's intent or purpose behind the query.
Entity Extraction: Extracting specific pieces of information from the user's query.
Context Management: Maintaining context throughout the conversation for more coherent responses.

3. Data preprocessing and Cleaning :

This block involves tasks like stemming, stop-words removal, etc. to prepare the dataset for training.
Data cleaning ensures high-quality input for the model.

4. Model Training and Evaluation :

Training the model using Artificial Neural Networks (ANN) and Natural Language Processing (NLP). Evaluating the model using parameters like accuracy, loss function, etc.

5. Integration with Large Language Models :

Creating a large language model. This block involves integrating with external services or APIs to leverage the model's capabilities.

6. Query from user :

This is the user-facing component where users interact with the chatbot.
It includes elements like chat windows, input fields, buttons, and menus for user interaction.

7. Dialogue Management :

Dialogue management controls the flow of the conversation and decides what responses to generate.
It may involve a rules-based system, a decision tree, or more advanced techniques like reinforcement learning for dialogue optimization.

8. Response Generation :

This block creates the chatbot's responses based on the user's query and the knowledge base.
Techniques may include template-based responses, dynamic content insertion, or even natural language generation using large language models.

9. Final Response to the user :

This block involves tracking the performance of the chatbot, including user interactions, errors, and user satisfaction.
Analytics help improve the chatbot's capabilities over time.

Output in the desired format :
Output for the image analysis and the chatbot in the proper format is presented to the users via the mobile and web applications.

Helpline numbers and further assistance if required :

This block focuses on the fact that some of the queries cannot be solved by our chatbot hence further assistance is required. In case of disease detection, for further purchase of pesticides, fertilisers, etc requires contact information.

C. *Requirements*

<u>Software Specifications</u>

Language: Python, Java, XML
IDE: Jupyter Notebook, Google Colab, Atom, Visual Studio Code, PyCharm

To build the application: Android, Java
Python basic stack:

1. Scikit learn
2. NumPy
3. Matplotlib
4. Pandas
5. Keras
6. TensorFlow

V. PROPOSED ALGORITHM

1. NLP Pipeline

The different phases of the NLP pipeline include - Tokenization, converting to lower case, stemming, excluding punctuation characters and applying the Bag Of Words approach. Each phase in detail is as follows -

- Tokenization : The entire query which was passed as a string is converted into different tokens.
- Lower case + Stemming : The tokens are converted to lowercase and then stemming is performed using the PorterStemmer Algorithm.
- Excluding punctuation characters : Punctuation characters such as "?,:,;,,,*,<,>" are excluded for smooth functioning and for the bag of words approach.
- Bag of Words : Bag of Words Approach is used to and after using softmax the final probability is calculated. If the probability is greater than 0.9 then the query is classified into that tag and the appropriate response from the response array is returned.

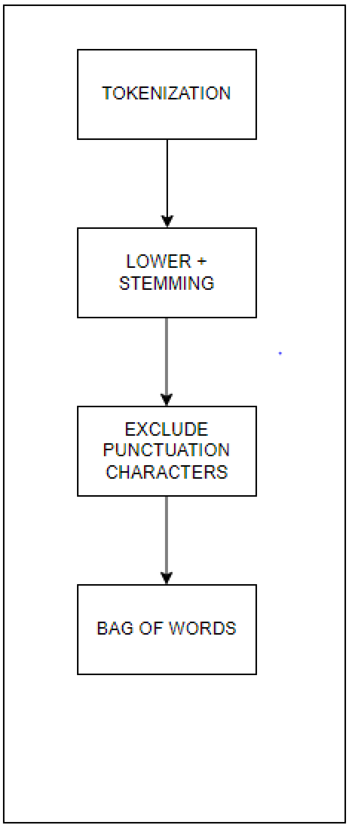

Figure 4 : NLP Pipeline

2. Chatbot Pipeline

The different phases of the chatbot pipeline is Training data, Train-Test Split, Bag of Words Approach, Feed Forward Neural Network, Output. Each phase in detail is as follows -

- Training data : The training data is in the json format. The json file has several tags along with the query and the responses. This is used for training the ANN model.
- Train-Test Split : The queries and responses are used as the input data and the tags are the predicted output.
- Bag Of Words : Bag of Words Approach is used to and after using softmax the final probability is calculated. If the probability is greater than 0.9 then the query is classified into that tag and the appropriate response from the response array is returned.
- Feed Forward Neural Network : The neural network used has 3 layers. Then softmax is used for calculating the probabilities and the final output.
- Output : The output is the tag if the probability calculated is greater than 0.9.

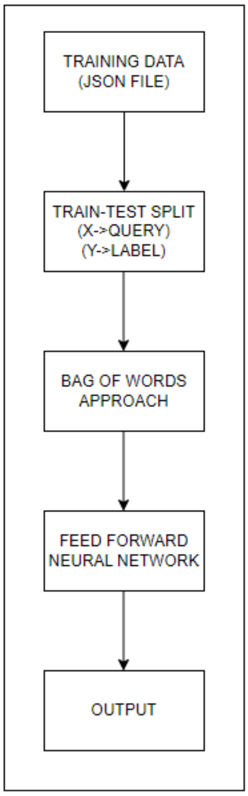

Figure 5 : Chatbot Pipeline

3. Feed Forward Neural Network

- Model Architecture:
  The model has three linear layers :
  input_size: The size of the input features.
  hidden_size: The size of the hidden layer.
  num_classes: The number of output classes.
- Activation Function:
  The Rectified Linear Unit (ReLU) activation function is applied after each linear layer.
  ReLU introduces non-linearity to the model, allowing it to learn complex patterns in the data.
- Forward Pass:
  In the forward method, the input tensor x is passed through the linear layers and ReLU activation functions.
  The output of the final linear layer is returned without applying any activation function.
- Softmax:
  Softmax activation is typically used in the final layer for multi-class classification tasks to obtain class probabilities.

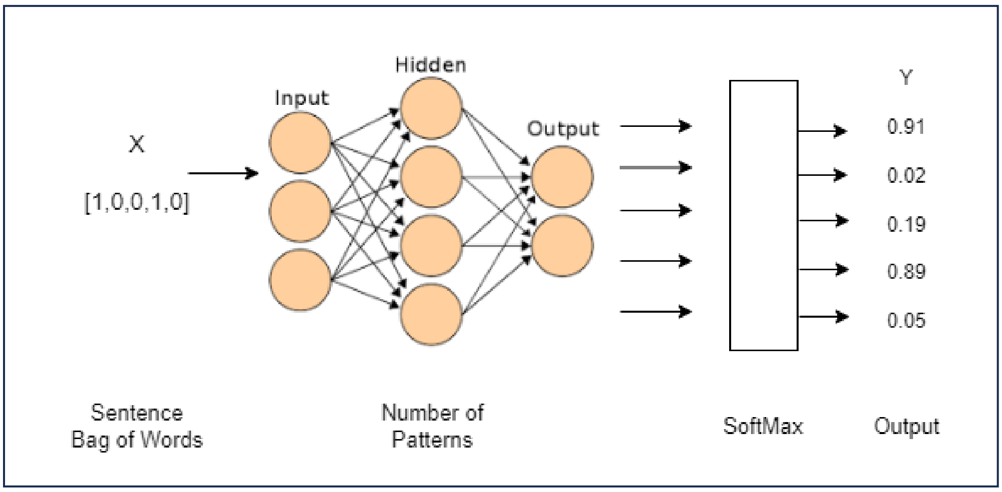

Figure 6 : Feed Forward Neural Network

VI. RESULTS AND EVALUATIONS

A. *Confusion matrix:*

Confusion Matrix is a machine learning method used to measure a classifier's performance. It helps to visualise and summarise the performance of a classification algorithm. It plots actual vs. predicted i.e., actual classes vs. the classes predicted by the model.

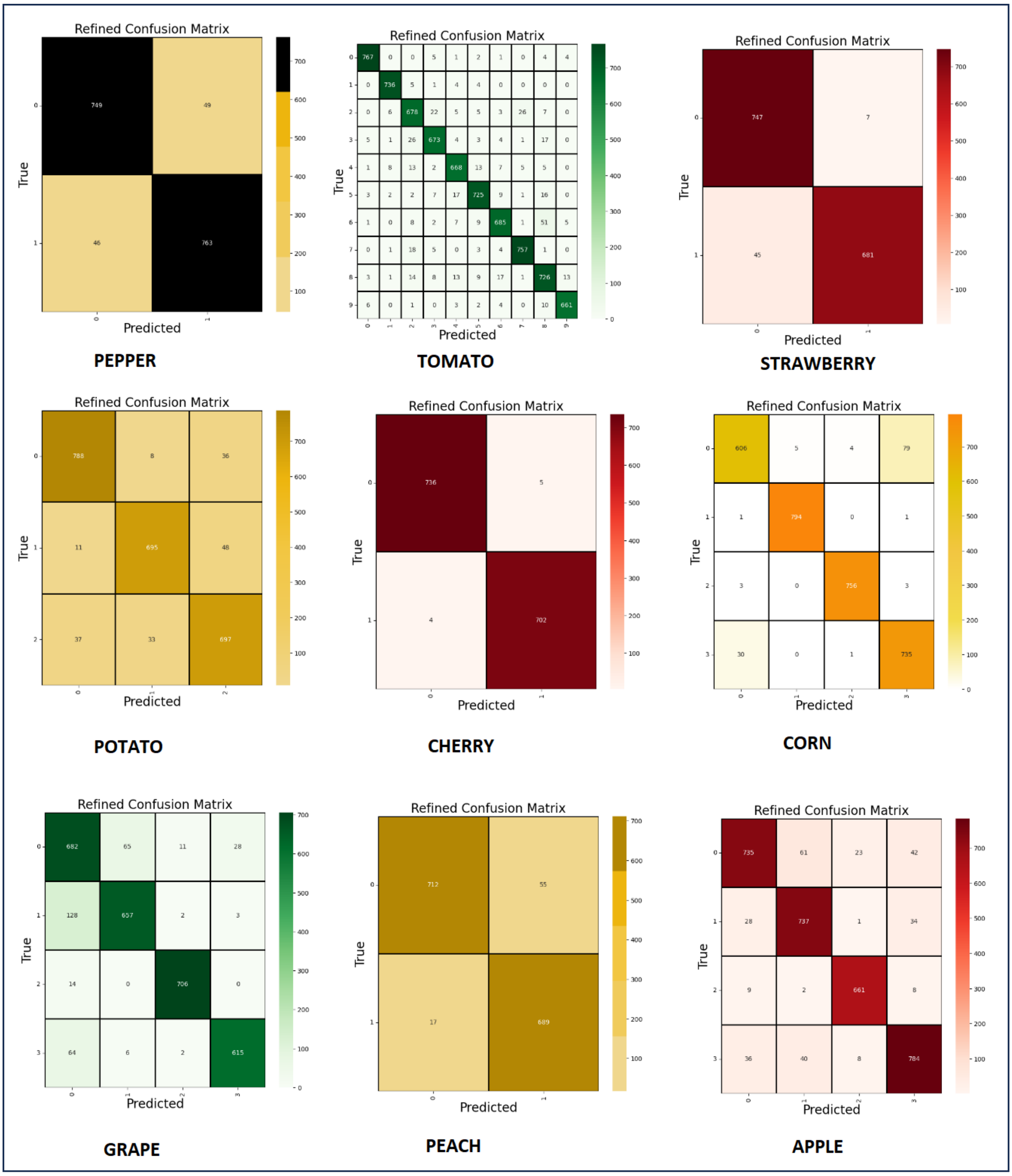

Figure 7 : Confusion Matrix

B. *Chatbot Loss Function:*

- Number of Epochs: 1000 epochs are used for training the chatbot. Epochs refer to the number of times the entire training dataset is passed through the neural network. Typically, multiple epochs are used to allow the model to learn patterns in the data.

- Loss Function: The loss function is a measure of how well the model is performing. A loss of 0.0000 is ideal and suggests that the model has perfectly learned the training data. However, in real-world scenarios, achieving exactly 0.0000 loss is rare and could potentially indicate overfitting to the training data.

```
272 227
Epoch [100/1000], Loss: 0.1749
Epoch [200/1000], Loss: 0.0011
Epoch [300/1000], Loss: 0.0001
Epoch [400/1000], Loss: 0.0000
Epoch [500/1000], Loss: 0.0000
Epoch [600/1000], Loss: 0.0000
Epoch [700/1000], Loss: 0.0000
Epoch [800/1000], Loss: 0.0000
Epoch [900/1000], Loss: 0.0000
Epoch [1000/1000], Loss: 0.0000
final loss: 0.0000
training complete. file saved to data.pth
```

Figure 8 : Chatbot Performance

VII. CONCLUSION AND FUTURE WORK

The proposed project aims to address the pressing issue of plant diseases and their detrimental impact on farmers in India. By leveraging deep learning tools and natural language processing techniques, the system endeavours to provide quick and accurate analysis of plant health based on leaf images, assisting farmers in combating diseases and preventing losses. The incorporation of regional languages, starting with Marathi and Hindi, demonstrates a commitment to catering to the vernacular audience and ensuring accessibility for farmers across diverse linguistic backgrounds. The implementation of a user-friendly website and android application with a multilingual chatbot further enhances the platform's usability and effectiveness. Overall, this project's objective is to provide better agricultural assistance to farmers in their regional languages, significantly contributing to reducing losses caused by disease proliferation and empowering farmers with vital information for improved crop management.

Future Work

- To begin with the Chatbot implementation, focus will primarily be on Marathi and Hindi and later the scope will be broadened to different state languages.
- Extra Features such as information regarding the price and demand of certain crops in the available markets will be provided.
- Quick access to various government schemes and facilities available for the farmers will be provided.
- The disease predicting model will be expanded to more species of plants and other types of vegetations.

VIII. ACKNOWLEDGEMENT

We would like to express our special thanks to our mentor **Dr. Mrs. Sujata Khedkar** for giving us this golden opportunity to do this project on the topic **Plant Disease Detection.** The valuable guidance and suggestions of our mentor helped us in various phases of the project.

This project helped us in enhancing our concepts and research new streams of technology.

We are very thankful to our HoD **Dr. (Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair** for their guidance and support throughout this project.

IX. REFERENCES

[1] P. Y. Niranjan, V. S. Rajpurohit and R. Malgi, "A Survey on Chat-Bot system for Agriculture Domain," 2019 1st International Conference on Advances in Information Technology (ICAIT), Chikmagalur, India, 2019, pp. 99-103, doi: 10.1109/ICAIT47043.2019.8987429.

[2] D. Sawant, A. Jaiswal, J. Singh and P. Shah, "AgriBot - An intelligent interactive interface to assist farmers in agricultural activities," 2019 IEEE Bombay Section Signature Conference (IBSSC), Mumbai, India, 2019, pp. 1-6, doi: 10.1109/IBSSC47189.2019.8973066.

[3] U. Kiruthika, S. K. S. Raja, V. Balaji and C. J. Raman, "E-Agriculture for Direct Marketing of Food Crops Using Chatbots," 2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India, 2020, pp. 1-4, doi: 10.1109/ICPECTS49113.2020.9337024.

[4] S. V. Militante, B. D. Gerardo and N. V. Dionisio, "Plant Leaf Detection and Disease Recognition using Deep Learning," 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), Yunlin, Taiwan, 2019, pp. 579-582, doi: 10.1109/ECICE47484.2019.8942686.

[5] H. Andrianto, Suhardi, A. Faizal and F. Armandika, "Smartphone Application for Deep Learning-Based Rice Plant Disease Detection," 2020 International Conference on Information Technology Systems and Innovation (ICITSI), Bandung, Indonesia, 2020, pp. 387-392, doi: 10.1109/ICITSI50517.2020.9264942.

[6] N. Saranya, L. Pavithra, N. Kanthimathi, B. Ragavi and P. Sandhiyadevi, "Detection of Banana Leaf and Fruit Diseases Using Neural Networks," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 493-499, doi: 10.1109/ICIRCA48905.2020.9183006.

[7] R. D. Devi, S. A. Nandhini, R. Hemalatha and S. Radha, "IoT Enabled Efficient Detection and Classification of Plant Diseases for Agricultural Applications," 2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), Chennai, India, 2019, pp. 447-451, doi: 10.1109/WiSPNET45539.2019.9032727.

[8] Kirti and N. Rajpal, "Black Rot Disease Detection in Grape Plant (Vitis vinifera) Using Colour Based Segmentation & Machine Learning," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 2020, pp. 976-979, doi: 10.1109/ICACCCN51052.2020.9362812.

[9] J. S. H. Al-bayati and B. B. Üstündağ, "Artificial Intelligence in Smart Agriculture: Modified Evolutionary Optimization Approach for Plant Disease Identification," 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Istanbul, Turkey, 2020, pp. 1-6, doi: 10.1109/ISMSIT50672.2020.9255323.

[10] M. A. Rahman, M. M. Islam, G. M. Shahir Mahdee and M. W. Ul Kabir, "Improved Segmentation Approach for Plant Disease Detection," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-5, doi: 10.1109/ICASERT.2019.8934895.

[11] B. S. Kusumo, A. Heryana, O. Mahendra and H. F. Pardede, "Machine Learning-based for Automatic Detection of Corn-Plant Diseases Using Image Processing," 2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Tangerang, Indonesia, 2018, pp. 93-97, doi: 10.1109/IC3INA.2018.8629507.

[12] S. Y. Yadhav, T. Senthilkumar, S. Jayanthy and J. J. A. Kovilpillai, "Plant Disease Detection and Classification using CNN Model with Optimized Activation Function," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 564-569, doi: 10.1109/ICESC48915.2020.9155815.

[13] L. Li, S. Zhang and B. Wang, "Plant Disease Detection and Classification by Deep Learning—A Review," in IEEE Access, vol. 9, pp. 56683-56698, 2021, doi: 10.1109/ACCESS.2021.3069646.

[14] T. Maginga, J. Nsenga, P. Bakunzibake and E. Masabo, "Smallholder farmer-centric integration of IoT and Chatbot for early Maize diseases detection and management in pre-visual symptoms phase," 2022 IEEE Global Humanitarian Technology Conference (GHTC), Santa Clara, CA, USA, 2022, pp. 369-372, doi: 10.1109/GHTC55712.2022.9911047.

[15] V. Rajesh Kumar, K. Pradeepan, S. Praveen, M. Rohith and V. Vasantha Kumar, "Identification of Plant Diseases Using Image Processing and Image Recognition," 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), Puducherry, India, 2021, pp. 1-4, doi: 10.1109/ICSCAN53069.2021.9526493.

[16] Deepa, R. N and C. Shetty, "A Machine Learning Technique for Identification of Plant Diseases in Leaves," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 481-484, doi: 10.1109/ICICT50816.2021.9358797.