PLANT DISEASE DETECTION USING RESNET AND XGBOOST

Abstract-The Model is to apply Convolutional Neural Network (CNN) and XGBoost algorithms to diagnose plant sicknesses and remedy agricultural issues. Traditional manual examination is useless for prognosis, requiring automatic image analysis. Diseases reason good sized yield losses globally, affecting food security. Machine learning offers an answer, but challenges remain: heterogeneous leaf picture prevent model generalization, confined label facts make training difficult, and the computational depth of deep mastering models creates deployment challenges. Despite those obstacles, the usage of these technologies may want to revolutionize the control of plant sicknesses and conserve agricultural yields essential for international food protection.

Keywords: Machine Learning (ML), Deep Learning, Convolutional Neural Networks (CNN), XGBoost, Disease of Plant.

I. INTRODUCTION

Plant diseases jeopardize global food security, causing up to 40 percent annual crop yield losses per FAO estimates. This substantial impact necessitates robust disease detection methods. Leveraging Convolutional Neural Networks (CNN) and XGBoost, this project aims to develop an advanced system for accurate plant disease analyses through image analysis. However, challenges such as image variability, limited labeled data, and computational complexity pose obstacles to effective implementation. Overcoming these hurdles is crucial for enhancing agricultural sustainability and food security worldwide. Used the fusion of resnet 18 and XGBoost as the custom model for the web application and the plants names are apple, blueberry, cherry, corn, grape, orange, peach, potato and tomato.

II. OBJECTIVE

Create a numerous dataset along with images of healthful and diseased flowers from one-of-a-kind plants. Create an adaptive CNN to extract complicated features from image datasets. Integrate XGBoost for accurate sickness type and boom accuracy with ensemble gaining knowledge of. Refine the version through cautious assessment, validation and hyperparameter tuning. Ensure model robustness and generalization throughout plant species and sickness scenarios for robust identification competencies.

III. PROBLEM *STATEMENT*

Plant diseases significantly affect crop yield, resulting in economic losses and jeopardizing food security, particularly in developing nations. Current Traditional disease detection methods reliant on visual inspection by experts are time-consuming, error-prone, and inadequate for large-scale agricultural operations. Machine Learning Solution: Leveraging Convolutional Neural Networks (CNN) and XGBoost for disease detection introduces potential solutions but faces challenges Variability in Leaf Images, Limited Labeled Data, Computational Complexity.

IV. LITERATURE *SURVEY*

1. Plant Foliar Disease Detection Using Small Approach (2021).

This gets rid of the limitations of People's creativity experimenting with expensive lab techniques. Analyzes SVM and Random Forest algorithms and compares accuracy charges to help farmers reduce disorder detection time and expenses. Proposed Automated Plant Leaf Disease Detection System Using Machine Learning Algorithms Including SVM and Random Forest. It have a look at productiveness metrics and accurate results to assist farmers perceive diseases faster and more price-successfully, thereby improving farm productivity. A confined range of ML algorithms for disease detection may additionally forget other useful mechanisms. In addition, the dataset used inside the study won't all plant sicknesses, which limits the generalizability of the effects. Future employment opportunities for plant leaf ailment diagnosis using ML encompass deep studying fashions, dataset augmentation, multispectral imaging, real-time detection and transfer learning to know the use of drones.

2. Machine Learning Approach for Detection and Classification of Exudative Diseases in Rice (2022)

Diagnosis usage of image processing and SVM class methods. It detects 5 styles of rice sicknesses, demonstrating its potential for realistic use in agriculture. Using a restrained dataset for education and testing reduces generalizations between one of a kind types of rice illnesses. Expand the dataset to encompass greater types of sicknesses and environmental changes. Learning different superior devices and getting to know algorithms consisting of deep getting to know fashions can further improve accuracy.

3. Machine Learning XGBoost and Random Forest Based Methods for Crop and Fertilizer Prediction – (2022)

Face the challenges related to variable soil conditions, variety of crop types and converting climatic situations that restrict its practical software. The article also provides ailment detection the use of leaf images. There is no particular analysis of the impact of external factors on overall performance expectations. It is no manner covers the proposed actual-time data system. Extend the proposed system to handle exclusive soil kinds and regional differences. It is crucial to enhance the accuracy of disorder diagnosis and use actual-time weather data for dynamic forecasting.

4. Detection and Prevention of Plant Diseases Using Machine Learning (2022)

In the agricultural area, speedy and correct detection of plant sicknesses is crucial to reduce meals losses and boom crop yields. Population growth and climate exchange have led to the unfold of fungal and bacterial illnesses in flowers, posing a chief challenge to sustainable agriculture. A novel technique for early detection of plant sicknesses using progressed k-means clustering and coloration-primarily based segmentation is supplied. It explores image processing and facts mining techniques to diagnose and classify numerous sicknesses. There aren't enough in-intensity studies on the scalability and real-world troubles related to the implementation of the proposed method. It additionally does no longer keep in mind the impact of mild situations on the accuracy of disease diagnosis and variations between extraordinary plant species. Focus on making the proposed system extra bendy to special lighting fixtures situations and plant species. Accuracy can be improved via incorporating extra advanced deep gaining knowledge of strategies.

5. Detection of Multiple Crop Diseases Using Image Processing Techniques - (2021)

It proposes a photograph processing primarily based approach for automated multi-crop sickness detection in cucumber, guava, groundnut, pumpkin and cowpea leaves. Provides a green method for automated crop disorder diagnosis the usage of picture processing strategies to discover numerous leaf sicknesses of cucumber, guava, groundnut, pumpkin and cowpea. The small length of the dataset influences the generalizability of the proposed approach to a huge range of sicknesses in real time, the usage of agricultural drones or clever farming structures, it's miles feasible to combat crop diseases efficiently and timely. Additionally, gadget accuracy can be stricken by adjustments in lighting conditions and image first-rate. To improve the generalization of the model, you may cognizance on increasing the scale of the dataset. Diagnostic integration into plants.

V. EXISTING SYSTEM

Existing plant ailment analysis answers regularly rely upon guide monitoring and diagnosis by means of agricultural specialists, which is time-ingesting, subjective and impractical for huge-scale tracking. Some computerized systems use conventional machine getting to know algorithms, however these may additionally lack the robustness and accuracy had to comprehensively diagnose diseases throughout distinct plant species and environmental conditions.

VI. PROPOSED SYSTEM

The proposed answer makes use of CNN and XGBoost feature extraction for accurate disease classification in plant pictures, correctly distinguishing wholesome from diseased flora. The essential targets are to collect a diverse dataset, teach a sturdy CNN for characteristic extraction, and integrate XGBoost for advanced categories. Furthermore, improving the model through cautious hyperparameter estimation, validation and tuning guarantees its robustness and generalizability to exclusive plant species and disease occasions, promising a complete and efficient disorder diagnosis system.

VII. SYSTEM REQUIREMENTS

Hardware Requirements

Procesor: intel i3 3rd gen

CPU: Nvidia geforce gtx 650 or AMD raedon rv200

Ram: 4GB

Hard Disk: 4 GB

Software Requirements

Python environment with required libraries (TensorFlow, Keras, scikit-learn, Numpy, Pandas, torch), flask, cuda software

VIII. BLOCK DIAGRAM

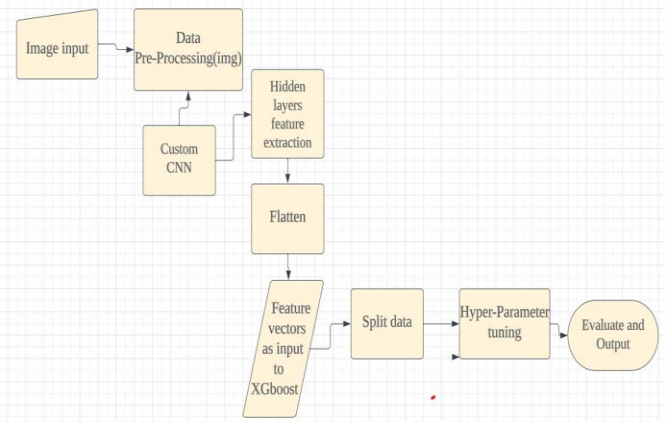


fig (i) Block Diagram of the output

Data Preprocessing

In this step, the leaf pictures are pre-processed to enhance the performance of the CNN model. These consist of resizing pictures, normalizing pictures values, and casting off noise.

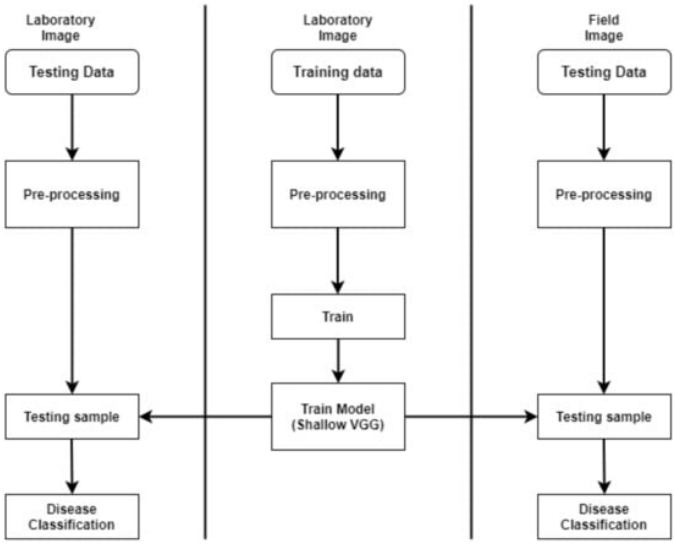


fig (ii) Data preprocessing of the model

Classification

A CNN model is used to extract features from pre-processed leaf photographs. These features are high-level representations of photograph content material vital for disorder type. The extracted capabilities are used to teach the XGBoost classifier. This classifier learns to understand the relationship between symptoms and disorder labels. After the classifier is trained, pictures of sparkling leaves may be used to categories healthful and diseased varieties.

IX. PROPOSED ALGORITHMS

1. Convolutional Neural Networks (CNNs)

A Convolutional Neural Network (CNN) is a form of gadget getting to know model, a type of deep learning set of rules mainly suitable to classifying visual data. Although arrays are mostly used for photo processing, they also can be adapted to work with information and other sign statistics.

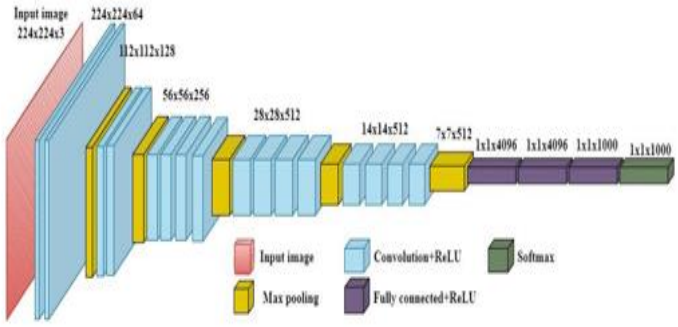
Architecture is based totally at the communique equipment of the human mind, specifically the visual cortex, which performs a key role in the belief and processing of visible stimuli. An artificial neuron in CNN is designed to effectively procedure visible records, permitting these fashions to method entire snap shots. Because arrays are so beneficial in identifying objects, they are regularly utilized in imaginative and prescient for image recognition and object detection. Although arrays are frequently used for image processing, they also can be adapted to work with audio information and other signal records. The convolution operation paperwork the idea of any convolutional neural networks.

- Input Leaf image (PNG or JPG or JPEG)
- Output Disease Name
- Procedure Image Pre-processing (Leaf Images, Output Directories)
- for each class directory in the input dataset do
- for each image file in the class directory do
- Load image using OpenCV library
- Resize image to (256, 256) using bilinear interpolation
- center crop to 224 and random horizontal flip
- transform using tensor and normalize using the mean and std deviation
- separate the dicts for image paths and labels
- return image and label using cv2 and OS
- create data loaders using the parameters: batch size=32, shuffle, workers=4 in train, validation
- create feature extractor class using resnet18
- create empty dicts for features and labels and evaluate models
- for inputs, label batch in data loader do
- send inputs, labels batch and append them using NumPy
- return vertical stack and horizontal stack

CNNs, predominant in recognition and classification, outperform hand crafted methods by learning robust features directly from input images. Popular models like VGG, ResNet18, and Dense Net excel in plant disease identification. Research demonstrates their superior accuracy, with CNNs comprising convolutional, pooling, and fully connected layers for effective classification.

Prerequisites

- Python environment with the OpenCV and torch libraries installed.
- Python libraries such as TensorFlow, keras.



fig(iii) Model of Convolutional Neural Network (CNN)

2. *XGBoost*

XGBoost is a gradient boosting algorithm widely utilized in statistics science. It is an implementation of gradient boosting this is designed to be especially efficient, bendy, and compact. XGBoost became originally evolved through Tianqi Chen as a development of the GBM algorithm. XGBoost is rapid and efficient. It can without difficulty deal with big records sets and has been proven to be faster than different algorithms. XGBoost is greater accurate. When you have a variety of observations inside the schooling statistics. Number of observations < Number of observations within the education records. It works properly while the facts includes numeric and categorical statistics or numeric information only. When you keep in mind a metric to build a model.

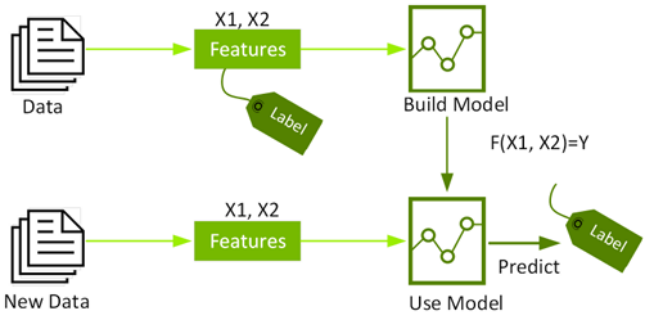
XGBoost or Extreme Gradient Boosting is an ensemble device getting to know set of rules. This technique is used to display gaining knowledge of difficulties which include retardation and segmentation. XGBoost iteratively combines the predictions of several awesome models to shape a predictive model, often decision timber.

- Input feature vectors from the returned feature matrix
- Output Trained CONRXG Model and Evaluation Metrics
- Procedure Xgboost training (Preprocessed Images, Labels, Hyperparameters)
- Split the dataset into training, validation, and test sets
- initialize the Xgboost classifier and fit the train features and train labels
- Specify hyperparameters like learning rate, optimizer, and batch size based on optimization trials
- Define a learning rate scheduler function for dynamic adjustment during training
- Compile the model using categorical loss and F1 score
- Implement early stopping and model checkpoint to monitor validation loss and save the best model
- Train the model on the training set with specified hyperparameters and callbacks
- Evaluate the trained model on the test set and compute accuracy, F1 score, precision, recall, and confusion matrix
- Visualize training and validation metrics, including loss, accuracy, and F1 scoreal-culate distance maps using cosine distance metric
- using joblib, save the model as a pkl file so that it will be easy for the web application to use
- end procedure.

The feature matrix, derived from images, serves as XGBoost's input, each row representing an image and columns representing features. The binary classification (0,1) denotes healthy or diseased status. The dataset undergoes division into 8 train, validation, and test sets. Hyperparameters like learning rate, maximum depth, and estimators are tuned using grid search for optimal performance. Finally, the model undergoes evaluation and testing to assess its efficiency.

Prerequisites

- Preprocessed image and corresponding labels.
- Python environment with necessary libraries installed, including TensorFlow, Keras, and scikit-learn, torch, os, cuda.



fig(iv) Model of XGBoost

Web Application Using Flask:

Flask is a small and lightweight Python web framework that provides useful tools and features that make creating web applications in Python easier. It gives developers flexibility and is a more accessible framework for new developers since you can build a web application quickly using only a single Python file. Flask is also extensible and doesn't force a particular directory structure or require complicated boilerplate code before getting started.

Algorithms used - CNN and XGBoost

1) VGG19 in CNN and Xgboost.

2) RESNET 18 in CNN and Xgboost.

We tried both 1 and 2 and algo 1 got 75% and algo 2 got 90% accuracy so our new custom algorithm is resnet18 (CNN) x Xgboost and is named as CX.

VGG 19

1. Data Preparation:

Imports libraries for data manipulation, model building, and evaluation. Lists files in the dataset directory (optional for checking structure). Creates data augmentation generators for training and validation sets (optional): Applies random transformations (zoom, shear, flip) to training images for data augmentation. Applies preprocessing (normalization) to both training and validation images. Loads training and validation data using Image Data Generator. flow_from_directory (): Resizes images to a fixed size (256x256 pixels). Batches images for efficient training.

2. Model Building:

Loads the pre-trained VGG19 model (excluding the final classification layers) for feature extraction. Freezes the weights of the pre-trained model to prevent overfitting. Adds a new dense layer with a softmax activation for multi-class classification (predicting disease type). Compiles the model with: Adam optimizer for training. Categorical cross entropy loss function for multi-class classification. Accuracy metric to monitor training performance.

3. Training with Regularization:

Implements early stopping to prevent overfitting: Stops training if validation accuracy doesn't improve for a set number of epochs. Implements model checkpointing: Saves the

model with the best validation accuracy achieved during training. Trains the model on the prepared training data with validation data for performance monitoring.

4. Evaluation and Prediction:

Loads the best model saved during training. Evaluates the model's accuracy on the validation data. Defines a function prediction(path) to predict disease class for a given image: Loads and preprocesses the image. Makes a prediction using the trained model. Returns the predicted disease class based on a class index mapping. Optionally iterates over test images, predicts their disease classes, and displays results.

RESNET 18

ResNet18 is a seventy two-layer structure 18 layers deep. The structure of this network is designed for green operation with a massive number of convolutional layers. However, adding too many deep layers to the network often effects in distorted output. This is referred to as the vanishing gradient trouble, wherein neural networks learn by way of back propagation in gradient descent, averting the loss characteristic to locate the least weight. With extra layers, the multiplication aspect slope decreases and will become smaller, as a consequence "reducing off", which leads to saturation of the community overall performance or reduced performance.

The core concept of ResNet is the usage of shipping hyperlinks, usually called quick links or identification hyperlinks. These connections basically paintings via switching among one or more layers, growing shortcuts between those layers. The cause of introducing those short links is to resolve the principle trouble of the vanishing gradient of that deep facial community. Quick links dispose of the hassle of the slide disappearing again the usage of the previous layer's movements.

1. Data Preparation:

Defines directories containing training images categorized by disease type. Creates functions to preprocess images including resizing, random cropping, flipping, converting to tensors, and normalization. Creates a custom dataset class to load images and labels based on folder structure. Creates data loaders to manage batches of images and labels for training and validation.

2. Feature Extraction using a Pretrained CNN:

Defines a Feature Extractor class that utilizes a pre-trained ResNet18 convolutional neural network (CNN) model. Freezes the pre-trained model weights to avoid retraining them. Extracts features from the data using the feature extractor, resulting in a reduced representation of the images.

3. Train XGBoost Model:

Trains an XGBoost model, a machine learning algorithm for classification, using the extracted features as input and disease labels as targets.

4. Model Evaluation:

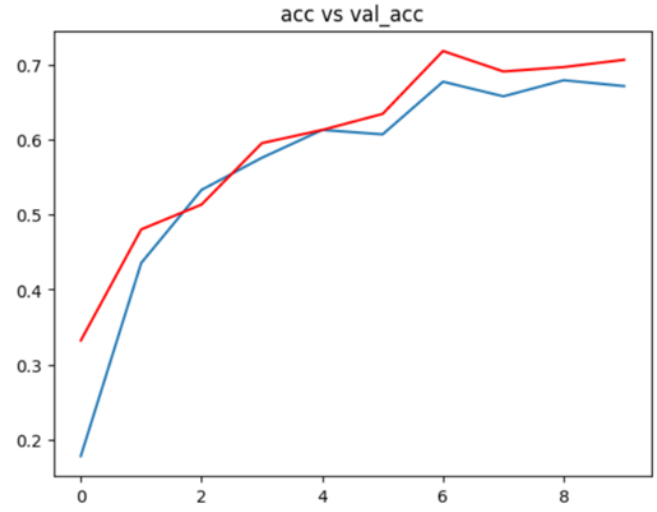
Evaluates the trained XGBoost model on unseen validation data. Calculates and prints various performance metrics including accuracy, F1 score, recall, and precision.

5. Save the XGBoost Model:

Saves the trained XGBoost model for future use. Overall, this code trains a machine learning model to classify crop diseases in images by first extracting features using a pre-trained CNN and then training an XGBoost model on those features.

X. RESULTS

The record graph is colored blue for educated and purple for examine depending on the train and test accuracy.

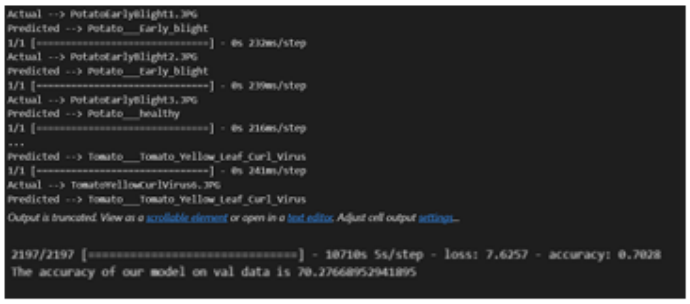


fig(v) Graph 1 of train and test accuracy

Based on train losses and survey losses, a historical map is created with purple coloration for survey and blue for train.

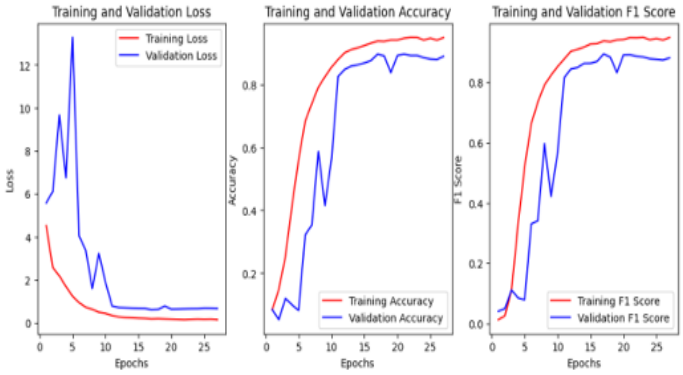
fig(vi) Graph 2 of loss

Implemented an adaptive version with CNN for feature extraction and Xgboost for training and testing. We finished 70% accuracy on nearby and small datasets with a don't forget rating of 90.18.

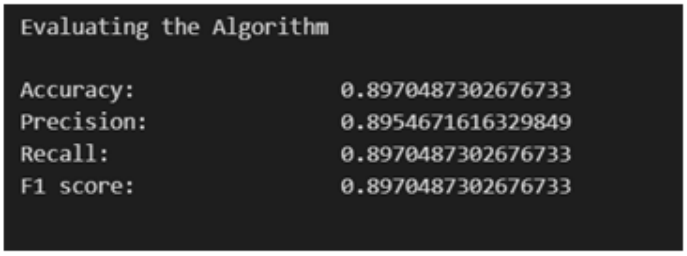


fig(vii)Adaptive Version of CNN and XGBoost

A custom version turned into constructed with the following hyperparameters: RMS prop op, gaining knowledge of rate: 0.0005857, and block size: 64.



fig(viii) Custom Version



Fig(ix) Resnet algorithm with 90% accuracy

The resnet algorithm that got 90% accuracy with this model. Residual networks, or ResNets, study residual features representing the enter layer to research capabilities without uncertainties. Instead, we hope that each few layers correspond at once to the preferred subscript mapping, and the relaxation of the community allows those layers to correspond to the rest of the mapping. RESNET stands for residual community and named with the aid of He et al. ResNet18 is 72-layer structure 18 layers deep. ResNets are common neural community architectures used for deep studying computer vision programs which includes object detection and picture segmentation.

XI. CONCLUSIONS

CX, a revolutionary plant disease detection approach, combines the blessings of XGBoost and CNN classifiers in this project. CX gives farmers and agronomists with an automated, correct and efficient tool with a view to revolutionize plant sickness control. Potential consequences encompass elevated agricultural production, improved meals security and reduced crop losses. Our studies are presently restricted by means of gathering and validating extra facts to increase the number of detectable illnesses under exceptional environmental situations and enhance efficiency. The usability of CX can be in addition more suitable by means of integrating it into actual-time and mobile applications.

XII. FUTURE SCOPE

Crop sicknesses purpose tremendous yield losses, lessen crop high-quality and reason financial problem to farmers. Early detection and accurate diagnosis of plant sicknesses are critical to prevent unfold of illnesses and take timely and targeted measures to lessen crop loss. Crops like rice take numerous months to develop, so early detection of the sickness is crucial to keep away from devastating losses. The assignment consists of studies on crop illnesses and improvement of a gadget gaining knowledge of application. The purpose of this application is to prevent farmer suicides and shield families from the troubles resulting from them.

XIII. REFERENCES

1. Plants Diseases Prediction Framework: An Image-Based System Using Deep Learning Publication Type : Conference Paper Publisher : IEEE Source : 2022 IEEE World Conference on Applied Intelligence and Computing (AIC), 2022, pp. 307-313, doi: 10.1109/AIC55036.2022.9848899. Url: https://ieeexplore.ieee.org/document/9848899
2. Machine Learning and Deep Learning Methods in Heart Disease (HD) Research Publication Type : Journal Article Publisher : International Journal of Pure and Applied Mathematics Source : International Journal of Pure and Applied Mathematics Volume 119 No. 18 2018, 1483-1496 ISSN: 1314-3395 Url : https://acadpubl.eu/hub/2018-119-18/2/116.pdf
3. Analysis on Deep Learning methods for ECG based Cardio,vascular Disease prediction Publication Type : Journal Article Publisher : Scalable Computing: Practice and Experience Source : Scalable Computing: Practice and Experience;Volume 21, Issue 1, pp. 127–136 Url : https://www.researchgate.net/publication/340041238_Analysis_on_Deep_Learning_methods_for_ECG_based_Cardiovascular_Disease_prediction Keywords : Deep learning, Python, CVD, ECG
4. Analysis of Ensemble of Machine Learning Algorithms for Detection of Parkinson's Disease Publication Type : Conference Paper Publisher : 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC) IEEEb Source : 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022, pp. 354-361, doi: 10.1109/ICAAIC53929.2022.9793048 Url : https://ieeexplore.ieee.org/document/9793048
5. Analysis of Deep Learning Models using Convolution Neural Network Techniques Publication Type : Journal Article Publisher : IJEAT Source : International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 -8958, Volume-8, Issue-3S, February 2019 Url : https://www.ijeat.org/wp-content/uploads/papers/v8i3S/C11220283S19.pdf

6. A Comparative Study on Machine Learning Algorithms for Predicting the Placement Information of Under Graduate Students Publication Type : Conference Paper Publisher : Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) Source : 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) (2019) Url : https://ieeexplore.ieee.org/document/9032654
7. S.K. Lokesh Naik; M Silpa Raj; Plant Disease Detection using Image Processing with Machine Learning IEEE[2023]. https://ieeexplore.ieee.org/document/10192986[2023]
8. P. Chaitanya Reddy; Sarat Chandra; Detection of Plant Leaf-based Diseases Us ing Machine Learning Approach,IEEE[2022] https://ieeexplore.ieee.org/document/9683020
9. M S Praveen Kumar; Machine Learning-Based Approach to Detect and Classify Rice Blast Disease, IEEE[2021]. https://ieeexplore.ieee.org/document/9972605/authorsauthors
10. Dipra Mitra; Shikha Gupta; Plant Disease Identification and its Solution using Machine Learning, IEEE[2023]. https://ieeexplore.ieee.org/document/9853136
11. Akanksha Soni; Jeetendra Kumar Soni; Detection of Multiple Crop Diseases using Image Processing Techniques,IEEE[2022]. https://ieeexplore.ieee.org/document/9622812.
12. Robert G. de Luna, Elmer P. Dadios, Argel A. Bandala, (2019)"Automated Image Capturing System for Deep Learning-based Tomato Plant Leaf Disease Detection and Recognition," International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)
13. Suma VR Amog Shetty, Rishab F Tated, Sunku Rohan, Triveni S Pujar,(2019) "CNN based Leaf Disease Identification and Remedy Recommendation System," IEEE conference paper
14. Marwan Adnan Jasim and Jamal Mustafa ALTuwaijari , "Plant Leaf Diseases Detection and Classification Using Image Processing and Deep Learning Techniques", 2020 International Conference on Computer Science and Software Engineering, IEEE 2020.
15. Poojan Panchal, Vignesh Charan Raman and Shamla Mantri , "Plant Diseases Detection and Classification using Machine Learning Models", IEEE 2019.
16. Jiten Khurana and Anurag Sharma ," An Integrated Deep Learning Framework of Tomato Leaf Disease Detection", International Journal of Innovative Technology and Exploring Engineering (IJITEE),2019.
17. Melike Sardogan, Adem Tuncer and Yunus Ozen, "Plant Leaf Disease Detection and Classification Based on CNN with LV Algorithm", IEEE 2018.
18. Gaurav Langar, Purvi Jain and Nikhil, "Tomato Leaf Disease Detection using Artificial Intelligence and Machine Learning", International Journal of Advance Scientific Research and Engineering Trends, 2020.