Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 26, Issues 1, pp. 96–113, DOI 10.12694/scpe.v26i1.3826

IOT ENABLED SMART AGRICULTURE SYSTEM FOR DETECTION AND CLASSIFICATION OF TOMATO AND BRINJAL PLANT LEAVES DISEASE

ROHIT KUMAR KASERA *, SWARNALI NATH, BIKASH DAS, ANIKET KUMAR, AND TAPODHIR ACHARJEE

Abstract. Internet of Things (IoT) assisted smart farming techniques are gradually being used efficiently for identification and classification of vegetable plant diseases. Detection and classification of diseases in these plant families like Solanaceae are still problematic using DCNN due to variations in environmental conditions, genome variation, type of disease, etc. In this paper, two methods for spotting and diagnosing diseases of brinjal and tomato plants leaves named as Optimal Environmental Traversing Alert (OETA) and Optimum diagnosis of Solanaceae leaf diseases (ODSLD) respectively have been proposed. The OETA machine learning (ML) based method is used first to detect the disease, and then the ODSLD deep convolutional neural networks (DCNN) method is used to classify it. An analysis of the proposed method experiments showed that OETA disease detection for brinjal plant (eggplants) was 97.81 percent and for tomato plants was 99.03 percent. For disease classification by ODSLD method, the VGG-16 for brinjal plant and ResNet-50 for tomato plants outperformed other existing DCNN computer vision methods.

Key words: Smart farming, Disease detection, VGG19, DenseNet121, Edge computing, Raspberry pi pico

1. Introduction. India and several other nations rely substantially on agriculture as their primary source of income. It has an enormous influence on the economic development of many nations [1]. In traditional farming, farmers in remote areas face obstacles in following and assessing meteorological conditions, soil quality, water availability, insect control, disease identification, regular monitoring of farming field, and other factors [2]. The majority of leaflets get struck with multiple diseases before and during harvest, which diminishes crop quality and yields. Every year a lot of plants and crops are destroyed due to various causes, including fungal microorganisms pH imbalances in the soil, severe temperatures, alterations in atmospheric moisture or humidity, an inadequate volume of nutrients in the soil, and other aspects [3]. This work of disease detection research focuses on two species of plants from the nightshade family (Solanaceae) [4] namely tomato and brinjal (Eggplant). The diseases that mostly affect tomato and brinjal crops are given in Fig. 1.1 and 1.2.

We have selected tomato and brinjal in our study because both are cultivated simultaneously by farmers across India in dry seasons and and are highly susceptible to diseases [5]. So if a single system can be used to predict and classify disease in the plants, it will be quite a cheap software for farmers. Using precision farming technology with the Internet of Things (IoT), farmers can effortlessly recognise the category of diseases that harms a leaf of tomato and brinjal, thereby improving yield and production [6]. The automated identification and categorization of diseases is an increasingly prominent area of research in smart agriculture today. Various of research have been been pulled off for the recognition and categorization of crop illnesses by employing predictive modelling or Deep convolutional neural networks (DCNN) approaches involving Support vector machines (SVM), Random forest (RF) [7], Artificial neural networks (ANN) [8], Decision trees (DT) [9], Convolutional neural networks (CNN), Visual Geometry Group-16 (VGG16), Inception v3, DenseNet-121, and Residual Network 50 (ResNet 50), U-Net [10, 11]. Sophisticated smart farming systems could be enhanced by combining IoT, machine learning (ML), and DCNN methods. Precisely anticipating and classifying crop health metrics for boosting yield in agriculture is one of the main obstacles. Due to lack of accurate, proficient data and efficient predictive models, genetic factors of plants, and variability in leaf images it is challenging to develop a real-time decision system. Farmers often have difficulty in determining well-informed decisions concerning crop monitoring, pest control, and irrigation without real-time disease detection systems, involving soil nutrient data and a breakdown of tomato and brinjal crop disease types. As a consequence, reliable and

*Department of Computer Science & Engineering, Triguna Sen School of Technology, Assam University, Silchar, Assam, India. (rohitkumar.kasera@aus.ac.in)

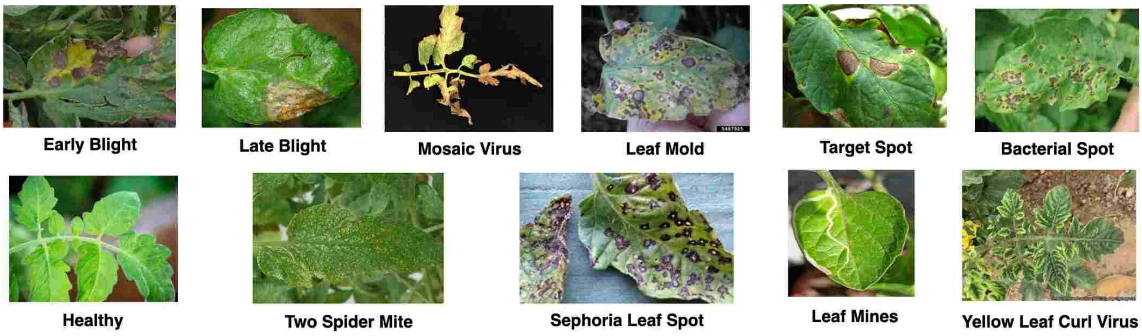

Fig. 1.1: Segmentation of tomato ailments into classes

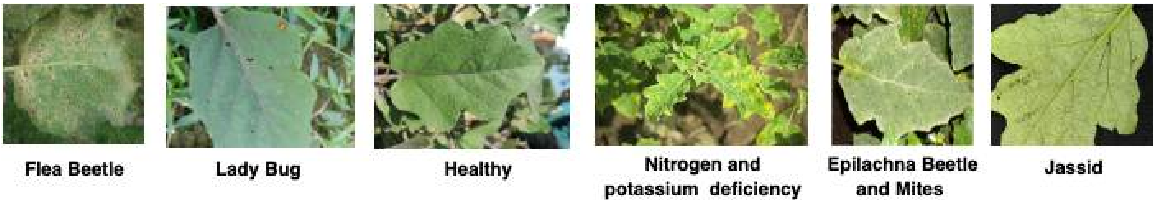

Fig. 1.2: Segmentation of brinjal plant ailments into classes

Table 1.1: Optimal Ranges values for Tomato and brinjal plant cultivation

Plant	**Temp** [$^{\circ}C$]	**Hum** [%]	**SM** [%]	**pH**	**N** [mg/Kg]	**P** [mg/Kg]	**K** [mg/Kg]
Tomato	10 -28	60 - 80	60 - 90	5 - 8	40 - 120	30 - 110	30 - 100
Brinjal	20 - 28	60 - 80	65 - 90	4 - 7	60 - 120	50 - 80	50 - 90

scalable hybrid Edge IoT and AI-enabled alternatives have to exist for the ability to effectively detect illnesses, forecast crop conditions, and allocate resources in intelligent agriculture systems [26]. Recent research indicates the optimal temperature (Temp), humidity (Hum), soil moisture (SM), pH, nitrogen (N) phosphorus (P) and potassium (K) values ranges for growing cultivating tomatoes [12, 13] and brinjal [14] are shown in the Table 1.1. N, P, and K scales on the soil can be used to gauge tomato and Brinjal plant growth. It is possible to avoid both inadequate nutrient supply and over fertilization by anticipating nutrient data. Consequently, farmers can make informed decisions based on real-time temperature, humidity, soil moisture, pH, N, P, K levels and determine the exact amount of fertilizer required for robust plant growth [15], and nutrient level information will help the ML methods to decide the optimal disease prediction.

The intended effect of this research is to assist low-income farmers by predicting and recognizing types of diseases in the early stage through the adoption of an amended IoT edge AI-based methods as Optimal Environmental Traversing Alert (OETA) and Optimum diagnosis of Solanaceae leaf diseases (ODSLD). The proposed system can be employed with some initial cost (specially season cost) in first years but following years only minimal maintenance costs are to be borne by the farmers. This research's key contributions are outlined below:

1. Real-time data through sensor including camera were collected from tomato and brinjal plants fields using an LPWAN communication method. These data are used to develop a disease monitoring system that can detect and recognize diseases.
2. This article adopts an innovative approach utilizing an IoT Edge-based disease observation module called the OETA prognosis method. It employs a modified stacked ensemble learning method. This

method collects data from the surrounding environment (temperature and humidity) and soil (Moisture, pH, N, P, and K) of tomato and brinjal growing fields in real-time to predict the illness and soil health status.

3. The ODSLD hybrid DCNN computer vision technique was designed to classify the disease found on brinjal and tomato plants.
4. An extensive comparative analysis of the hybrid disease testbed framework has been accomplished through experimental results and discussions to validate its performance.

This article is structured into several sections, the Sect. 1 describes the problem domain, possible discrepancies in existing research, objectives, and contributions to strengthen the IoT-enabled illness detection system. Sect. 2 covers in-depth literature surveys along with their limitations and research deficiencies; Sect. 3 demonstrates the proposed methodology; Sect. 4 unveils experiment results and an assessment of the proposed methodology with comparative analysis; and Sect. 5 conveys a summary of the proposed system with future research.

2. Literature Review. In [16] researchers proposed an IoT monitoring system to analyze the minimal and extreme ranges of every environmental factor to envision four types of plant diseases. For classification, they used ANNs and achieved more than 98% accuracy, but the system is not fully automated and adaptive and has very few sensors.

A disease recognition framework for tomato and brinjal plant has been put forward using SE-Inception in [18]. The model utilizes a multi-scale mining module to boost its efficiency and a batch normalization layer to accelerate network convergence. 98.29% accuracy was accomplished with this particular strategy. The assessment of this methodology's flaws indicates floating point arithmetic can impact the mobile device's efficiency although real-time data hasn't been utilized during any kind of system validation.

In [17] the authors put forward a two-stream classification strategy using Inception V3 and inference with recognition using "CNN-softmax". It is observed that the system is not fully automated and precision is low.

Research in [19] indicates utilizing OPNN-based plant disease diagnosis to identify brinjal leaf problems early on. RGB transformation, pre-processing with a median filter, feature extraction with ideal weight values, and ARMKFCM segmentation of impacted areas are all-encompassed. The main flaw of the prevailing approach is its inadequate datasets.

In the article [21], an apparatus is utilized to grab an image and transmit it to a neural network for categorization. Considering brinjal plant of sixteen classes, they adopted CNN transfer learning ("DenseNet201, Xception, ResNet152V2"). DenseNet 201's effectiveness reached 99.06%. The suggested approach has a few limitations, such as skipping the dataset scale constraint. Transfer learning causes an overfitting problem during validation.

Tomato plants are allowed to acquire environmental data with the assistance of an IoT module in [20]. Models like Random Forest, SVM, and K-means are used for assessing the health of plants for three sickness classes and one healthy class. Vanilla Architecture, VGG16, and VGG19 were used to gauge how well the suggested models performed. VGG16 obtained a 92.08% accuracy rate. The outcomes can be upgraded by collaborative prediction, and the efficacy of the model is not validated.

Incorporating IoT and machine learning strategies in [22], IQWO-PCA is an enhanced quantum whale optimization practice that foresees plant epidemics in farming. The system has a few limitations, such as a lack of information on the type of IoT devices used to set up the experiment testbed. This model cannot classify many types of tomato diseases because the system has been tested on a single disease type.

In [23], a CNN-modified imitation has been conceived to predict tomato leaf diseases from a 50k dataset of 14 crops with a performance at 91.2%. In its entirety, CNN triumphed over other pre-trained models, notably VGG16, InceptionV3, and MobileNet, in contrast to their contributions. A few limitations are the inadequate performance and deficient adaptability of the suggested model.

The aforementioned survey emphasizes the ongoing research gap concerning the performance involving various computer vision-based illness detection methods, gradient vanishing, system responsiveness, recognizing between healthy and damaged plants, adoption of IoT devices for real time system, Long range data transmission, sensor data processing, and automated computational tasks.

Table 3.1: Unprocessed sensor data

Time	Temp	Hum	SM	pH	N	P	K	Light
14/12/2023 10:30:00 AM	20.23	74	62	6.05	120	61	92	278
08/02/2024 7:30:00 PM	15.23	73	81	4.23	112	56	87	112
13/03/2024 12:45:00 PM	30.8	69	52	5.62	117	47	67	338
13/03/2024 01:30:00 PM	30.01	72	57	4.02	117	47	67	356

3. Methodology. Considering the research gaps which are addressed in Sect. 2, a hybrid IoT edge-enabled disease diagnosis approach is put forward in this current work. The two methods compose the overall testbed of the system. The first predicts disease using the Optimum Environmental Traversing Alert (OETA) method based on IoT edge and ML, and the second classifies types of diseases using image processing and DCNN named as optimal diagnosis of Solanaceae leaf diseases (ODSLD). The comprehensive architecture for automated disease detection in tomatoes and brinjal plant is portrayed in Fig. 3.1.

The disease recognition framework in Fig. 3.1 is an Internet of Things (IoT) edge-based architecture composed of sensing layer unit (SLU), IoT edge gateway unit (IGU), cloud and application layer unit (ALU). The SLU processes real-time sensor data on temperature (Temp), humidity (Hum), soil moisture (SM), pH, nitrogen (N), phosphorus (P), potassium (K), and light through Raspberry Pi Pico WH microcontroller from an open tomato and brinjal field. Sensor data is transmitted in real-time via the LoRa SX1278 433Mhz module to an IGU. Real-time transmission over a low power wide area network (LPWAN) at long range establishes a point-to-point (P2P) connection between SLU and IGU. The Raspberry Pi 4 module is used as an IoT edge gateway (IGU) for entire system testing. The accumulated data from SLU is processed under IGU. The data are preprocessed and then trained using the OETA classification method, which is a modified hybrid machine learning (ML) ensemble classifier. Hybrid ensemble approaches assess the abilities of more than one robust ML model in a classification task. It reduces overfitting, maximizes true positives, and contributes to an imbalanced dataset. An "OV7670" camera module is deployed under the SLU. If the OETA method predicts that the tomato or brinjal crop has a disease based on the current state of the real-time sensor data, IGU will instruct SLU to activate the camera to take a picture of the leaf. SLU will transmit the captured image to IGU via LoRa SX1278 to recognize the type of disease. The transmission of image data through the LoRa module is accomplished based on the existing multi-packet LoRa transmission protocol [24] using a lightweight Joint Photographic Experts Group (JPEG) coder [25]. The accumulated captured leaf images are stored in the local storage of the Raspberry Pi 4 to recognize the type of leaf disease that occurred. The stored leaf images are preprocessed to remove outliers through image processing. Following preprocessing, these leaf image data are trained using a hybrid deep convolutional neural networks (DCNN) model named ODSLD for tomato and brinjal disease classification under IGU. Once the ODSLD model has been learned, it is tested and validated. Validation involves evaluating the predicted disease classification, which recognizes the particular tomato or brinjal leaf disease name. Farmers Solanaceae disease dashboard provides daily, next-day, and disease prediction alerts which can be accessed through mobile or laptop.

3.1. IoT edge based disease detection system. As shown in Fig. 3.2 (a), a LoRa enabled SLU prototype setup is placed near tomato and brinjal plants to accumulate real-time sensor data, capture images, and process them. Fig. 3.2 (b) shows the LoRa-enabled IoT edge gateway (IGU) setup which is used to receive the tomato and brinjal crop sensor data for further processing.

The sensors are mounted in an array to grab the data. For the day, the sensor reveals the data values at three-minute intervals. Eight traits have been acquired for tomatoes: Temperature (Ft_1), Humidity (Ft_2), Soil moisture (Ft_3), pH (Ft_4), N (Ft_5), P (Ft_6), K (Ft_7), and light (Ft_8). Regarding brinjal, the eight amenities that are collected are Temperature (Fb_1), Humidity (Fb_2), soil moisture (Fb_3), pH (Fb_4), N (Fb_5), P (Fb_6), K (Fb_7); and light (Fb_8). The sensor data in .csv format are extracted to create the dataset for IoT edge-based illness training and prediction. It is a three-minute-long real-time dataset with various attributes. A total of 20532 entries of data accumulated for tomato and 10236 for brinjal plant between 20-11-2023 to 25-03-2024. Table. 3.1 shows raw data of both crops based on sensor readings.

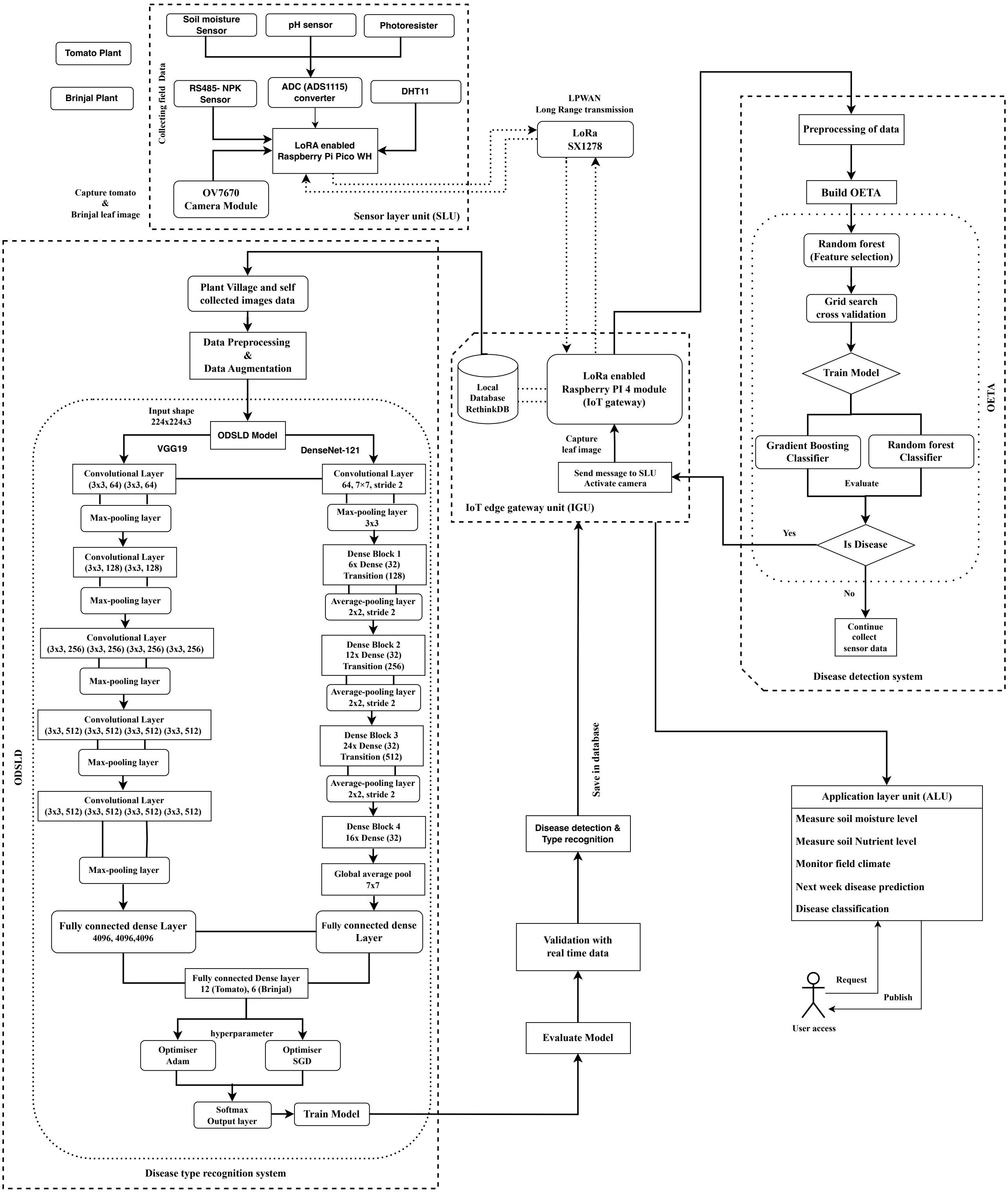

Fig. 3.1: Framework for IoT-based disease detection and classification

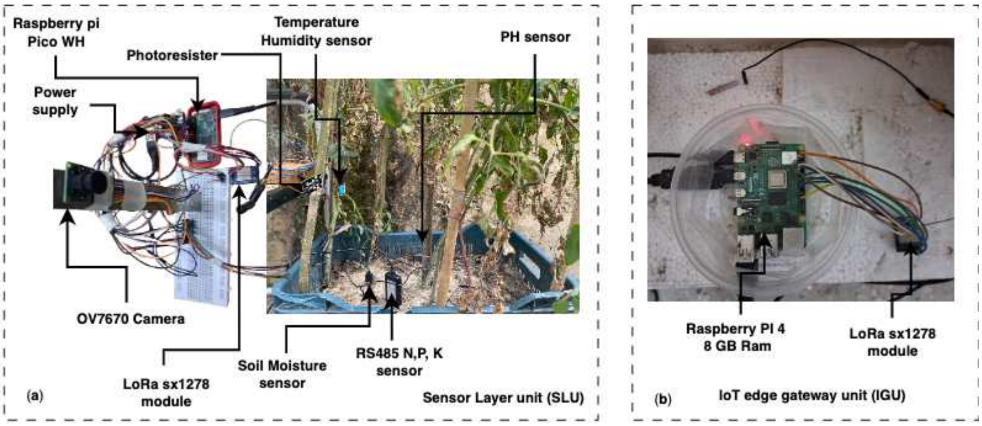

Fig. 3.2: Prototype setup of IoT Edge based disease detection module

The unprocessed data are first filtered by labeling, normalizing, and wiping missing and negative values. The data have been normalized by incorporating five minutes for all features. As discussed in Sect. 1, the optimal value range for each feature has been considered to set the label data $t_{response}$ as 0 or 1. A label data row with '1' indicates that any of the features is outside the range (diseased), while '0' indicates that it is in between or equal the range (Non disease). OETA ML is a converged method of Random Forests (RF), gradient boosting (GB), and stacked classifiers. It is used as a base model for training OETA. This hybrid method aims at filling the shortcomings of individual ML methods. As a result, overfitting is reduced and large and small datasets are handled efficiently. In this necessary features are selected based on predicted RF feature scores and make predictions through grid search and parameter optimization of each model. As part of the process of building the OETA method, the following steps are listed with mathematical representations.

- A Z-score normalization is achieved using Eq. (3.1). This method standardizes data by scaling fx_i features.

$$Zx_i \rightarrow \frac{fx_i - M\mu_i}{S\sigma_i} \tag{3.1}$$

  In the Eq. (3.1), Zx_i is the normalized feature, $M\mu_i$ is the intermediary, and $S\sigma_i$ is the standard error of the fx_i feature.
- An RF classifier is used to estimate important features score for the feature fx_i in Eq. (3.2).

$$pivotal(fx_i) \geq threshold \tag{3.2}$$

  In Eq. (3.2), the threshold is the average of all features' importance scores, and pivotal (fx_i) is expressed as importance. Table. 3.2 summarizes the statistical evaluation of the real-time sensor data after normalization and feature extraction.
- The base RF and GB model is converged by assembling to build and train the OETA model.

$$\hat{O}_{mf} \rightarrow \frac{1}{K}\sum_{l=1}^{K} H_l(fx) \tag{3.3}$$

  In the Eq. (3.3), a random decision tree is concatenated to enhance predictive performance. Where prediction $\hat{O}_{mf}$ is the mean prediction of K trees, and $H_l(fx)$ is the prediction of l^{th} tree.

$$\hat{O}_{gb} \rightarrow \sum_{k=1}^{K} n * w_k(fx) \tag{3.4}$$

Table 3.2: Statistical summary of feature variable

Feature variable	**Mean**	**Standard Deviation**
Temperature	28.60	1.76
Humidity	85.29	12.30
Soil moisture	314.24	74.58
pH	3.34	0.40
N	117.70	37.24
P	42.86	25.44
K	60.0	27.72

A tree is formed iteratively using GB classifier to optimize prediction performance and minimize error. In Eq. (3.4), $\hat{O}_{gb}$ describes the final prediction of the model, n depicts the learning rate, $w_k(fx)$ defines the prediction of the k_{th} iteration, and K expresses frequency of boasting loops.

$$\hat{O}_{converge} \rightarrow U_{ultimate}([\hat{O}_{mf}, \hat{O}_{gb}]) \tag{3.5}$$

Eq. (3.5) shows base model is converged after evaluating and its prediction is used as a feature to build and train the OETA classifier model, where OETA ultimate prediction output as $\hat{O}_{converge}$, and $U_{ultimate}$ is the OETA classifier which incorporates all of the predictions.

- Each parameter in the base model and OETA classifier is optimized by grid search cross-validation using Eq. (3.6).

$$score_{hyperparameter} \rightarrow crossVal_{score}(U_{converge}, S_{feature}, t_{response}) \tag{3.6}$$

In Eq. (3.6), $S_{feature}$ is the selected features evaluated using Eq. (3.2) can contain the arrays of feature as Ft_1, Ft_2,...,Ft_8 for tomatoes and Fb_1, Fb_2,...,Ft_8 for brinjal plant, $t_{response}$ denotes the target data which contains two classes, and $U_{converge}$ denotes the ultimate ensemble classifier.

- The soil nutrient recommendation is based on current nutrient levels and specified thresholds.

$$\text{If } fx_i < \text{limit}_{\text{low}}, \text{then Increase } fx_i$$

$$\text{If } fx_i > \text{limit}_{\text{high}}, \text{then Decrease } fx_i$$

Several evaluation metrics, namely mean squared error (MSE), mean absolute error (MAE), and root mean square error (RMSE), seemed applied to assess the efficiency of the OETA model. Furthermore, the efficiency of the suggested approach was also validated using additional ML classifiers, notably Support vector machine (SVM), Logistic regression (LR), Decision tree (DT), and RF.

3.2. Disease classification and recognition system. The plant leaf diseases are recognised using the hybrid computer vision deep convolutional neural networks (DCNN) approach named as Optimum diagnosis of Solanaceae leaf diseases (ODSLD) for tomato and brinjal plants. Existing CNN-based models for tomato and leaf disease classification have various challenges and shortcomings. These include Precise feature extraction, effectiveness of model learning, diminishing gradients, intricate datasets, reliability of the model, overfitting challenges, dense connectivity, and maximal illness classification. To overcome these challenges, the ODSLD model improves in-depth and reusable feature extraction from images. It strengthens gradient flow, which means there is less chance of vanishing gradients, and efficiently manages small and complex datasets. Overfitting is less probable as features are reused with fewer parameters, maximizing the prediction performance and model learning efficiency using multiple optimizers. The ODSLD DCNN model is a hybrid convolutional neural network (CNN) derived from VGG19 [27] and DenseNet121 [28]. The ODSLD model architecture is depicted in Fig. 3.1 for categorizing tomato and brinjal disease types. The model consists of 137 convolutional layers, 9 pooling layers, global pooling layer, 4 fully connected dense layers, and 1 output softmax layer. Data processing and training of ODSLD models are explained in the following steps.

- Kaggle - Tomato Leaves Dataset [29] is the origin of the dataset for the tomato plant, whereas Kaggle - Brinjal Diseases Detection [30] is the origin of the dataset for the brinjal plant.
- The tomato plant dataset contains 12 classes of images, with 5024 images and 95 megabyte (MB) data size, representing 11 types of diseases. The diseases are Early Blight, Late Blight, Septoria Leaf Spot, Tomato Yellow Leaf Curl Virus, Bacterial Spot, Target Spot, Leaf Mold, Tomato Mosaic Virus, Spider mites Two-spotted, Powdery Mildew and a new disease class, Leaf Mines along with a healthy class. This Leaf Mines disease is caused by the larvae of Tuta absoluta [31] that causes white splotches on the tomato leaves.
- The brinjal plant dataset contains 1235 images of 115 MB data size consisting of six classes. Among the six classes, one is healthy, and the other five are diseased: Epilachna Beetle and Mites, Flea Beetle, Jassid, Potassium, and Nitrogen Deficiency, and one new species is included known as Ladybugs.
- In furtherance of pre-existing datasets, real-time datasets were additionally gathered through visits to outdoor tomato and brinjal agricultural fields in Assam, India, to test and validate the system.
- Preliminary processing entails working on the gathered image data to eradicate undesired images (blurred and out-of-frame images). It entails scaling the photos of the brinjal and tomato leaves to 256x256. Flip and rotation at 90, 180, and 270 degrees were additionally applied to raise the amount of images and zooming in or out is a possibility up to 20%. 10% of the dataset has been split for the validation set. Once the data is adequately collected and preprocessed, it is fed into the ODSLD model for training.
- We used two dense layer hyperparameters. First, dense layer hyperparameters have a minimum value of 128 and a maximum value of 512, the second has a minimum value of 64 and a maximum value of 256 with a RELU activation function.

The custom random search method configures the Keras tuner and optimizes the model hyperparameter using training and validation data. Data have been encapsulated in a custom method to assemble images and labels. In the ODSLD model tuner search, an optimizer is automatically selected during training between ADAM and SGD, and the learning rate is altered correspondingly. In this way, it diminishes the catastrophic over-fitting issue in image-driven plant diagnosis tasks and improve the ODSLD model accuracy. A hyperparameter tuning process determines the best model and evaluates its performance. The reliability of the ODSLD model has been validated using metric factors, which include reliability using Eq. (3.7), true positive rate (TPR) using Eq. (3.8), precision using Eq. (3.9), and F1-score is measured using Eq. (3.10) [32]. Were AO is the actual optimistic, AU is actual unfavourable, DO is deceptive optimistic, and DU is deceptive unfavourable.

$$reliability \rightarrow \frac{AO + AU}{AO + AU + DO + DU} \tag{3.7}$$

$$TPR \rightarrow \frac{AO}{AO + DO} \tag{3.8}$$

$$precision \rightarrow \frac{AO}{AO + DU} \tag{3.9}$$

$$F1score \rightarrow \frac{2 * TPR * precision}{TPR * precision} \tag{3.10}$$

A Raspberry Pi pico camera captures the image of the tomato or brinjal plant leaf and transmits it to the IoT edge gateway in real-time for further processing. The received leaf images are preprocessed into 224x224 size and entered into the ODSLD model to predict the disease class. The predicted disease class will be stored in the real-time rethinkDB database to access. IoT edge-based diagnosis system uses a lightweight message-queued telemetry transport protocol (MQTT) to transmit data. When the real-time disease system receives subscribed disease notifications, it will publish them to ALU through the MQTT broker. Overall, the system is automated and runs computations in real time. The proposed intelligence disease framework assists farmers in preventing infections in tomato and brinjal farms, hence increasing the production of brinjal and tomatoes.

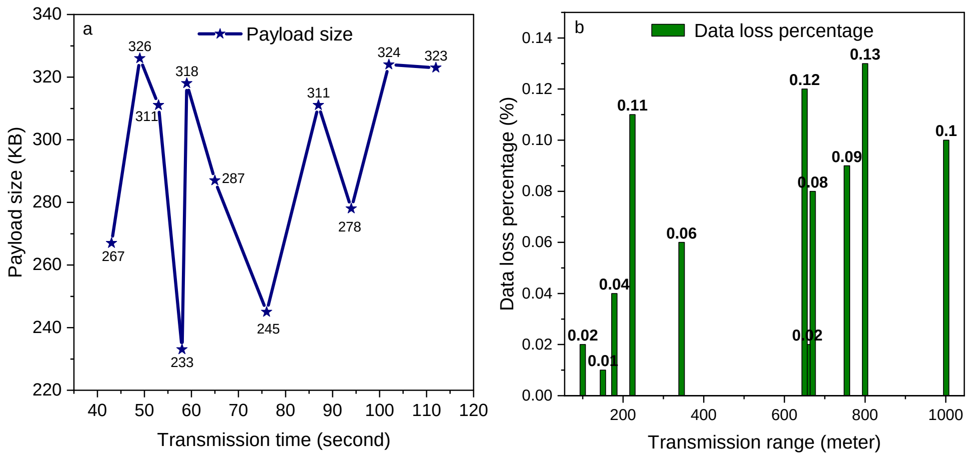

Fig. 4.1: Tomato and brinjal field transmission measurement between SLU to IGU

4. Experiment Results and Discussion. In the Raspberry Pi IoT edge environment, overall system frameworks have been built in Arduino, C++ and Python. Fig. 4.1 (a) and (b) shows the long range wide area network (LoRaWAN) transmission duty cycle of tomato and brinjal crop sensor data. Fig. 4.1 (a) shows the sensor data transmission analysis between SLU and IGU based on payload size (KB) and transmission time (seconds). Fig. 4.1 (b) shows the analysis of data loss percentage based on transmission range. It can be seen from Fig. 4.1 that LoRaWAN transmission between SLU and IGU performs better in terms of payload size, long-range transmission, and lower loss rate. The average transmission time takes up to 61 seconds to transmit the sensor data with a 0.17% loss rate from SLU to IGU.

4.1. Disease detection performance analysis. Real-time tomato and brinjal crop data collected from the open field are normalized as discussed in 3.1. A random forest (RF) ML method was applied to identify essential feature variables. A heatmap matrix shown in Fig. 4.2 depicts the correlation between the feature variables. Fig. 4.2 correlation matrix demonstrates how favorably all feature variables correlate with each other. There is a strong correlation between the variables temperature, humidity, soil moisture, and potassium. There is a positive correlation among the variables potassium, phosphorus, nitrogen, and pH. The strongest correlation is between nitrogen and phosphorus variables.

OETA ML model was trained on 70% training data and 30% testing data. Across the training phase of OETA models of tomato and brinjal crops, a K-split cross-validation at a 5-fold cross is used from which it optimized the mean value, alleviated bias and inconsistency, balanced the data and enhanced data utilization owing to every data point utilized to assess its model's efficiency. The OETA model parameters are the gradient boosting (GB) and random forest (RF) convergence hyperparameters, configured as OETA model-tuning hyperparameters. The optimal hyperparameters on which OETA model performance improves are gb___learning_rate 0.1, gb___max_depth 3, gb___n_estimators 100, rf___min_samples_leaf 1, rf___min_samples_split 2, and rf___n_estimators 100. Using the best hyperparameters, the OETA model achieved 98.86% testing accuracy for brinjal plant and 99.23% accuracy for tomato plant. The ROC of the OETA model has been evaluated as 0.97 for the brinjal plant and 0.99 for the tomato plant. Fig. 4.3 (a) and (b) show a plotted graph of the ROC curve based on the X-axis as the False positive rate and the Y-axis as the True positive rate.

To validate the OETA ML model using real-time sensor data, the model has been dumped using Joblib. Table. 4.1 and 4.2 display the overall validation classification report of both crop from various ML and OETA method for disease detection using real time sensor data.

The classification score analysis revealed in Table. 4.1 and 4.2 that the OETA ML model edges out another ML model in terms of disease prediction for tomato and brinjal plants. The estimations show that tomato plants had an RMSE of 0.08 and brinjal plants had an RMSE of 0.12 in disease prediction. Table. 4.3 shows

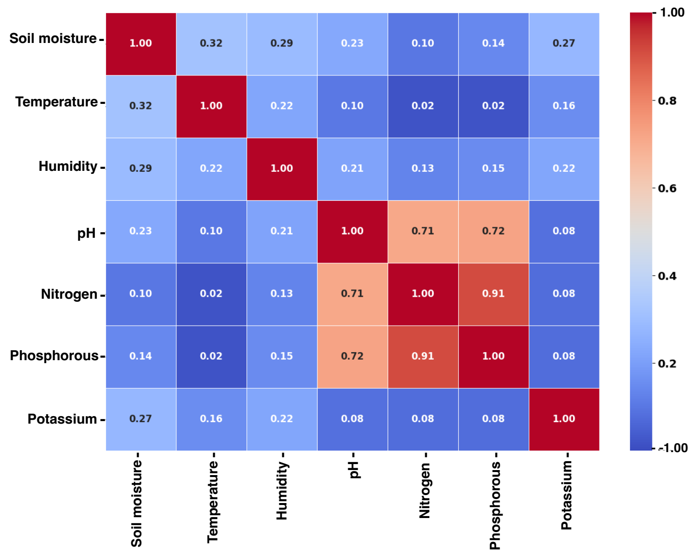

Fig. 4.2: Correlation coefficients matrices between the set of selected features variable

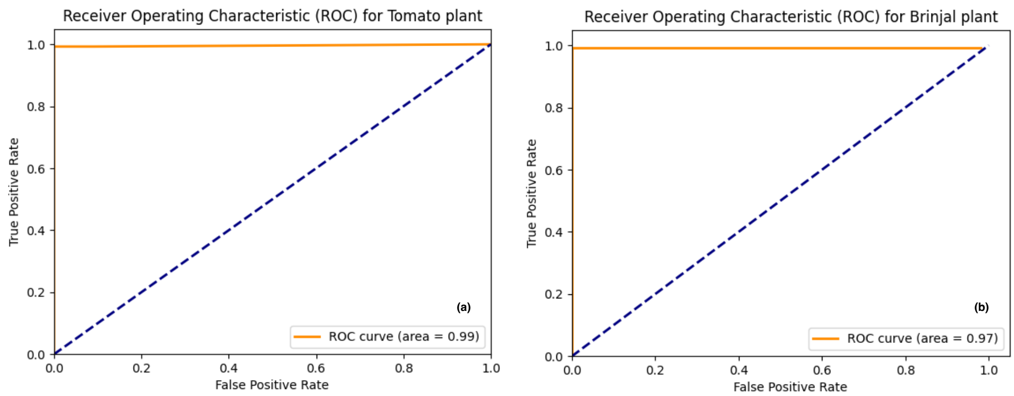

Fig. 4.3: OETA model ROC curve analysis for Tomato and Brinjal plant disease detection

Table 4.1: OETA and other ML model performance report for brinjal plant disease detection

ML Model	**Accuracy**	**F1-score**	**Precision**	**Recall**	**MSE**	**MAE**
OETA	98.56	98.90	98.50	97.90	0.09	0.08
SVM	88.12	89.61	90.13	89.11	3.78	3.23
RF	95.21	95.55	95.16	95.97	3.56	3.65
DT	94.07	94.55	95.16	94.97	2.88	2.53
LR	88.75	89.18	90.12	88.26	4.34	3.67

the disease prediction performance of the OETA method with other existing ML methods through real-time data.

Table. 4.3 shows that whenever the climate conditions, soil moisture, and nutrients are in the range of disease (1) or not disease (0), the proposed OETA method performs better prediction than the other existing

Table 4.2: OETA and other ML model performance report for tomato plant disease detection

ML Model	Accuracy	F1-score	Precision	Recall	MSE	MAE
OETA	99.07	99.16	99.32	99.11	0.07	0.08
SVM	87.15	86.17	86.71	86.59	4.12	3.43
RF	94.26	95.49	95.56	95.18	2.78	2.59
DT	93.15	94.21	95.11	94.35	3.14	3.08
LR	85.47	88.13	85.62	89.16	5.21	5.28

Table 4.3: Comparative prediction performance analysis of OETA with existing ML technique

ML Model	Temp	Hum	SM	N	P	K	Actual	Predicted
OETA	29.80	61.70	113	134	97	76	1	1
SVM	31.32	86.17	634	124	66	83	1	0
RF	30.04	78.17	546	115	123	118	1	0
DT	20.04	85.17	213	103	90	95	0	1
LR	21.14	67.17	278	111	91	94	0	1

Table 4.4: Summary of GPU processor used for training ODSLD method

Plant leaf	Memory usage	Average GPU usage	Total training time
Tomato	70230 MB	48%	12 hours
Brinjal	65216 MB	36%	3 hours

Table 4.5: Comparative performance analysis of DCNN classification model for tomato plant

Method	Test Loss	Validation Loss	Learning rate	Optimizer
ODSLD	0.032	0.017	0.0098	SGD
ResNet-50	2.41	2.36	0.0045	Adam
Inception-V3	3.13	3.22	0.0098	Adam
VGG16	2.11	1.88	0.0034	Adam
CNN	5.26	5.38	0.0098	Adam
VGG19+ResNet-50	1.26	1.38	0.0076	Adam

machine learning methods. When analyzing the prediction performance of conventional RF methods, we observe that the prediction value suggests no disease, and the actual value suggests a disease. The RF method accuracy for tomato and brinjal plants is around 94%, as shown in Table. 4.1 and 4.2.

4.2. Disease classification method performance analysis. Optimum diagnosis of Solanaceae leaf diseases (ODSLD) using deep neural networks model is trained using TensorFlow 2.17, Keras tuner random search, and NUMPY. Data is split into 80% training, 10% testing and 10% validation using Keras' preprocessing module to introduce randomness and shuffle. The model has been trained at 50 epochs throughout each trial using various combinations of hyperparameters for a maximum of 10 trials. This approach optimizes model learning by training the model on each hyperparameter with batch sizes of 16 respectively. The ODSLD model's total parameter size is 176.87 megabytes (MB), the trainable parameter size is 176.55 MB, and the non-trainable parameter is 326.75 kilobytes (KB). A summary of Graphics processing unit (GPU) resources used for constructing and training the ODSLD DCNN model is shown in Table. 4.4. The implementation has been done in a GPU with 81920 MB memory and 70350 MB processing capacity in 12.08 hours. The performance of ODSLD models with existing DCNN models for the tomato plant is validated in Fig. 4.4 and Table. 4.5.

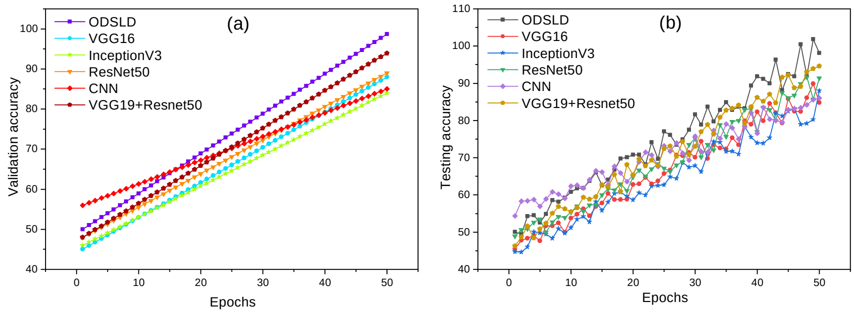

Fig. 4.4: Testing and validation performance curve for tomato plant using real time data

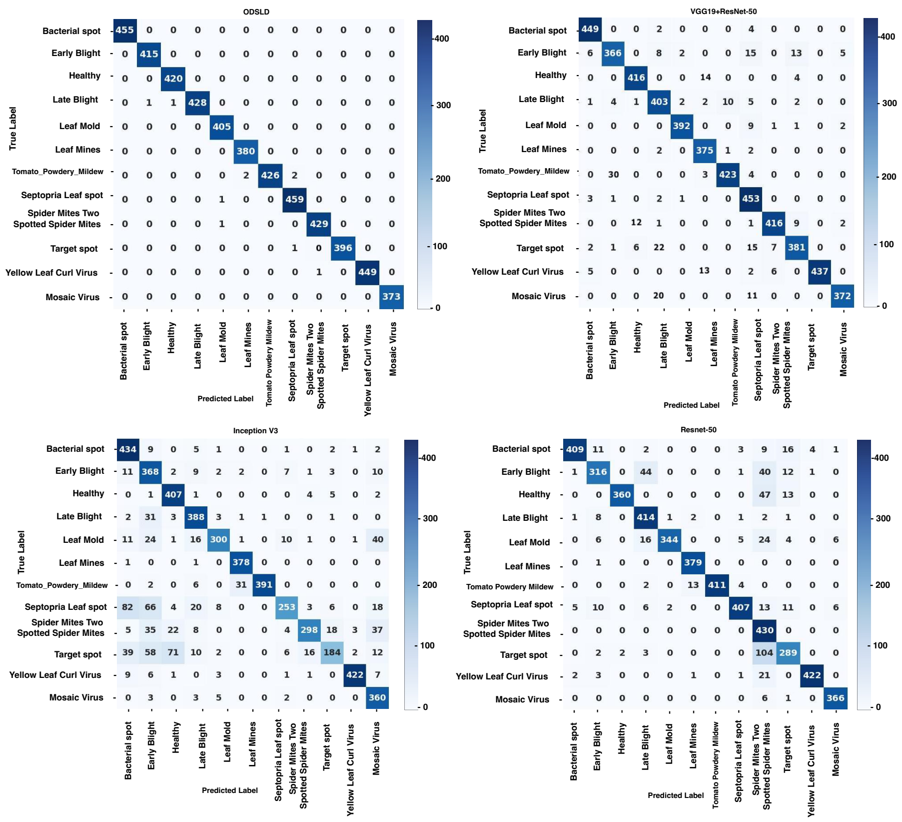

Fig. 4.5: Consequence matrices of ODSLD and existing DCNN model for tomato plant disease

The statistical performance metrics presented in Table. 4.5, show that the validation and testing loss score of the VGG19+ResNet-50 DCNN model is close to the ODSLD model compared to other DCNN models. The ODSLD and existing DCNN model validation and testing preciseness graph leveraging real-time images are portrayed in Fig. 4.4 (a) and (b) shows that the existing DCNN model learning accuracy is not improving because of the large unbalanced dataset, overfitting issue, and hyperparameter not optimised. Despite this, ODSLD model accuracy continues to improve with an increase in the number of testing and validation epochs. The ODSLD method performance validation is justified with the existing DCNN model based on the confusion matrix report depicted in Fig. 4.5.

Table 4.6: Comparative performance analysis of DCNN classification model for brinjal plantbrinjal

Method	Test Loss	Validation Loss	Learning rate	Optimizer
ODSLD	0.028	0.028	0.00080	SGD
ResNet-50	2.09	1.98	0.00045	Adam
Inception-V3	1.13	1.67	0.00073	Adam
VGG16	1.67	1.88	0.0034	Adam
CNN	2.79	2.38	0.0098	Adam
VGG19+ResNet-50	1.06	1.18	0.0076	Adam

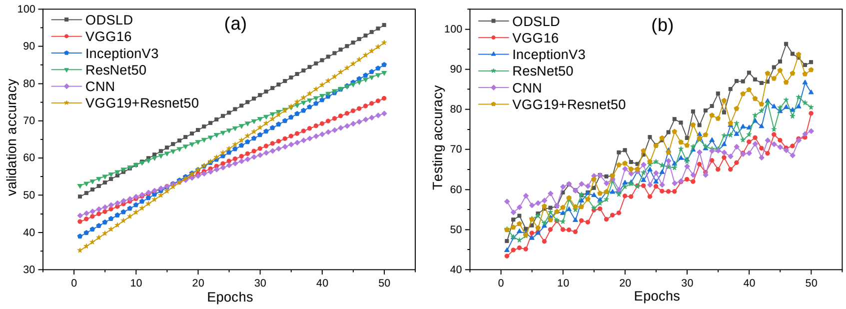

Fig. 4.6: Testing and validation accuracy and loss for brinjal plant using real time data

Inception V3, VGG19+ResNet-50 works better with large datasets [33], whereas proposed ODSLD model prediction is more accurate when tested and validated using large and small real-time image data. According to the confusion matrix shown in Fig. 4.5, ODSLD model able to recognise disease better than other existing DCNN models on real-time datasets. Despite this, ODSLD has the following advantages over other DCNN models. First of all, ODSLD produces a model with a comparatively lighter weight of 160 MB, the model can learn with customized large and small datasets than the other DCNN models while reducing the overfitting problem. Based on the loss-lessening feature of the ODSLD model, it is possible to classify the images more accurately with an average of 98.23% accuracy as shown in Fig. 4.4. The confusion matrices in Fig. 4.5 highlight that although VGG19+ResNet-50 and other DCNN model efficiently classify only 5-6 classes, ODSLD accurately classifies 12 classes, including novel illness classifications of Leaf Mines disease.

The ODSLD model for the brinjal plant has been established with analogous DCNN training and tuning parameters discussed above, and the ODSLD model performs exceedingly well compared to preceding DCNN classification models. Statistical evaluations of the ODSLD models for brinjal are presented in Table. 4.6.

The ODSLD model validation and testing loss are found to be better than the other DCNN models based on Table. 4.6 evaluation metrics. Fig. 4.6 depicts the ODSLD model validation and testing precisions and loss graph using real-time images, and Fig. 4.7 depict the prediction confusion matrix report of ODSLD and other DCNN models for the brinjal plant.

In Fig. 4.6 (a) and (b), the ODSLD model has the highest validation and testing accuracy of 95.67% and an average loss score of 0.028 shown in Table. 4.6 for the classification of six classes of brinjal plant diseases using real-time data. Despite the brinjal plant dataset's smaller size, the ODSLD model is more effective at classifying disease types than VGG19+ResNet-50 or other DCNN models. The confusion matrices in Fig. 4.7 emphasize that existing DCNN model classifies only 2-3 classes, and ODSLD precisely classifies all 6 types of disease classes, including novel illness classifications of Ladybugs caused diseases.

To recognize the disease in real-time, the ODSLD model has been dumped using the Keras API package. On the edge gateway, the ODSLD model is loaded. As soon as the OETA predicts a disease, the ODSLD model will recognize the disease type for both plant in real time and transmit the disease status to the cloud layer application through MQTT publish / subscribe protocol for monitoring so that farmers can take appropriate

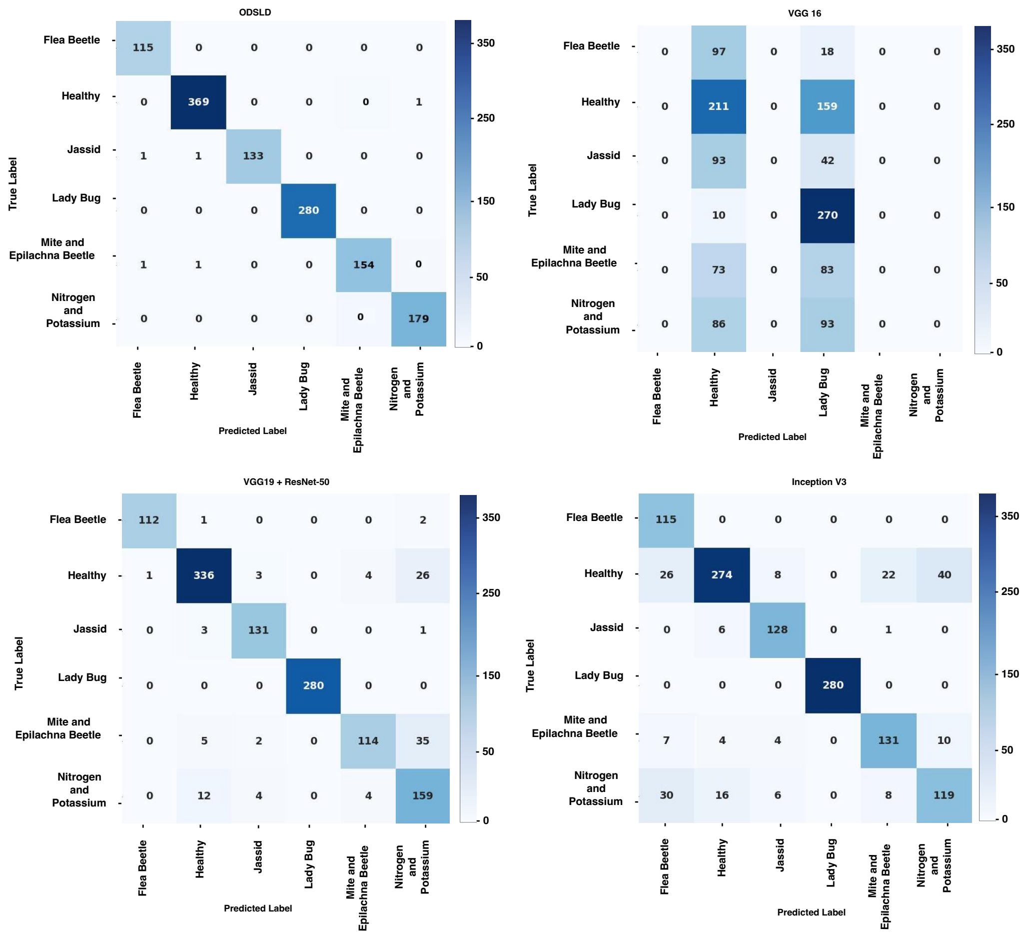

Fig. 4.7: Consequence matrices of ODSLD and existing DCNN model for brinjal plant disease

action as shown in the Fig. 4.8. Fig. 4.9 demonstrates the both plant disease detection monitoring graph. Date-wise disease detection status is displayed on the graph according to temperature, humidity, soil moisture, pH, nitrogen, phosphorus, and potassium.

Based on the aforementioned test outcomes and analysis, it seems to be assessed that the lightweight IoT edge-based hybrid (OETA + ODSLD) model is effective and efficient for detecting and diagnosing diseases for both leaves by overcoming the problem of existing smart leaves diseases method. Traditional pre-trained models, such as VGG16, ResNet50, and Inception v3, have many challenges in recognizing specific plant leaf diseases. Specifically, it is ineffective at capturing fine-grained data on both plant pathologies. To identify leaf disease, the DCNN model is hampered by its model size and inability to analyze comprehensive data. For example, variation in light, image quality, angles, resolution, or background noise. OETA is the best fit for disease prediction, considering data on soil nutrients and environmental conditions in Solanaceae plant fields to make an optimal decision. Proposed hybrid ODSLD method for both plant are the most suitable models for classifying disease type. The outcomes show that the most impoverished models for identifying illnesses in tomatoes are VGG16, ResNet50, and Inception V3 and for brinjal are inceptionV3, CNN and VGG16. In Table. 4.7, the overall proposed approach is compared with other existing approaches.

The comparison of the approaches shown in Table. 4.7 leads one to the finding that the suggested model for finding illnesses and categorization is automated, adaptable, and capable of determining the best decisions.

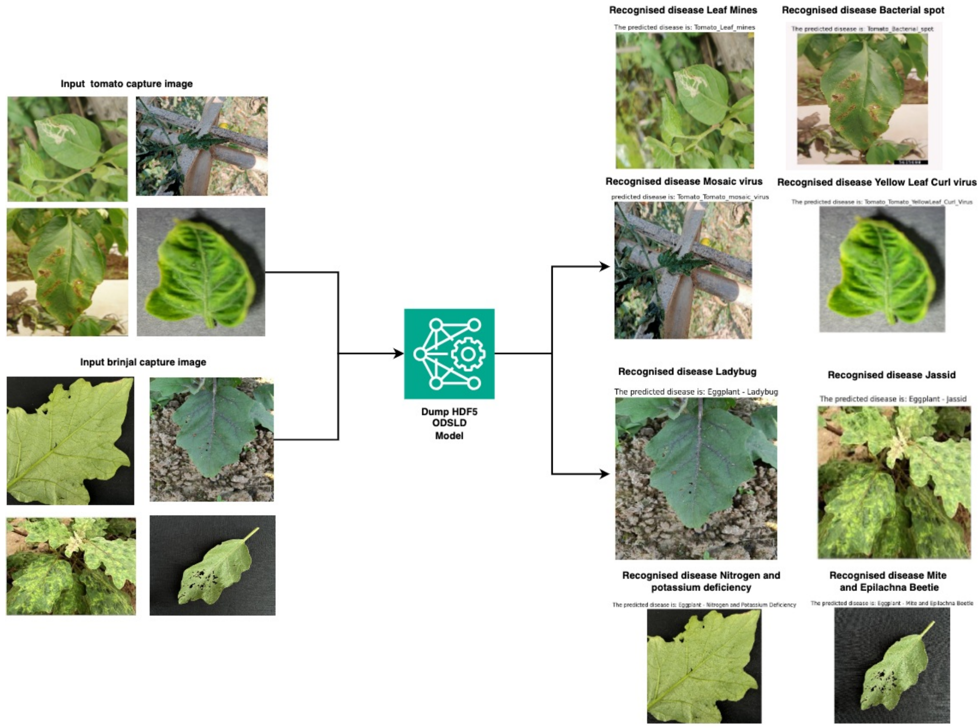

Fig. 4.8: Input captured image of detected disease with recognise output for Tomato and Brinjal plant

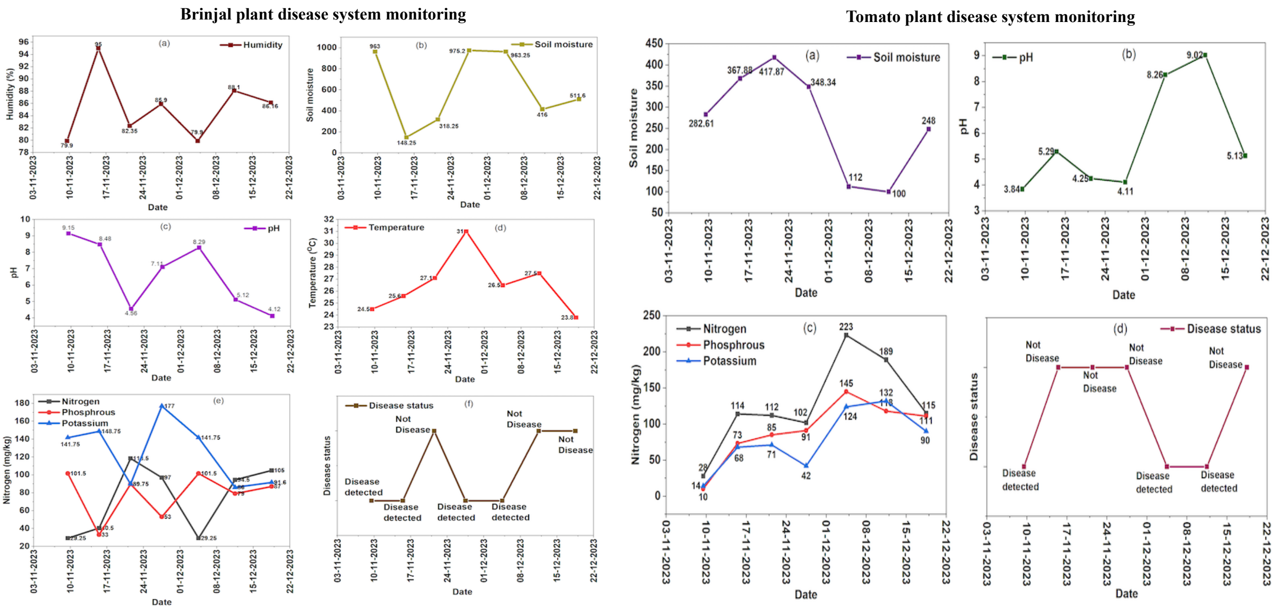

Fig. 4.9: Date wise disease detection status for both plant

Table 4.7: Comparative analysis summary

Reference	Method	Model performance	Customized system
[18]	SE inception DCNN based Disease recognition	Few number of disease are recognized	Large model size and absence of automated image capture, and disease identification system.
[20]	IoT ML	Disease detection depends on soil moisture content and climate only, and identifies fewer disease categories with lower predictive accuracy.	For automating the operation, the trained ML DCNN model is not dumped and absence of long range transmission.
[22]	IoT enabled with disease recognition	Since absence of real-time envision testing, the efficiency of the custom DCNN model is inadequate and non-adaptive.	Not deployed to evaluate model size and not fully automated and customized application.
Proposed framework	IoT Edge (OETA + ODSLD)	Detection of illness with 98.56% accuracy using various soil properties (SM, N,P, and K) and validation of disease recognition using real-time data, encompassing novel illnesses for both plant. Predict the disease each day and next days and wwek. Measure climate condition and soil health status.	To automate the system, the trained model is dumped with a size of 160 MB data for an brinjal and tomato plant. The system is able to communicate at long range wirelessly without any internet connectivity between the SLU to IGU.

5. Conclusion. This work presents models with implementation results for disease prediction, detection, and classification in tomato and brinjal plants. OETA framework is proposed for disease detection and prediction. ODSLD framework is proposed for classifying both plant leaf diseases. The LoRaWAN network transmits the tomato and brinjal field sensor data to an edge gateway. The OETA model normalizes the sensor data and detects and predicts disease. OETA's framework performs better for detecting diseases with an RMSE of 0.08 for tomatoes and 0.12 for brinjal plants. Whenever a disease, a real-time camera module captures the image of leaves and transmits it via LoRaWAN to the edge gateway. ODSLD pre-processes the captured image and classifies the disease. ODSLD model validation accuracy for tomato plants is 98.23% and for brinjal plants, it is 95.67%. The proposed method outperformed existing ML and DCNN models in a comparative analysis. This complete system will immensely assist small farmers in predicting, detecting, and classifying tomato and brinjal diseases. It will enable farmers to control diseases and alert them when to apply water, fertilizer, and pesticides to improve their harvest. In the future, other diseases of tomato and brinjal plant may be included in the proposed DCNN model to enhance model performance and provide assistance for automatically identifying disease remedies.

REFERENCES

[1] Bhattacharya, A., De, D, *AgriEdge: Edge Intelligent 5G Narrow Band Internet of Drone Things for Agriculture 4.0*, *In: Krause, P., Xhafa, F. (eds) IoT-based Intelligent Modelling for Environmental and Ecological Engineering. Lecture Notes on Data Engineering and Communications Technologies*, Springer, Cham, 67, 2021, doi: 10.1007/978-3-030-71172-6_3

[2] Mohanty, M.N., *Machine intelligence techniques for agricultural production*, *Smart Agriculture: Emerging Pedagogies of Deep Learning, Machine Learning and Internet of Things*, CRC press ,175, 2021, doi: 10.1201/b22627-13.

[3] Demilie, W.B, *Plant disease detection and classification techniques: a comparative study of the performances*, *J Big Data*, 11(5), 2024, doi: 10.1186/s40537-023-00863-9.

[4] W. L. Morris and M. A. Taylor,*The Solanaceous Vegetable Crops: Potato, Tomato, Pepper, and Eggplant*, *Encyclopedia of Applied Plant Sciences*, Elsevier, 55–58, 2017, doi: 10.1016/B978-0-12-394807-6.00129-5.

[5] Ghosh, Sunil Kumar, *Eggplant (Solanum melongena L.) and climate resilient agricultural practices*, *Climate Change Dimensions and Mitigation Strategies for Agricultural Sustainability (Volume II) edited by Suborna Roy Choudhury & Chandan Kumar Pand*, 2, 1-24, 2022.

[6] V. Sharma, A. K. Tripathi, and H. Mittal, *Technological Advancements in Automated Crop Pest and Disease Detection: A Review & Ongoing Research*, *2022 International Conference on Computing, Communication, Security and Intelligent*

Systems (IC3SIS), Kochi, India, IEEE, 1–6, 2022, doi: 10.1109/IC3SIS54991.2022.9885605.

[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*, IEEE, 447-451, 2019, doi: 10.1109/WiSPNET45539.2019.9032727.

[8] Gupta, A., Nahar, P, *Classification and yield prediction in smart agriculture system using IoT*, *J Ambient Intell Human Comput*, 14, 10235–10244, 2023, doi: 10.1007/s12652-021-03685-w.

[9] B. Charbuty and A. Abdulazeez, *Classification Based on Decision Tree Algorithm for Machine Learning*, JASTT, 2(1),20 - 28, Mar. 2021, doi: 10.38094/jastt20165.

[10] Q. H. Cap, H. Uga, S. Kagiwada, and H. Iyatomi, *LeafGAN: An Effective Data Augmentation Method for Practical Plant Disease Diagnosis*, *IEEE Trans. Automat. Sci. Eng*, 19(2), 1258–1267, Apr. 2022, doi: 10.1109/TASE.2020.3041499.

[11] C. Wang, P. Du, H. Wu, J. Li, C. Zhao, and H. Zhu, *A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net*, *Computers and Electronics in Agriculture*, 189, 106373, Oct. 2021, doi: 10.1016/j.compag.2021.106373.

[12] Rossy Nurhasanah, Lira Savina, Zul Mahadi Nata, and Ilham Zulkhair, *Design and Implementation of IoT based Automated Tomato Watering System Using ESP8266*, *Journal of Physics: Conference Series*, 1898, 012041, 2021, doi: 10.1088/1742-6596/1898/1/012041.

[13] Shamshiri, Redmond R., James W. Jones, Kelly R. Thorp, Desa Ahmad, Hasfalina Che Man and Sima Taheri, *Review of optimum temperature, humidity, and vapour pressure deficit for microclimate evaluation and control in greenhouse cultivation of tomato: a review*, *International Agrophysics*, 32, 287 - 302, 2018, doi: 10.1515/intag-2017-0005.

[14] Radicetti, Emanuele, Riccardo Massantini, Enio Campiglia, Roberto Mancinelli, S. Ferri, and Roberto Moscetti, *Yield and quality of eggplant (Solanum melongena L.) as affected by cover crop species and residue management*, *Scientia Horticulturae*, 204, 161-171, 161-171, 2016, doi: 10.1016/j.scienta.2016.04.005.

[15] Reichardt, K., Timm, L.C, *How Plants Absorb Nutrients from the Soil*, *In: Soil, Plant and Atmosphere*, Springer, Cham, 313–330, 2020, doi: 10.1007/978-3-030-19322-5_16.

[16] M. Kumar, A. Kumar and V. S. Palaparthy, *Soil Sensors-Based Prediction System for Plant Diseases Using Exploratory Data Analysis and Machine Learning*, *IEEE Sensors Journal*, 21(16), 17455-17468, 2021, doi: 10.1109/JSEN.2020.3046295.

[17] Haque MR, Sohel F, *Deep Network with Score Level Fusion and Inference-Based Transfer Learning to Recognize Leaf Blight and Fruit Rot Diseases of Eggplant*, *Agriculture*,12(8), 1160, 2022, doi: 10.3390/agriculture12081160.

[18] Li, Zhenbo, Yongbo Yang, Ye Li, RuoHao Guo, Jinqi Yang, and Jun Yue, *A solanaceae disease recognition model based on SE-Inception*, *Computers and Electronics in Agriculture*, 178, 2020, 105792, doi: 10.1016/j.compag.2020.105792.

[19] Jayanthi, M. G., Shashikumar, D. R, Preethi. S, *Eggplant leaf disease detection and segmentation using adaptively regularized multi Kernel-Based FuzzyC-Means and Optimal PNN classifier*, *Indian Journal of Computer Science and Engineering*, 13(5), 1542-1558, 2022, doi: 10.21817/indjcse/2022/v13i5/221305073.

[20] Saiqa Khan and Meera Narvekar, *Disorder detection of tomato plant(solanum lycopersicum) using IoT and machine learning*, *Journal of Physics: Conference Series*, IOP Publishing, 1432, 012086, 2020, doi:10.1088/1742-6596/1432/1/012086.

[21] Saad, Izazul Haque, Md Mazharul Islam, Isa Khan Himel, and Md Jueal Mia, *An automated approach for eggplant disease recognition using transfer learning*, *Bulletin of Electrical Engineering and Informatics*, 11(5), 2789-2798, 2022, doi: 10.11591/eei.v11i5.3575.

[22] Sowmiya, M., and S. Krishnaveni, *IoT enabled prediction of agriculture's plant disease using improvedπ quantum whale optimization DRDNN approach*, *Measurement: Sensors*, 27, 100812, 2023, doi: 10.1016/j.measen.2023.100812.

[23] Agarwal, Mohit, Abhishek Singh, Siddhartha Arjaria, Amit Sinha, and Suneet Gupta, *ToLeD: Tomato leaf disease detection using convolution neural network*, *Procedia Computer Science*, 167, 293-301, 2020, doi: 10.1016/j.procs.2020.03.225.

[24] T. Chen, D. Eager and D. Makaroff, *Efficient Image Transmission Using LoRa Technology In Agricultural Monitoring IoT Systems*, *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA*, pp. 937-944, 2019, doi: 10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00166.

[25] Dong, M., Sha, H., Lu, M., Ma, Z., *ULIC: Ultra Lightweight Image Coder on Wearable Devices In: Zhai, G., Zhou, J., Ye, L., Yang, H., An, P., Yang, X. (eds) Digital Multimedia Communications. IFTC 2023. Communications in Computer and Information Science. Springer, Singapore*, 2067, 2024, doi: 10.1007/978-981-97-3626-3_8.

[26] Y. -H. Tsai and T. -C. Hsu, *An Effective Deep Neural Network in Edge Computing Enabled Internet of Things for Plant Diseases Monitoring,2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), Waikoloa, HI, USA*, 695-699, 2024, doi: 10.1109/WACVW60836.2024.00081.

[27] Nguyen T-H, Nguyen T-N, Ngo B-V. *A VGG-19 Model with Transfer Learning and Image Segmentation for Classification of Tomato Leaf Disease*, *AgriEngineering*, 4(4), 871-887, 2022. doi: 10.3390/agriengineering4040056.

[28] G. Huang, Z. Liu, G. Pleiss, L. v. d. Maaten and K. Q. Weinberger, *Convolutional Networks with Dense Connectivity*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12), 8704-8716, 2022, doi: 10.1109/TPAMI.2019.2918284.

[29] Motawani A, Khan. Q *Tomato Leaves Dataset*, *Kaggle*, Access on November 2023 https://www.kaggle.com/datasets/ashishmotwani/tomato?resource=download

[30] Dutta. G *Eggplant Diseases Detection*, *Kaggle*, Access on December 2023 https://www.kaggle.com/datasets/gauravduttakiit/eggplant-diseases-detection

[31] Loyani, Loyani, *Segmentation-based Quantification of Tuta absoluta's Damage on Tomato Plants*, *Smart Agricultural Technology*, 7, 100415, 2024, doi: 10.1016/j.atech.2024.100415.

[32] Shahinfar, Saleh, Paul Meek, and Greg Falzon, *How many images do I need?" Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring*, *Ecological Informatics*, 57, 2020, 101085, doi: 10.1016/j.ecoinf.2020.101085.

[33] Cao, J., Yan, M., Jia, Y. et al, *Application of a modified Inception-v3 model in the dynasty-based classification of ancient murals*, *EURASIP J. Adv. Signal Process*, 49, 2021, doi: 10.1186/s13634-021-00740-8.

Edited by: Vadivel Ayyasamy
Special issue on: Internet of Things and Autonomous Unmanned Aerial Vehicle Technologies for Smart Agriculture Research and Practice
Received: Apr 29, 2024
Accepted: Oct 8, 2024