



DESIGN AN EFFICIENT DISEASE MONITORING SYSTEM FOR PADDY LEAVES BASED ON BIG DATA MINING

Suresh. K¹, S.Karthik², M.Hanumanthappa³

¹Research Scholar, Research and Development Centre, Bharthiar University, Coimbatore-641046, India, mailsuresh.k@gmail.com

²Professor, S.N.S. College of Technology, Coimbatore-641035, India, profskarthik@gmail.com

³Professor, Bangalore University, Bangalore-560056, India, hanu6572@gmail.com

ABSTRACT: With the progressions in Information and Communication Technology (ICT), the innumerable electronic devices (like smart sensors) and several software applications can proffer notable contributions to the challenges that are existent in monitoring plants. In the prevailing work, the segmentation accuracy and classification accuracy of the Disease Monitoring System (DMS), is low. So, the system doesn't properly monitor the plant diseases. To overcome such drawbacks, this paper proposed an efficient monitoring system for paddy leaves based on big data mining. The proposed model comprises 5 phases: 1) Image acquisition, 2) segmentation, 3) Feature extraction, 4) Feature Selection along with 5) Classification Validation. Primarily, consider the paddy leaf image which is taken as of the dataset as the input. Then, execute image acquisition phase where 3 steps like, i) transmute RGB image to grey scale image, ii) Normalization for high intensity, and iii) preprocessing utilizing Alpha-trimmed mean filter (ATMF) through which the noises are eradicated and its nature is the hybrid of the mean as well as median filters, are performed. Next, segment the resulting image using Fuzzy C-Means (i.e. FCM) Clustering Algorithm. FCM segments the diseased portion in the paddy leaves. In the next phase, features are extorted, and then the resulted features are chosen by utilizing Multi-Verse Optimization (MVO) algorithm. After completing feature selection, the chosen features are classified utilizing ANFIS (Adaptive Neuro-Fuzzy Inference System). Experiential results contrasted with the former SVM classifier (Support Vector Machine) and the prevailing methods in respect of precision, recall, F-measure, sensitivity accuracy, and specificity. In accuracy level, the proposed one has 97.28% but the prevailing techniques only offer 91.2% for SVM classifier, 85.3% for KNN and 88.78% for ANN. Hence, this proposed DMS has more accurate detection and classification process than the other methods. The proposed DMS evinces better accuracy when contrasting with the prevailing methods.

Keywords: Alpha-trimmed mean filter, Fuzzy C-Means (FCM) algorithm, Multi-Verse Optimization (MVO) and Adaptive Neuro-Fuzzy Inference System (ANFIS).

1. INTRODUCTION

Long-ago, incessant studies have been made to evince the interactions betwixt plant immune response and pathogens. Massive data has been attained as of those researches owing to the incredible advances in proteomics and also genomics [1]. The naked eye observation of specialists is the central strategy undertaken for detecting and recognizing the insect pests and plant diseases [2]. Nevertheless, it needs continual monitoring of specialists which might be expensive in huge farms. Furthermore, in certain countries, farmers go longer distances to interact with specialists, this is costly and time-consuming [3]. Monitoring of health, as well as diseases, on plant comprises an imperative role in triumphant cultivation of crops in the farm. In former days, the monitoring and examination of plant diseases were done manually by the proficient person in that field [4]. Automated detection of existent plant diseases is an indispensable research domain as it may confirm benefits in observing huge crop fields by automatically spotting the symptoms of disease and insect pests at the moment they appear on plant leaves [5].

An accurate estimation of diseases and pest in plants rests as a challenge in the scientific society. Diseases and pest in plants are generated by several causes and show diverse variations throughout their infection status. Insects, viruses, fungus, and Bacteria may bring plant diseases and also damages [6]. Lately, 'Internet of things' (IOT) is globally a topmost topic, and various researchers have started examining how to utilize IOT in agriculture [7, 8]. IOT would be a great revolution in modern agriculture domain. Several categories of various sensors are developed for sensing agricultural objects, like crops, animals, etc; simultaneously the sensed data is delivered to the Internet

via a wireless network. IOT renders requisite changes in the remote monitoring of the plant disease as well as insect pests. Agricultural IOT technology is a sort of network, like the i) RFID technology (i.e. radio frequency identification), ii) laser scanner, iii) infrared sensors, and iv) GPS (i.e. global positioning system), an information sensing equipment. As per the agreement, it integrates the Internet with any products to employ information communication and exchange, and to comprehend intelligent detection along with location monitoring, tracking and management [9].

Wireless Sensor Network (WSN) is utmost helpful in certain applications that need real-time data monitoring. Precision agriculture is the utmost imperative one amongst those applications [10]. WSN is delineated as a compilation of wireless sensor which is employed in the application domain and centered on the requisite of the data the sensor may vary, i.e., when the node is employed, the network is structured. Afterward, the nodes will gather the data and transfer it to the centralized node to process the data as per the user needs [11].

Health science domain faces numerous problems that are addressed with big data technologies, like i) recommendation scheme in healthcare, ii) internet centered epidemic surveillance, iii) sensor centered food safety and health condition monitoring, iv) GWAS (Genome-Wide Association Studies), v) inferring air quality utilizing big data, vi) eQTL (expression Quantitative Trait Loci) and vii) ionomics and metabolomics for nutritionists [12].

Big data analytics is the process of analyzing large volume of data. It can be processed with software tools such as data mining, predictive analysis, text analysis. Data mining extract significant information from large data sets and this technique are used in agriculture field. It extracts important and useful information for large data sets. Here classification technique used to classify the unknown values by classified samples using information and clustering used to split unknown samples, it deals with finding the structure in collection of unlabelled data.

Data mining technique helps to predict various diseases of leaf and several types of leaf. Data mining technologies are applied for disease prediction and monitoring and with the help of image processing segmentation is performed to identify the symptoms of leaf disease. Moreover, data mining in this implementation is used to extract information from the data set and transform it into an understandable structure.

CONTRIBUTION OF THE PAPER

The entire contribution of this paper is to propose a monitoring system for paddy leaves. By the consideration of paddy leaf image, the input is taken and this work comprises through five phases they are, image acquisition, segmentation, feature extraction, feature selection and classification. After these executions the experimental results are contrasted with prevailing methods and finally we verified the accuracy level of the developed model. The performance of this DMS is evaluated and the results are discussed in the subsequent sections.

The draft structure of the paper is systematized as: Section 2 surveys the associated works regarding the proposed method, Sections 3 briefly discuss the proposed methodology, Section 4 scrutinizes the Investigational outcome and section 5 proffers the conclusion of this paper.

Table 1: Abbreviations and Expansions

| Abbreviations | Expansions |
|---------------|---|
| DMS | Disease Monitoring System |
| ATMF | Alpha-trimmed mean filter |
| FCM | Fuzzy C-Means |
| MVO | Multi-Verse Optimization |
| ANFIS | Adaptive Neuro-Fuzzy Inference System |
| SVM | Support Vector Machine |
| GMM | Gaussian Mixture Model |
| WEP | Wormhole Existence Probability |
| TDR | Travelling Distance Rate |
| KNN | K-Nearest Neighbor |
| ANN | Artificial Neural Network |
| ESMO | Exponential Spider Monkey Optimization |
| R-FCN | Region-centered Fully Convolutional Network |
| FR-CNN | Faster Region-centered CNN |
| SSD | Single Shot Multi-box Detector |

2. RELATED WORK

Bratislav *et.al* [13] delineated software equipment for the feasible prediction of fruits' disease infection. A solution for predicting the peril of fruit infection was grounded on data that symbolize data pathogens and weather/meteorological conditions. This strategy comprises data mining centered tool for earlier prediction of fruit pathogenic infection was executed. It was modeled to be utilized by the consumer who accesses the weather

information as of meteorological stations and also information about pathogenic spore activities. A well-educated farmer or specified employee from a certified service could act as the software user.

Ayman *et.al* [14] propounded the relevant methodology in the zone of food security for forecasting the requisite production quantity and imports to include the requisites of the planned crops for Egyptians associated to populace growth. KDD would be begun via i) data selection, ii) pre-processing, iii) processing and employing data mining methodologies and iv) visualizing the outcomes to make the apt decision. Artificial Neural Network (ANN) is utilized in designing the requisites of the needed quantity of food as per the populace growth rate. The decision takers could comprehend the outcomes by visualization interface and consequently choose a decision.

Konstantinos [15] introduced Convolutional Neural Network (CNN) prototypes which were to execute plant disease diagnosis utilizing leaves images of healthier and diseased plants, via deep learning prototypes. The training was made utilizing an openly existent database of 87,848 photographs, pictured in the lab and real conditions in cultivation areas. The data encompasses 25 plant varieties on 58 disparate classes of [plant, disease] combinations, encompassing certain healthy plants. Innumerable model architectures got trained by showing the topmost performance attaining a success rate of 99.530% in recognizing the corresponding combination [plant, disease] (or healthier plant).

Sandeep *et.al* [16] propounded a feature selection strategy utilizing ESMO (Exponential Spider Monkey Optimization) for detecting plant disease. This strategy has 3 steps: 1st step utilizes SPAM for feature extortion as of the compilation of leaf images. In 2nd step chooses disparate features by the propounded feature selection strategy utilizing ESMO, and in 3rd the classifier was utilized to divide the leaf images into healthier and diseased leaf images. A most notable step in image examination was feature extortion. The extorted features always decide the victory of a classifier.

Yang *et.al* [17] recommended identification approach of rice diseases centered on deep CNN methodologies. Primarily, CNN was initially employed to the issue of the discovery of rice diseases. The recommended CNN-centered model can effectually categorize 10 familiar rice diseases via images recognition. Experiential outcomes evinced that the CNN methodology could augment the convergence speed and attain higher recognition accurateness than other prototypes. The rice disease recognition application evinces that the recommended CNN could precisely and effectually recognize rice diseases via image recognition.

Kuldeep *et.al* [18] proffered the image processing methodologies to recognize and categorize pea's fungal rust disease. The target of this methodology was to spot and recognize the earlier symptoms of rust disease via the microscopic images. The phases like a) preprocessing, b) feature extortion and c) classification were executed and their respective performance was appraised on microscopic images. Lastly, SVM classifier was utilized to spot the disease of Pea. This methodology could successfully spot and scrutinized disease with accurateness of 89.60%.

Harshadkumar *et.al* [19] propounded a framework system for spotting and classifying rice diseases centered on the images of infected rice plants. This framework intended to spot 3 rice diseases like Leaf smut, Brown Spot along with Bacterial leaf blight. Subsequently, capture infected rice plants' images as of a field utilizing a digital camera. Assess 4 methods of background elimination and 3 techniques of segmentation. To facilitate accurate extortion of features, and then suggest centroid feeding centered K-Means Clustering (KMC) for segmenting the disease portion as of a leaf image. To augment the KMC output, eradicate the green pixels on the diseased portion. Lastly, extort several features under 3 categories: texture, shape, and color.

Alvaro *et.al* [20] proffered a deep-learning-centered approach to spot pests and diseases of tomato plants utilizing images taken by a camera with disparate resolutions. This scheme introduces a practical solution for spotting the class as well as the location of diseases existent in tomato plants, which in fact symbolizes a chief comparable difference with other methodologies for plant diseases categorization. This approach regards 3 chief detectors: i) FR-CNN (Faster Region-centered CNN), ii) R-FCN (Region-centered Fully Convolutional Network), and iii) SSD (Single Shot Multi-box Detector), which for this work were termed 'deep learning meta-architectures'. Experiential outcomes evinced that this approach could effectually recognize 9 disparate categories of pests and diseases, with the competency to handle complex scenarios as of a plants' surrounding area.

3. PROPOSED METHODOLOGY

Automated intelligent systems are now becoming more indispensable to evade plant diseases. Presently, WSNS, cloud computing, RFID and IoT have been employed in modern agriculture. This paper proposes an effectual paddy leaves DMS centered on big data mining. The proposed model comprises 5 phases: 1) Image acquisition, 2) segmentation, 3) Feature extraction, 4) Feature Selection along with 5) Classification Validation. Primarily, take the paddy leaves' image as the input as of the dataset. Then, execute image acquisition phase where 3 steps like, i) transmute RGB image to grey scale image, ii) Normalization for high intensity, and iii) preprocessing utilizing ATMF through which the noises are eradicated and its nature is the hybrid of the mean as well as median filters, are performed. Then, the selected features are extracted. Then, the extorted features are picked by utilizing MVO algorithm. After that, the chosen features are categorized utilizing ANFIS algorithm. The classifier effectively

classifies centered on the type of disease found in the paddy leaves grounded on the disease color, size, and shape features. Then, to segment the resulted image FCM Clustering Algorithm is utilized. Segmentation means parting of images into various parts of some features or comprising some similarity. The block diagram for the proposed plant disease detection monitoring system is exhibited below in fig.1,

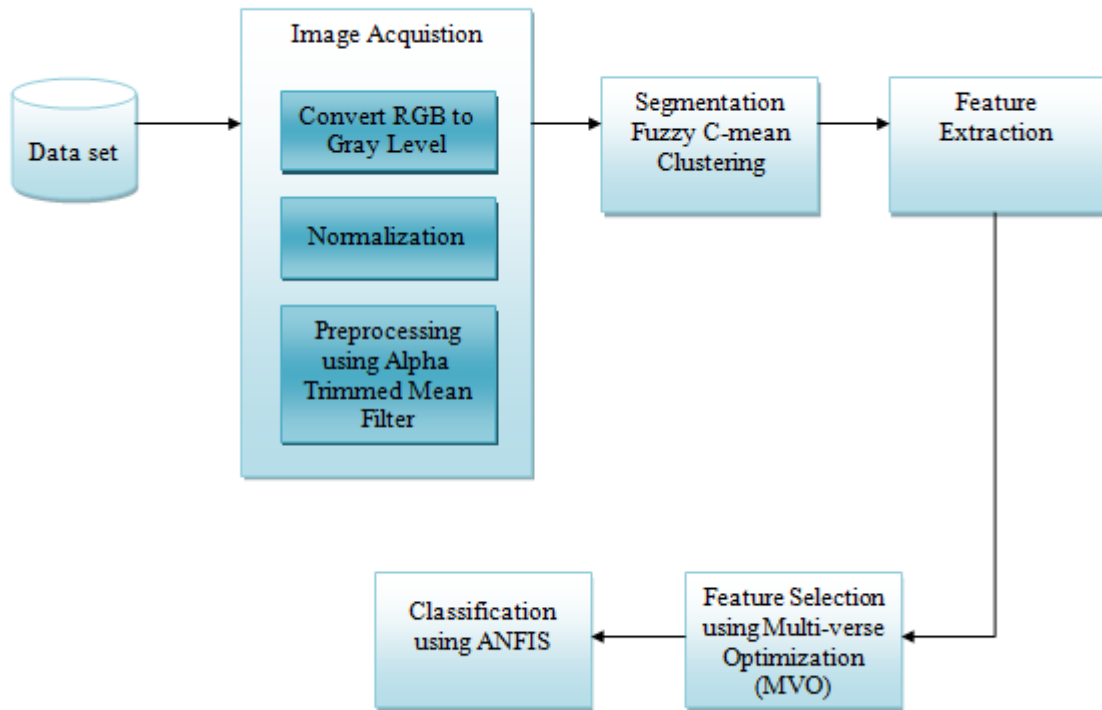


Figure 1: Block diagram for proposed system

3.1 Data Collection

The used dataset is real collected data using cameras in the form of images. Here considered 4 types of diseases (i.e), “Helminthosporium, Magnaporthe grisea, Xanthomonas and Pyricularia grisea”. These diseases are the paddy leaves diseases and the images are attained from the publically existent dataset.

3.2 Image Acquisition

It transforms an optical image (Real World Data) to a collection of numerical data. This phase is developed by utilizing a device like a camera and a microcontroller with wireless connectivity. The output of this phase is a count of captured plant leaves’ images. In this proposed method the image acquisition phase comprises 3 steps namely i) transmute RGB to gray level, ii) Normalization and iii) preprocessing which are delineated below,

3.2.1 Convert RGB to Gray Level

Initially, let’s take the paddy leaf as an input image I_r , and here the RGB image is transmuted to gray image. This process of RGB to Gray level conversion is expressed as,

$$G_r = ((\alpha * R) + (\beta * G) + (\gamma * B)) \tag{1}$$

Where, G_r symbolizes the grayscale image, α signifies the contribution of red R color, β designates the contribution of green G color, and γ indicates the contribution of blue B color.

3.2.2 Normalization

It modifies the range of pixel intensity values. Sometimes it is termed as histogram or contrast stretching. It is expressed as,

$$N_i = (A - Min) \frac{newMax - newMin}{Max - Min} + newMin \tag{2}$$

Where, N_I symbolizes the normalized image with the intensity in the range (Min, Max) , and $(newMin, newMax)$.

3.2.3 Preprocessing

The chief aim of the pre-processing is an augmentation of the input data (image) that control undesired distortions. In this proposed work, the input image noises or other objects are eliminated utilizing ATMF which is the hybrid of the mean as well as median filters. It is employed when an image comprises the long and short-tailed categories of noise (e.g. both a) Gaussian and b) salt and pepper noise). Let $\{x(i), x(i-1), \dots, x(i-n+1)\}$,

Where $n = 2N + 1$ indicates a collection of N sample signal values perceived on a window, W_i . If those values are structured in ascending centered on their amplitude, the order statistics outcome becomes,

$$x_{(1)}(i) \leq x_{(2)}(i) \leq \dots \leq x_{(n)}(i) \quad (3)$$

Where, $x_{(1)}(i)$ denotes the minimum, $x_{(n)}(i)$ is the maximum, and $x_{(N+1)}(i)$ symbolizes the median of the above set of signal values. The output of the ATMF is,

$$A_T = \frac{1}{N - 2[\delta N]} \sum_{j=[\delta N+1]}^{N-[\delta N]} x_{(j)}(i) \quad (4)$$

Where, δ indicates the percentage of the trimmed samples.

3.3 Segmentation using Fuzzy C-Means Clustering Algorithm

Fuzzy C-means algorithm is a clustering method which allows one piece of data to belong two or more clusters and it works by assigning membership to each data point corresponding to cluster center on the basis of distance between the center and data point. FCM is popularly used for segmentation and it based on region growing method to classify pixels into appropriate segments it uses membership grades of pixels. In this phase, the preprocessed image is to be segmented utilizing FCM Clustering Algorithm. FCM Clustering executes clustering by iteratively exploring a set of fuzzy clusters and the related cluster centers that signify the structure of the data as good as possible. Here, the disease affected paddy leaves are segmented using this clustering algorithm. The algorithm relies on the user to specify the count of clusters in the present input paddy leaves image data. The FCM initially partition the data into c fuzzy clusters by diminishing the within group sum of squared error objective function as follows:

$$L_m(J, C) = \sum_{k=1}^n \sum_{i=1}^c (J_{ik})^m \|l_k - c_i\|^2, \quad 1 \leq m \leq \infty \quad (5)$$

Where, $L_m(J, C)$ indicates the sum of squared error for the set of fuzzy clusters signified by the membership matrix J , and the related set of cluster centers C . $\|\cdot\|$ symbolizes some inner product induced norm. $\|l_k - c_i\|^2$ indicates the distance betwixt the data l_k and the cluster center c_i . The squared error is utilized as a performance index that gauges the weighted sum of distances betwixt cluster centers and the elements in their respective fuzzy clusters. The number m controls the effects of membership grades on the performance index. The necessary conditions for above eq.(5) to acquire its minimum are,

$$J_{ik} = \left(\sum_{j=1}^c \left(\frac{\|l_k - c_i\|}{\|l_k - c_j\|} \right)^{2/(m-1)} \right)^{-1} \quad \forall i, \forall k \quad (6)$$

And

$$c_i = \frac{\sum_{k=1}^n (J_{ik})^m l_k}{\sum_{k=1}^n (J_{ik})^m} \quad (7)$$

In every iteration of the FEM algorithm, a matrix J is computed utilizing eqn.(6) and the related cluster centers are computed as eq.(7). Subsequently, evaluate the square error in eq.(5). The algorithm ends when either the error is below a specific tolerance value or its augmentation over the former iteration is below a specified threshold.

3.4 Feature Extraction

After preprocessing, the feature extraction is executed. The features include GMM (Gaussian Mixture Model), Cluster Shade, Cluster prominence, SFTA (Segmentation centered Fractal Texture Analysis), LBP (Local Binary Pattern), HM (Harmonic Mean), color, Gabor feature and shape which are delineated below,

3.4.1 Cluster Shade

Cluster shade measures skewness of gray-level co-occurrence matrix and here higher value indicates asymmetry. The skewness of matrix (or) inadequacy of symmetry is delineated by the cluster shade feature. When it is elevated, the image becomes non-symmetric.

$$C_S = \sum_{i,j=0}^N (i - \alpha_x + j - \alpha_y)^3 P(i, j) \quad (8)$$

3.4.2 Cluster Prominence

Cluster Prominence measure asymmetry of the gray-level co-occurrence matrix and here the higher value indicates more asymmetry and lower value indicates peak of the distribution centered on the mean. When it is lower, the co-occurrence matrix values elevates to a maximal point around the mean values. For the image, it means that there is extremely less variation in gray-scales.

$$C_P = \sum_{i,j=0}^N (i - \alpha_x + j - \alpha_y)^4 P(i, j) \quad (9)$$

3.4.3 Gaussian Mixture Model (GMM)

A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. Gaussian mixture model used to partition pixels into similar segments for analysis. And it is said to be a parametric probabilistic density function which is signified as a weighted summation of Gaussian component densities as proffered by the equation,

$$P(X/\lambda) = \sum_{i=0}^M w_i g\left(X/\mu_i, \sum_i\right) \quad (10)$$

Where, X symbolizes a D-dimensional continuous value data vector (which means the measurement of features). $w_i, i = 1, \dots, M$, indicates the mixture weights, and $g\left(X/\mu_i, \sum_i\right), i = 1, \dots, M$ denotes the component densities.

3.4.4 Segmentation based Fractal Texture Analysis (SFTA)

A segmentation-based fractal texture analysis is used to classify different image textures. The Extraction algorithm is mainly used to decompose the input image into set of binary images since fractal dimensions resulting regions are computed to describe segmented texture patterns. STFA is utilized to classify disparate image textures. The Extraction algorithm is chiefly utilized to decompose the inputted image into binary images set since fractal dimensions resulting regions are evaluated to delineate segmented textures. The classification is executed through collective hierarchical cluster tree with Inner squared distance methodology.

3.4.5 Local Binary Pattern (LBP)

The LBP operator tags the image pixels by decimal numbers, termed LBP codes, which encodes the local structures around every pixel. Formally, provided a pixel at (x_c, y_c) , the resulted LBP is evinced in decimal form as:

$$LBP(x_c, y_c) = \sum_{p=0}^{p-1} S(i_p - i_c) 2^p \quad (11)$$

Where, i_c and i_p - respectively symbolizes grey-level values of the central pixel and P nearby pixels in the circle with R as radius.

3.4.6 Harmonic Mean (HM)

It indicates an "average" when the numbers are provided in relation to a certain unit. The HM is expressed as,

$$H_M = \frac{n}{\sum_{j=1}^n \frac{1}{x_j}} \quad (12)$$

3.4.7 Color Feature

The preliminary step to extort color feature is to choose a relevant color space. Several color spaces are available, they are RGB, CMYK, HSV, CIE L*u*v*, and CIE L*a*b*. Nevertheless, RGB is not perceptually uniform, which symbolizes that 2 colors with larger distance could be perceptually more identical than another 2 colors bearing smaller distance, i.e., the color distance in RGB does not signify the effective color distance

3.4.8 Shape Feature

In shape-centered Feature Extraction Methodologies, the visible features of objects are termed as the visual features or shape characteristics. Example: circular, triangular or other specific shapes of objects, the border diameter, perimeter boundary of the objects, etc. The visual features evinced instinctively are all regarded as shape features. Shape features are divided as 2 categories, i) centered on the characteristics of the border, and ii) centered on regional characteristics. Grounded on the understandability and intuitiveness of shape features, objects in images could be exactly identified. In this proposed method, the 4 categories of disease are extorted centered on width and length.

Identified the disease, as the Helminthosporium type of disease has size of 1000 μm length, and 500 μm to 2000 μm width, the Pyricularia grisea disease has 500 μm to 1500 μm length, and 300 μm to 500 μm width, the MAgnaporthe grisea disease has 1000 μm to 1500 μm length and 300 μm to 500 μm width, the Xanthomonas disease has 400 μm to 1000 μm width and 1200 to 3000 μm length.

3.5 Feature Selection

Here, the extorted features are chosen by utilizing MVO. Feature selection is technically a very crucial component in a Data Scientist's workflow. When proffered higher dimensionality data, the models choke since training time elevates exponentially with the count of features, and models have more risk of over-fitting with increasing count of features. Feature selection assists with those problems by diminishing the dimensions with no loss in information. It also assists to value the features and also its importance. This proposed method utilizes the MVO for feature selection which is the latest evolutionary meta-heuristic algorithm, which apes the rules in the theory of multi-verse. This algorithm's chief inspiration is the theory of the existence of multi-universes and their interrelations via white, black, as well as wormholes. This algorithm is a population-centered stochastic algorithm and estimates the global optima for optimization problems with a compilation of solutions.

Multi-verse optimizer algorithm (MVO) is nature inspired optimization technique. The MVO algorithm simulates the multi-verse theory, where three verses play main part in algorithmic procedure are white holes, black holes, and wormholes. It mainly balanced in terms of exploitation and exploration and it is applied to solve many real world problems. This algorithm obtained competitive results when compared to meta-heuristic optimization algorithm.

Primarily, the 2 parameters like WEP (Wormhole Existence Probability) and TDR (Travelling Distance Rate) should be computed for updating the solutions. Those parameters delineate how often and also how much the solutions alter whilst optimization and it is defined as:

$$WEP = u + i \times \left(\frac{v - u}{I} \right) \quad (13)$$

Where, u indicates the minimum, v symbolizes the maximum, i designates the current iteration, and I signifies the maximal acceptable count of iterations.

$$TDR = 1 - \frac{i^{1/p}}{I^{1/p}} \quad (14)$$

Where p symbolizes the exploitation accuracy.

The chief parameter of TDR is p . The exploitation is emphasized proportional to this parameter value. After calculating WEP and TDR, the position of solutions is updated utilizing the succeeding equation:

$$x_i^j = \begin{cases} \left\{ \begin{aligned} &x_j + TDR + ((mv_j - nv_j) * r_4 + nv_j) && \text{if } r_3 < 0.5 \\ &x_j + TDR + ((mv_j - nv_j) * r_4 + nv_j) && \text{if } r_3 < 0.5 \end{aligned} \right\} && \text{if } r_2 < WEP \\ x_R^j && \text{if } r_2 \geq WEP \end{cases} \quad (15)$$

Where, x_j is the j^{th} element of the best individual, WEP, TDR are coefficients, mv_j and nv_j are the upper as well as lower bounds of the j^{th} element, r_2, r_3, r_4 are arbitrarily generated numbers drawn as of the interval of $[0, 1]$, x_i^j signifies the j^{th} parameter in i^{th} individual, and x_R^j is the j^{th} element of a solution picked by the mechanism of roulette wheel selection. This above equation (15) evinces that the position of solution can be updated in respect of the current best individual attained utilizing the WEP. WEP stands for Wormhole Existence Probability. The Multi-verse optimizer algorithm make use of three concepts such as namely white hole, black hole and worm hole, which helps to improve the random solution iteratively. White and black holes used to explore the search space and moreover the wormhole is used to exploit the search space, where each solution of this concept plays an important role. Meanwhile, the variables are the components of the solution and the objects are the components.

To balance betwixt exploitation and exploration, WEP and TDR should be altered adaptively utilizing eq.(13) and (14). The MVO algorithm initially creates a set of arbitrary solutions and computes their respective objectives. The position of those solutions is constantly updated utilizing Eq. (15) until the end condition is satisfied. Moreover, the random parameter (r_2, r_3, r_4), TDR and WEP are updated for every solution.

3.6 Classification Utilizing ANFIS Algorithm

ANFIS generally utilizes a hybrid learning algorithm which is the combination of gradient descent and least square method to adapt the parameters in adaptive network. The maximized features F_1, F_2, F_n are extricated from the optimization of MVO and are classified utilizing the classifier termed ANFIS which comprises 5 layers of nodes. In the 5 layers, the 1st and 4th layer comprise adaptive nodes whilst, the 2nd, 3rd, and 5th layer holds fixed nodes. The structural design of ANFIS is proffered in figure 2. Utilizing the ANFIS, the features are categorized. The Rule fundamental ANFIS is of the form:

If F_1 is A_i , F_2 is B_i and F_n is C_i

$$Rules_i = a_i * F_1 + b_i * F_2 + c_i * F_n + f_i \quad (16)$$

In which, F_1, F_2, F_n symbolizes the inputs, $A_i, B_i,$ and C_i indicates the fuzzy series, $Rules_i$ is the throughput amidst the specified fuzzy region through the fuzzy rule, $a_i, b_i,$ and c_i signifies the designing parameters which are determined via the training methodology. The structural design of ANFIS is elucidated in figure 2.

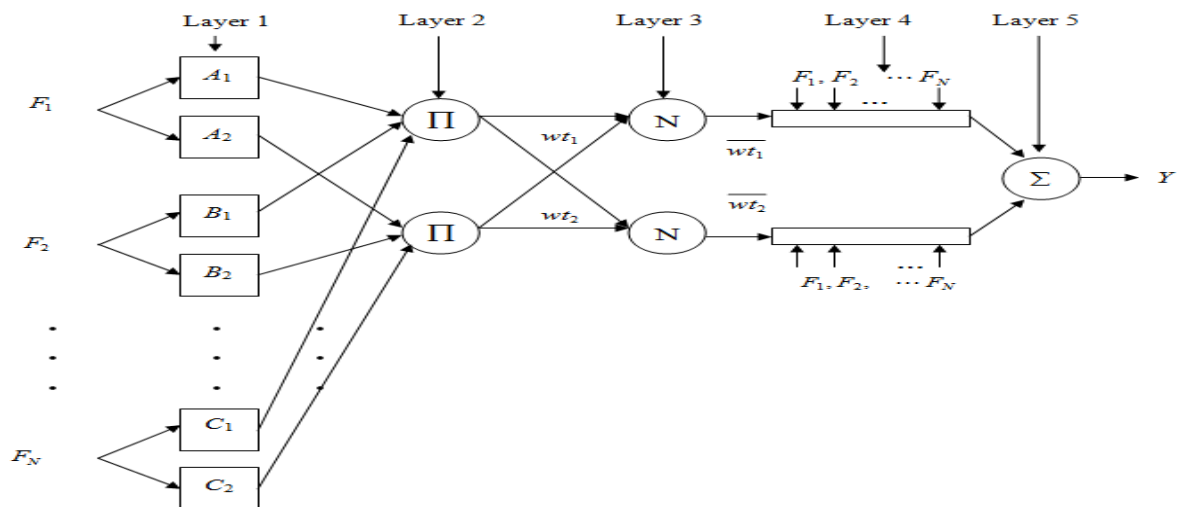


Figure 2: Architecture of ANFIS

Layer-1: Its each node i contains a square node and also a node function.

$$O_{1,i} = \mu_{Ai}(F_1), O_{1,i} = \mu_{Bi}(F_2), O_{1,i} = \mu_{Ci}(F_n) \quad (17)$$

Typically $\mu_{Ai}(F_1)$, $\mu_{Bi}(F_2)$, $\mu_{Ci}(F_n)$ are selected by utilizing bell-shaped with optimal parallel to 1 and minimal equal to 0 and are defined as,

$$\mu_{Ai}(F_1) = \mu_{Bi}(F_2) = \mu_{Ci}(F_n) = \frac{1}{1 + \left[\left(\frac{x - o_i}{p_i} \right)^2 \right]^{q_i}} \quad (18)$$

Where, o_i , p_i , q_i symbolizes the parameter set which is mentioned here as basis parameters.

Layer-2: Its nodes comprise a circle node labeling that multiplies the received signals and the product is sent out. i.e,

$$O_{2,i} = wt_i = \mu_{Ai}(F_1) \times \mu_{Bi}(F_2) \times \mu_{Ci}(F_n), \quad i = 1, 2 \quad (19)$$

Each node throughput denotes the firing strength (FS) of a rule.

Layer-3: Its nodes hold circle node termed N . The i^{th} node evaluates the ratio of the i^{th} rules FS to the total of every rule's FS:

$$O_{3,i} = \overline{wt}_i = \frac{wt_i}{(wt_1 + wt_2)}, \quad i = 1, 2 \quad (20)$$

Layer-4: Its i^{th} nodes hold a square node and a node function

$$O_{4,i} = \overline{wt}_i \cdot Rules_i \quad i = 1, 2 \quad (21)$$

Where wt_i indicates the 3rd layer's output and a_i, b_i, c_i, f_i symbolizes the parameter set. Here, the parameters of this layer are signified as subsequent parameters.

Layer-5: The single node of this layer holds a circle node termed Σ that measures the complete throughput as the summary of all receiving signals:

$$O_{5,i} = \sum_i \overline{wt}_i Rules_i = \frac{\sum_i wt_i Rules_i}{\sum_i wt_i} \quad (22)$$

$$n(o_i^f) = \overline{wt} Rules_1 + \overline{wt} Rules_2 \quad (23)$$

Accordingly, the attained feature is classified by the ANFIS and then the classified feature is signified as $n(O_i^f)$. Next, ω symbolizes the predetermined threshold value whilst (Y) signifies the result of the neural network. If the neural network output Y is superior to the threshold value ω it indicates that the provided input image is acknowledged and if Y is lesser than ω , it symbolizes that the image is never acknowledged.

4. RESULT AND DISCUSSION

The proposed DMS for paddy leaves system centered on big data mining methodologies is deployed in the working platform of MATLAB with machine configuration as:

Processor: Intel core i7
 CPU Speed: 3.20 GHz
 Operating System: Windows 7
 RAM: 4GB

4.1 Performance Analysis

Accuracy: Accuracy measures how close the classified image is to the query image.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (24)$$

Sensitivity: Sensitivity computes the proportion of images which are pertinent to query images which are successfully classified.

$$Sensitivity = \frac{TP}{TP + FN} \tag{25}$$

Specificity: Specificity estimates the proportion of images which are pertinent to the query images that are not correctly classified.

$$Specificity = \frac{TN}{FP + TN} \tag{26}$$

Precision: Precision is the fraction of images classified which are pertinent to the query image.

$$Precision = \frac{TP}{TP + FP} \tag{27}$$

TP Means True Positive, *TN* refers to the True Negative value of the image, *FP* means False Positive and *FN* signifies False Negative.

Recall: Recall determines the fraction of images which are pertinent to the query images which are successfully classified.

$$recall = \frac{TP}{TP + FN} \tag{28}$$

F-Measure: It is simply a weighted HM of both the recall and precision.

$$FMeasure = 2 \frac{Precision \times recall}{Precision + recall} \tag{29}$$

4.2 Comparative Analysis

The performance analysis values of the proposed ANFIS classifier with the prevailing classification methods namely, SVM, ANN and K-Nearest Neighbor (KNN) is elucidated in below table.

Table 2: Performance Table for classification

| Performance measure | Proposed ANFIS | Existing SVM | Existing ANN | Existing KNN |
|---------------------|----------------|--------------|--------------|--------------|
| Accuracy | 97.28% | 91.2% | 85.3% | 88.78% |
| Precision | 92.68% | 86.81% | 85.7% | 86.3% |
| Sensitivity | 88.7% | 80.2% | 79.3% | 78.4% |
| Specificity | 90.65% | 85.3% | 60.2% | 68.4% |
| Recall | 88.7% | 72.4% | 70.3% | 69.7% |
| F-Measure | 91.38% | 86.2% | 83.4% | 80.2% |

Discussion: Table.2 delineates the comparison of the proposed ANFIS classification system with the prevailing classification techniques in respect of recall, F-Measure, precision, sensitivity, accuracy, and specificity. The proposed one has higher accuracy when contrasted to the prevailing system. Thus, it was deduced that the proposed ANFIS proffer the better result when contrasted with the prevailing system. Graphical interpretation of the table values is evinced in fig.3.

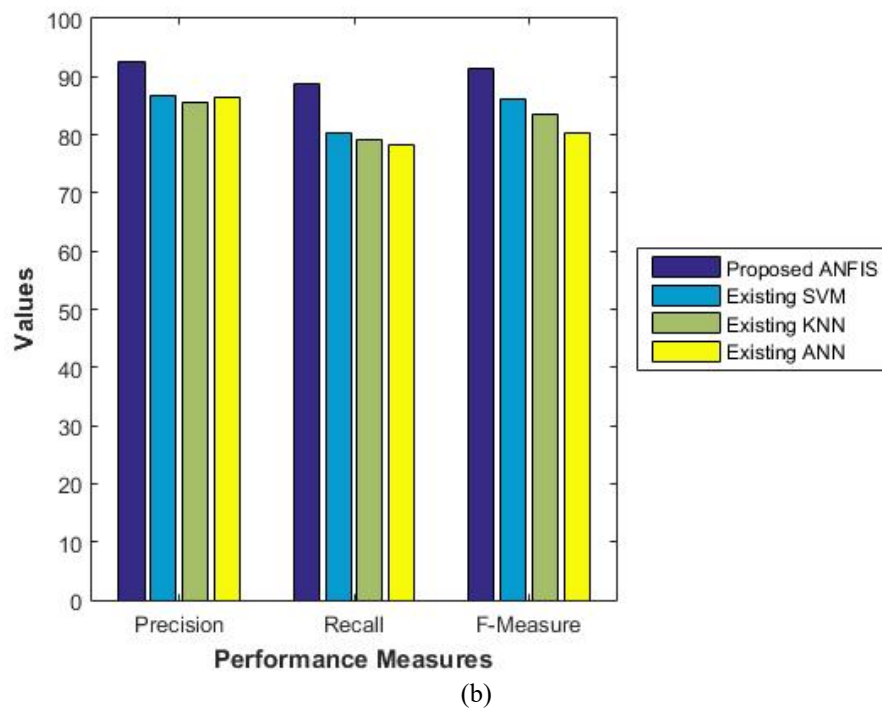
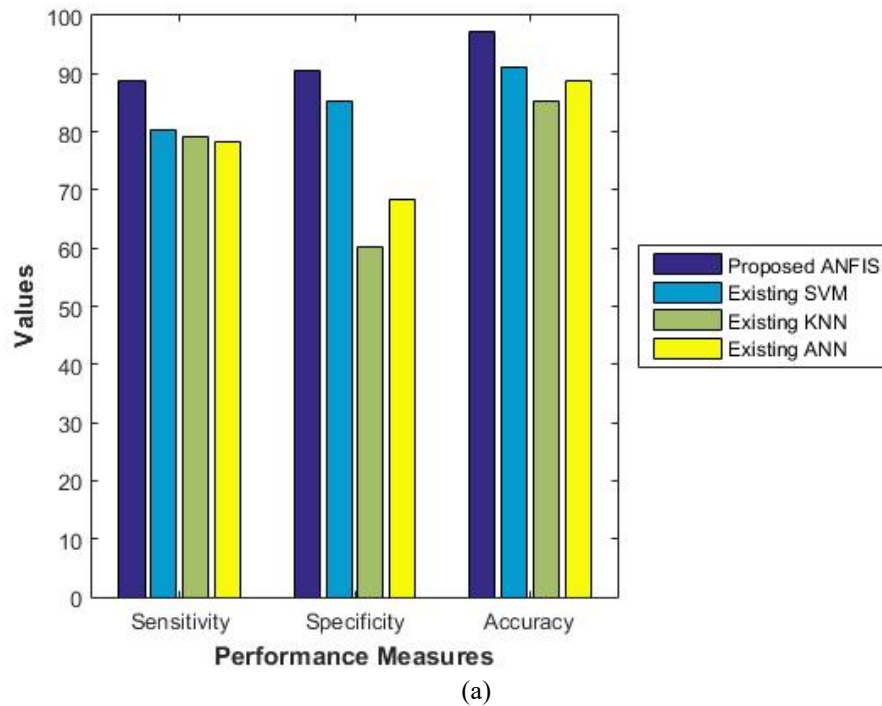


Figure 3: Compared the performance of the proposed ANFIS classifier with the existing classifier techniques in terms of (a) sensitivity, specificity, accuracy, and (b) precision, recall, F-Measure

Discussion: Fig.3 delineates the comparison of the proposed classification system with the prevailing classification system centered on performance metrics. Fig.3 (a) proposed ANFIS offers 88.7% sensitivity, 90.65% specificity, and 97.28% accuracy, but the prevailing system achieves lesser values of sensitivity, accuracy, and specificity. Fig.3 (b) ANFIS system scores 92.68% precision, 88.7% recall, and 91.38% F-Measure. When contrasted to the prevailing system the proposed classification system proffer better achievements.

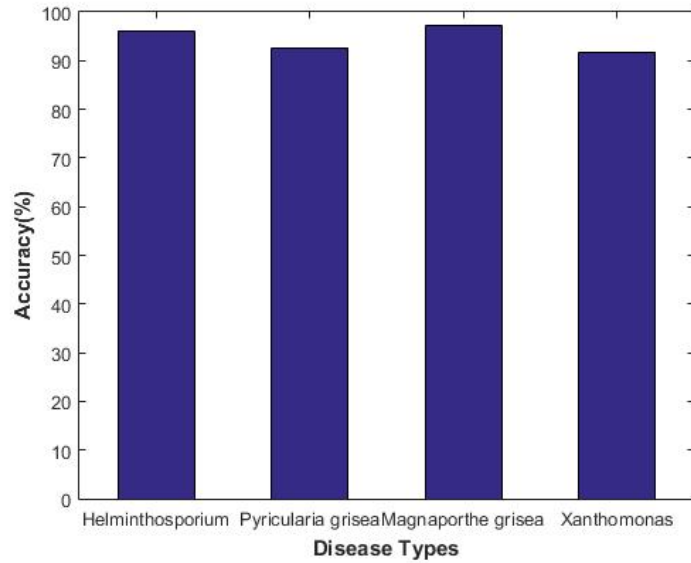


Figure 4: Paddy Leaves Disease Wise Accuracy for proposed system

Discussion: Fig.4 evinces the disease wise accuracy achievements of the proposed system. The ANFIS system achieves 96.2% accuracy for Helminthosporium disease classification, 92.7% accuracy for pyricularia grisea disease classification, 97.1% accuracy for Magnaporthe grisea disease classification, and 91.8% accuracy for Xanthomonas disease classification. Thus, it is deduced that the proposed one attains topmost performance.

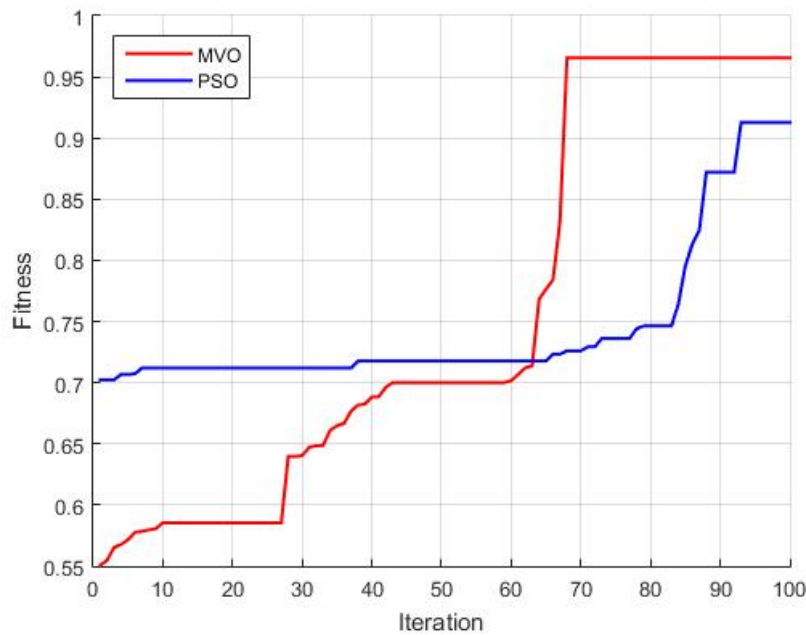


Figure 5: Demonstrate the convergence value of the proposed MVO with the existing Particle Swarm Optimization (PSO)

Discussion: Fig.5 delineates the fitness value of the proposed MVO and it is contrasted with the prevailing optimization PSO. From the figure, the Fitness value is differed depending on the number of iterations. Here, the proposed MVO proffers the better performance.

5. CONCLUSION

In this work, an effectual DMS for paddy leaves was designed centered on big data mining. The proposed system’s performance was analyzed utilizing the real collected data as of a camera. The performance analysis has exhibited that the proposed monitoring system has proffered an incredible rate of sensitivity, accuracy, and specificity. The comparison outcome delineates that the proposed system has higher sensitivity, accuracy, and specificity than the prevailing methods. This proposed system spots and classifies 4 paddy leaves diseases, encompassing Helminthosporium, Magnaporthe grisea, Pyricularia grisea, and Xanthomonas. They are identified

by the disease affected area size centered on width and length. In accuracy level, the proposed one has 97.28% but the prevailing techniques only offer 91.2% for SVM classifier, 85.3% for KNN and 88.78% for ANN. Hence, this proposed DMS centered on big data mining methodologies has more accurate detection and classification process than the other methods.

REFERENCES

- [1] Xin Yang, & Guo, Tingwei, "Machine learning in plant disease research", *European Journal of BioMedical Research*, Vol. 3, 2017
- [2] Al-Hiary, H., S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. ALRahamneh, "Fast and accurate detection and classification of plant diseases", *Machine Learning*, Vol. 14, no. 5, 2011.
- [3] Yin Laiwu, Chen Deyun, Li Changcheng, and Chen Dong, "A novel recognition algorithm based on optimal wavelet packet and non-negative matrix factorization for extracting pathologic features of plant image", *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol. 6, No. 5, pp. 89-100, 2013.
- [4] Sachin D. Khirade, and A. B. Patil, "Plant disease detection using image processing", In *Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on*, IEEE, pp. 768-771, 2015.
- [5] Yun Shi, Zhen Wang, Xianfeng Wang, and Shanwen Zhang, "Internet of things application to monitoring plant disease and insect pests", In *International Conference on Applied Science and Engineering Innovation (ASEI 2015)*, pp. 31-34, 2015.
- [6] Alvaro F. Fuentes, Sook Yoon, Jaesu Lee, and Dong Sun Park, "High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank", *Frontiers in Plant Science*, Vol. 9, 2018.
- [7] Raghavendra V. Kulkarni, Anna Forster, and Ganesh Kumar Venayagamoorthy, "Computational intelligence in wireless sensor networks: A survey", *IEEE Communications Surveys & Tutorials*, Vol. 13, No. 1, pp. 68-96, 2011.
- [8] Siquan Hu, Haiou Wang, Chundong She, and Junfeng Wang, "AgOnt: ontology for agriculture internet of things", In *International Conference on Computer and Computing Technologies in Agriculture*. Springer, Berlin, Heidelberg, pp. 131-137, 2010.
- [9] Wang, X. F., Z. Wang, S. W. Zhang, and Y. Shi, "Monitoring and discrimination of plant disease and insect pests based on agricultural IOT", In *International Conference on Information Technology and Management Innovation (ICITMI 2015)*, pp. 112115, 2015.
- [10] Saeed Azfar, Adnan Nadeem, and Abdul Basit, "Pest detection and control techniques using wireless sensor network: a review", *Journal of Entomology and Zoology Studies*, Vol. 3, No. 2, pp. 92-99, 2015.
- [11] Subramania Ananda Kumar, and Paramasivam Ilango, "The impact of wireless sensor network in the field of precision agriculture: a review", *Wireless Personal Communications*, Vol. 98, No. 1, pp. 685-698, 2018.
- [12] Huang T. et al., "Promises and challenges of big data computing in health sciences", *Big Data Research*, 2015.
- [13] Bratislav Predic, Milos Ilic, Petar Spalevic, Slavisa Trajkovic, Srdjan Jovic, and Andrija Stanic, "Data mining based tool for early prediction of possible fruit pathogen infection", *Computers and Electronics in Agriculture*, Vol. 154, pp. 314-319, 2018.
- [14] Ayman E. Khedr, Mona Kadry, and Ghada Walid, "Proposed framework for implementing data mining techniques to enhance decisions in agriculture sector applied case on food security information center ministry of agriculture, Egypt", *Procedia Computer Science*, Vol. 65, pp. 633-642, 2015.
- [15] Konstantinos P Ferentinos, "Deep learning models for plant disease detection and diagnosis", *Computers and Electronics in Agriculture*, Vol. 145, pp. 311-318, 2018.
- [16] Sandeep Kumar, Basudev Sharma, Vivek Kumar Sharma, Harish Sharma, and Jagdish Chand Bansal, "Plant leaf disease identification using exponential spider monkey optimization", *Sustainable Computing: Informatics and Systems*, 2018.
- [17] Yang Lu, Shujuan Yi, Nianyin Zeng, Yurong Liu, and Yong Zhang, "Identification of rice diseases using deep convolutional neural networks", *Neurocomputing*, Vol. 267, pp. 378-384, 2017.
- [18] Kuldeep Singh, Satish Kumar, and Pawan Kaur, "Support vector machine classifier based detection of fungal rust disease in Pea Plant (*Pisam sativam*)", *International Journal of Information Technology*, pp. 1-8.
- [19] Harshadkumar B. Prajapati, Jitesh P. Shah, and Vipul K. Dabhi, "Detection and classification of rice plant diseases", *Intelligent Decision Technologies*, Vol. 11, No. 3, pp. 357-373, 2017.
- [20] Alvaro Fuentes, Sook Yoon, Sang Cheol Kim, and Dong Sun Park, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition", *Sensors*, Vol. 17, No. 9, pp. 2022, 2017.