

Paddy Leaf Disease Recognition System using Image Processing Techniques and Support Vector Machine

Hayatun Syamila Mamat¹, Nur Azwani Zaini² and Suhaila Abd Halim³

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

*corresponding author: ³suhaila889@uitm.edu.my

ARTICLE HISTORY

Received
19 September 2020

Accepted
3 December 2020

Available online
30 December 2020

ABSTRACT

Paddy is a crucial agroculture sector since rice is the staple food for the majority of the world's population. However, the production of paddy is slower and less productive since many factors have affected the growth of the paddy. The existence of disease in paddy component affects the quality of rice produced. Hence, the recognition of the disease at the beginning stage is crucial as the initial approach for prevention purposes. In this study, a system is developed to detect the paddy leaf disease such as bacterial leaf blight, brown spot and leaf smut. All the processes involved are implemented and compiled using MATLAB R2020a. A set of 105 image data with disease is converted to binary image using thresholding. 6 features from all the data are extracted and divided to testing and training set before the classification process. A cubic support vector machine is used for the classification process. Lastly, accuracy, precision, and misclassification for each disease are calculated for performance evaluation. Results show that the average performance of the diseases on accuracy, precision, and misclassification are 88.57%, 82.97%, and 11.43% respectively. The use of the processes act as assistance to the paddy farmer to identify the existence of the paddy leaf disease. This could improve the quality of the paddy produced by reducing the process of manual disease checking.

Keywords: MATLAB; Cubic Support Vector Machine; binary image; image processing; paddy leaf disease.

1. INTRODUCTION

Paddy also known as rice is the starchy seeds of an annual Southeast Asian cereal grass (*Oryza sativa*) that is cooked and used for food. Rice is an important food crop worldwide and is the staple food for most of the world's population. However, due to pests and diseases, the paddy productions have become slower and less productive. This is because of the limited resources and technology to determine the paddy diseases at the early stage. The paddy disease diagnosis is also limited by human visual capabilities because the first symptoms are too small or microscopic. Even though the diseases are being determined, the error might still occur due to the false diagnosis.

For the past few years, the paddy disease had been detected manually by the farmers. A paddy farmer is usually trained at Malaysia Agricultural Research and Development Institute (MARDI) to be able to analyze the type of paddy diseases. Once the paddy disease has been diagnosed the suitable treatment is done to cure the diseases. However, paddy farmers have to

spend a lot of time and energy to detect the disease and it is hardly possible to accurately estimate the infected areas seriously in large-scale farming. By using manual observation, incorrect disease identification might occur as the symptom of one disease is similar to the others. The implementation of technology is essential to accurately detect the diseases and the prevention can be taken directly to prevent further spread. As such, the digital image processing is one of the new diagnostic processes that can be embedded with the method to classify paddy diseases through its leaf, stem or soil.

The use of technology has successfully minimized the work of paddy farmers in detecting type of disease from paddy leaf. Hence, in this study, sets of image processing techniques are applied and compiled using MATLAB. There are plenty of algorithms that exist for classification, but the selection of classifier is still an open problem as they offer from low to complex algorithms that are able to increase the accuracy of the model.

The application of classification offers several classification techniques in the field of agriculture such as paddy leaf [1], chilli leaf [2], brinjal leaf [3], crop leaf [4]. Back-Propagation Neural Network (BPNN) is one the mostly used neural network techniques. besides support vector machine (SVM). Although Swathi and Barathi [5] proposed the use of ANN on real time image, they also suggested that more features are required and the application of SVM would be able to gain a better performance in future. Convolutional Neural Network (CNN) performed better than logistic regression, K-Nearest Neighbours (K-NN) and SVM on leaf of crops such as apple, potato, tomato, grape, strawberry and corn [4]. Deep learning is suitable in this case as the data used is about 20,000. [6] evaluated the performance of multiple classification techniques that are BPNN, Radial Basis Function Neural Network (RBFN), General Regression Neural Network (GRNN) and SVM with RBFN. The modification of SVM with RBFN offers the highest accuracy rate as compared to the other techniques.

In this study, SVM is used to detect the paddy disease. The SVM is used since the techniques are simple, stable, easy to program, and it is a standard method that generally works well. This study aims to develop a system using MATLAB which can be used to classify the paddy leaf disease type and suggest the prevention action for the disease for the farmer reference.

2. LITERATURE REVIEW

Image processing plays important task in manipulating and processing the image data before further process. For classification purposes, SVM is one of the processes that can be applied.

2.1 Image Processing

Research and applications of image processing in detecting the diseases have arisen drastically due to the evolvement of computing technology. The colour-to-grayscale conversion process to convert the RGB image to Gray-scale image colour for further processing the image that aims for different purpose [7]-[9]. Median filter is commonly used to remove image noise as it has a good edge preserving properties and robust as it is able to remove outliers without blurring effects [10]. Histogram Equalization (HE) is used to enhance the contrast of the acquired image after resizing and noise restoration [11]. In addition to that, contrast enhancement and RGB to L*a*b transformation is one of the main processes in the pre-processing stage [12]. Besides, colour-to-grayscale conversion method and the Rayleigh distribution function also aims to improve the image contrast [6]. After pre-processing, image segmentation is required to

highlight the object. Thresholding is an image segmentation that converts grayscale to binary image. The technique is simple to implement and it works well as compared to global thresholding.

2.2 Support Vector Machine

The SVM is a supervised machine learning algorithm which can be used for both classification and regression. The SVM constructs an optimal hyperplane to find a separating line between data of two classes and maximize the margin of the data. There are several kernel functions in SVM as shown in Figure 1. Each kernel gives different effects as it depends on user requirement.

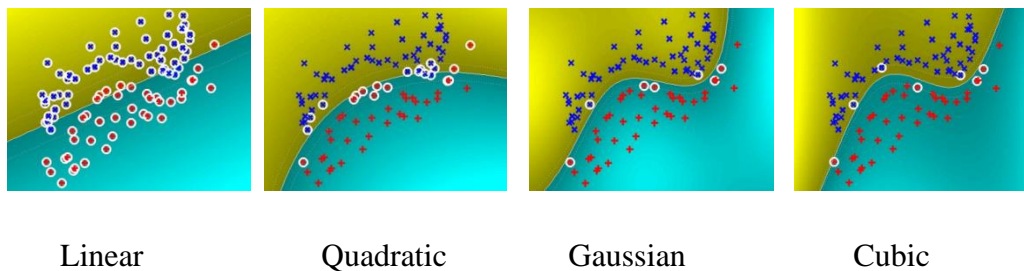


Figure 1. Different kernel functions in SVM, [13]

In the SVM, the kernel trick is a technique used to maximize the margin hyperplane and to transform the data. After the transformations, an optimal boundary between possible outputs is found. The kernel trick allows operating in the original function space without calculating the data coordinates in a higher dimensional space [14]. The SVM constructs its solution in terms of a subset of the training input (Awad and Khanna [15]). The system is used to determine the type of disease with the process of training and testing to reduce the number of errors in the network. Cubic kernel is represented as in Equation (1).

$$k(x_i, x_j) = (x_i^T x_j + 1)^3 \quad (1)$$

In this study, the cubic SVM is used to detect the paddy leaf disease. It is effective in high dimensional spaces. According to Khan et al. [16], a cubic SVM is a type of classifier where the classifier's kernel function is cubic given as $k(x_i, x_j)$ as in Equation (1).

One advantage of cubic SVM is that it gives the highest or gives a high accuracy. The study from Jain et al. [17] and Umar et al. [18] indicated that the cubic SVM classifier gives a high accuracy compared to other classifier. Mohan and Balasubramaniam [7] apply SVM to recognize the category of the diseases. The decision boundary hyperplane between two classes is required to classify the diseases [19].

3. METHODOLOGY

This project aims to develop a tool for paddy leaf disease detection by using image processing that is implemented using MATLAB R2020a. The image undergoes several processes and

techniques until the image has been classified into different disease types. The flow chart of the process of the system is illustrated in Figure 2.

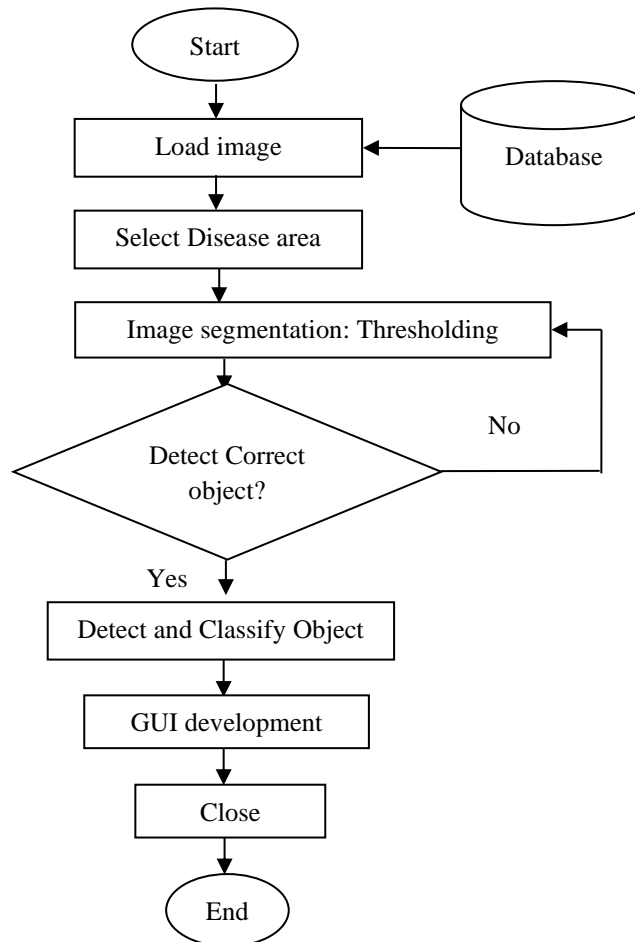


Figure 2: Flow Chart of Methodology

Figure 2 shows the flow chart of the processes involved in the methodology. A system is developed using MATLAB R2020a for paddy leaf diseases analysis. A total of 105 data from UCI database of Rice Leaf Diseases [20] for each disease is tabulated in Table 1.

Table 1: Total data of Paddy Leaf Disease

Disease	Total Data
Bacterial Leaf Blight	39
Brown Spot	33
Leaf Smut	33

Figure 3 shows the three types of paddy leaf disease samples that are bacterial leaf blight, brown spot and leaf smut are taken from the database.



Figure 3. Sample of Paddy Leaf Diseases

In order to reduce processing time, region of interest (ROI) for each image is required to be determined by manually selecting the region of disease area. Thresholding is applied for image segmentation, the image is converted gray-scale to a binary image using the global image threshold using the Otsu method. This method is widely used since it is easy to apply and time saving. This particular method segments the image into light region, T_0 and dark region, T_1 . Region T_0 is a set of intensity levels from 0 to t or in set notation $T_0 = \{0, 1, \dots, t\}$ and region $T_1 = \{t, t+1, \dots, l-1, l\}$ where t is the threshold value, l is the maximum gray level. In the implementation, otsu thresholding in MATLAB is used. After thresholding, the required object highlighting the disease is highlighted.

If the disease is not correctly detected, then the thresholding process will be repeated using different threshold value. In disease detection, a set of features such as area, perimeter contrast, correlation, energy, and homogeneity are calculated.

The area and perimeter was obtained by using blob measurement from the blob function. After the outlines are created, the area and perimeter were calculated based on the region. The other four features that contrast, correlation, energy, and homogeneity are extracted from the Gray-Level Co-occurrence Matrix (GLCM).

$$\text{Contrast} = \sum_{i=0, j=0}^N p_{i,j} (i-j)^2 \quad (2)$$

$$\text{Correlation} = \sum_{i=0, j=0}^N p_{i,j} \frac{(j-\mu_j)(i-\mu_i)}{\sigma_i \sigma_j} \quad (3)$$

$$\text{Energy} = \sum_{i=0, j=0}^N (p_{i,j})^2 \quad (4)$$

$$\text{Homogeneity} = \sum_{i=0, j=0}^N \frac{p_{i,j}}{1+(i-j)} \quad (5)$$

where N represents the total number of pixels, $p_{i,j}$ represents the probability of pixel values, μ represents the mean, and σ represents the standard deviation.

After all the features for the dataset is collected, they are used for training and testing using cubic SVM. According to [21], it is appropriate way to split the dataset into training and testing. In this study, all the data are separated into two tables which are 70% for the training while the

other 30% for the testing [22]. The idea is to provide more training data which in turn makes the classification model better.

Then, the performance of the cubic SVM is evaluated using accuracy, precision and misclassification based on confusion matrix. Accuracy gives the percentage of correct predictions made by the system. Then, precision shows the percentage of positive predictions made by the system. Next, misclassification indicates the percentage of wrong predictions made by the system. The formulas for accuracy, precision and misclassification are as in (6), (7) and (8) respectively.

$$accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (6)$$

$$precision = \frac{TP}{TP + FP} \quad (7)$$

$$misclassification = 1 - accuracy \quad (8)$$

where,

True Positives (TP): predicted yes for the paddy disease, and the paddy do have the disease.

True Negatives (TN): predicted no for the paddy disease and the paddy do not have the disease.

False Positives (FP): predicted yes for the paddy disease, but the paddy actually does not have the disease.

False Negatives (FN): predicted no for the paddy disease, but the paddy actually having the disease.

The last step is to develop a system that compiled all the processes. Figure 4 shows the design layout of the system of paddy leaf disease detector.

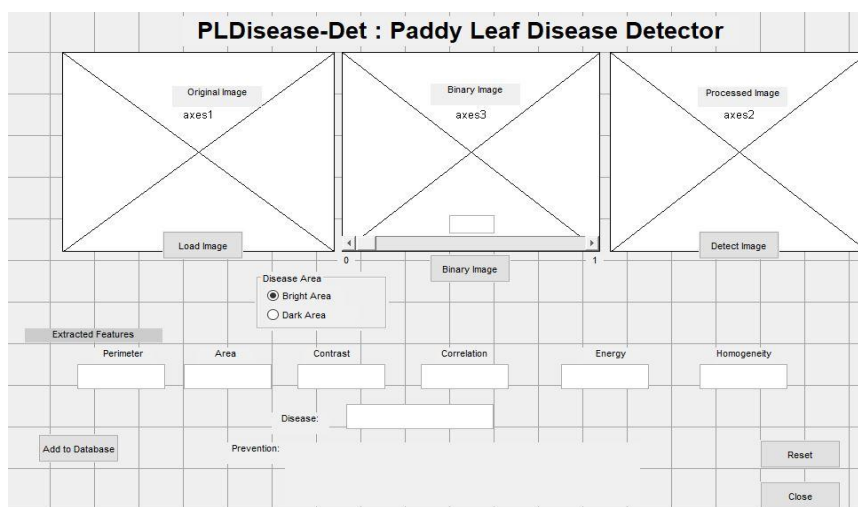
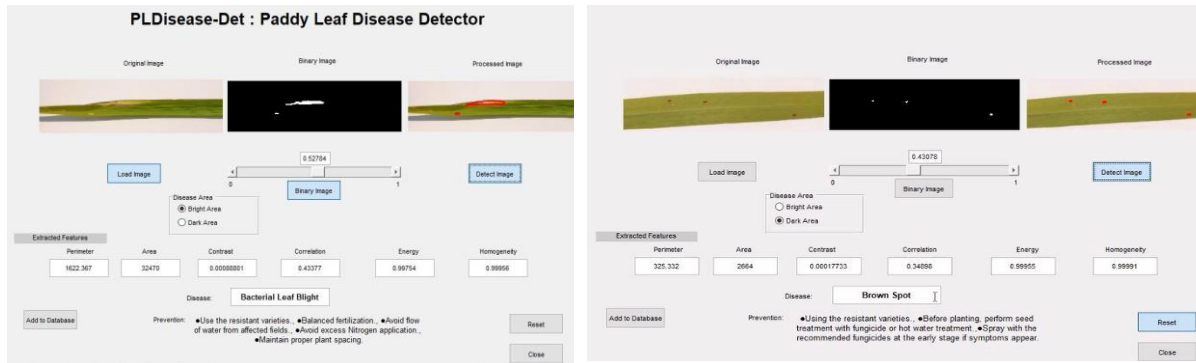


Figure 4: The design layout GUI for paddy leaf disease detector

The user needs to load the image of the paddy leaf. Then, the user is required to select the ROI of the disease before the image is converted to binary image. The value of threshold is automatically calculated but it can be also manually adjusted accordingly. The cubic SVM is applied to train the dataset from all the calculated features.

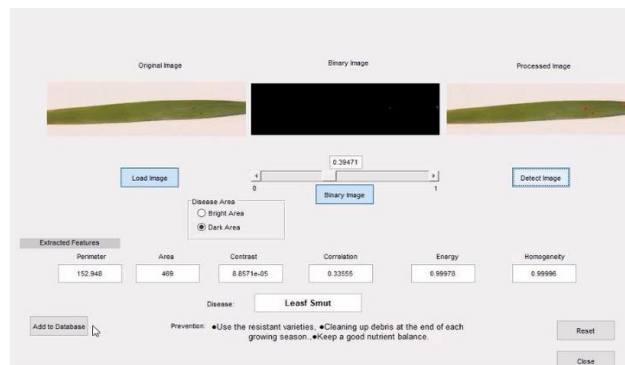
3. RESULT AND DISCUSSION

Figure 5 shows the result of three samples of (a) bacterial leaf blight, (b) brown spot and (c) leaf smut.



(a) bacterial leaf blight data

(b) brown spot data



(c) leaf smut data

Figure 5: The System of paddy disease detection

The confusion matrix is used to calculate the accuracy, precision, and misclassification for each of the disease. Figure 6 below shows the confusion matrix for the paddy leaf disease detection.

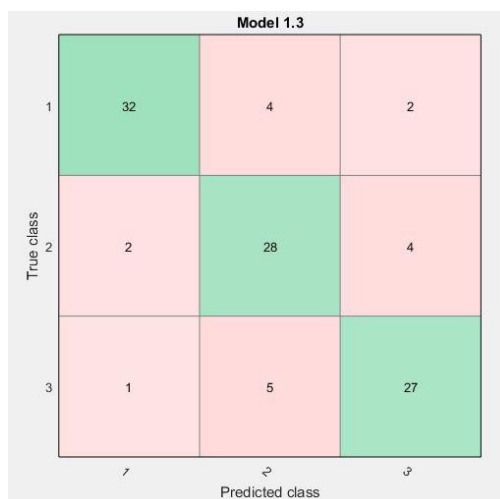


Figure 6. The confusion matrix for paddy leaf disease detection

From the confusion matrix, the values of TP , TN , FN and FP can be identified. The values are shown in Table 2.

Table 2. The calculation for accuracy for each of the paddy leaf disease

Bacterial Leaf Blight	Brown Spot	Leaf Smut
$TP = 32$	$TP = 28$	$TP = 27$
$TN = 64$	$TN = 62$	$TN = 66$
$FN = 6$	$FN = 6$	$FN = 6$
$FP = 3$	$FP = 9$	$FP = 6$

Table 3 shows that the percentage of accuracy, precision, and misclassification for each of the paddy leaf disease using the values in in Table 2.

Table 3: The Percentage of Accuracy, Precision, and Misclassification for Each of Paddy Leaf Disease

Measurement \ Disease	Bacterial Blight	Leaf Brown Spot	Leaf Smut	Average
Accuracy	91.43%	85.71%	88.57%	88.57%
Precision	91.43%	75.67%	81.81%	82.97%
Misclassification	8.57%	14.29%	11.43%	11.43%

As for the result, it can be observed that the accuracy and the precision of bacterial leaf blight are the same at 91.43%. Furthermore, the brown spot and the leaf smut have the accuracy of 85.71% and 88.57% respectively. The precision for brown spot and leaf smut are calculated and the percentages for each disease are 75.67% and 81.81% respectively. Meanwhile, the misclassification for bacterial leaf blight, brown spot, and leaf smut are 8.57%, 14.29%, and 11.43% each. The average of the accuracy is 88.57% and the average of precision is 82.97% while, the average for misclassification is 11.43%. From this result, it is observed that the

proposed method can identify and classify the paddy leaf disease at more than 80% in terms of accuracy and precision while misclassification at less than 15%.

3. CONCLUSION

The set of processes developed in this study successfully detects the diseases using all the features value based on the sample image. The system can be used to assist the farmers and also the paddy researchers to detect the paddy disease through the leaf easily and they can take an action to control the diseases from getting widespread. This system is capable to detect the diseases earlier as soon as it occurs on leaf which can help avoid the loss and reduce the dependency on the export of some degree. For the improvement, less technical part of the system should be applied for the paddy farmer to use the system. A fixed value of threshold value is applied so that the user can get the output after inserting the image without the need to adjust the threshold value. Besides that, other part of paddy can be used for recognition purposes.

ACKNOWLEDGEMENT

The authors would like to express our gratitude to Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Shah Alam for the technical support. Besides that, special thanks also goes out to those who have been associated/ involved directly or indirectly in this project and to the organizations for supporting us throughout our study.

REFERENCES

- [1] N. Manohar and K. J. Gowda, "Image Processing System based Identification and Classification of Leaf Disease: A Case Study on Paddy Leaf," *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 451-457, 2020.
- [2] A. Patil and K. Lad, "Review of Diseases Detection and Classification for Chilli Leaf using Various Algorithms," *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, , pp. 1-4, 2019.
- [3] R. Anand, S. Veni and J. Aravinth, "An application of image processing techniques for detection of diseases on brinjal leaves using k-means clustering method," *2016 International Conference on Recent Trends in Information Technology (ICRTIT)*, Chennai, pp. 1-6, doi: 10.1109/ICRTIT.2016.7569531, 2016.
- [4] P. Sharma, P. Hans and S. C. Gupta, "Classification Of Plant Leaf Diseases Using Machine Learning And Image Preprocessing Techniques," *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, pp. 480-484, doi: 10.1109/Confluence47617.2020.9057889, 2020.
- [5] D. Swathi and A. Bharathi, "Disease analysis of paddy leaves using classification," *South Asian Journal of Engineering and Technology*, vol. 2, no. 9, pp. 1-5, 2016.
- [6] T. Gayathri, and P. Neelamegam, "Paddy leaf disease detection using SVM with BFN classifier," *International Journal of Pure and Applied Mathematics*, vol. 117, no. 15, pp. 699-710, 2017.
- [7] K. J. Mohan and M. Balasubramanian, "Recognition of paddy plant diseases based on Histogram Oriented Gradient Features," *International Journal of Advance Research in Computer and Communication Engineering*, vol. 5, pp. 3-6, 2016.
- [8] P. R., Narmadha, and G. Arulvadiyu, "Disease Symptoms Identification In Paddy Leaf Using Image Processing," *The Journal of Advances in computational Sciences and Technology*, vol. 10, no. 8, pp. 2213-2223, 2017.

- [9] A. A. Joshi and D. B. Jadhav, "Monitoring and controlling rice diseases using image processing technique," *2016 International Conference on Computing, Analytics and Security Trends (CAST)*, pp. 471-476, 2016.
- [10] G. Jayanthi, K. S. Archana and A. Saritha, "Analysis of automatic rice disease classification using image processing techniques". *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 3, pp. 1-6, 2019.
- [11] M. A. Bakar, A. H. Abdullah, N. A. Rahim, H. Yazid, S. N. Mismam and M. J. Masnan, "Rice leaf blast disease detection using multi-level colour image thresholding," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 10, pp. 1-6, 2018.
- [12] G. T. Devi, A. Srinivasan, S. Sudha and D. Narasimhan, "Web enabled paddy disease detection using compressed sensing". *Journal of Mathematical Biosciences and Engineering: MBE*, vol. 16, no. 6, pp. 7719-7733, 2019.
- [13] S. Dreiseitl, Support vector machines. Retrieved 30 June, 2020, from: <http://dsg.harvard.edu/courses/hst951/Spring02/951-12.pdf>, 2002.
- [14] G. Zhang, (2018). *What is the kernel trick? Why is it important?*. Retrieved June 30, 2020, from <https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important-98a98db0961d>.
- [15] M. Awad and R. Khanna, "Support Vector Machines for Classification," In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designer*, pp. 39-66. Apress, Berkeley, CA, 2015.
- [16] M. Khan, T. Goskula, M. Nasiruddin and R. Quazi, "Comparison between k-nn and svm method for speech emotion recognition," *International Journal on Computer Science and Engineering*, vol. 3, no.2, 607-611, 2011.
- [17] U. Jain, K. Nathani, N. Ruban, A. N. J. Raj, Z. Zhuang and V. G. Mahesh, "Cubic SVM classifier based feature extraction and emotion detection from speech signals," *2018 International Conference on Sensor Networks and Signal Processing (SNSP)*, pp. 286-391, 2018.
- [18] A. A. Umar, I. M. Saaid and A. A. Sulaimon, "An SVM-Based Classification and Stability Analysis of Synthetic Emulsions Co-Stabilized by a Nonionic Surfactant and Laponite Clay," *Science and Technology Behind Nanoemulsions*, vol. 111, 2018.
- [19] N. Mangala, P. B. Raj, S. G. Hedge and R. Pooja, "Paddy leaf disease detection using image processing and machine learning," *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, vol. 7, no. 2, pp. 97-99, 2019.
- [20] H. B. Prajapati, J. P. Shah, and V. K. Dabhi, "Detection and classification of rice plant diseases". *Journal: Intelligent Decision Technologies, IOS Press*, vol.11, no. 3, pp. 357-373, 2017.
- [21] R. Draelos, (2019). *Best Use of Train/Val/Test Splits, with Tips for Medical Data*. Retrieved 16 July, 2020, from <https://glassboxmedicine.com/2019/09/15/best-use-of-train-val-test-splits-with-tips-for-medical-data/>
- [22] K. Korjus, M. N. Hebart, and R. Vicente, "An efficient data partitioning to improve classification performance while keeping parameters interpretable". *PloS one*, vol. 11, no. 8, e0161788, 2016.