Research Design 2

Katherine Attard

June 27, 2023

1 Chosen Research

1.1 Description of Theme and Topic Rationale

Manual identification of plant diseases can be time-consuming and inaccurate. There is a growing need for automating the process of identifying and treating plant diseases to maintain agricultural productivity. The application of image processing and AI can provide fast and accurate detection, which will enhance disease management strategies. In this paper image processing and artificial intelligence are used for the purpose of detecting and identifying plant diseases.

1.2 Positioning and Research Onion

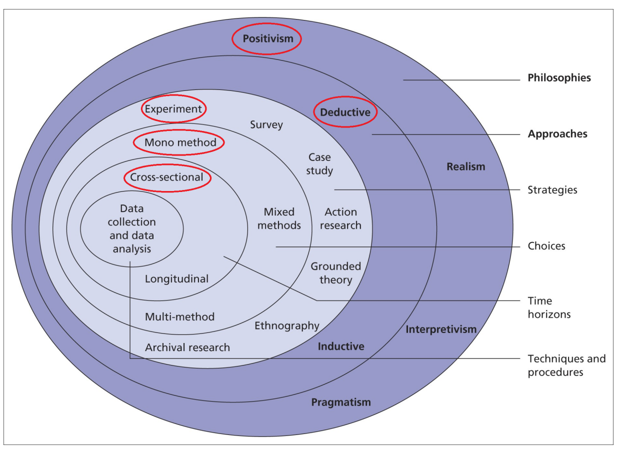

Figure 1: Research Onion

The approach taken in this study is one with a positivist philosophy with a deductive approach. The study has a positivist approach as it relies on empirical, observable data for research. It uses quantifiable measures to evaluate the accuracy of the image recognition system. The project relies on an objective approach to identify plant diseases and develop techniques to uncover their presence without inclusion of subjective personal opinions. The research uses a deductive approach where a hypothesis was derived from the existing literature and then tested using empirical evidence. The hypothesis was that the accuracy of disease detection using image processing would be better than random.

The study uses an experimental design as it tests the hypothesis using a controlled method. The experiment involved applying image recognition and processing algorithms to a specific dataset of tomato plants and calculating the accuracy of disease detection.The research can be considered a mono method study as it primarily uses one method of data collection, which is image processing and recognition on a pre-existing dataset. It is also cross-sectional since the study examines a particular phenomenon (disease in tomato plants) at a specific point in time, using a specific dataset. Changes are not tracked over time, and longitudinal effects are not examined.

1.3 Background to this research theme

In the context of increasing food demand, ensuring plants are healthy and increasing productivity is vital. Traditional plant disease detection relies on expert knowledge which is not always available, especially in remote or underprivileged areas. With advances in AI and image processing technologies, automatic plant disease detection has become an area of intense study. This research theme originates from this need for improved, accessible disease identification and management tools.

1.4 Hypothesis

The underlying hypothesis of this study was that using Histogram of Oriented Gradients (HOG) and k-Nearest Neighbor (k-NN) significantly improves plant disease detection.

H0: There is no significant improvement in plant disease detection using HOG and k-NN.

H1: There is a significant improvement in plant disease detection using HOG and k-NN.

1.5 Research Aim and Purpose statement

The aim of this research is to evaluate the effectiveness of image processing and AI techniques in detecting and identifying plant diseases. The purpose is to understand the potential of these technologies in improving the accuracy and speed of disease detection, and how they could be integrated into existing agricultural practices for better yield and improved food security.

2 Review of Research Methodology

2.1 Undertake a short literature review about the methodologies used in other studies

There are a variety of causes of plant diseases - these include infectious causes (biotic factors) or factors caused by for example nutrient deficiency and soil composition and irrigation issues (abiotic factors). Studies on plant disease identification generally rely on visible changes in the leaf such as colour differences, the presence of holes or changes in edges or white powders on the leaf. These make it possible for image processing to detect and identify plant diseases. Image processing generally includes image acquisition, image pre-processing, image segmentation, feature extraction and classification. Although several imaging techniques could be applied such as thermal imaging, fluorescence imaging, multispectral and hyperspectral imaging [2],image acquisition in the visible range using a digital camera eliminates the need for additional costly equipment. Furthermore there is a growing number of public databases with images from the visible spectrum which serve as a reference of healthy and diseased plants such as PlantVillage (with over 50,000 images on some 14 crops and 26 diseases) [3], Image Database for Plant Disease Symptoms (over 2300 images of a large number of plant diseases) [4] or the American Pathological Society (APS) database (over 7000 images from various plants, https://imagedatabase.apsnet.org/search.aspx). Others are described in [5]. During preprocessing the image is resized, enhanced,, with colour converted to grayscale, noise removal and histogram equalization [6,7,8,9,10]. In segmentation, which may be local or global if the whole image is used, similar parts of the image are grouped together [11,12,13,14,15] using one techniques such as k-means clustering segmentation [16,17,18,19], active contour model segmentation [20], thresholding based segmentation [21,22,23] or Fuzzy C-Means Clustering segmentation [24]. First various types of noise generated during transmission of images such as Gaussian noise or Speckle noise [30,31] is reduced using filters such as Mean Filter, Guassian Filter, Weiner filter, Kuan filter and others [29,32]. Critical features such as such as leaf colour, texture, morphology, contours and edges [25] are then selected in feature extraction and the features that are retained are used to classify the data into clusters in a supervised or unsupervised manner. Techniques used for this may include artificial neural network [26,27], decision tree [27], fuzzy measure [27,28] or support vector machine [27,28].

Extensive research has been carried out on detecting various plant diseases, with many focusing on powdery mildew and blight. . The techniques used generally include k-Means Clustering Segmentation and SVM for image processing [33,34,35,36].

Whilst the accuracy of image detection differs widely, from 68.1% to 97.2%, depending on the algorithm and plant, (reviewed in [37]), the highest accuracy (97.2%) was obtained using Supported Vector Machines (SVM), a supervised learning algorithm, which classifies several objects into two categories by plotting features as coordinates on a 2d or 3d plane and draws a hyperplane with the maximal distance from the nearest coordinates for each category. However accuracy to detect disease can also vary using SVM [37]

2.2 Distinguish between academic and non-academic material

This review relies on academic sources of information. Academic material usually refers to scholarly works that are peer-reviewed and published in academic journals. These works are written by researchers and professionals in a specific field and follow a structured format, including an abstract, methodology, results, and references. Non-academic materials, on the other hand, are those that are typically written for a general audience, such as news articles, blog posts, and opinion pieces, which may not have gone through a rigorous peer-review process. The plant images may be from non-academic sources.

2.3 Contextualised literature and research material

The selected literature provides a comprehensive overview of the current methodologies and techniques used in the field of image [Find reference?] processing and AI for plant disease detection. It sets the context for understanding how different researchers have approached this issue, the technologies they have employed, and the efficacy of these methods.

2.4 Add a good element of critical literature arguments

There is a general agreement in the literature that image processing and AI have the potential to significantly improve plant disease detection. However, there is a contrast in the specific techniques used, such as different machine learning algorithms (as evidenced by I don't know what reference to use here yet). The main knowledge gap appears to be how these techniques can be best implemented at scale in diverse agricultural settings and for a wide range of plant diseases.

3 Reflection on the Chosen Methodology

3.1 Research Questions

Is it possible to use HOG and k-NN to detect diseases in plants? Does a data set need to be created, or are there publicly available datasets?

3.2 Objectives

Identify suitable existing data sets of images of diseased and healthy tomato plants. Propose a suitable model by reviewing existing algorithms. Evaluate the proposed model's effectiveness using real data.

3.3 Understanding of Research Philosophies, Approaches, and Main Research Paradigms

The research described here related to AI and image processing for plant disease detection is empirical, objective and largely positivist using these computational tools to gain understanding on the observable world as opposed to an interpretivist research philosophy which is largely subjective, pragmatist research philosophy, or realistic research philosophy. The approach is largely quantitative, relying on measurable data (like the features extracted from the images) to develop and validate the models. This aligns with the post-positivist paradigm, which maintains that knowledge is conjectural and that our understanding of the world can be refined through observation and measurement. The approach is deductive - based on a hypothesis which is tested by collecting data and using the quantitative result of the accuracy of disease identification using the AI and image processing tools compared to random as opposed to inductive approach where for example an observation could result in some recognition of a pattern from which general conclusions can be drawn. In the deductive approach used here the accuracy comparison will be used to determine whether to reject the initial hypothesis or accept it.

3.4 Chosen Methodology

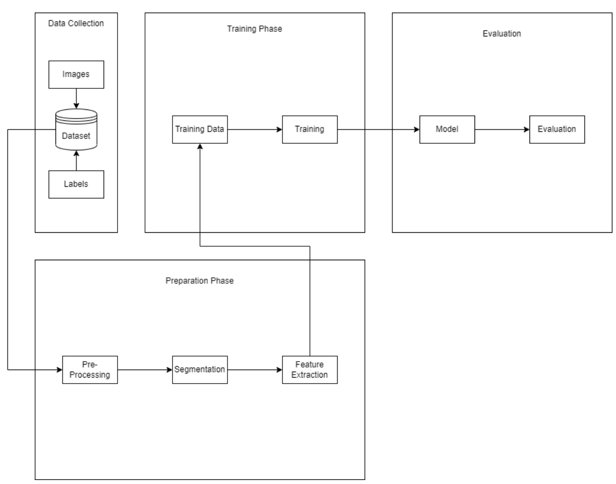

Figure 2: Research Pipeline

se. Based on literature reviewed, many other papers use k-Nearest Neighbor for classification. 18300 images were acquired from a publicly available dataset on Kaggle (https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset). Pre-processing will be converting the image to grayscale and 72x72 pixel resolution. The images will be segmented into cells of 16x16 pixels. HOG will extract features and k-NN shall classify the images.

3.5 Description of Chosen Research Methodology, Experiment Design, and Method of Analysis

The proposed research will utilize a combination of computer vision and machine learning techniques to classify plant diseases from a large dataset of images. The methodology encompasses several stages. The initial step in the methodology is pre-processing. The images will be converted to grayscale to reduce the computational complexity. After converting to grayscale, the images will be resized to a resolution of 72x72 pixels. This step ensures that all images have the same dimensions. The next step involves segmenting the resized images into non-overlapping cells of 16x16 pixels. By breaking down the images into smaller parts, the algorithm can capture local features more effectively. For each cell, we will extract features using the HOG descriptor. HOG is a popular technique for object recognition tasks, as it captures the shape information in an image by quantifying the orientations of the gradients within the cell. The extracted features will be used to classify the images using the k-NN algorithm. k-NN is a type of instance-based learning that predicts the class of a new sample based on the classes of its k closest samples in the feature space.

The dataset will be split into training and testing sets, with a split of 80% for training and 20% for testing. The training set will be used to train the k-NN model, and the test set will be used to evaluate the performance of the model on unseen data. The performance of the classifier will be evaluated using accuracy.

3.6 Ethical Considerations

There are a few ethical considerations to be aware of in this research. First, if the image data is collected from third-party sources, it is crucial to ensure that the data is used ethically and with permission. Secondly, it's essential to acknowledge the limitations of the developed AI model and not overstate its capabilities, as doing so could potentially lead to misapplication and harm (for instance, misdiagnosis of a plant disease could lead to incorrect treatment and economic loss). Lastly, the research should strive to be inclusive and consider a wide variety of plant diseases, not just those that impact economically significant crops, to avoid biases.

4 Results, Analysis, and Discussion

4.1 Present your results

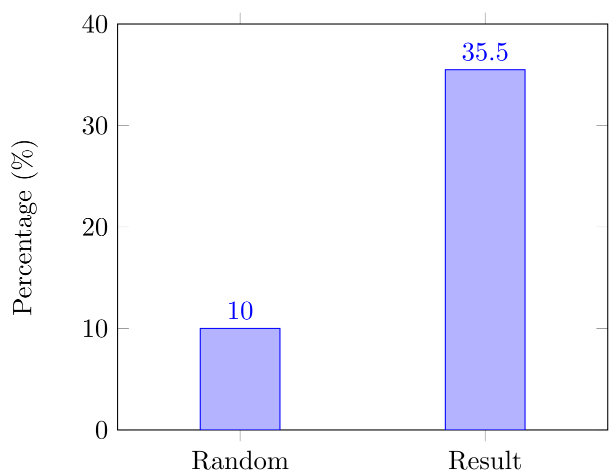

Figure 3: Percentage accuracy of plant disease identification obtained with HOG and K-NN compared to random

Using the AI and image processing resulted in 35.5% accuracy as compared to random accuracy of 10%. This is shown graphically in Figure 3. The random result of 10% was calculated by dividing the total number of images in the database, 18300 images into 10 roughly equal categories, one category for healthy and one for each disease included; namely Healthy, Bacterial spot, Early blight, Late blight, Leaf mould, Septoria leaf spot, Spider mites two spotted spider mite, Target Spot, Tomato mosaic virus and Tomato Yellow Leaf Curl Virus.

4.2 Analyse and interpret results

The approach adopted here determined the correct answer 35.5% of the times when compared to the random approach which would be 10%. Thus the approach described here was 3.6 fold (35.5% compared to 10%) times better compared to random guessing.

4.3 Compare, contrast and comment on different results

4.4 Discuss results in relation to the original hypothesis and other studies where appropriate

The model used here improved plant disease identification 3.5 fold. Thus the hypothesis that AI and image processing can improve disease identification can be accepted. However the proposed pipeline has a much lower level of accuracy than those reported in the literature of between 68.1% to 97.2% [37]. The accuracy may depend on the plant type and disease. The colour and shape of leaves might have an impact, as well as the effect of the plant disease on the leaf, making some diseases harder to detect than others. In further work the influence of disease type on the accuracy could be determined. Despite the influence of plant disease, the method could benefit from improvements. Convolutional Neural Networks (CNN) are a more complex feature extraction method and would be expected to yield better results than HOG. In the literature SVM was more successful for clustering than other approaches used [37-39]. De-noising the images using for example the Guassian or Mean filter is also recommended to remove or minimise the effect of artefacts from data transfer [29,32]. This was not attempted here and could very well be the largest limitation of the study. It would need to be added to the pipeline for more successful results that could then be applied in the field. A limitation of this approach in general is that to detect plant disease based on differences observable in the leaf requires that some damage has already occurred. An approach that includes prevention before that stage would be better. This could be done by monitoring pH and nutrients in the soil substrate or in water for hydroponics for example. The image processing described here should therefore not be seen alone, but it could be combined with other technological tools involved in this monitoring, gathering information on climate changes and pests in the region from satellite images or other data that is collected regionally would be a stronger approach to reach food sustainability. When seeing if this approach can be applied to real life situations, one needs to consider the influence of climate and cloud cover on lighting and therefore on image quality. Improving the image preprocessing steps would be crucial to overcome some of these effects. There are an increasing number of image databases with larger numbers of images. However the variety of crops included is still limited and till now regional differences due to for example different weather and climatic conditions and different plant varieties are generally not taken into account. More is required to gather images from real-life settings as opposed to research on images collected in a research or lab setting. Thus it is advisable to expand datasets taking these factors into account. The effect of a plant disease can vary depending on the plant life cycle. Whilst this and most studies focus on the leaf, the flowers, fruit, fruit colour, shape and even the roots could also be important for a successful crop. Leaf health is an important first step. In future one can expect other parts of the plant to be studied using image recognition. Moreover a plant can have combinations of diseases, or the effect of more than one disease and adverse conditions. There is a lack of research on such combined effects which are expected to make it harder to recognise the plant's ailments and it is recommended that more is done in this regard. It would be expected that this technology can be applied using a mobile app since mobiles have very good cameras and are not within reach of most farmers.

5 Conclusion

5.1 State the main conclusion/s of your research

AI and image processing can indeed help improve plant disease identification. The pipeline successfully proved that results are better than random. In this study accuracy was improved 3.5 fold however this can be improved further to reach levels reported in the literature or better. Further work includes investigating the effect of plant disease on accuracy, studying combined effects of multiple diseases and/or the effect of substrate and environmental influences, research using images from real-life situations not in a lab setting is required, the effect of climate and cloud cover. A further improvement could be the development of a mobile app.

5.2 Make use of conclusion/s to address your research questions and hypothesis

A suitable dataset was publicly available and used for this research. AI and image processing are capable of improving plant disease identification. The pipeline successfully proved that results are better than random. In this study accuracy was improved 3.5 fold however this can be improved further to reach levels reported in the literature or better.

5.3 Identify any shortcomings in methodology

The model's performance is significantly influenced by the quality and diversity of the images used for training. Poor image quality or a lack of diversity in leaf images can limit the model's ability to generalize to real-world conditions. Additionally, the methodology did not generate a confusion matrix, precision, F1-score or recall.

5.4 Suggest ideas for further research

The addition of various de-noising filters such as a Gaussian filter [29,32] may improve accuracy. The application of more complex feature extraction algorithms instead of HOG such as CNN is also recommended. SVM gave high accuracies in the literature [37-39]. Obtaining results more in-depth than accuracy such as confusion matrix, precision, F1-score, and recall, could allow for further and more thorough analysis. Future research could aim to improve the model's ability to handle varying image quality and further increase its robustness to real-world conditions. This could involve collecting a more diverse dataset, incorporating additional data augmentation techniques, or exploring other deep learning architectures. It would also be worthwhile to investigate how the model performs in a practical field environment and how it could be integrated into a user-friendly tool for farmers or agricultural experts. Further research could also explore the potential of multi-label classification for identifying instances where leaves are afflicted with more than one disease.

References

[1] A. Sinha and R. Shekhawat, "Review of Image Processing Approaches for detecting Plant diseases," *IET Image Processing*, vol. 14, June 2020.

[2] V. Singh, N. Sharma, and S. Singh, "A review of imaging techniques for plant disease detection," *Artificial Intelligence in Agriculture*, vol. 4, pp. 229–242, Jan. 2020.

[3] D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing," *ArXiv*, Nov. 2015.

[4] J. Barbedo, L. Koenigkan, B. Halfeld-Vieira, R. Costa, K. Nechet, C. Godoy, M. Lobo Junior, F. Patrício, V. Talamini, L. Chitarra, S. Oliveira, A. Keiko, J. M. Fernandes, T. Santos, F. Cavalcanti, D. Terao, and F. Angelotti, "Annotated Plant Pathology Databases for Image-Based Detection and Recognition of Diseases," *IEEE Latin America Transactions*, vol. 16, pp. 1749–1757, June 2018.

[5] J. Liu and X. Wang, "Plant diseases and pests detection based on deep learning: a review," *Plant Methods*, vol. 17, p. 22, Feb. 2021.

[6] M. G. Kaur, S. Kaur, and A. Kaur, "Plant Disease Detection: a Review of Current Trends," *International Journal of Engineering & Technology*, vol. 7, no. 3.34, pp. 874–881, 2018. Number: 3.34.

[7] K. Gavhale, U. Gawande, and K. Hajari, "Unhealthy region of citrus leaf detection using image processing techniques," *2014 International Conference for Convergence of Technology, I2CT 2014*, Apr. 2015.

[8] P. S. Marathe, "Plant Disease Detection using Digital Image Processing and GSM," 2017.

[9] K. Khairnar and R. Dagade, "Disease Detection and Diagnosis on Plant using Image Processing A Review," *International Journal of Computer Applications*, vol. 108, pp. 36–38, Dec. 2014.

[10] T. N. Tete and S. Kamlu, "Plant Disease Detection Using Different Algorithms," in *Annals of Computer Science and Information Systems*, vol. 10, pp. 103–106, 2017. ISSN: 2300-5963.

[11] D. Kaur and Y. Kaur, "Various image segmentation techniques: A Review," *International Journal of Computer Science and Mobile Computing*, vol. 3, pp. 809–814, May 2014.

[12] Patel Jigna. J, "Various Segmentation Techniques in Image Processing.pdf," *International Journal of Engineering Science and Computing*, vol. 7, no. 6, pp. 12781–12783, 2017.

[13] S. Yuheng and Y. Hao, "Image Segmentation Algorithms Overview," July 2017.

[14] N. M. Zaitoun and M. J. Aqel, "Survey on Image Segmentation Techniques," *Procedia Computer Science*, vol. 65, pp. 797–806, Jan. 2015.

[15] L. Lalaoui, T. Mohamadi, and A. Djaalab, "New Method for Image Segmentation," *Procedia - Social and Behavioral Sciences*, vol. 195, pp. 1971–1980, July 2015.

[16] P. B. Padol and A. A. Yadav, "SVM classifier based grape leaf disease detection," *2016 Conference on Advances in Signal Processing (CASP)*, pp. 175–179, June 2016. Conference Name: 2016 Conference on Advances in Signal Processing (CASP) ISBN: 9781509008490 Place: Pune, India Publisher: IEEE.

[17] P. P. Warne and S. Ganorkar, "Detection of Diseases on Cotton Leaves Using K-Mean Clustering Method," *International Journal of Engineering and Technology(IJET)*, vol. 2, no. 4, 2015.

[18] S. Naikwadi and N. Amoda, "Advances in image processing for detection of plant diseases," *International journal of Application or Innovation in Engineering & Management (IJAIEM)*, vol. 2, no. 11, 2013.

[19] R. Young, P. Birch, F. Fina, J. Obu, B. Faithpraise, and C. Chatwin, "Automatic plant pest detection and recognition using k-means clustering algorithm and correspondence filters,"

[20] P.R. Rothe and R .V. Kshirsagar, "A study and implementation of active contour model for feature extraction: with diseased cotton leaf as example," *International Journal of Current Engineering and Technology*, vol. 4, no. 2, pp. 812–816, 2014.

[21] S. Chouhan, A. Koul, D. U. Singh, and S. Jain, "Bacterial Foraging Optimization Based Radial Basis Function Neural Network (BRBFNN) for Identification and Classification of Plant Leaf Diseases: An Automatic Approach Towards Plant Pathology," *IEEE Access*, vol. PP, pp. 1–1, Feb. 2018.

[22] V. Dabhi, H. Prajapati, and Bhumilka S.Prajapati, *A survey on detection and classification of cotton leaf diseases*. International Conference on Electrical, Electronics and Otimization Techniques(ICEEOT), Mar. 2016.

[23] Prof. R. N. kadu, S. Kangane, S. Vikhe, R. Pandita, and V. Inamke, "Leaf disease detection using Arm7 and image processing," *International Journal of Engineering Research and Applications (IJERA)*, vol. 5, no. 2, pp. 68–71, 2015.

[24] A. Kumari, S. Meenakshi, and S. Abinaya, "Plant Leaf Disease Detection Using Fuzzy C-Means Clustering Algorithm," *International Journal of Engineering Development and Research*, vol. 6, no. 3, pp. 2321–9939.

[25] P. J.k and R. Kumar, "Feature Extraction of diseased leaf images," *Journal of Signal and Image Processing*, vol. 3, pp. 60–63, Mar. 2012. Publisher: Bioinfo Publications.

[26] Anand.H.Kulkarni and Ashwin Patil R. K., "Applying image processing technique to detect plant diseases," *International Journal of Modern Engineering Research (IJMER)*, vol. 2, no. 5, pp. 3661–3664, 2012.

[27] P. P. Naswale and P. E. Ajmire, "Image Classification Techniques- A Survey," *International Journal of Application or Innovation in Engineering & Management*, vol. Volume 5, Issue 2, March - April 2016.

[28] Santosh Kumar Sao and Sandeep B. Patil, "A survey on classification techniques for plant disease detection using image processing," *International Journal for Scientific Research & Development*, vol. 3, no. 4, pp. 1122–1125, 2015.

[29] S. Das, J. Saikia, S. Das, and N. Goni, "A comparative study of different noise filtering techniques in digital images," vol. 3, no. 5, 2015.

[30] Er. Amita Kumari and Er .Pankaj Dev Chadha, "A survey on filtering technique for denoising images in digital image processing," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 8, pp. 612–614, 2014.

[31] Mohd Awais Farooque and Jayant S. Rohnakar, "Survey on various noises and techniques for denoising the color image," *International Journal*

of application or Innovation in Engineering and Management(IJAIEM), vol. 2, no. 11, pp. 217–221, 2013.

[32] Govindaraj. V and Sengottaiyan. G, "Survey of image denoising using different filters," *International Journal of Science Engineering and Technology Research(IJSETR)*, vol. 2, no. 2, pp. 344–351, 2013.

[33] G. Tripathi and J. Save, "An image processing and neural network based approach for detection and classification of plant leaf diseases," *International Journal of Computer Engineering & Technology (IJCET)*, vol. 6, no. 4, pp. 14–20, 2015.

[34] H. Wang, G. Li, Z. Ma, and X. Li, *Application of neural networks to image recognition of plant diseases*. May 2012. Journal Abbreviation: 2012 International Conference on Systems and Informatics, ICSAI 2012 Pages: 2164 Publication Title: 2012 International Conference on Systems and Informatics, ICSAI 2012.

[35] S. Sannakki, V. Rajpurohit, V. Nargund, and P. Kulkarni, *Diagnosis and classification of grape leaf diseases using neural networks*. July 2013. Journal Abbreviation: 2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013 Pages: 5 Publication Title: 2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013.

[36] Rakesh Chaware, Rohit Karpe, Prithvi Pakhale, and Prof. Smita Desai, "Detection and recognition of leaf disease using image processing," *International Journal of Engineering Science and Computing (IJESC)*, vol. 7, no. 5, pp. 11964–11967, 2017.

[37] Preet Kaur and Parminder Kaur, "Plant Disease Detection using Image Processing: A Review," *Electronics and Communication Engineering Department, Punjab Technical University*, vol. 6, no. 11, pp. 1079–5131, 2019.

[38] S. Phadikar and J. Sil, *Rice disease identification using pattern recognition techniques*. Proceedings Of 11th International Conference On Computer And Information Technology, Jan. 2009. Pages: 423.

[39] Geng Ying, Li Miao, Yuan Yuan, and Hu Zelin, "A Study on the Method of Image PreProcessing for Recognition of Crop Diseases," 2008.

[1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39],