# Tomato Leaf Disease Detection Based on Image Processing and Machine Learning Techniques



By Arfan Shah

In partial fulfillment of the requirement for the degreeBachelor of Science in Computer Science

Department of Computer Science, School of Arts & Science University of Central Asia,

Naryn campus, Kyrgyz republic (2023)

#### **DECLARATION BY AUTHOR**

I/we certify that this work has not been accepted in substance for any degree and is not concurrently being submitted for any degree other than that of Bachelor of Science in Computer Science being studied at the Department of Computer Science, School of Arts & Science, University of Central Asia, Kyrgyz Republic. I/we also declare that this work is the result of my/our own findings and investigations except where otherwise identified by references and that I/we have not plagiarized another's work.

Arfan Shah

Arfan Shah

## **DECLARATION BY SUPERVISOR**

I, the undersigned hereby certify that I have read this project report and finally approve it with recommendation that this report may be submitted by the authors above to the final year project evaluation committee for final evaluation and presentation, in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science at the Department of Computer Science, School of Arts & Sciences, University of Central Asia, Kyrgyz Republic.



Dr. Muhammad Fayaz

#### **ABSTRACT**

The world is facing food crises and the situation in developing countries is getting worst day by day. People do not get enough food which raises questions about their health. Fruits and vegetables are a major source of food for various regions especially in developing countries. Diseases in fruits and vegetables result to reduction of food. To increase the production of fruits and vegetables farmers should be able to identify the disease in time to avoid further damages. This research aims to identify and detect diseases in tomato leaves using Image Processing and Machine Learning techniques. There are six stages from the beginning to the end of this research. The stages include image acquisition, pre-processing, image segmentation, feature extraction, classification, and disease identification. Image cropping, image resizing, and image enhancement will be used to pre-processing the images. For image segmentation various segmentation algorithms will be used, such as Edge Based Segmentation, Threshold Based Segmentation, Region-Based Segmentation, and Cluster-Based Segmentation. Gray Level Co-occurrence Matrix (GLCM), Principal Component Analysis (PCA), and Independent Component Analysis (ICA) will be utilized for feature extraction. Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest, Decision Tree, Logistic Regression, Artificial Neural Network (ANN), and Naïve Bayes will be used for classification. The disease will be classified into four classes namely healthy leaf, early blight, late blight, and Septoria leaf spot.

**Keywords:** Tomato leaf disease, Leaf disease, Image Processing, Machine Learning.

# **ACKNOWLEDGEMENTS**

In this research I have received a generous support from my supervisor Dr. Muhammad Fayaz. I would like to say thank you for the support. In addition, I would also like to express my gratitude to my friends who helped me in this project.

# **CONTENTS**

# **Table of Contents**

| ABSTRACT  | 3  |
|---|----|
| CHAPTER 1: INTRODUCTION                         | 6  |
| CHAPTER 2: LITERATURE REVIEW                    | 8  |
| CHAPTER 4: PROPOSED METHODOLOGY                 | 11 |
| 3.1 PREPROCESSING                               | 13 |
| 3.1.1 Filters:                                  | 13 |
| 3.2 IMAGE ENHANCEMENT                           | 15 |
| 3.3 IMAGE SEGMENTATION                          | 19 |
| 3.4 FEATUER EXTRACTION                          | 21 |
| CHAPTER 5: SIMILAR APPLICATION COMPARISON TABLE | 23 |
| CHAPTER 6: NOVELTY OF THIS RESEARCH             |    |

#### **CHAPTER 1: INTRODUCTION**

Food is one of the most basic needs of human being. Humans eat to survive. As population of the earth increases, food demand also increases and to fulfill the demand, the growth and production of eatables must also increase. The issue of malnutrition is common in developing countries, such as Pakistan, Kyrgyzstan, Tajikistan etc. According to United Nations approximately 37.5 million people in Pakistan are suffering from food crises (Imam, March 9, 2022). In Kyrgyzstan almost 12 percent of children are affected by malnutrition (Kopytin, June 25, 2022). According to United Nations International Children's Emergency Fund (UNICEF), in Tajikistan more than 8 percent of children are affected by lack of food (Ruziev, 2019). Most of the population suffering from lack of food in the mentioned countries belong to its mountainous regions where there is lack of facilities. This research is focused specifically on the mountainous regions of the mentioned countries. This research aims to detect the leaf diseases of various fruits while focusing on tomato in the beginning. In long term this research will work on detecting leaf disease of other fruits and vegetables which grow in the mountainous regions of Pakistan, Kyrgyzstan, Tajikistan, Kazakhstan, and other Central Asian countries.

This research proposes various Image Processing and Machine Learning techniques to detect leaf disease in plants. The research primarily focuses on the tomato leaf diseases and in long run it aims to detect leaf diseases in all sorts of plants. The topic chosen for this research is unique and noble in Central Asian countries especially in Kyrgyzstan and Tajikistan. Most of the people belonging to mountainous areas like Gilgit-Baltistan are dependent on the fruits and vegetables they produce at home. The people belonging to remote regions cannot afford to buy everything from shops. It is convenient for them to grow vegetables and fruits at their homes. This will not only fulfill their own demands, rather they can sell and earn some money. The production of fruits and vegetables can be affected negatively due to various diseases. Yearly farmers suffer from the losses which occur due to plant diseases. If these diseases are identified at early stages, then curing it will be easy. Which can help to enhance vegetables and fruits production. Due to the lack of facilities farmers belonging to these remote regions suffer a lot from plant diseases. For instance, if tomatoes have some sort of disease, farmers cannot find a doctor in villages who can cure it. They cannot afford to go to city or call a doctor from city. To solve this issue, this paper proposes a simple solution using Machine Learning and Image Processing. The farmer can take a picture from the leaf and send it to the database. The database will sort out the disease and will send back a notification with the disease and its

remedy. The research aims to build a simple interface which can be easily used by common farmers. If time permits the research aims to build an app which can be run in both mobile and on web.

Various datasets would be used in this research. A dataset taken from Kaggle will be used for the training purpose. Another dataset used in this research is collected by some researchers at the University of Malakand, Pakistan. This dataset will be used for both training and testing purposes. Moreover, a local dataset collected from Gilgit-Baltistan, Pakistan may also be used for testing purpose.

#### **CHAPTER 2: LITERATURE REVIEW**

(Rahman et al., August 24, 2022) apply Image Processing methods to detect and treatment of diseases in tomato leaves. They have applied Support Vector Machine for classification of diseases. They collected images of healthy, early blight, late blight, and Septoria. The labelled image dataset is trained and testing using Image Processing techniques. The model achieved an accuracy of 100 percent for healthy leaves, 95 percent for early blight, 90 percent for Septoria, and 85 percent for late blight. They first collected images from local tomato plants in Dir region of Pakistan. They have made simple android app named as Leaf Disease Identifier (LDI). User can take a pic from the tomato plant and upload it into LDI. LDI sent image to its database, and it identifies the disease, and it sends back a notification with remedy for the disease. (Dr. Sarojadevi and Nagamani, 2022) published a research paper on detection of disease in tomatoes using deep learning. They have categorized tomato leaf diseases into six categories. They have applied three different Machine Learning algorithms, such as Fuzzy Support Vector Machine (Fuzzy-SVM), Convolutional Neural Network (CNN), and Region-based Convolutional Neural Network (RCNN). In order to train the images, they used color thresholding, image scaling, gradient local ternary patterns, flood filing approaches, and Zernike moments. R-CNN is used to classify images into different categories of diseases. Fuzzy-SVM and CNN are also used for diseases classification. Results of all the three classifiers are compared and picked R-CNN which performed best by giving an accuracy of 96.735%.

(Chowdhury et al., April 30, 2021) apply some state-of-the-art techniques based on Convolutional Neural Network (CNN) to detect tomato leaf disease. They have collected 18,162 images of tomatoes including healthy and affected images. The researchers first classified tomato leaves into healthy or unhealthy (binary classification) and then they categorized the leaves into 6 groups (healthy and different types of unhealthy leaves), and later they also classified into ten classes (healthy and various sorts of unhealthy diseases). They have applied various CNN algorithms, such as ResNet18, MobileNet, DenseNet201, and InceptionV3. InceptionV3 performed best with an accuracy of 99.2% on binary classification while DenseNet201 achieved an accuracy of 97.99% for six class classification and 98.05% accuracy for ten-class classification. (Kirange and Gadade, February 2020) build a module which automatically detect diseases in plant leaves. For feature extraction, they have used GLCM, Gabor and SURF. To classify disease, they have used various algorithms, such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN),

and Naïve Bayes. For this project they have used 500 images of tomato leaves with seven different diseases symptoms. The results showed that Gabor performed well for feature extraction while SVM outperformed in classifying disease.

(Dr. Sreelatha et al., 2021) have developed and designed a computer vision model which helps to build a system for image detection, feature extraction, and classification in real-time. They have used Deep Neural Network (DNN) to classify tomato images in real-time. They found that rates of classification increased compared to the existing models. (Trivedi et al., November 30, 2021) use Convolutional Neural Network (CNN) to effectively find and classify leaf diseases in tomato plants. They have used Google Colab throughout the experiment. They have collected 3000 images of tomato leaves and the images include a healthy leaf along with nine different diseases. They first pre-processed the images and the targeted regions are separated from the original images. Secondly, the images are further processed by using various hyper-parameters of CNN. Lastly, various characteristics, such as color, texture, edge etc are extracted using CNN. The results showed that the model achieved an accuracy of 98.49%.

(Shoaib et al., October 07, 2022) propose a model based on deep learning to detect leaf disease in tomatoes. Model is trained over 18,161 segmented and non-segmented images using various deep learning and Convolutional Neural Network (CNN) techniques. Two state-of-the-art: U-Net and Modified U-Net are used for detection and segmentation of regions of affected leaves. They have made two classes: six-level classification (one healthy and five different diseases) and ten-level classification (one healthy and nine various diseases). The modified U-Net segmentation performed well with an accuracy of 98.66 percent compared to other techniques used. InceptionNet1 achieved an accuracy of 99.12% for six-class classification. The results showed that the proposed model outperformed compared to the existing models. (Khasawnch, Faouri, and Fraiwan, August 24, 2022) apply deep transfer learning to find and classify nine different diseases in tomato plants. In this model they have used leaf images as an input that is given to CNN model for classification. Interestingly they have not done any pre-processing, feature extraction, or image processing. Models used on this research are based on the techniques of transfer learning of deep learning networks. The experiments in this research are repeated ten times to counter randomness. The results in this research are 99.3% precision, 99.2% F1 score, 99.1% recall, and 99.4% accuracy.

(Mohanty, Hughes, and Salathe, September 22, 2016) use 54,306 public images to classify disease prone and healthy leaves which have been collected under

controlled conditions. These images are used for training purpose using Convolutional Neural Network. Fourteen different crop species are used twenty-six various diseases are identified. The trained model achieved an accuracy of 99.35%. (Geetha et al., 2020) use four different stages to detect the type of disease in tomato plants' leaves. The four stages are pre-processing, image segmentation, feature extraction, and classification. Image segmentation is used to detect damaged or affected parts of leaf and pre-processing is used to remove noise. For classification and regression purpose, supervised Machine Learning algorithm, K-Nearest Neighbor (KNN), is used. The research also proposes remedy after detecting the disease in plants' leaves. (Andrew et al., October 03, 2021) used pre trained models based on Convolutional Neural Networks (CNN) for disease identification in plant leaves. Fine tuning of hyperparameters have been done for some popular pre-trained models, such as InceptionV4, ResNet-50, DenseNet-121, and VGG-16. A popular plant dataset from village having 54,305 is utilized for this research. Images are taken from various plant species and diseases are classified into 38 different classes. The research showed that DenseNet-121 outperformed with an accuracy of 99.81%. (Kulkarni et al., November 22, 2021) use a smart and efficient technique for plant disease detection based on Computer Vision and Machine Learning at Cornell University. The model is trained to classify 20 different diseases in five common plants. The model achieved an accuracy of 93%.

This proposed research work consists of six different steps. First, images are being collected from various sources, then the dataset is being pre-processed. After that image enhancement and image segmentation is done. Then features are extracted and then finally the images are being classified. The whole methodology is shown is shown below in Figure 3.1

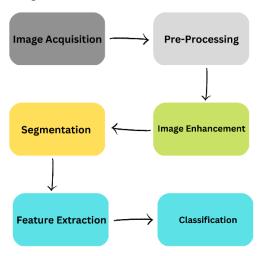


Figure 3.1 Model showing methodology

Below shown is a brief description of all work which will be done in this project. It starts from image acquisition and ends at the classification of image using various Machine Learning algorithms. All the steps are shown below in figure 3.2.

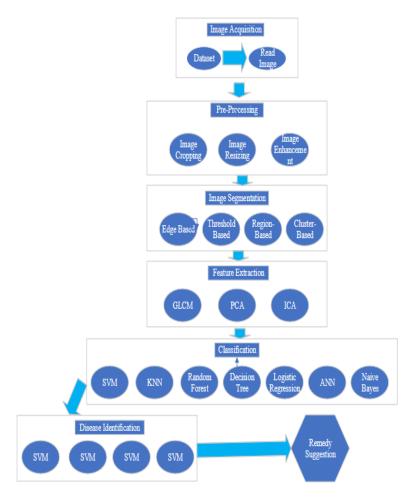


Figure 3.2 Diagram to show all the work.

## 3.1 PREPROCESSING

Pre-processing is the first step in both machine learning and image processing techniques. Before applying machine learning models, the data should be clean so that the predictions should be accurate and precise. Preprocessing in image processing refers to the steps that are performed on an image before it is fed into an algorithm or model for further processing. The goal of preprocessing is to prepare the image for analysis by making it more suitable for the specific task at hand.

Preprocessing can be an important step in the image processing pipeline, as it can help to improve the performance of downstream tasks and algorithms. Preprocessing techniques can be applied individually or in combination, depending on the specific needs of the task at hand.

## **3.1.1 Filters:**

In image processing, filters are used to modify the appearance of an image by applying a mathematical operation to the pixel values. Filters can be used to enhance the visual quality of an image, or to extract specific features or information from the image. Filters are often applied as part of the image processing pipeline, either as a preprocessing step or as a way to extract specific features from the image for further analysis. The choice of filter and the parameters used depend on the specific requirements of the task at hand.

Many filters have been applied in this project. Firstly, mean filter is being applied. A mean filter, also known as a box filter, is a simple and widely used image processing technique that applies a local averaging operation to an image. The basic idea behind a mean filter is to replace each pixel in the image with the average value of its neighboring pixels, within a specified window size. This has the effect of smoothing the image, by reducing the level of detail and noise present. Below in the figure 3.1.1 shown the image after applying a mean filter.

Original Mean filter 14



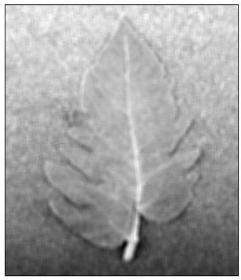


Figure 3.1.1 Original image: left side and Mean filter: right side.

Secondly, gaussian filter is applied to remove unnecessary noises. A Gaussian filter is a type of smoothing filter in image processing that is used to reduce noise and detail in an image. It works by applying a weighted average to the pixel values within a window, with the weights determined by a Gaussian function. The Gaussian function has the property of being symmetrical and bell-shaped, with the highest weight at the center and the weights decreasing as you move away from the center. This results in a smoothing effect that preserves more of the original image detail compared to other types of smoothing filters, such as mean filters. Gaussian filters are widely used in image processing due to their ability to preserve image detail while still reducing noise and smoothing the image. They are often used as a preprocessing step in image processing pipelines, or as a way to smooth images before applying other types of filters. Below in the figure 3.1.2 shown the results after applying gaussian filter.





Thirdly, median filter is applied. A median filter is a type of non-linear smoothing filter in image processing that is used to reduce noise and detail in an image. It works by replacing the value of each pixel with the median value of its neighboring pixels, within a specified window size. Median filters are effective at reducing noise and preserving edges in an image, as they do not blur the image as much as other types of smoothing filters such as mean filters. However, they are also slower to implement, as they require sorting the pixel values within the window. Median filters are often used as a preprocessing step in image processing pipelines, or as a way to smooth images before applying other types of filters. Below in figure 3.1.3 shown the results after applying median filter.

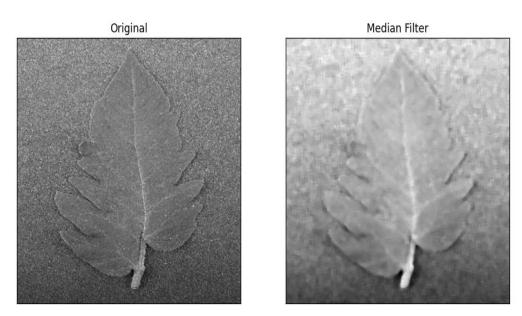


Figure 3.1.3 Original image: left side and Median filter: right side.

### 3.2 IMAGE ENHANCEMENT

After having done with pre-processing next comes image enhancement where different techniques are applied to enhance the quality of image. This is also done to reduce the negative impacts done when various filters are applied. Image enhancement is a process in image processing that is used to improve the visual quality of an image. It can be used to make an image more clear, crisp, and

visually appealing, or to highlight specific features or details in the image. Image enhancement techniques can be applied individually or in combination, depending on the specific needs of the task at hand. They are often used as part of the image processing pipeline to improve the visual quality of the image before it is analyzed or used for further processing.

For enhance the quality of image, image inverse is applied. An image inverse is a type of image enhancement technique that can be used to improve the visual quality of an image. It works by reversing the pixel values in the image, so that the lightest pixels become the darkest and vice versa. To apply an image inverse to an image, the pixel values of the image are first normalized to a range of 0 to 1. Then, the inverse of each pixel value is calculated using the formula: inverse = 1 – value, where "value" is the normalized pixel value. The inverse pixel values are then mapped back to the original pixel range of the image. Applying an image inverse can have the effect of increasing the contrast and visual appeal of an image, especially if the original image is relatively low contrast. It can also be used to highlight certain features or details in the image, by making them more visible against a darker or lighter background.

Below in figure 3.2.1 shown an image after applying image inverse.





Figure 3.2.1 Original image: left side and Image inverse: right side.

Log transformation is another method to improve the quality of an image. Log transformation is a type of image enhancement technique in image processing that is used to increase the contrast and visual appeal of an image. It works by applying a logarithmic function to the pixel values in the image, which has the effect of compressing the range of pixel values and increasing the contrast. The logarithmic function used for log transformation is typically of the form:

$$output = c * log(1 + input)$$

where "input" is the original pixel value, "output" is the transformed pixel value, and "c" is a constant that controls the amount of contrast enhancement. A larger value of "c" results in more contrast enhancement. To apply a log transformation to an image, the pixel values of the image are first normalized to a range of 0 to 1. Then, the logarithmic function is applied to each pixel value, and the transformed pixel values are mapped back to the original pixel range of the image. In figure 3.2.2 result after applying log transformation is shown.

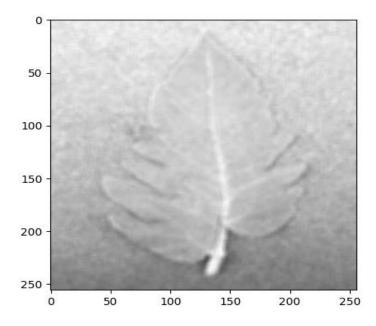
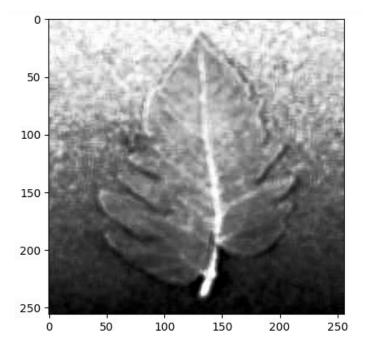


Figure 3.2.2 Image after log transformation.

Next, histogram equalizer is being used to improve quality of image so that the image can be used for proper prediction. Histogram equalization is a type of image enhancement technique in image processing that is used to increase the contrast and visual appeal of an image. It works by adjusting the distribution of pixel values in the image so that the intensity values are more evenly spread across the range. This has the effect of increasing the contrast and making the image appear more balanced and visually appealing. Histogram equalization is a widely used technique for improving the contrast and visual appeal of images, and is especially useful for images with a low contrast or a narrow range of pixel values. It is also commonly used in medical and scientific imaging to improve the visibility of features or details in the image. However, it can also result in a loss of detail in the highlights and shadows of the image, and may not be suitable for all types of images. In the figure 3.2.3 shown the image after applying histogram equalization.



**Figure 3.2.2** Image after Histogram Equalization.

#### 3.3 IMAGE SEGMENTATION

Image segmentation is the process of dividing an image into multiple segments or regions, each of which corresponds to a different object or background in the image. The goal of image segmentation is to partition the image into meaningful regions that can be analyzed or processed individually. Image segmentation is a fundamental step in many image processing pipelines, and is used in a wide range of applications including object recognition, image analysis, and computer vision. The choice of segmentation technique depends on the specific requirements of the task at hand and may involve a combination of different approaches.

There are many image segmentation techniques. In this research Otsu segmentation is being used. Otsu segmentation is a technique for image thresholding, which is a process of dividing an image into two regions based on the intensity values of the pixels. Otsu segmentation is named after Nobuyuki Otsu, who proposed the technique in 1979. The basic idea behind Otsu segmentation is to find the threshold value that maximizes the variance between the two regions of the image. The variance is a measure of how spread out the pixel values are within each region. By maximizing the variance, the resulting regions are more distinct and easier to distinguish. Figure 3.3.1 shows the result after applying otsu segmentation.

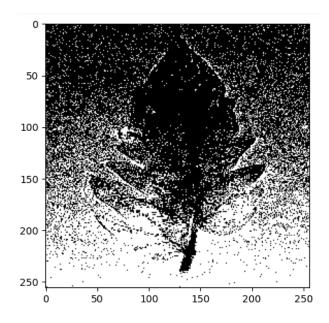


Figure 3.3.1 Image after Otsu Segmentation.

Watershed segmentation is a technique for image segmentation that is based on the idea of treating the image as a topographic map, with the intensity values representing the height of the terrain. The goal of watershed segmentation is to partition the image into regions or "basins" that correspond to the local minima or maxima of the image gradient. Watershed segmentation is a powerful technique for image segmentation, and is especially useful for images with clear boundaries or distinct objects. It is often used in medical and scientific imaging to segment objects or tissues of interest. Figure 3.3.2 shows result after applying watershed segmentation.

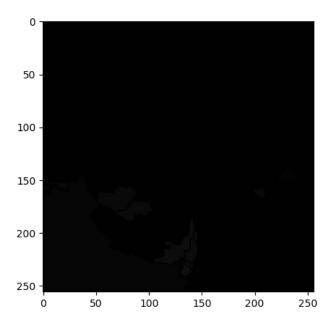
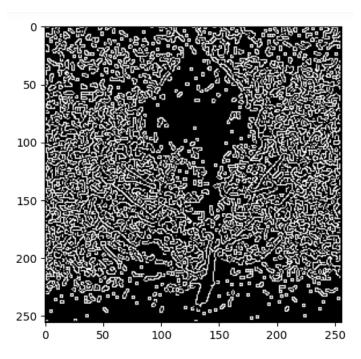


Figure 3.3.2 Image after Watershed Segmentation.

Feature extraction is the process of extracting relevant information or features from an image, in order to analyze or classify the image. The goal of feature extraction is to identify and extract the most important or distinctive aspects of the image, while ignoring irrelevant or redundant information.

There are many different techniques for feature extraction, depending on the specific requirements of the task. For instance, Edge detection: This technique involves identifying the boundaries or edges in the image, which can be used to extract shapes or contours. Feature detection: This technique involves identifying specific points or regions in the image that are distinctive or characteristic, such as corners or blobs. Texture analysis: This technique involves quantifying the spatial arrangement or distribution of the pixel values in the image, in order to capture the texture or pattern of the image. Color analysis: This technique involves quantifying the colors in the image, in order to capture the distribution or intensity of the colors. Feature extraction is a critical step in many image processing pipelines, as it allows the image to be analyzed or classified based on its relevant features rather than raw pixel values.

Canny edge detection is a technique for feature extraction in image processing that is used to identify the boundaries or edges in an image. It was developed by John Canny in 1986 and has since become a widely used method for edge detection. Figure 3.4.1 shows the result after applying canny feature extractor.



**Figure 3.4.1** Image after Canny feature extractor.

Moreover, many other features have been extracted using various techniques. Around 37 features have extracted using various methods such as Gabor, Roberts, Sobel, Prewitt. After extracting 37 features a dataframe has been generated and saved as csv file to be used for applying various machine learning models.

A brief description of the above-mentioned features extractors have been given below. Gabor filters are a type of linear filter that is used for texture analysis and edge detection in image processing. They are named after Dennis Gabor, who invented them in the 1940s. Gabor filters are used to detect features in images by analyzing the frequency and orientation content of the image. They are particularly useful for extracting texture information from images, as well as for identifying edges and lines in an image. Gabor filters can be used in a variety of applications, including object recognition, machine learning, and pattern classification.

The Sobel operator is a mathematical method for detecting edges in digital images. It is based on the concept of calculating the gradient of the image intensity at each point, and then finding the direction of the highest rate of change. The Roberts cross operator is a simple edge detection algorithm that is used to identify the edges in an image. It works by calculating the gradient of the image intensity at each point using a 2x2 kernel. The Prewitt operator is a edge detection algorithm that is used to identify the edges in an image. It works by calculating the gradient of the image intensity at each point using a 3x3 kernel.

# **CHAPTER 5: SIMILAR APPLICATION COMPARISON TABLE**

| Researchers                          | Year | Segmentation<br>Techniques  | Feature<br>Extraction<br>Techniques | Classifier   | Dataset              | Disease   | Accuracy                    |
|--------------------------------------|------|---|-------------------------------------|--|----------------------|---|-----------------------------|
| Rahman et al.                        | 2021 | Otsu method   | GLCM                                | SVM  | 400 images.          | Healthy and three diseases  | 100%,<br>95%,<br>90%, 85%   |
| Sarojadevi<br>and<br>Nagamani        | 2022 | Image<br>Scaling, Color<br>Thresholding,<br>Flood Filling<br>Approaches | GLTP,<br>Zernike<br>Moments         | R-CNN  | 735<br>images        | One healthy and six different diseases                            | 96.735%                     |
| Chowdhury et al.                     | 2021 | -"  | -                                   | ResNet18,<br>MobileNet,<br>DenseNet201,<br>InceptionV3 | 18,162<br>images     | One healthy, and nine different diseases                          | 99.2%,<br>97.99%,<br>98.05% |
| Gadade and<br>Kirange                | 2020 | -   | GLCM,<br>Gabor,<br>SURF             | SVM, KNN,<br>Naïve Bayes,<br>Decision Tree             | 500 images           | One<br>healthy<br>and<br>seven<br>various<br>kinds of<br>diseases | 73.39%                      |
| Dr. Sreelatha et al.                 | 2021 | -   | GLCM                                | LSTM, DNN,<br>ANN                                      | 1500<br>images       | -   | 98.81%                      |
| Trivedi et al.                       | 2021 | -   | -                                   | CNN  | 3000 images          | One<br>healthy<br>and nine<br>various<br>diseases                 | 98.49%                      |
| Shoaib et al.                        | 2022 | U-Net<br>architecture,<br>K-Fold cross<br>validation                    | -                                   | CNN, U-Net,<br>Modified U-<br>Net                      | 18,161<br>images     | One healthy and nine different kinds of diseases                  | 98.66%,<br>99.12%           |
| Khasawnch,<br>Faouri, and<br>Fraiwan | 2022 | -   | -                                   | CNN, Deep<br>Learning<br>Networks                      | 18,160<br>images     | One<br>healthy<br>and nine<br>various<br>diseases                 | 99.4%                       |
| Arfan Shah                           | 2023 | Otsu Method   | GLCM,<br>Gabor                      | Random<br>Forest, KNN,<br>Decision Tree<br>and more    | More<br>than<br>2000 | One healthy and nine diseases                                     | >99%                        |

No previous research is found which is specifically focusing on tomato leaf disease detection in the mountainous regions of Gilgit-Baltistan, Kyrgyzstan, and Tajikistan. The pathogens which cause diseases in tomatoes is usually favored by cool and wet weather. In mountainous areas the weather is usually cool which encourages tomato disease (Meadows and Quesada-Ocampo, 2019). Hopefully this will be the first research which is focusing on tomato leaf disease detection in the mentioned areas and that is the novelty of this research. In addition, previous studies either used one, two, or three Machine Learning algorithms for classification purpose, but this research aims to apply more than five algorithms and compare their results and chose the one which performs the best. This study aims to follow specific stages, such as pre-processing, image segmentation, feature extraction etc. on the other hand, all these stages are not used at once in the previous studies. Furthermore, previous research uses one or no technique for these stages, but this study aims to use multiple techniques for each stage. Moreover, there is not any mobile app or web facility for common farmers in Gilgit-Baltistan, Tajikistan, and Kyrgyzstan which can help to identify the disease and suggest a remedy for the disease. In the future, hopefully this research will help the farmers in the mountainous regions to enhance their fruits and vegetables production.

- 1. Imam, S. (March 09, 2022). Friday Times. Malnutrition in Pakistan-Searching for Solutions. Retrieved from https://www.thefridaytimes.com/2022/03/09/malnutrition-in-pakistan-searching-for-solutions/
- 2. Kopytin, Y. (June 25, 2022). 24KG. 12 percent of children in Kyrgyzstan are stunted due to malnutrition. Retrieved from https://24.kg/english/238122\_12\_percent\_of\_children\_in\_Kyrgyzstan\_are\_st unted due to malnutrition/
- 3. Ruziev, M. (2019). UNICEF Tajikistan. Nutrition. Retrieved from https://www.unicef.org/tajikistan/nutrition
- 4. Rahman, S. U., Alam, F., Ahmad, N., & Arshad, S. (2022). Image processing-based system for the detection, identification and treatment of tomato leaf diseases. Multimedia Tools and Applications, 1-15.
- 5. Nagamani, H. S., & Sarojadevi, H. (2022). Tomato Leaf Disease Detection using Deep Learning Techniques. International Journal of Advanced Computer Science and Applications, 13(1).
- 6. Chowdhury, M. E., Rahman, T., Khandakar, A., Ibtehaz, N., Khan, A. U., Khan, M. S., ... & Ali, S. H. M. (2021). Tomato Leaf Diseases Detection Using Deep Learning Technique. Technology in Agriculture, 453.
- 7. Gadade, H. D., & Kirange, D. D. (2020). Machine Learning Approach towards
  Tomato Leaf Disease Classification. International Journal of Advanced
  Trends in Computer Science and Engineering, 9(1), 490-495.
- 8. Sreelatha, P., Udayakumar, M. S., Karthick, S., Ch, S. C., Kavya, K. C. S., & Madiajagan, M. (2021). Managing The Tomato Leaf Disease Detection Accuracy Using Computer Vision Based Deep Neural Network. Journal of Contemporary Issues in Business and Government Vol, 27(1).
- Trivedi, N. K., Gautam, V., Anand, A., Aljahdali, H. M., Villar, S. G., Anand,
   D., ... & Kadry, S. (2021). Early detection and classification of tomato leaf
   disease using high-performance deep neural network. Sensors, 21(23), 7987.
- Shoaib, M., Hussain, T., Shah, B., Ullah, I., Shah, S. M., Ali, F., & Park, S. H. (2022). Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease. Frontiers in Plant Science, 13, 1031748.
- Khasawneh, N., Faouri, E., & Fraiwan, M. (2022). Automatic Detection of Tomato Diseases Using Deep Transfer Learning. Applied Sciences, 12(17), 8467.

- 12. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in plant science, 7, 1419.
- Geetha, G., Samundeswari, S., Saranya, G., Meenakshi, K., & Nithya, M. (2020, December). Plant leaf disease classification and detection system using machine learning. In Journal of Physics: Conference Series (Vol. 1712, No. 1, p. 012012). IOP Publishing.
- Eunice, J., Popescu, D. E., Chowdary, M. K., & Hemanth, J. (2022). Deep Learning-Based Leaf Disease Detection in Crops Using Images for Agricultural Applications. Agronomy, 12(10), 2395.
- Kulkarni, P., Karwande, A., Kolhe, T., Kamble, S., Joshi, A., & Wyawahare,
   M. (2021). Plant disease detection using image processing and machine learning. arXiv preprint arXiv:2106.10698.
- 16. Meadows, I., Quesada-Ocampo, L. (January 17, 2019). NC State Extension. Tomato Late Blight: Vegetable Pathology Factsheets. Retrieved from https://content.ces.ncsu.edu/tomato-late-blight.