UNIVERSITY OF CENTRAL ASIA

School of Art and Sciences
Department of Computer Science
B.Sc. in Computer Science
COMP 4811: Final Year Project-1

Proposal Form

Title:	Tomato Leaf Disease Detection Based on Image Processing and Machine Learning Techniques	
Project author:	Arfan Shah	
Supervisor:	Muhammad Fayaz	
Co-supervisor (if applicable):		
Main subject Area(s):	Machine Learning, Deep Learning, Image Processing, and Artificial Intelligence	
Keywords:	Tomato Leaf Disease, Plants Leaf Disease, Machine Learning, Image Processing, Image segmentation, Feature extraction	
Project type:	Research Project	
Short project description:	Leaf disease detection in tomato plants using various Machine Learning, Deep Learning, and Image Processing techniques. This research aims to build a simple interface through people can take a picture of image and the system will detect the disease and the system will also notify the person with the possible remedy.	
Project Aim and Objective(s):	Aim: To help farmers in Mountainous regions of Pakistan and Central Asian countries to increase their fruits and vegetables production especially tomatoes production. Objectives: 1. Produce a research paper. 2. Help farmers in mountainous regions to enhance tomato and other fruits and vegetables production. 3. Apply various Image Processing and Machine Learning techniques to detect and classify leaf disease in tomatoes. 4. Making simple interface so that common farmers can use it. 5. Local dataset should be used for testing. 6. The model must detect leaf disease. 7. The model should also suggest a remedy for the identified disease. 8. If time permits the research aims to detect and classify disease for other fruits and vegetables.	

Equipment and critical resources required:	1. Good RAM 2. A good processor 3. Enough SSD 4. External Hard Disk 5. GPU 6. Dataset for training 7. Local dataset
Recommended pre-requisites / Knowledge required and Supporting 3rd Year Study units:	1. Programming in Python 2. Machine Learning 3. Artificial Intelligence 4. Data Science 5. Statistics 6. Software Engineering
Foreseeable Ethical issues and how these will be tackled (if applicable):	A good laptop having enough RAM and SSD is required to train the model properly. For that laptop needs to be upgraded. External disk is required to store data externally in case of computer crash. Moreover, professor may be annoyed due to any reasons, which can be tackled by having a friendly relation with professor. Meeting on regular basis can avoid such misconceptions.
Copyright note:	By signing the Student Declaration Form, it is agreed that all result, design or patent from the student project is under the Department of Computer Science, University of Central Asia.
Expected outcomes:	1. Research Paper 2. Report 3. Model capable to detect disease in tomato plants 4. Simple interface usable in mobile or web
Expected deliverables:	1. Project report: (original + pdf) soft copy + hard copy 2. Presentation (original + pdf) 3. GitHub repository downloaded zip file 4. Installation file or link to front-end and back-end 5. Supervisor review 6. External expert review
Methods:	Waterfall model will be used for this project as most of the things are pre-defined. Iterative model can also used as some of the things need to be tested periodically with the passage of time.
GitHub repository link:	https://github.com/arfan64shah/Arfan_FinalYearProject.git
Programming language(s):	Python
Framework (if applicable):	NumPy, Pandas, and Scikit-learn, Matplotlib, Seaborn, Ggplot, Plotly, React, Django, Tailwind CSS, Bootstrap, and many more.
External libraries:	OpenCV, Scikit-Image, Tkinter, SciPy, and many more.

Chair department Dr. Ayman Aljarbouh signature:

Department committee decision:

Approve / Reject

Contents

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.

Key Words: Tomato leaf disease, Leaf disease, Image segmentation, Feature extraction, Image Processing

Project Description including References and keywords (~300 words):

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.

Literature Survey:

(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%.

Similar applications comparison table:

Researchers	Year	Segmentation Techniques	Feature Extraction Techniques	Classifier	Dataset	Disease	Accuracy
Rahman et al.	2021	Otsu method	GLCM	SVM	400 images. 120 for training and 280 for testing	Healthy, Early blight, Late blight, Septoria leaf spot	100%, 95%, 90%, 85%
Sarojadevi and Nagamani	2022	Image Scaling, Color Thresholding, Flood Filling Approaches	GLTP, Zernike Moments	R-CNN	735 images	One healthy and six different diseases	96.735%
Chowdhury et al.	2021	-	-	ResNet18, MobileNet, DenseNet201, InceptionV3	18,162 images	One healthy, and nine different diseases	99.2%, 97.99%, 98.05%
Gadade and Kirange	2020	-	GLCM, Gabor, SURF	SVM, KNN, Naïve Bayes, Decision Tree	500 images	One healthy and seven various kinds of diseases	73.39%
Dr. Sreelatha et al.	2021	-	GLCM	LSTM, DNN, ANN	1500 images	-	98.81%
Trivedi et al.	2021	-	-	CNN	3000 images	One healthy and nine various diseases	98.49%
Shoaib et al.	2022	U-Net architecture, K-Fold cross validation	-	CNN, U-Net, Modified U-Net	18,161 images	One healthy and nine different kinds of diseases	98.66%, 99.12%
Khasawnch, Faouri, and Fraiwan	2022	-	-	CNN, Deep Learning Networks	18,160 images	One healthy and nine various diseases	99.4%

Technical specification of the project:

One of the top priorities of this project is to produce a research paper. This project aims to apply various Machine Learning and Image Processing techniques to identify leaf disease in plants specifically leaf diseases in tomato plants. Dataset for this project is taken from various sources. One dataset which includes pictures of various plants' leaves is taken from Kaggle. Another dataset is taken from research done at University of Malakand, Pakistan. This project will use the mentioned datasets for training the model. The model will be trained using the available datasets and then the trained model will be used to test for the data which is collected from Gilgit-Baltistan and Kyrgyzstan. This project also plans to use the local datasets for training purposes.

Firstly, the collected datasets will go through pre-processing stage where different techniques will be applied to preprocess and clean data. For instance, image cropping will be used to crop the desired part from leaves pictures. Image resizing will be used to reduce size of images, so that it will be easier to train and test model. Image enhancement technique will be used to enhance the quality of images. Good quality images will be helpful while doing feature extraction. Secondly, the preprocessed dataset will be used for image segmentation. Image segmentation is crucial as it helps to find out the important parts from an image. Images are divided into sub-groups called segments which are used to collect further details. Image segmentation algorithms such as Threshold Based Segmentation, Edge Based Segmentation, Region Based Segmentation, Cluster Based Segmentation, and many others will be used for image segmentation. Thirdly, features will be extracted from the segmented dataset. Statistical features such as contrast, variance, root mean square (RMS), smoothness, kurtosis, interdepartmental major (IDM), comlation, energy, homogeneity, mean, standard deviation, entropy, skewness, and correlation will be extracted. Feature extraction algorithms will be used to extract statistical features from images. Gray Level Co-occurrence Matrix (GLCM) will be the primary algorithm to extract statistical features. In addition, other algorithms such as Principal Component Analysis (PCA), Independent Component Analysis (ICA) and many others will also be used to extract features from dataset. The dataset achieved from the previous processes will be categorized into labels. The labels include healthy, early blight, late blight, and Septoria leaf spots. First model will be trained and tested on the labelled data. Later the model will be tested on random datasets collected from mountainous regions of Pakistan, Kyrgyzstan, and Tajikistan. Various Machine Learning classification algorithms such as Support Vector Machine SVM, K-Nearest Neighbor (KNN), Random Forest, Decision Tree, Logistic Regression, Artificial Neural Network (ANN), Naïve Bayes, Stochastic Gradient Descent algorithm, and Kernel Approximation algorithm will be used to classify leaves into the mentioned categories. After disease identification the farmer or user will also be notified with remedy for the disease. Moreover, a simple interface will be developed so that common farmers could use it.

Functional Requirements:

- Research paper
- Simple and friendly user interface
- Takes image in any format
- Pre-process the image
- Image segmentation occurs
- Feature extraction
- Training and Testing
- Classification
- Detect leaf disease
- Recommends required remedy
- Notification to user
- Functions within few minutes

Non-Functional Requirements:

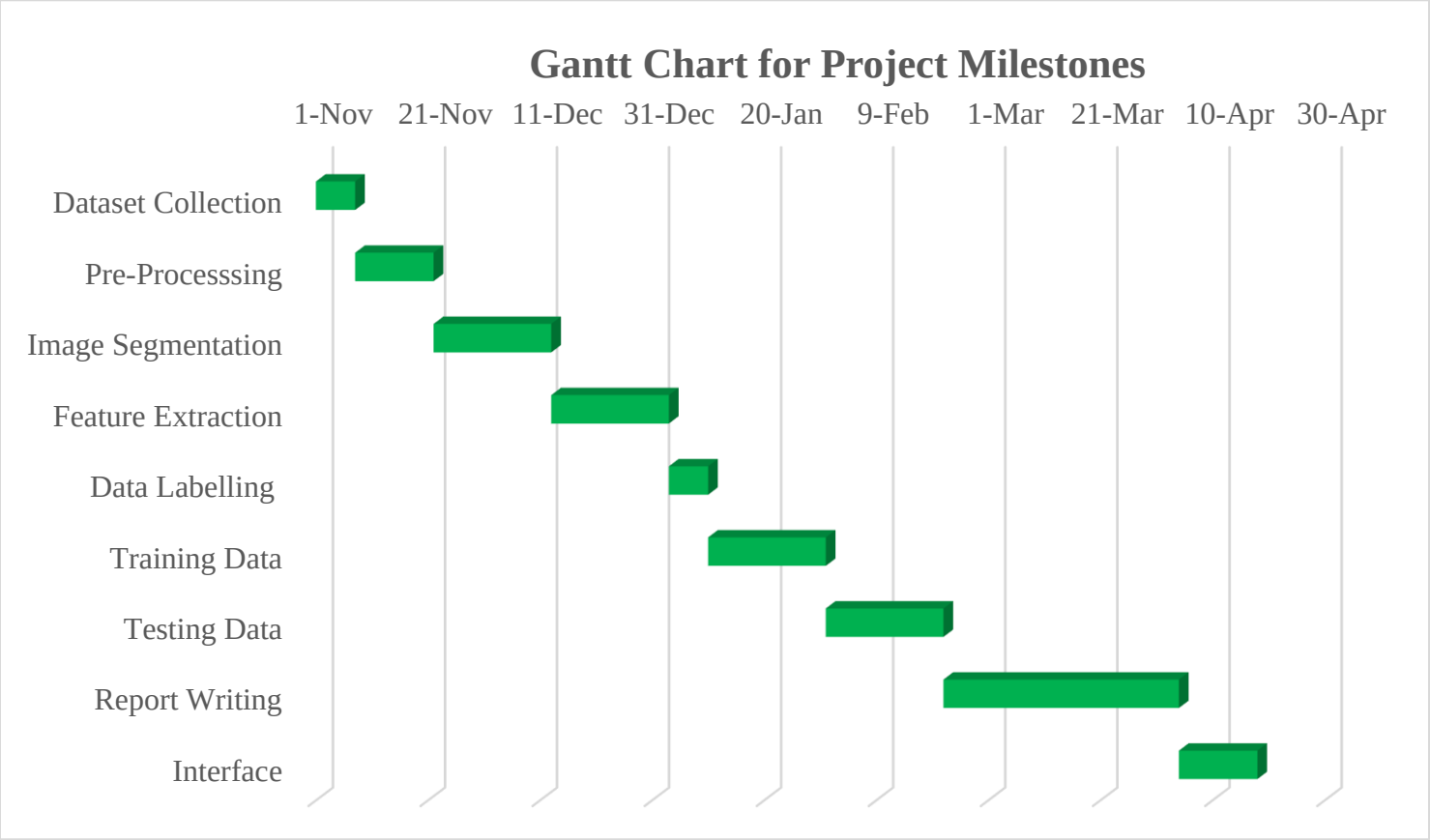
- System will function all time
- Application will take few minutes from capturing image to recommend a remedy
- Easy to use
- Freely available to everyone

Novelty of this research:

No previous research is found which is specifically focusing on tomato leaf disease detection in the mountainous regions of Gilgit-Baltistan, Kyrgyzstan, and Tajikistan. 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.

Project plan and schedule:

Task	Duration	Number of Weeks
Dataset Collection	01 Nov-08 Nov 2022	1 Week
Pre-Processing	08 Nov-22 Nov 2022	2 Weeks
Image Segmentation	22 Nov-13 Dec 2022	3 Weeks
Feature Extraction	13 Dec 2022-03 Jan 2023	3 Weeks
Data Labelling	03 Jan-10 Jan 2023	1 Week
Training Data	10 Jan-31 Jan 2023	3 Weeks
Testing Data	31 Jan-21 Feb 2023	3 Weeks
Report Writing	21 Feb-04 April 2023	6 Weeks
Interface	04 April-18 April 2023	2 Weeks



Risk management plan:

Risks are part of life and when a person is careless then the risks might turn into a huge loss. While doing projects, risks are highly likely to happen, but careful and mindful actions can reduce the intensity of risks. Generally, there are two types of risks: technical and non-technical risks.

Technical Risks:

Technical risks are the risks which can be occur due to the hardware or software which are being used in the project. Laptop can be crashed while doing the project. This can result to loss of the data and other important things. This risk can be handled by constantly pushing/uploading the work on GitHub or any other online server. This risk can also be handled by saving the data in an external hard disk. Moreover, the algorithms used in the project are complex. If the laptop or PC has less RAM and ROM, the training and testing process might not be efficient. To tackle this risk the laptop or personal computer must be upgraded and should have GPU to process swiftly. In other words, laptop or PC should have enough space and random-access memory (RAM) to function properly. Furthermore, processing images need a good processor.

Non-Technical Risks:

Non-technical risks are the risks which can hinder the project development other than software or hardware. For instance, supervisor can leave the university in the middle of the project. This risk can be managed by choosing a co-supervisor. If the supervisor does not continue due to any reason co-supervisor can help with the project. Moreover, supervisor can be annoyed due to any reason, like not visiting him/her regularly or not asking for help etc. This risk can be handled by visiting the supervisor and co-supervisor on daily or weekly basis. Continuous hard-work and showing progress to the supervisor can minimize this risk. Keeping a professional and friendly relation with my supervisor is must to avoid any misunderstandings. Furthermore, when I will submit my project the external committee and instructors can think that the project is not done by me. To avoid this misconception, I must push everything to GitHub on daily or weekly basis. This will show that the work belongs to me.

Bibliography:

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_stunted_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).
9. 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.
10. 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.
11. 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.
13. 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.
14. 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.
15. 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*.