How to Cite:
Raigonda, M. R., Terdal, S. P., & Raigond, B. (2022). Detection of the viral disease on the potato foliar and tubers using a machine learning approach. *International Journal of Health Sciences*, *6*(S4), 9336–9354. https://doi.org/10.53730/ijhs.v6nS4.10740

Detection of the viral disease on the potato foliar and tubers using a machine learning approach

Megha Rani Raigonda
Research Scholar, Department of Computer Science and Engineering, Poojya Doddappa Appa College of Engineering, Kalaburagi, Karnataka
Corresponding author email: megharaigond@gmail.com

Sujatha P Terdal
Professor, Department of Computer Science and Engineering, Poojya Doddappa Appa College of Engineering, Kalaburagi, Karnataka

Baswaraj Raigond
Senior Scientist, Department of Virology, ICAR- Central Potato Research Institute, Shimla, Himachal Pradesh

***Abstract*---**Agriculture is the backbone of the country as it feeds the human race. India's top third crop cultivation is the potato. The potatoes are propagated by vegetative methods where several pathogens tend to assemble over succeeding generations causing a reduction in the crop yield and quality of the crop. The yield of the potato production is most likely affected by the virus compared to the bacteria and fungus and there is a need to develop a model which diagnoses the disease at the initial stage. Long ago scientists had found the disease caused to the potato tuber, and leaf by the virus by laboratory methods Reverse transcription-polymerase chain reaction (RT-PCR), enzyme-linked immunosorbent assay (ELISA), etc., but it needs large infrastructure, labor management, and more time to diagnose the disease. This leads to intensive research in the field of virology. The potato-growing countries have suffered due to the disease-causing agents from the viruses which decreased the crop yield. To address these problems had proposed a framework using the machine learning such as Support Vector Machine (SVM), Logistic Regression (LR), KNN, and Decision Tree (DT) to find the diseases on leaves like leaf curl, Mosaic, Potato Leaf Roll Virus (PLRV) and tuber-like Potato spindle tuber viroid (PSTVD), Potato virus Y(PVY) tuber cracking. The DSLR camera is used to capture the images of the diseased leaves and tubers and prepared the own dataset for tuber with 717 high dimensional images and 1486

Manuscript submitted: 27 April 2022, Manuscript revised: 18 June 2022, Accepted for publication: 9 July 2022

images of leaf these images cover all the diseases stated above. Out of the above classifiers, the decision tree had achieved the highest accuracy of 81.05% in detecting the viral disease. Once the disease is identified the model suggests the management practices to control the disease and improve the crop yield.

***Keywords*---**potato viral disease, leaf curl, PSTVD, PVY, mosaic leaf, machine learning.

Introduction

Rice and wheat are the important crops of the country as it is the major consuming crops. Next to these crops, the potato (*Solanum tuberosum*) is the third-largest consuming food. The growth of the potato has considerably increased as compared to ten years ago due to an increase in the demand for the countries located in tropical and sub-tropical regions [1]. The energy generated per unit of potato is also good enough to maintain the appetite and generates the revenue as it is the commercial plants [21]. Among pathogens, Viral diseases are of serious concern in the production system as they may cause degeneration of seed potatoes and may decline the yield of the crop. The spreading of infection rate from the virus increases with an increase in temperature. Hence there is a need for an automatic system that can detect the virus disease, as the virus can able to destroy fifty percent of the total yield [2]. Few viruses are very harmful that cause severe yield and economic loss all over the world such as PVY, Potato Leaf Roll Virus (PLRV) which may cause up to 70-80% of yield loss. The potato yield will be good only in the regions where temperatures are less. In India, "The viruses like *Potato virus Y* (PVY), *Potato leaf roll virus* (PLRV), *Tomato leaf curl New Delhi virus* (ToLCNDV) and *Groundnut bud necrosis virus* (GBNV) is a major threat to the production of potatoes" [22].

Evolution of diagnostics in Potato virus

The author in [21] discusses the history of potato viral diseases. Earlier, the most common method to determine the virus activity is through the visual estimation which is well known in the traditional methods but this inspection depends on the years of experience by the expertise or the traditional practices of the farmer. But the major constraint is that the symptoms can be the same or variable due to biotic/abiotic factors. In some viruses PVX, PVA and PVM the symptoms may be the same in certain varieties but up to a certain extent, this can be controlled by the indicator host. But the above technique is not suitable for handling a large number of samples in a short period. There are several bio-medical methods to determine the disease on the foliar, tuber, and stem of the plant which are as follows, in 1970 the serological-based techniques and histochemical tests were standardized for detection of viral diseases. Enzyme-linked immunosorbent assay (ELISA) has come up as the most suitable serological approach and become a standard laboratory-based virus testing method. But due to some drawbacks such as it needs a large workspace, and requires a large sample volume. In the same period, Transmission Electron microscopy (TEM) in life sciences was introduced which also plays a vital role in virus detection but it is very expensive and laboratories cannot confer. Further, the polymerase chain reaction (PCR) and real-time reverse-transcription (RT- PCR) assay has been developed for almost all types of viruses. In PCR assay the DNA can be extracted to test for infected or healthy leaf or tuber [4] and requires a well-equipped laboratory and well-qualified manpower. Later on, nanotechnology has been introduced in the field of diagnostics in agriculture like Lateral Flow Immunoassay (LIFA) for rapid and on-field disease detection. With the advancement in science and technology new molecular techniques were introduced such as isothermal-based amplification technologies Loop-mediated isothermal amplification (LAMP) assay and Recombinase Polymerase Assay to overcome the drawbacks of traditional laboratory methods. The spectrometers can be employed on the foliar of the plants to determine whether it is unhealthy or healthy [3]. But the above methods are tedious, computational costs along with time will be high and it is requiring a more labor intensive and controlled lab structure to carry out the experiment [5].

To reduce the above stated problem the easiest method employs the computer vision systems, artificial intelligence methods, and deep learning methods to detect the viral diseases in the foliar, tuber, and stem of the plants rapidly. By adopting the new technology and continuous monitoring, the diseases can be identified at the initial stage that can be avoided and limit the yield loss to a greater extent. The machine learning algorithms provide higher results compared to the traditional methods [6] because the farmer or expert can validate the disease in the field itself and take the control measures accordingly.

To detect the disease-causing agents on foliar, stem, and tuber many systems had been proposed for several crops. The classifier model is fed a bunch of features by scholars on different types of diseases and different crops [6-8]. In our proposed work the prediction model was in which five potato viral diseases were considered. Above ground the disease on foliar like PLRV, Mosaic and Leaf curling and below ground disease on tuber-like PSTVD, PVY strain tuber cracking. we use machine learning algorithms like SVM, decision tree, random forest, KNN, etc., for accurately classifying the viral disease. The remaining document is structured in the following sections as section II describes the research carried out by the research scholars ranging from 2015 to 2021. Section III describes the proposed methodology and mathematical model involved in designing the classifiers and the last two sections describes the experimental results and conclusion.

Related Work

The survey says that the authors didn’t concentrate on a single crop but carried the research on varieties of crops [9-11]. Most of the researchers had employed a Plant village [12] dataset of potatoes to carry out their research but these datasets were generated by capturing the potato grown in the USA but the size, the shape of foliar, tuber, and stem of potato will vary compared to the crops grown in different regions [13]. The most significant root and tuber crop in the world is potatoes. A sick seed potato may yield around ten potato tubers, and the disease can spread throughout the seed potato production cycle. For the production and distribution of seed potatoes, the Japanese government set up a propagation system. The average precision (AP) for detection was 78.2 percent [14].

Plant infections have arisen as one of the important worldwide threats to food security because of their overwhelming effect on yields and vegetables. There are more than 50 different viruses and one viroid in [15] Potato virus Y(PVY) and PLRV are one of the most damaging viruses worldwide. Potato Spindle tuber disease caused by a viroid, PSTVd likewise weakens tuber quality notwithstanding direct yield loss. The commonly caused disease on the potato plant is as shown in the below table [16]

Table 1: Virus disease causing agent

Disease name	Symptoms	Causing agent
Leaf Curl	The leaf of the potato will roll	Potato Leafroll virus
Mosaic	Mild Mosaic	Potato virus X
Mosaic	Yellow color is spotted on the surface of the leaf, Necrotic Rings	Potato aucuba Mosaic Virus
Mosaic	Rugose Mosaic	Potato Virus Y
Early mosaic	Like mosaic but mild	Potato Virus A
Potato Virus S	Rugosity with lesser	Potato Virus S
Potato Virus M	Crinkling is going to happen on foliar	Potato Virus M

To identify the early blight fungal disease in potatoes [17] author uses three convolutional neural networks (CNNs),namely GoogleNet, VGGNet, and EfficientNet, which were trained using the PyTorch framework. Results suggested that the performance accuracy of EfficientNet and VGGNet was higher in terms of disease detection at different stages when compared with GoogleNet. GoogleNet was observed to have the lowest inference time compared with EfficientNet and VGGNet. In [18] author integrates image processing and machine learning to diagnose leaf disease in potatoes. The dataset taken was from the publicly available dataset 'plant village' segmentation approach used was thresholding and for feature extraction, gray level Co-occurrence matrix(GLCM) was used where the author extracts 10 features(Color and texture) for every leaf in the dataset. Finally for classification Support vector machine is used and achieved an accuracy of 95%. In [19] author considers fungal diseases like Early blight and Late blighusesnd use pre-trained models like VGG19 for transfer learning to extract the pertinent features from the dataset. Then, with the help of multipleclassifiers results the logistic regression outperformed others by a substantial margin of classification accuracy obtaining 97.8%. The images of the foliar and tuber when the virus affects it is as shown in the below fig,

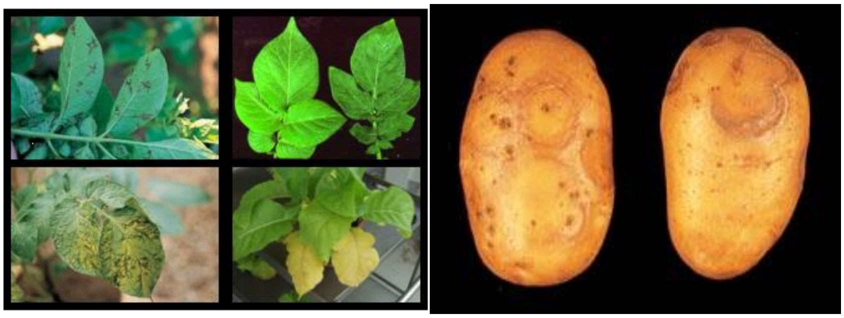


Figure 1: Affected region by Viruses (Potato virus Y)[21]

The seeds for growing the potato are collected from the certified seed centres located in Shimla, the diseased leaves are captured from their field and the tuber images are captured by taking expert advice from the agriculture university in Dharwad. The instance of the images is shown in the below table

Table 2: Instance of the captured images from the field.

Name	Captured image
Healthy leaf	
Healthy potato image	
Leaf curl	
Mosaic	
PLRV leaf	
PSTVD	

PVY tuber cracking	

This paper provides an overview of several forms of plant diseases as well as machine learning classification algorithms for finding diseases in different plant leaves. Plant leaves are classified based on their morphological characteristics and attributes, such as color, intensity, and size [20]. From the above literature, it is clear that more work was done on fungal diseases and very less work was done in the field of viral diseases in potato crops using machine learning anddeepp learning.

Proposed Methodology

The visual inspection to determine the disease needs years of practice and expertise. For the farmer without any sophisticated equipment or hand-held lens, it's hard to detect. To cover all the fields by the experts in the period is almost impossible as potato cultivation span is three months i.e., 90 days starting from seeding to yield generation. Hence an automatic system is essential to detect the disease by observing the image. The job of the farmer is to capture the image and feed it to the model. It needs to provide the decision of which viral disease the crop is affected with and should suggest management practices to avoid the disease. The details of the proposed framework are as follows, the proposed methodology has the two sub-divisions such that to identify the disease in the foliar region of potato and in tuber which has a data collection module, enhancement, feature extraction and ensemble classification of three classifiers (Decision tree, Logistic regression, Support Vector Machine) the below fig show the block diagram.

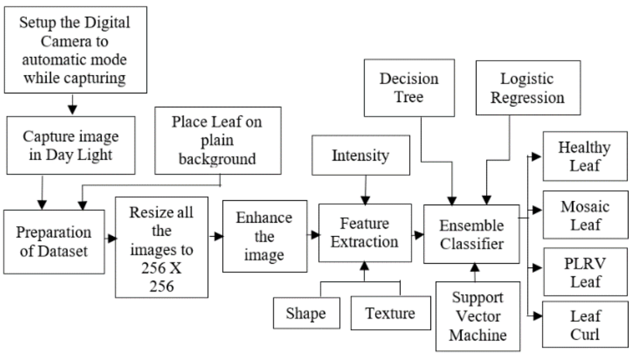


Figure 2: A proposed system for foliar disease detection

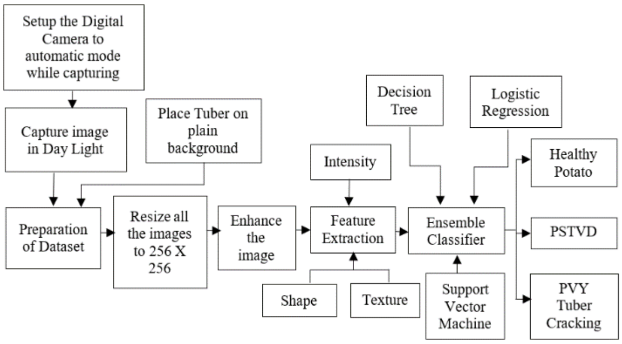


Figure 3: A proposed Tuber disease detection module.

Step 1: Dataset preparation

The Images of potato Foliar affected with the viral disease were provided by a scientist/ advisor from the CPRI Shimla and with the assistance of experts from the university of agriculture sciences, Dharwad diseased leaves and tubers were collected by visual investigation from the field. Once the image data is collected, those had been placed on a plain background and during the daylight, the images were captured. To identify the disease on the foliar and tuber the above dataset is employed in this proposed work. The employed background for capturing the image is as shown in the below figure

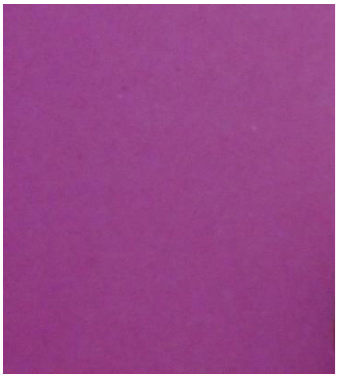


Figure 4: Employed background for capturing the image

To capture the images from the field the Nikon d90 is employed to capture the input image by placing the camera at 90 degrees facing to object of interest. All the images were captured in the daylight with a 35 mm prime lens and the ten to twenty mm wide zoom angle is set to capture the dataset with the dimension of 2400 x 3200, with the ISO speed of ISO-500, exposure time of 1/160 seconds. The below figure shows the instance images of the affected regions of the foliar, tuber of the potato

Figure 5: Instance of the captured images of the tuber and foliar of the potato.

The below table summarizes the total number of images captured from the cultivated field

Table 3:Summary of the collected dataset

Disease Name	Number of images captured
Healthy tuber	44
Healthy foliar	135
Cracking tuber (Type-2)	553
Leaf curl	154
Mosaic leaf	688
PLRV leaf	530
PSTVD tuber	85
PVY cracking	35

The total number of images captured was 2224 out of which leaf images are 1507 and the rest are tuber images of size 9GB which covers the virus disease. As the literature survey on the standard dataset for virus detection is very less hence the created dataset can be useful for future researchers who like to study the nature of potatoes grown in the south region of Dharwad and the Northern region of Shimla.

Step 2: Resizing the image

The original image's dimensions are 2400x3200 which will take more computation time to finish the training of the model. To reduce this overhead had reduced the dimension of the image from 2400x3200 to 256x256, this will achieve a reduced training time and lesser memory to process the data.

Step 3: Contrast enhancement

To enhance the quality of the image contrast plays an important role. The contrast of the image is enhanced by using the open-cv library. In the library the enhancing factor is provided to increase, decrease or remain the same by changing the enhancing factor, the conditions of the enhancing factor are as shown in the below table,

Table 4: Contrast enhancement parameters

Enhancing factor	Contrast

1	Remains same
Greater than one	The contrast of the image improves
Lesser than one	The contrast of the image decreases.

Less contrast image = enhance factor < 1
Enhance of contrast of image = enhance factor > 1.
Eq. 1

When the enhancing factor goes towards zero the color image turns to a grayscale image. For the feature extraction process, the datatype of the image needs to be in unsigned integer 8 (uint8). The data type of the image is converted to uint8 from double.

Step 4: Feature extraction

The feature extraction is the dominating step in determining the pattern of the disease by referring the several images. By selecting the dominant features, the training time and storage can be reduced because instead of storing all the features only the dominant features are stored. The feature extraction module has the sub-modules namely shape feature, texture, and intensity extraction and the detailed explanation is as follows

Shape features

Steps to extract the features of the shapes are as follows,
Step 1: Convert the color image to a grayscale image
Step 2: Convert the grayscale image to a binary image.
Step 3: apply the edge detection method (canny) to obtain a binary image from step 2.
Step 4: Find the blobs in the image: It is a collection of linked pixels in the binary image that has a similar characteristic. It is computed by using the connected components. To find the linked pixels the 4 or 8 connected components are used. The 4 connected components check the current pixels in two directions whereas 8 connected components check the connected components in three directions. The direction checks are shown in the below figure.

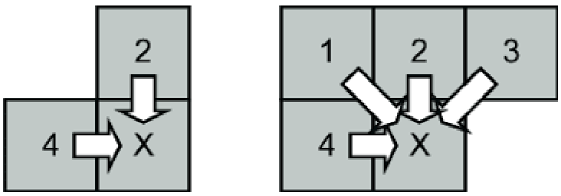


Figure 6: Connected components illustration

An analysis of the shape of object contours is very much useful.
For finding the contours in the binary image the cv. RETR_TREE which returns all the contours as per the size (area)

Step 4: Obtain the moments from the generated contours: Image Moment is a weighted average of image pixel intensities that may be used to determine certain image attributes such as radius, area, and centroid.
The below equations are employed to determine the centroid

$Cx = \frac{M_{10}}{M_{00}}$

$Cy = \frac{M_{01}}{M_{00}}$ Eq. 2

Where M denotes the moment of the image Cx, Cy is the centroid coordinates in x and y-axis is.

Center = [Cx , Cy] Eq. 3

Compute the area and perimeter of every shape and return the values.

Texture features

In the texture feature extraction, the two textures are extracted first is Haralick and the second one is the Gray-Level Co-Occurrence Matrix (GLCM).

Haralick features

It generates the two forms of the features namely global and local features. It is one of the texture characterizing values collected based on the adjacent matrix from the various position of the image p(x,y). It finds the number of times the pixels take the value x next to the pixel value of y. From the model, the first thirteen features are extracted and the rest are ignored as they are unstable. The equation for extracting the haralick is provided below,

Intensity feature

The intensity of the images is characterized by computing the histogram of the image.
Step 1: Convert the image into the grayscale.
Step 2: Compute the histogram of the image
Step 3: Normalize the histogram and convert it to the vector and return it to the main function.

GLCM features

The Gray-level co-occurrence matrix (GLCM), also known as the Gray-level spatial dependency matrix, is a statistical approach to assessing texture that takes into account the spatial connection of pixels. The GLCM functions calculate how frequently pairs of pixels with given values and in a defined spatial relationship appear in an image, create a GLCM, and then extract statistical measures from this matrix to describe the texture of an image. The features extracted from the GLCM are contrast, dissimilarity, homogeneity, and energy. All the features are collected and prepared CSV file which has both the labels and features of the respective images which is fed to the classification module.

Step 5: Classification stage

The last step in the proposed system is the classification phase, where it takes the shape, texture, and intensity features to perform the differentiation between the classes. In the classification phase, there are two stages firstly training where one will allow the model to learn through input images and output labels. For the training thousand, nine hundred images were reserved only for the training phase, and the rest four hundred images were left for the testing purpose. To enhance the accuracy of the detection had considered the three different machine learning algorithms such as SVM, LR, DT, and KNN.

Support Vector Machine (SVM)

The SVM is widely employed for the classification purpose for the linear and non-linear data. The SVM has the two types linear and kernel types. Straight-line hyperplanes are used in linear SVM, while non-linear hyperplanes are used in kernel-based SVM [23].

Model parameters for Linear and Radial Basis Function (RBF) kernels

There are two parameters for the RBF kernels to design with the first one is the 'c' and the second one is the gamma parameter. 'c' value is the trade-off between the right classification and the margin of the decision boundary. If c is small then the margin will be less and otherwise, it will be more. The inverse of the radius of effect of samples picked by the model as support vectors is the gamma parameter. The parameters for the detection of the foliar and tuber disease using RBF[22].

Table 5:Model parameters for RBF kernels

Disease type	Parameter	
	C	Gamma
Foliar disease	1.5	0.1
Tuber disease	1	0.15

Parameters for the linear kernels for both the disease identification is as shown in the below table

Table 6: Linear parameters for SVM

Disease type	Parameters			
	Penalty	Loss	C	Multiclass
Foliar disease	L2	Square hinge	1	One-versus-Rest (Ovr)
Tuber disease	L1	Square hinge	1.5	crammer_singer

The details of the parameters L1, L2 and C are as follows, L1 provides the co-efficient vectors L2 normalization of vectors. Loss uses the square hinge which computes the square of the hinge loss. C is the regularization parameter. The multi-class parameter is ovr means trains one-versus-rest of the class and crammer trains all versus all class. This algorithm solves the multi-class problem as a single optimization problem instead of breaking down the multi-class problem into a series of binary problems.

Linear regression

It is the fundamental linear classification method. It is fast and less complex and provides good results for binary classification.
As it is the linear classification the equation for the linear function is given below,
$F(x) = b_0 + b_1x_1 + \dots + b_nx_n$ Eq. 4
Where b0 to bn are the regression co-efficients.
The weights are computed by using the below equation
$P(x) = 1/1+e^{(-f(x))}$ Eq. 5
Where p(x) is the weights and f(x) is the sigmoid functions for logistic regression.
The parameters for the linear regression are the same as the linear SVM which is formulated in table 3.

Decision tree

The idea behind the decision tree is, to split the records based on the attributes, based on the attribute breakdown of the dataset into subsets. Repeat the process till the minimum samples are obtained. Parameters for the foliar and tuber detection are tabulated in the below table,

Table 7:Parameters for decision tree

Disease type	Parameters		
	Criterion	Splitter	Max features
Foliar disease	Gini	Best	Sqrt
Tuber disease	Entropy	Best	Log2

K-Nearest Neighbourhood (KNN)

The principle behind nearest neighbor methods is to find a predefined number of training samples closest in distance to the new point and predict the label from these. The number of samples can be a user-defined constant (k-nearest neighbor learning), or vary based on the local density of points (radius-based neighbor learning). The distance can, in general, be any metric measure: standard Euclidean distance is the most common choice and the below equation provides the same,

instance $D=\sqrt{(x_2 - x_1)^2 - (y_2 - y)^2}$

Eq.6

Choosing the accurate k value decides the accuracy of the model hence we had taken the k value ranging from 1 to 10 and performed the training and k = 8 gave the highest result for foliar detection and k = 1 for the tuber disease detection.

Result and Discussion

The existing dataset for detecting the potato disease is Plant Village has been captured in the USA region which is not effective in detecting the crop disease for the local regions. Hence there is a need to create the dataset by taking the assistance of the Certified seed breeder ICAR-CPRI Shimla and from the potato fields of the university of agriculture sciences, Dharwad. The diseased foliar and tuber images are captured using the Nikon D90 DSLR camera and prepared a dataset of 2300 images which includes the viral disease caused by foliar and tuber. The paper proposes a framework to predict the viral diseases for the local dataset which is useful for the local farmers to identify and detect the disease. The accuracy of the model achieved is 81.005% which is higher than the 2021 [14] model. The management practices for the diseases are as follows, For the mosaic disease, there is no control once the disease is present, but there are ways to reduce the risk of potato diseases. Use seed certified free from viruses or with a very low incidence of infected tubers. Disinfect all field equipment by dipping in or washing them either with trisodium phosphate or calcium hypo chloride (1%) solution [24] [25].

For the Potato Leafroll virus (PLRV), Potato tubers can be washed regularly with the water having a temperature of 29-to-55-degree C. The region of infection from the foliar and tuber need to be removed as soon as possible to avoid the spread of the PLRV [24][25]. For Potato Leaf Curl, the seeds which are going to be planted should be checked thoroughly to ensure that they are free from viruses. Placing the traps for every height of 60 cm^2. Once the harvesting every crop there is a need to destroy the residues of the crop. Burn the regions of infection caused by the host plant and pluck the weeds and unwanted plants [24][25]. For Potato Spindle Tuber Viroid Disease, there is no cure for avoiding the spread of PSTVD hence the only way to avoid this disease is by choosing the healthy seeds. The tools which are used to remove the diseased region of the plant then need to be properly treated with solutions.

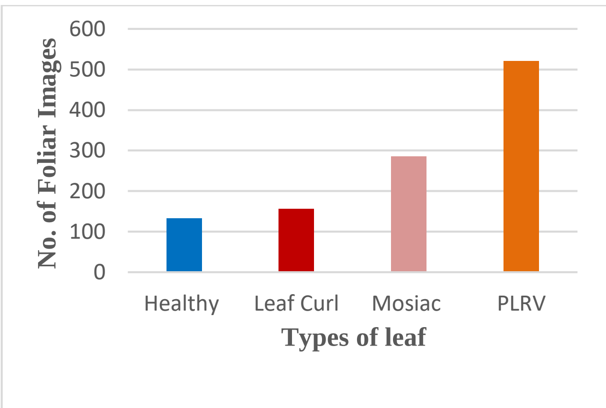
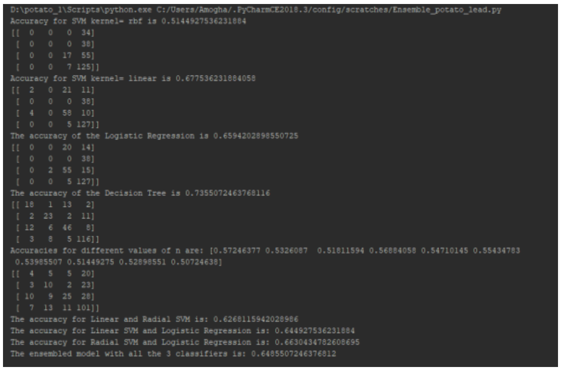


Figure 7: Number of images employed for training and testing

The accuracy of the KNN achieved for all the values of ‘K’ is as shown in the below table

Table 8: KNN accuracy for every value of K

Kth value	Accuracy
1	57.2
2	53.2
3	51.8
4	56.8
5	54.7
6	55.8
7	53.6
8	51.4
9	52.89
10	50.72



```
D:\potato_1\Scripts\python.exe C:/Users/Amogha/.PyCharmCE2018.3/config/scratches/Ensemble_potato_lead.py
Accuracy for SVM kernel= rbf is 0.5144927536231884
[[  0   0   0  34]
 [  0   0   0  38]
 [  0   0  17  55]
 [  0   0   7 125]]
Accuracy for SVM kernel= linear is 0.677536231884058
[[  2   0  21  11]
 [  0   0   0  38]
 [  4   0  58  10]
 [  0   0   5 127]]
The accuracy of the Logistic Regression is 0.6594202898550725
[[  0   0  20  14]
 [  0   0   0  38]
 [  0   2  55  15]
 [  0   0   5 127]]
The accuracy of the Decision Tree is 0.7355072463768116
[[ 18   1  13   2]
 [  2  23   2  11]
 [ 12   6  46   8]
 [  3   8   5 116]]
Accuracies for different values of n are: [0.57246377 0.5326087  0.51811594 0.56884058 0.54710145 0.55434783
 0.53985507 0.51449275 0.52898551 0.50724638]
[[  4   5   5  20]
 [  3  10   2  23]
 [ 10   9  25  28]
 [  7  13  11 101]]
The accuracy for Linear and Radial SVM is: 0.6268115942028986
The accuracy for Linear SVM and Logistic Regression is: 0.644927536231884
The accuracy for Radial SVM and Logistic Regression is: 0.6630434782608695
The ensembled model with all the 3 classifiers is: 0.6485507246376812
```

Figure 8: Accuracy of the model

Table 9: Accuracy comparison for foliar disease detection

Method	Accuracy (%)
Linear SVM	51.44
RBF kernel SVM	67.75
Logistic Regression	65.94
Decision Tree	74.64
KNN	57.24
Linear and RBF SVM	64.13
Linear SVM and Logistic Regression	64.13
Rbf SVM and Logistic Regression	66.30
Ensemble all classifiers	64.85

A total of 1300 images of healthy, leaf curl, mosaic leaf, and PLRV leaf are considered. The decision tree had achieved the highest accuracy of 74.64% in detecting the disease in the foliar

Tuber detection

The potato tuber diseased (Potato Spindle Tuber Viroid, PVY Cracking) images and healthy tuber images are taken to train the model and get the accuracy of disease detection. The results of screen accuracy and confusion matrix are shown in fig.10 and a comparison table of different methods is also shown in table 11. The number of images taken for the implementation is shown below in fig. 9.

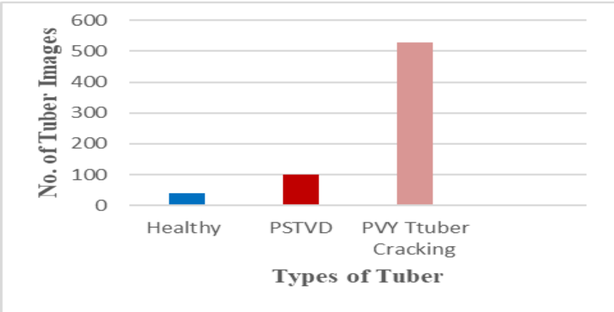


Figure 9: Number of tuber images taken for training and testing.

The accuracy of the KNN for every value of K in the detection of the tuber is shown in the below table.

Figure 10: Accuracy of KNN for different K values

Kth value	Accuracy
1	74.2
2	77
3	72
4	77.5
5	77.1
6	78.2
7	78
8	78.9
9	75.5
10	77

The confusion matrix and accuracy of the models are shown in the below table and fig.

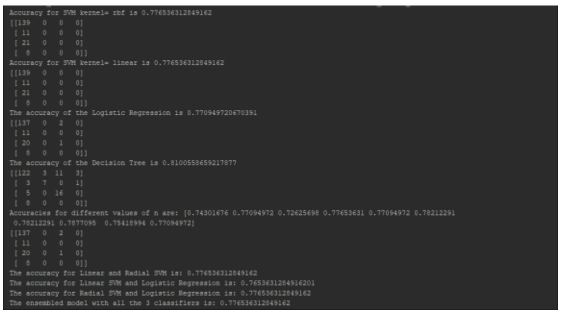


Figure 11: Accuracy and confusion matrix for the model

Table 10: Accuracy Comparison for tuber disease detection

Model employed	Accuracy (%)
SVM kernel RBF	77.65
SVM linear	77.65
Logistic Regression	77.09
Decision Tree	81.005
KNN (n=8)	78.77
Linear and Rbf SVM	77.65
Linear SVM and Logistic Regression	76.36
Rbf and Logistic Regression	77.65
Ensemble classifiers	77.65

The maximum accuracy achieved in the detection of the tuber is 81.05% by the decision tree. The disease management for the diseases of tuber and foliar suggested by the proposed model is as shown in the below fig.

Figure 12: Disease management for the foliar and tuber.

The disease prediction of the disease infected by the virus to the foliar and tuber by the proposed model is as shown in the below fig.

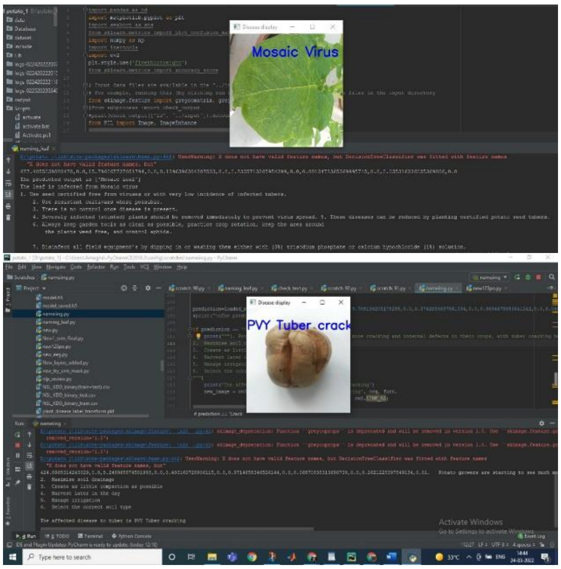


Figure 13:Predicted diseases of the foliar and tuber along with the management practices

The comparison for the detection of foliar using the three machine learning approaches is shown in the below figure

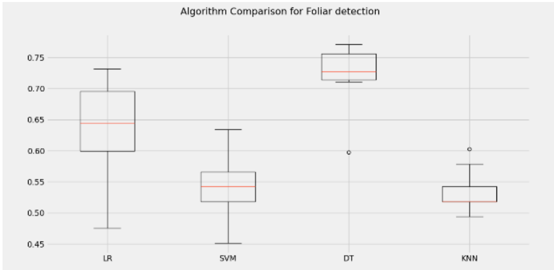


Figure 14: Comparison of accuracy of different models.

Conclusion

Disease prediction and Identification Play a vital role in the field of agriculture. Different Image processing and Machine Learning algorithms are used for early prediction to avoid huge crop and economic loss to the farmers. The dataset is created by taking images of foliar as well as a tuber from ICAR-CPRI Shimla and UAS Dharwad. Three machine learning techniques are applied to the above dataset to detect and identify the viral diseases on foliar and tuber. The result analysis shows that the accuracy of the proposed model is high by the decision tree when compared to other methods. The Image dataset is collected from the Central Potato Research Centre, Shimla, and the University of Agricultural Sciences, Dharwad. The Total images of diseased foliar and tuber collected are two thousand and three hundred. The collected dataset image is pre-processed and enhanced by increasing the contrast of the image and extracting the shape, intensity, and texture features. These features are fed to the three classifiers and the ensemble classifiers. The achieved accuracy in tuber detection is 81.05%. The future scope of the work can be extended by applying deep learning models to enhance the accuracy of the model.

Author Contributions

The paper conceptualization, collected the image data from the potato field, Pre-processing, feature extraction, Classification, methodology, software, validation, formal analysis, investigation, resources, writing—original draft preparation and editing, visualization, have been done by 1st author. The supervision and project administration, have been done by 2nd author

Acknowledgment

I kindly acknowledge Director, ICAR-CPRI Shimla for extending support and data sharing as per the MoU between ICAR-CPRI, Shimla, and PDACE, Kalaburagi. My special thanks to Dr. T R Shashidhar, Professor & Head of AICRP(Potato), Monitoring Agricultural Resources (MARS), University of Agriculture Sciences, Dharwad for assisting me in collecting diseases tubers from the potato field. My thanks to Dr. Venkatesh Kulkarni, Scientist (Plant Pathology), UAS, Dharwad for assisting and guiding me in collecting and capturing diseased leaves in the potato field. My sincere thanks to My thanks to Dr. Sandhya P, Associate Professor, VTU CPGS Mysuru for support in carrying out my research work.

References

[1] Afzaal H, Farooque AA, Schumann AW, Hussain N, McKenzie-Gopsill A, Esau T, Abbas F, Acharya B. Detection of a Potato Disease (Early Blight) Using Artificial Intelligence. Remote Sensing. 2021; 13(3):411. https://doi.org/10.3390/rs13030411.

[2] Azimi, Hamed, Hossein Bonakdari, and Isa Ebtehaj. "Design of radial basis function-based support vector regression in predicting the discharge coefficient of a side weir in a trapezoidal channel." Applied Water Science 9.4 (2019): 1-12.

[3] Baker, N.T.; Capel, P.D. Environmental Factors that Influence the Location of Crop Agriculture in the Conterminous United States; US Department of the Interior, US Geological Survey: Reston, VA, USA, 2011

[4] D. Tiwari, M. Ashish, N. Gangwar, A. Sharma, S. Patel and S. Bhardwaj, "Potato Leaf Diseases Detection Using Deep Learning," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 461-466, doi: 10.1109/ICICCS48265.2020.9121067.

[5] Devaux A, Kromann P, Ortiz O (2014) Potatoes for sustainable global food security. Potato Res 57:185–199.

[6] Geetharamani, G.; Pandian, A. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. Comput. Electr. Eng. 2019, 76, 323–338.

[7] Harahagazwe D, Condor B, Barreda C et al (2018) How big is the potato (Solanum tuberosum L.) yield gap in Sub-Saharan Africa and why? A participatory approach. Open Agriculture 3(1):180–189. https://doi.org/10.1515/opag-2018-0019

[8] Henson, J.M.; French, R. The polymerase chain reaction and plant disease diagnosis. Annu. Rev. Phytopathol. 1993, 31, 81–109. [CrossRef]

[9] http://vegetablemdonline.ppath.cornell.edu/NewsArticles/Potato_Virus.htm

[10] https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

[11] Hughes, D.; Salathé, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv 2015, arXiv:1511.08060.

[12] Kamal, K.; Yin, Z.; Wu, M.; Wu, Z. Depthwise separable convolution architectures for plant disease classification. Comput. Electron. Agric. 2019, 165, 104948.

[13] Khamparia, A.; Saini, G.; Gupta, D.; Khanna, A.; Tiwari, S.; de Albuquerque, V.H.C. Seasonal Crops Disease Prediction and Classification Using Deep Convolutional Encoder Network. Circuits Syst. Signal Process. 2019, 39, 818–836. [CrossRef]

[14] Koo, C.; Malapi-Wight, M.; Kim, H.S.; Cifci, O.S.; Vaughn-Diaz, V.L.; Ma, B.; Kim, S.; Abdel-Raziq, H.; Ong, K.; Jo, Y.K.; et al. Development of a real-time microchip PCR system for portable plant disease diagnosis. PLoS ONE 2013, 8, e82704. [CrossRef]

[15] Kreuze, J.F., Souza-Dias, J.A.C., Jeevalatha, A., Figueira, A.R., Valkonen, J.P.T., Jones, R.A.C. (2020). Viral Diseases in Potato. In: Campos, H., Ortiz,

O. (eds) The Potato Crop. Springer, Cham. https://doi.org/10.1007/978-3-030-28683-5_11

[16] Kumar, Ravinder & Jeevalatha, Arjunan & Raigond, Baswaraj & Tiwari, Rahul. (2020). Viral and Viroid Diseases of Potato and Their Management.

[17] L. Sherly Puspha Annabel, T. Annapoorani, and P. Deepalakshmi. "Machine Learning for Plant Leaf Disease Detection and Classification–A Review." 2019 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2019.

[18] Lopez, M. M. L., Herrera, J. C. E., Figueroa, Y. G. M., & Sanchez, P. K. M. (2019). Neuroscience role in education. *International Journal of Health & Medical Sciences*, *3*(1), 21-28. https://doi.org/10.31295/ijhms.v3n1.109

[19] LP, Awasthi, and H. N. Verma. "Current status of viral diseases of potato and their eco-friendly management-A critical review."

[20] Lu, J.; Hu, J.; Zhao, G.; Mei, F.; Zhang, C. An in-field automatic wheat disease diagnosis system. Comput. Electron. Agric. 2017, 142, 369–379. [CrossRef]

[21] M. Islam, Anh Dinh, K. Wahid and P. Bhowmik, "Detection of potato diseases using image segmentation and multiclass support vector machine," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 2017, pp. 1-4, DOI: 10.1109/CCECE.2017.7946594.

[22] Oishi, Yu, et al. "Automated abnormal potato plant detection system using deep learning models and portable video cameras." International Journal of Applied Earth Observation and Geoinformation 104 (2021): 102509.

[23] Raigond, B., Verma, G., Kumar, R., Tiwari, R.K. (2022). Serological and Molecular Diagnosis of Potato Viruses: An Overview. In: Chakrabarti, S.K., Sharma, S., Shah, M.A. (eds) Sustainable Management of Potato Pests and Diseases. Springer, Singapore. https://doi.org/10.1007/978-981-16-7695-6_13.

[24] Sasaki, Y.; Okamoto, T.; Imou, K.; Torii, T. Automatic diagnosis of plant disease-Spectral reflectance of healthy and diseased leaves. IFAC Proc. Vol. 1998, 31, 145–150. [CrossRef]

[25] Singh, U.P.; Chouhan, S.S.; Jain, S.; Jain, S. Multilayer Convolution Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease. IEEE Access 2019, 7, 43721–43729. [CrossRef]

[26] Widana, I.K., Sumetri, N.W., Sutapa, I.K., Suryasa, W. (2021). Anthropometric measures for better cardiovascular and musculoskeletal health. *Computer Applications in Engineering Education*, *29*(3), 550–561. https://doi.org/10.1002/cae.22202

[27] Zhang, K.; Xu, Z.; Dong, S.; Cen, C.; Wu, Q. Identification of peach leaf disease infected by Xanthomonas campestris with deep learning. Eng. Agric. Environ. Food 2019, 12, 388–396. [CrossRef]