Cotton Crop Disease Detection Using SWIN Transformers and Attention-based CNN

Galla Yagnesh[1], Mare Jagapathi[2] and Ebenezer Jangam[3]

[1,2,3] Department of Information Technology, Velagapudi Ramakrishna Siddhartha Engineering College, Vijaywada, Andhra Pradesh, India
[1] gallayagnesh@gmail.com
[2] jagapathi4321@gmail.com
[3] jebenezer@vrsiddhartha.ac.in

Abstract. Cotton is a vital cash crop globally, and diseases can cause significant damage, leading to reduced yields and quality. Early detection enables farmers to take necessary measures to curb the spread of diseases, enhancing crop quality and yields. Typically, cotton diseases have been identified by experts via visual inspection, which is costly, time-consuming, and not always precise. However, recent advances in deep learning and computer vision provide new opportunities for automated disease detection. This research paper presents a model for detecting cotton crop diseases using the SWIN transformer architecture and Attention-Based CNN model. The proposed system was trained on the large and small datasets of images of healthy and diseased cotton crops, covering three different diseases named Bacterial Blight, Curl Virus, and Fusarium Wilt. The Swin Transformer architecture was chosen for larger datasets and Attention-based CNN for smaller datasets due to their superior performances compared to other state-of-the-art deep learning models as it achieved better accuracies. The ultimate goal is to provide farmers with an effective tool for early detection and prevention of cotton crop diseases, improving crop yield and reducing economic losses. The success of this project could lead to further advancements in deep learning solutions in agriculture, ultimately enhancing global food security.

Keywords: Disease Detection, Deep Learning, Computer Vision, SWIN Transformer, Attention-Based CNN, Bacterial Blight, Curl Virus, Fusarium Wilt, Global Food Security.

1 Introduction

India, the greatest cotton grower in the world, has long relied on cotton as a major crop. Cotton is a cash crop for millions of farmers and a substantial source of income for small and marginal farmers. The cotton business is also vital to India's textile industry, which is a significant employer in the nation. The exports and foreign exchange profits of India are strongly influenced by the textile sector. The cotton business offers a variety of jobs, from picking to ginning and spinning. Considering cotton's economic importance, controlling crop diseases becomes essential. For farmers and workers, illnesses can result in significant financial losses. If diseases caused by viruses, bacteria, fungus, and pests are not identified and handled in a timely manner, they can spread quickly among cotton crops. It is crucial for farmers to identify cotton infections early in order to take the appropriate steps and reduce crop damage. In the past, diagnosing diseases required expensive and time-consuming visual inspection of each plant by qualified professionals. Since certain illnesses don't show noticeable signs until later stages, by which point they may have spread throughout the crop, this method might not always be reliable.

Automated illness detection systems have been developed as a result of developments in deep learning techniques, particularly in computer vision and image analysis. These systems use image recognition models and machine learning algorithms to swiftly and precisely identify crop diseases based on visual signals [6]. This makes it possible for farmers to take preventive steps to stop the spread of illness and minimize crop loss. Farmers can avoid wasting time and money on human inspections and diagnosis by automating the disease detection process. Frequent agricultural monitoring made easier by automation enables quick response to emerging disease outbreaks and prevents significant crop losses.

Automated disease identification has been made possible by recent developments in deep learning and computer vision. Automated disease identification is now possible because to models like convolutional neural networks (CNNs), recurrent neural networks (RNNs), and the SWIN Transformer architecture that can recognize patterns in pictures of healthy and diseased crops [11]. The 2021-debuted SWIN Transformer design has demonstrated promising performance in image classification tests. It makes use of hierarchical feature representation and multi-scale windows to capture global context data while maintaining high resolution. This makes it suitable for challenges involving image-based categorization that call for both fine-grained and comprehensive feature extraction [9].

2 Literature Survey

There are numerous models that have been put out by researchers that can be used to categorize leaf diseases in cotton crops using different frameworks. This section concentrates on the resources that improved our knowledge of Image Processing and SWIN Transformer Techniques, which aided in the development of the research field. The study articles go into great detail about the various categorization frameworks and algorithms that may be employed. We studied these publications and found that deep learning algorithms outperform machine learning methods. Furthermore, training a model requires more time because to the complexity of the used frameworks. The results show the need for a methodology that is quicker to teach and a framework that is simpler to develop. We found that using the previously suggested techniques requires more time to train a model. We gathered various methods for shortening the model's training time and improving its accuracy.

In [1], the authors introduced SWIN Transformer, a brand-new vision Transformer that serves as a flexible foundation for computer vision. However, due to variations in scale and sharpness, translating ideas from language to perception causes difficulties. The research suggested a hierarchical Transformer that computes representations using shifted windows to solve these disparities. This method improved efficiency by restricting self-attention computation to non-overlapping local windows and allowing for cross-window connection. The computational cost of the hierarchical architecture is linearly correlated with the size of the image and can simulate at different scales.

In [2], the GoogleNet and ResNet50 models are both helpful for categorizing different kinds of leaves, the study's findings show. The accuracy achieved by the GoogleNet CNN model is marginally (by around 2%) lower than that of healthy leaves. The average background class difference for other classes while using GoogleNet is 14%, whereas when using ResNet50, it is just 5%. The ResNet50 model surpasses GoogleNet on average by 1.67%, 2.49%, and 2.40% in terms of overall accuracy, sensitivity, and F-Score. According to these results, ResNet50 is a more reliable model for categorizing plant leaves.

In [3], for the purpose of this work, the researchers developed a model employing meta-deep learning to precisely identify several cotton leaf diseases. They took 2385 photographs of both healthy and diseased cotton leaves in the field to achieve this. To expand the dataset, they used a data augmentation approach. The dataset was then used to train the suggested Meta-Deep Learning leaf disease identification model as well as Custom CNN, VGG16 Transfer Learning, ResNet50, and other models. They improved the accuracy and generalizability of the proposed model using a meta learning technique. Overall, the results of this study demonstrate that Meta-Deep Learning is capable of correctly identifying and classifying cotton leaf diseases.

In [4], Using computer vision, quick crop disease identification might minimize monetary losses brought on by widespread ailments like anthracnose. The well-liked convolutional neural network (CNN) might not be good at spotting minute variations in crop damage, though. PAST-Net (Path Aggregation SWIN Transformer Network), a novel approach, has been suggested as a solution to problem. To extract features from input photos, PAST-Net uses adaptive feature pooling and the SWIN Transformer. It uses the mask branch for lesion segmentation and the box branch for classification and bounding box regression. Large lesions with consistent forms are best detected with this technique.

In [5], Cotton diseases have a negative impact on fibre quality and output. By combining feature extraction, deep learning-based techniques can increase the accuracy of illness identification. This study suggests an automated method for collecting cotton imaging data using a quadruped robot. The advantages of transformers and convolutional neural networks (CNNs) are combined in the ConvNeXt architecture. The multiscale spatial pyramid attention (MSPA) module allows ConvNeXt to concentrate on key feature map areas. On one competition dataset and two cotton datasets, ConvNeXt achieves recognition scores of 97.2%, 99.7%, and 100.0%, respectively, with the MSPA module and with a negligible increase in inference time. This demonstrates the quick detection speed of the suggested model.

In [6], Vision Transformers (ViTs) are useful for a variety of applications because of their outstanding performance in self-supervised learning for global and local representations. We suggest a novel self-supervised learning framework for medical picture interpretation that is motivated by these findings. Our framework consists of (i) a brand-new 3D transformer-based model called SWIN UNETR, with a hierarchical encoder for self-supervised pre-training, and (ii) proxy tasks that have been especially created to teach human anatomy patterns. With the use of 5050 freely available computed tomography (CT) scans of several organs, we were able to successfully pre-train the model. We show the efficacy of our approach by optimizing the pre-trained models on the Beyond the Cranial Vault Segmentation Challenge and the Medical Segmentation Decathlon datasets. On the public test leaderboards for the MSD and BTCV datasets, our model comes in first place.

In [7], The authors offer a self-supervised learning approach based on Vision Transformers termed MoBY. In order to obtain high accuracy on ImageNet-1K linear assessment, MoBY integrates elements of MoCo v2 and BYOL. Using DeiT-S and SWIN-T, respectively, with 300-epoch training, MoBY achieves top-1 accuracy of 72.8% and 75.0%. Performance-wise, it is marginally superior to more contemporary efforts like MoCo v3 and DINO, which employ less complex techniques but share a similar structural foundation. Our SWIN Transformer backbone enables evaluation on object detection and semantic segmentation tasks, in contrast to other ViT/DeiT techniques that only report results on linear evaluation. This offers a thorough analysis of Transformer design self-supervised learning methods.

[8], The SWIN Transformers and residual convolutional networks ensemble is introduced in this study to enhance the classification and detection of plant diseases. At various scales, SWIN Transformers provide effectiveness and adaptability, and potent Residual networks extract in-depth essential point properties. On PlantVillage Kaggle datasets, the proposed model is assessed using a variety of performance metrics, including accuracy, precision, recall, specificity, and F1-score. The suggested model is compared to existing architectures in order to show its superiority to techniques like FCN-8, CED-Net, SegNet, DeepLabv3, Densenets, and Centernets.

In [9], the authors of this research used a novel method for identifying rice diseases based on a SWIN-transformer with sliding window operation and hierarchical design. Several prevalent rice illnesses were identified by photographing them in a field setting as the research subject. The accuracy of the suggested SWIN-transformer model is 93.4%, which is higher than the accuracy of the 4.1% traditional machine learning model. The system has a particular reference for other agricultural products, and experiments have shown that it may effectively enhance the accuracy of rice disease diagnosis when compared to traditional machine learning.

[10], To solve weed recognition problems that might be used by automated robots, the authors used Attention-based Transformer models. Using a collection of 1006 photos divided into 10 different marijuana classes, they created deep learning-based semantic segmentation models. To provide a sizeable sample set for Transformer models, the dataset was increased. SegFormer, one of the examined Transformer topologies, had 3.7 M parameters and a mean accuracy (mAcc) and mean intersection of union (mIoU) of 75.18% and 65.74%, respectively. The dataset was used to evaluate the SWIN Transformer, SegFormer, and Segmenter architectures.

3 Methodology

3.1 Attention-Based CNN

This approach emphasizes spatial attention while incorporating attention mechanisms into the CNN's convolutional layers. These are the crucial actions:

1. *Convolutional Layers:* To extract hierarchical characteristics, input images are passed through convolutional layers.
2. *Attention Maps:* A distinct branch creates attention maps to draw attention to essential areas in the feature maps.

The importance of each spatial location in the feature maps is determined by attention weights, which are computed using attention maps. The attended feature maps are given by

$$F\_attended = A \odot F \qquad (1)$$

3. where ⊙ represents element-wise multiplication between the attention map A and the feature maps F. This operation applies the attention weights to each spatial location in the feature maps, emphasizing the informative regions and suppressing the less relevant ones.
4. *Attended Features:* To highlight informative regions, attention weights are added to feature maps using element-wise multiplication.
5. *Additional Processing:* To improve representations and extract higher-level features, attended feature maps can go through additional processing using layers like pooling, convolution, or fully linked layers.
6. *Regression or classification:* To get the desired results, refined feature maps are flattened or aggregated and run through layers of regression or classification.

The attention-based CNN architecture uses attention processes to increase discriminative capability and capture fine-grained picture features by selectively focusing on interesting regions [5].

3.2 SWIN Transformers

There is a proposed method for SWIN Transformer which introduces a novel shifted window mechanism to reduce the computational complexity of self-attention. There are several operations involved in this transformer method. They are,

1. *Patch Division:* The input image is divided into non-overlapping patches. Each patch is represented by a vector of pixel values. Let's assume we have an image of size H x W pixels, and each patch has a size of P x P pixels [18]. The number of patches is given by

$$N = (H/P) \times (W/P) \tag{2}$$

2. *Linear Projections:* Using a linear projection matrix W0, each patch is projected into a lower-dimensional representation, yielding projected patch vectors X0 = W0 X.
3. *Shifted Windows:* The image is split into shifted windows, and the patches inside each window are reorganized to create a new matrix X1. In order to effectively collect local and global contextual information, spatial changes are introduced into the patch ordering. [19]
4. *Self-Attention:* The self-attention output matrix Y1 is created by applying self-attention to the rearranged patch matrix X1. This approach uses query, key, and value projections to determine the relative relevance of each patch within a single window.
5. *Reverse Rearrangement:* The attended features in Y1 are put back in the correct sequence for the patches, restoring the patches' spatial structure.
6. *Hybrid Tokenization:* To create the hybrid tokenized representation X2, which captures both local and global contextual information, the patch matrix X0 from linear projections and the attended features from reverse rearrangement are joined.
7. *Transformer Layers:* X2 goes through several layers of the Transformer design, including feed-forward networks, layer normalization, and multi-head self-attention. These layers develop rich representations and hierarchical dependencies [12].

The above steps are repeated for multiple iterations to refine the representations, and finally, the output of the last Transformer layer is used for downstream tasks such as image classification or object detection [1].

4 Architecture

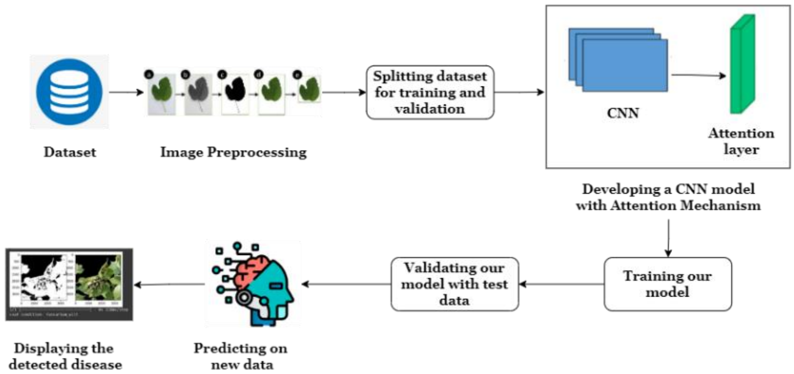

Fig. 1. Architecture diagram of Attention based CNN

In Fig. 1, the Attention-based CNN architecture uses a convolutional neural network as a feature extractor and an attention mechanism to learn which features to focus on for classification [22]. The attention mechanism takes the output of the CNN and learns a set of attention maps, which highlight the most informative regions of the image for the task at hand. These attention maps are then used to weight the feature maps produced by the CNN, emphasizing the most relevant features. The resulting weighted feature maps are then flattened and fed into a fully connected layer for classification. By using attention to focus on the most informative parts of an image, the attention-based CNN architecture can achieve better performance on image classification tasks with complex or cluttered backgrounds [5].

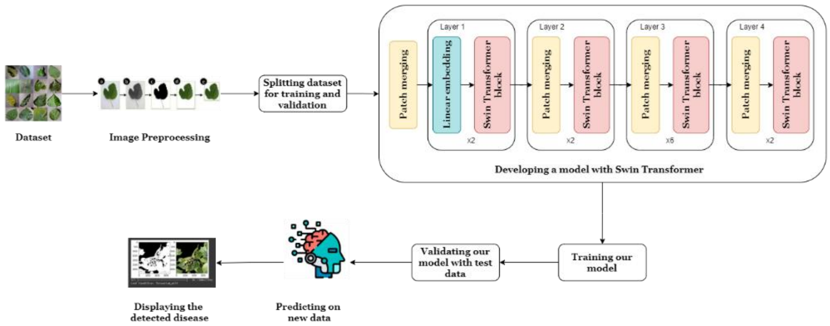

Fig. 2. Architecture diagram of SWIN Transformer

In Fig. 2, the SWIN Transformer architecture can be divided into two main components: the patch embedding stage and the hierarchical transformer stage.

In the patch embedding stage, the input image is first divided into non-overlapping patches of a fixed size. Each patch is then linearly projected into a lower-dimensional vector, which is passed through a small feedforward network to generate an initial feature representation. In the hierarchical transformer stage, the network processes the patches in a hierarchical manner using multiple stages of transformer blocks. Each stage contains a certain number of transformer blocks, and each block consists of two layers: a SWIN layer and a Shift layer [17]. The SWIN layer applies a multi-head self-attention mechanism to the input patches, allowing the network to capture global relationships between them. Specifically, the SWIN layer applies a self-attention operation to each patch, and then uses another set of linear projections to combine the information across all patches. This operation is similar to the self-attention mechanism used in the original Transformer architecture. The Shift layer shifts the positions of the patches in a structured manner, which helps to reduce computational complexity and improve efficiency. Specifically, the Shift layer first divides the patches into groups, and then shifts the positions of the patches within each group. This operation creates overlapping regions between adjacent patches, which helps to capture more local information about the input image. The output of each stage is then passed through a downsampling layer, which reduces the resolution of the patches and prepares them for processing in the next stage. This process is repeated for several stages until the final output, which is a vector of probabilities indicating the predicted class of the input image [20].

5 Dataset Description

There are two datasets in which larger dataset have 5816 images and smaller dataset have 1791 images. Here the images of Cotton crop are divided into four categories which are Fusarium Wilt, Curl Virus, Healthy, and Bacterial Blight classes.

Table 1. Larger dataset

Category	Number of Images
Healthy	848
Bacterial Blight	838
Curl Virus	1078
Fusarium Wilt	3052

This larger dataset was obtained from the Mendeley website and consists of 5816 images. We used 4652 images to train the model and 1164 images to validate the model. As per Table there is certain imbalance with the categories in which one category i.e., Fusarium Wilt has more images compared to other images because it is most occurring local disease in cotton crop and also cannot be ignored as well.

Table 2. Smaller dataset

Category	Number of Images
Healthy	506
Bacterial Blight	448
Curl Virus	418
Fusarium Wilt	419

This smaller dataset was obtained from the kaggle website and consists of 1791 images of Cotton crop with varying numbers of images in each class, providing a diverse and balanced representation of plant health conditions and diseases. We used 1433 photos to train the model and 358 images to validate the model.

6 Implementation and Results

6.1 Model Training

While beginning the model training, the both datasets need to be pre-processed. Therefore, resizing, normalization, data augmentation techniques like rotation range, width shift range, height shift range, shear range, zoom range, horizontal flip and sharpness are applied to the images to improve the model performance [25].

Both the datasets are divided into training and validation at 80 and 20 ratio. The first training stage focuses on Dataset A, the larger dataset. Mainly, the model's performance is monitored on the validation set, and hyperparameters, such as learning rate and regularization, are adjusted as needed. The proposed models are trained under both the training datasets.

6.2 Model Evaluation

After training the model, it is tested on the validation dataset to ensure that it can generalize to new data and not just memorize the training data.

To evaluate the models, confusion matrix and classification are used to define each model's accuracy under two different datasets.

Table 3. Confusion matrix of Attention Based CNN for larger dataset

	Healthy	Bacterial Blight	Curl Virus	Fusarium Wilt
Healthy	275	7	8	12
Bacterial Blight	7	101	8	6
Curl Virus	4	3	248	4
Fusarium Wilt	7	6	3	465

From the table 3, we can see that the model performs well in general, with high numbers on the diagonal indicating correct predictions. However, there are some misclassifications, particularly between Bacterial Blight and Curl Virus, as well as between Bacterial Blight and Fusarium Wilt. These misclassifications may suggest that the features used by the model to distinguish between these classes may not be strong enough, or that there may be some overlap in the features of these classes.

Table 4. Confusion matrix of SWIN Transformer for Larger Dataset

	Healthy	Bacterial Blight	Curl Virus	Fusarium Wilt
Healthy	283	5	4	8
Bacterial Blight	5	109	7	6
Curl Virus	1	3	252	3
Fusarium Wilt	2	2	3	483

From the table 4, However, there are some improvements in classifications, particularly between Bacterial Blight and Curl Virus, as well as between Bacterial Blight and Fusarium Wilt. The misclassifications, features overlapping, etc., occurred in Attention based CNN model are reduced comparatively.

Table 5. Confusion matrix of Attention Based CNN for smaller dataset

	Healthy	Bacterial Blight	Curl Virus	Fusarium Wilt
Healthy	93	2	1	2
Bacterial Blight	1	89	2	1
Curl Virus	1	2	81	1
Fusarium Wilt	1	1	2	79

From the table 5, by examining the distribution of values across the matrix, the model's performed good in classifying different labels can be identified.

Table 6. Confusion matrix of SWIN Transformer for smaller dataset

	Healthy	Bacterial Blight	Curl Virus	Fusarium Wilt
Healthy	89	3	2	3
Bacterial Blight	3	85	3	2

Curl Virus	2	4	77	2
Fusarium Wilt	2	3	2	76

From the table 6, by examining the distribution of values across the matrix, the model's performance decreased compared to table 5.2.2 in classifying different labels can be identified.

7 Results and Observations

Table 7. Comparison results of accuracies

Model Name	Larger Dataset	Smaller Datatset
SWIN Transformer	96.821	91.340
Attention Based CNN	93.556	95.530

The table 7, compares the accuracies of two different models, namely SWIN Transformer and Attention-based CNN, on both larger and smaller datasets. The accuracies are presented as percentages.

On the larger dataset, the SWIN Transformer model outperformed the Attention-based CNN model, achieving a greater accuracy of 96.821%, surpassing its performance which is 7.6% greater than [2]. This suggests that the SWIN Transformer model, which took advantage of the larger dataset's sophisticated architecture to capture complicated patterns and features contained in the data, was more successful in accurately identifying the data instances.

The performance of the models varied, though, when assessed using the smaller dataset. The accuracy of the SWIN Transformer model was 91.340%, which was somewhat less accurate than the accuracy of the Attention-based CNN model, which was 95.530% which is 6.3% greater than [2]. This shows that the Attention-based CNN model, as compared to the SWIN Transformer model, was better able to capture the specific patterns and features inside the smaller dataset.

These findings demonstrate the significance of choosing a model while taking the dataset's peculiarities into account. The SWIN Transformer model performed exceptionally well while processing the larger dataset, taking advantage of its capacity to record intricate patterns and broad contextual data. The Attention-based CNN model, on the other hand, displayed greater performance on the smaller dataset, demonstrating its capacity to efficiently learn and use local features and patterns. Achieving the best results in classification jobs requires selecting the right model based on dataset size.

8 Deployment

As per Fig. 3, flask module is used to deploy the model for image classification. Comparing both the model as per table 7 the model which gained more accuracy i.e., SWIN transformers is used for the deployment. Therefore, saved the SWIN transformer model and loaded the model at the time of deployment.

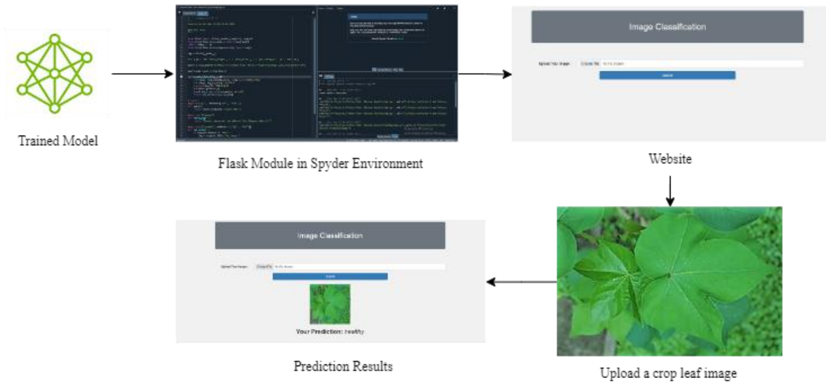

Fig. 3. Architecture diagram of model deployment

8.1 Test case Results

Test case-1
Expected output: Healthy
Actual output:

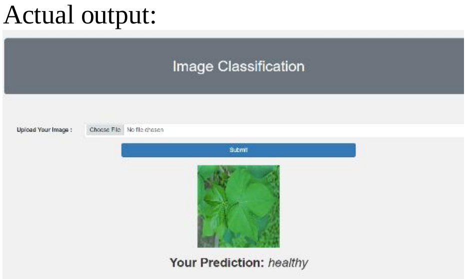

Fig. 4. Healthy

Test case-2
Expected output: Bacterial Blight
Actual output:

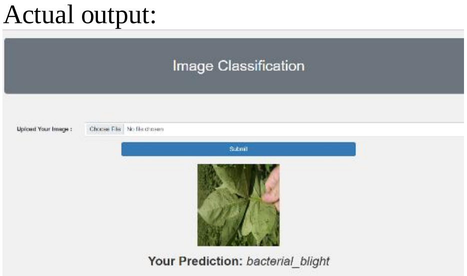

Fig. 5. Bacterial blight

Test case-3
Expected output: Curl Virus
Actual output:

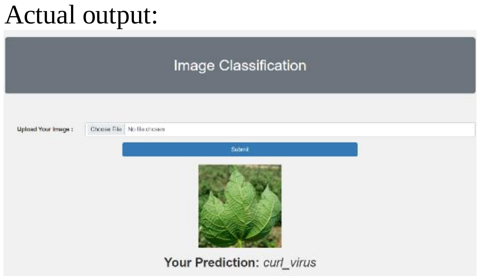

Fig. 6. Curl virus

Test case-4
Expected output: Fusarium wilt
Actual output:

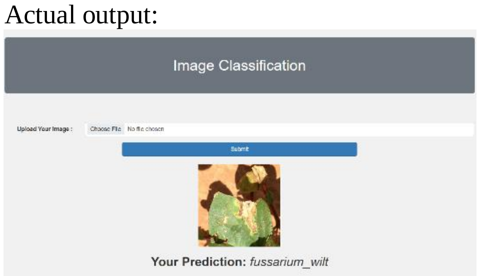

Fig. 7. Fusarium wilt

9 Conclusion

SWIN Transformers offer advantages over conventional approaches and show potential in the detection of crop illnesses, notably in cotton fields. They make it possible to analyse big datasets quickly and precisely, giving a thorough assessment of crop health. Early diagnosis and intervention in unhealthy cotton crops can be accomplished by utilising the high-level feature representations learned by SWIN Transformers, potentially saving farmers money. However, more study is required to verify these findings and determine whether it is feasible to use this technology in practical settings. The effectiveness of the model is also influenced by variables including data quality, amount, hyperparameters, and training technique. SWIN Transformers have a lot of potential for cotton crop disease detection and should be investigated further in agricultural research.

References

1. Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. "SWIN transformer: Hierarchical vision transformer using shifted windows." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012-10022. 2021.
2. Caldeira, Rafael Faria, Wesley Esdras Santiago, and Barbara Teruel. "Identification of cotton leaf lesions using deep learning techniques." *Sensors* 21, no. 9 (2021): 3169.
3. Memon, Muhammad Suleman, Pardeep Kumar, and Rizwan Iqbal. "Meta Deep Learn Leaf Disease Identification Model for Cotton Crop." *Computers* 11, no. 7 (2022): 102.
4. Wang, Yanxue, Shansong Wang, Weijian Ni, and Qingtian Zeng. "PAST-net: a SWIN transformer and path aggregation model for anthracnose instance segmentation." *Multimedia Systems* (2022): 1-13.
5. Tao, Yu, Fangle Chang, Yuhang Huang, Longhua Ma, Lei Xie, and Hongye Su. "Cotton Disease Detection Based on ConvNeXt and Attention Mechanisms." *IEEE Journal of Radio Frequency Identification* 6 (2022): 805-809.
6. Tang, Yucheng, Dong Yang, Wenqi Li, Holger R. Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. "Self-supervised pre-training of SWIN transformers for 3d medical image analysis." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20730-20740. 2022.
7. Xie, Zhenda, Yutong Lin, Zhuliang Yao, Zheng Zhang, Qi Dai, Yue Cao, and Han Hu. "Self-supervised learning with SWIN transformers." *arXiv preprint arXiv:2105.04553* (2021).
8. Ponugoti, Kalpana, R. Anandan, and Laith Abualigah. "Plant Disease Recognition Using Residual Convolutional Enlightened SWIN Transformer Networks." (2023).
9. Zhang, Zhenghua, Zhangjie Gong, Qingqing Hong, and Ling Jiang. "SWIN-transformer based classification for rice diseases recognition." In *2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI)*, pp. 153-156. IEEE, 2021.
10. Jiang, Kan, Usman Afzaal, and Joonwhoan Lee. "Transformer-Based Weed Segmentation for Grass Management." *Sensors* 23, no. 1 (2023): 65.
11. Ramacharan, D. S. "A 3-stage method for disease detection of cotton plant leaf using deep learning CNN algorithm." *Int. J. Res. Appl. Sci. Eng. Technol.* 9 (2021): 2503-2510.
12. Shereesha, M., C. Hemavathy, Hasthi Teja, G. Madhusudhan Reddy, Bura Vijay Kumar, and Gurram Sunitha. "Precision Mango Farming: Using Compact Convolutional Transformer for Disease Detection." In *International Conference on Innovations in Bio-Inspired Computing and Applications*, pp. 458-465. Cham: Springer Nature Switzerland, 2022.
13. Yan, Jingkun, Tianying Yan, Weixin Ye, Xin Lv, Pan Gao, and Wei Xu. "Cotton leaf segmentation with composite backbone architecture combining convolution and attention." *Frontiers in Plant Science* 14 (2023): 1111175.

14. Sharmila, S. K., R. Bhargavi, R. Yamini Anusha, K. Anusha, and B. Divya. "Cotton Leaf Disease Detection using Convolutional Neural Networks (CNN)." In *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)*, pp. 1179-1185. IEEE, 2023.
15. Huang, Jiahao, Yingying Fang, Yinzhe Wu, Huanjun Wu, Zhifan Gao, Yang Li, Javier Del Ser, Jun Xia, and Guang Yang. "Swin transformer for fast MRI." *Neurocomputing* 493 (2022): 281-304.
16. Wang, Fengyi, Yuan Rao, Qing Luo, Xiu Jin, Zhaohui Jiang, Wu Zhang, and Shaowen Li. "Practical cucumber leaf disease recognition using improved Swin Transformer and small sample size." *Computers and Electronics in Agriculture* 199 (2022): 107163.
17. Li, Li-Hua, and Radius Tanone. "Disease Identification in Potato Leaves using Swin Transformer." In *2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, pp. 1-5. IEEE, 2023.
18. Shamshad, Fahad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, and Huazhu Fu. "Transformers in medical imaging: A survey." *Medical Image Analysis* (2023): 102802.
19. https://towardsdatascience.com/a-comprehensive-guide-to-swin-transformer-64965f89d14c
20. Bi, Chunguang, Nan Hu, Yiqiang Zou, Shuo Zhang, Suzhen Xu, and Helong Yu. "Development of deep learning methodology for maize seed variety recognition based on improved swin transformer." *Agronomy* 12, no. 8 (2022): 1843.
21. Yang, Bin, Zhulian Wang, Jinyuan Guo, Lili Guo, Qiaokang Liang, Qiu Zeng, Ruiyuan Zhao, Jianwu Wang, and Caihong Li. "Identifying plant disease and severity from leaves: A deep multitask learning framework using triple-branch Swin Transformer and deep supervision." *Computers and Electronics in Agriculture* 209 (2023): 107809.
22. Jenifa, A., R. Ramalakshmi, and V. Ramachandran. "Cotton leaf disease classification using deep convolution neural network for sustainable cotton production." In *2019 IEEE international conference on clean energy and energy efficient electronics circuit for sustainable development (INCCES)*, pp. 1-3. IEEE, 2019.
23. Kukadiya, Hirenkumar, and Divyakant Meva. "Automatic cotton leaf disease classification and detection by Convolutional Neural Network." In *International Conference on Advancements in Smart Computing and Information Security*, pp. 247-266. Cham: Springer Nature Switzerland, 2022.
24. Manavalan, R. "Towards an intelligent approaches for cotton diseases detection: A review." *Computers and Electronics in Agriculture* 200 (2022): 107255.
25. Kumari, Ch Usha, S. Jeevan Prasad, and G. Mounika. "Leaf disease detection: feature extraction with K-means clustering and classification with ANN." In *2019 3rd international conference on computing methodologies and communication (ICCMC)*, pp. 1095-1098. IEEE, 2019.