2021 International Conference on
Artificial Intelligence (ICAI)
Islamabad, Pakistan, April 05-07, 2021

Wheat Disease Recognition through One-shot Learning using Fields Images

Hamza Mukhtar
Alkhawarizmi Institue of Computer Science
University of Engineering and Technology
Lahore, Pakistan
hamza.mukhtar@kics.edu.pk

Muhammad Zeeshan Khan
Alkhawarizmi Institue of Computer Science
University of Engineering and Technology
Lahore, Pakistan
zeeshan.khan@kics.edu.pk

Muhammad Usman Ghani Khan
Alkhawarizmi Institue of Computer Science
University of Engineering and Technology
Lahore, Pakistan
usman.ghani@kics.edu.pk

Haseeb Younis
Alkhawarizmi Institue of Computer Science
University of Engineering and Technology
Lahore, Pakistan
haseeb.younis@kics.edu.pk

2021 International Conference on Artificial Intelligence (ICAI) | 978-1-6654-3293-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICAI52203.2021.9445266

***Abstract*— Accurate disease recognition from wheat plants is essential to mitigate the effects and to stop the spread of the diseases. State-of-the-art algorithms have developed through deep learning which recognizes the disease from field images. Although such methods produce highly accurate results, they require tons of labeled images, and work only on the classes which are involved in the training phase. Thus, to get the recognition against the different classes, the model is required to retrain. This article has proposed a wheat disease recognition network based on one-shot learning which not only needs a small number of images for training, but also can accommodate new categories because it can be trained even on few images of a new type. So, farmers can retrain the network just by giving a few images of plants containing the concerned disease, and start testing immediately. We have used the MobileNetv3 network as a feature extractor which is extremely fast and accurate classification. This network is fine-tuned on the PlantVillage dataset, while the last two dense layers are fine-tuned on the plant images of 11 wheat disease, those images are taken from the CGIAR Crop Disease dataset and Google images. The whole One-shot network is trained on 440 images having 40 images of each class. Siamese networks are used for producing the encodings of input images, then the absolute difference is calculated between encodings, and similarity scores are determined through the Sigmoid unit. It assigns 1 score to similar images, while 0 to dissimilar images. Our Mobilenetv3 model has achieved around 98% training and 96% validation, while the whole one-shot network has achieved more than 92% accuracy, 84% precision, and 85 recall. Our proposed system requires just a few images of a new type for training, instead of retraining the network as in standard classification networks.**

Keywords— Continual learning, MobileNet, Siamese Network, Digital Agricultural, crop diseases.

I. Introduction

Recognition of disease is very crucial for the identification of pertinent treatment, and prevention of spread. Diseases have severe effects on the yield of the wheat crop, in turn, it has been enhancing food shortage around the globe. Different spectroscopic techniques can be used for the wheat diseases analysis, but can be time-consuming and expensive due to the involvement of various sensors. Moreover, these techniques are required laboratory and domain experts. Manual checking by the farmer can have the risk of error because most of the farmers are not disease experts especially in developing countries. Due to the wide utilization of camera-enabled smartphones, Computer Vision-based techniques have proved highly efficient and accurate in different visual tasks, such as image classification, object detection, and recognition. Therefore, crop disease recognition through field images of plants by applying deep learning techniques is growing in the age of digital agriculture, where the farmer is required to have a camera-enabled smart-phone [1].

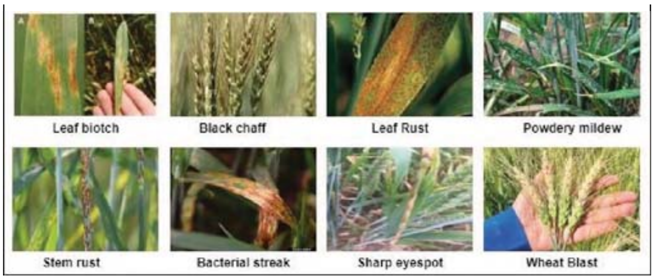

Fig. 1. Samples images from our dataset having various types of diseases.

At present, Convolutional Neural Networks (CNN) are outperforming all the traditional machine learning and computer vision techniques, where millions of learnable parameters are involved. Sardogan et. al. [6] used CNN with learning vector quantization algorithms for leaf disease classification, where CNN had 11 convolutional layers followed by max-pooling layers and 2 fully connected layers CNN with 9 convolutional layers was used by Khamparia et. al. [7] for seasonal crop diseased which achieved only 86% accuracy at test time. A much bigger network, VGG, was used for the classification of wheat diseases where the model was trained on 8 classes. Although achieved good accuracy, work is very slow due to the large CNN network [1]. For classifying the diseases of sugarcane, a CNN model having 4 convolutional layers was used [10].

CNN based classification frameworks [2, 3, 12, 9] are very accurate, however, these frameworks do not work when tested for diseases that were absent during training. Moreover, such frameworks require a huge number of images that are very difficult to acquire because the visual perception of diseases are not uniform, and vary according to the geographical areas.

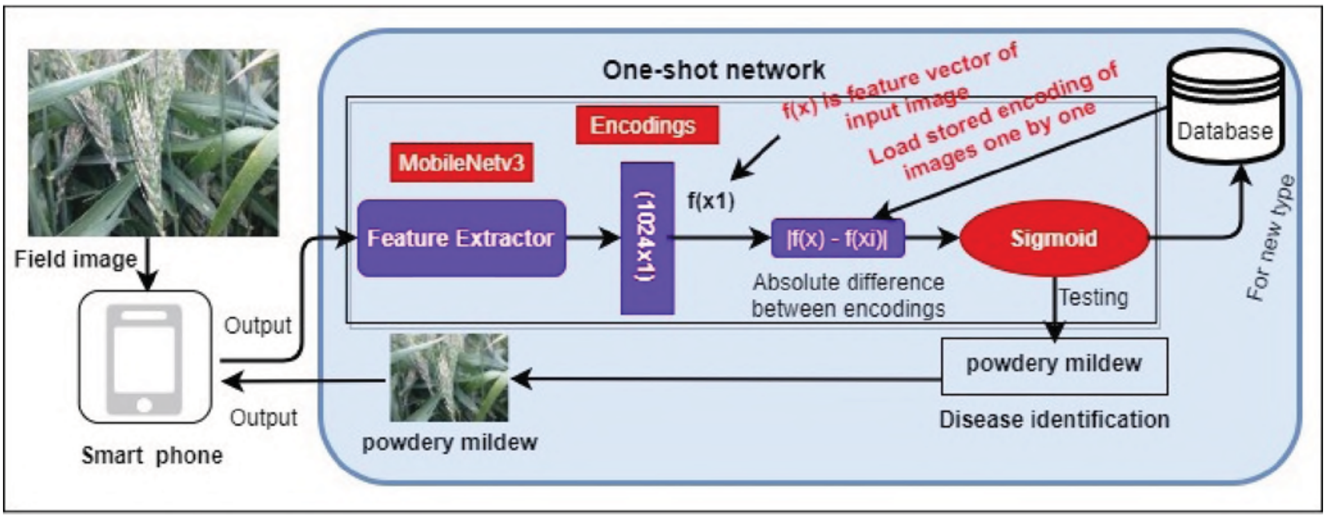

Fig. 2: Overall testing mechanism of our system.

So, the different and reshaped disease is required to retrain the whole network with large images of a new disease. Therefore, building a universal network with such state-of-the-art CNN for disease recognition is not possible which urges the need for such a system that can accommodate newer and varied diseases with slight effort.

A slightly different deep learning approach, called one-shot learning [4], can recognize the desired phenomenon in the input image, and accommodate images of different classes easily, without retraining the whole network. One-shot learning not only recognizes the new class accurately and efficiently, but also does it with few images. This approach also uses CNN as a backbone network for feature extraction, and builds the encoding of images, and determines the similarity of the input image by calculating the absolute difference between the encoding of the input image. For training, Siamese CNNs are used to extract the feature embeddings of both given images, and determine the similarity or dissimilarity by measuring the absolute difference by calculating the element-wise difference. A network is trained to learn the similarity function, and use this learning at test time where embeddings of image input are compared to the embeddings of all images that were involved in training [4]. This approach is different from template matching that compares image to image, works very slow with a quite low accuracy, and lacks robustness.

Standard CNN based classification is required to feed input images to a network of several layers and gives output in terms of probabilities against all the included classes. Mostly, the SoftMax function is used to get probabilities for all the classes. For example, if the network is trained on 4 classes then 4 probabilities will be given as output and the class with the highest probability will be assigned to the given input image. Now, trained models cannot work on the class that was not included at the training time of the network, and models need to retrain for different classes, but one-shot learning works even if at least one image is given for training.

A one-shot network for general image classification was applied which gave extremely good results where Siamese CNN was used as a feature extractor [4]. Argüeso et. al. [8] used few-shot learning which was comprised of inceptionv3 as feature extractor and SVM for the classification of 40 categories that achieved a reasonable performance. Few-shot approaches for classification involve model initialization, data hallucination, and metric learning.

We have proposed a disease recognition system by combining the one-shot learning with MobileNetV3 [5]. Our baseline system network is capable of recognizing the 11 commonly found diseases, but can be extended to work any new type of disease with very little effort. Few sample images of various diseases are shown in Fig 1. MobileNetV3 is used as a feature extraction network that efficiently produces the embedding of images. MobileNetV3 is an extremely fast, lightweight, and highly accurate classification network, developed for mobile devices [5]. For training, MobileNetV3 is involved in Siamese fashion [4] which receives two input images and output encodings of both inputs which are further used for similarity measurement. For testing, Farmer will capture the field image and upload it to the system, where it goes from feature extractor and encodings will be produced which are eventually used to calculate the absolute difference between the input encodings of an input image and the encodings stored in the database. The overall mechanism of our system for testing is illustrated in fig. 2. Main contributions of our proposed system:

- Can accommodate disease types due to continual learning mechanism
- No need for retraining for new types.
- Extremely efficient due to fast feature extractor.
- Need only a few images for one-shot training

In section 2 of this article, the methodology of our one-shot network is explained along with the used backbone network, while section 3 explains the dataset, training mechanism, evaluate and comparison performance with other approaches. In the end, section 4, concludes the article while giving a future path.

II. METHODOLOGY

Our system has three main components where input images feed to feature extraction units to obtain the encodings of the image. These encodings are then used to calculate the absolute difference (AD) between input images. In the end, one dense layer having a Sigmoid unit is used to generate similarity scores. For feature extraction, we have used MobileNetv3 which is very fast and accurate at the same time. Both Mobilenetv3 networks are the same, therefore, it called

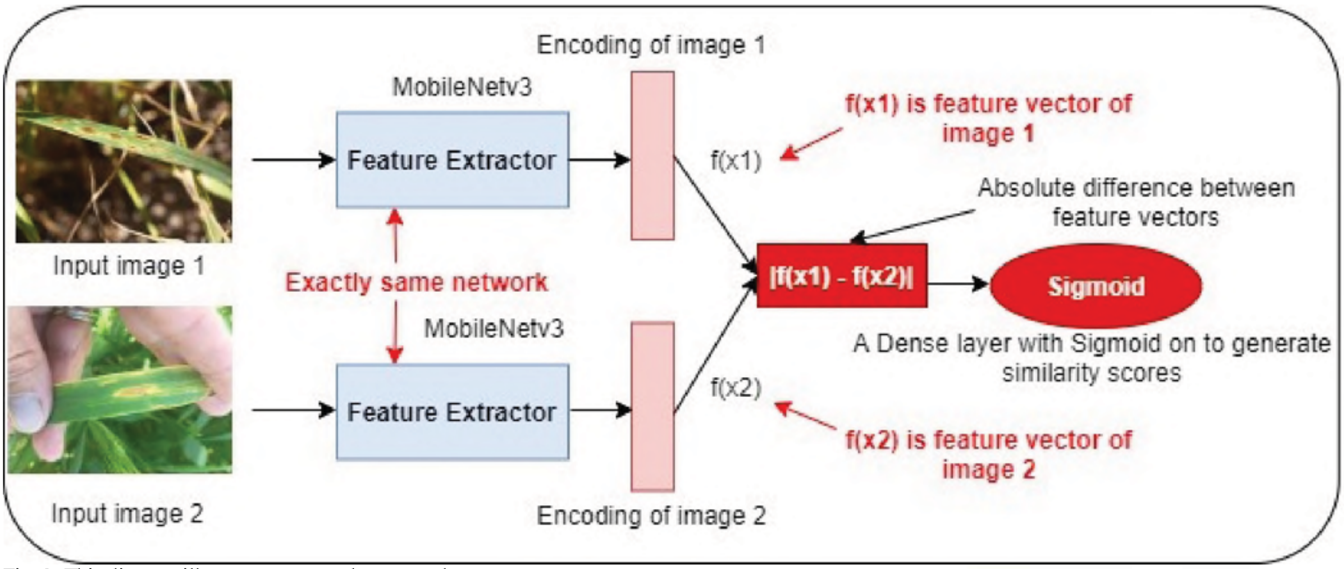

Fig. 3: This diagram illustrates our one-shot network

Siamese. For the training of one-shot networks, exactly the same feature extractors have been used to produce encodings. Encodings represent the feature vector of the given image which is further used for measuring the AD between the images. Fig. 3 is showing the proposed one-shot network.

We have mapped the disease recognition problem as a binary classification task. For example, our dataset is composed of in (X, Y) pairs, where X refers to all the input images and Y is the output against each input image pair. Against each pair, the system needs to give a similarity score between 0 to 1. Fig. 4 is showing this mapping.

Fig. 4: Mapping of the problem as binary classification.

A. MobileNetv3

In contrast with other classification models, it operates a single convolution on each depth of the input image instead of combining and flattening all the depths of input, achieves this through depth-wise separable convolutional. This depth-wise convolutional divides the convolutional process into two layers, one layer for filtering and one for combining.

This combination reduces the size of the model. MobileNetv3 [5] consists of 4 2D convolutional layers, 2 (112x122) bottlenecks layers, 2 (56x56) bottlenecks layers, 3 (28x28) bottlenecks layers, 7 (14x14) bottlenecks layers, 2 (7x7) bottlenecks layers where Swish and Relu is used as activations. One pooling layer (7x7) is used before two dense layers. Squeeze and excitation layers are also included which makes it faster and lightweight. This addition assigns unequal weights to channels while creating output feature maps. In the end, a dense layer with 1024 units is applied to get the feature vector. Specifications of the complete Mobilenetv3 model used are shown in fig. 5. which borrowed from [5].

Input	Operator	exp size	#out	SE	NL	s
$224^2 \times 3$	conv2d	-	16	-	HS	2
$112^2 \times 16$	bneck, 3x3	16	16	-	RE	1
$112^2 \times 16$	bneck, 3x3	64	24	-	RE	2
$56^2 \times 24$	bneck, 3x3	72	24	-	RE	1
$56^2 \times 24$	bneck, 5x5	72	40	✓	RE	2
$28^2 \times 40$	bneck, 5x5	120	40	✓	RE	1
$28^2 \times 40$	bneck, 5x5	120	40	✓	RE	1
$28^2 \times 40$	bneck, 3x3	240	80	-	HS	2
$14^2 \times 80$	bneck, 3x3	200	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	184	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	184	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	480	112	✓	HS	1
$14^2 \times 112$	bneck, 3x3	672	112	✓	HS	1
$14^2 \times 112$	bneck, 5x5	672	160	✓	HS	2
$7^2 \times 160$	bneck, 5x5	960	160	✓	HS	1
$7^2 \times 160$	bneck, 5x5	960	160	✓	HS	1
$7^2 \times 160$	conv2d, 1x1	-	960	-	HS	1
$7^2 \times 960$	pool, 7x7	-	-	-	-	1
$1^2 \times 960$	conv2d 1x1, NBN	-	1024	-	HS	1

Fig. 5: Complete MobileNetv3 network. Here SE represents the squeeze and excitation blocks, NL refers to nonlinear functions used. HS present Swish and RE for Relu activation unit.

B. Encodings

Encoding refers to the feature map of the given image. It's a 1024 size column vector. This vector is used to compute the absolute difference between images. At the end of all the convolutional layers, a dense layer having 1024 units is used to produce this encoding. These encodings are also stored in the database and used to compare input images at test time.

C. Absolute difference

This difference is actually an element-wise difference between the encodings of the two images. If the input images are of the same category, then the absolute difference between corresponding encodings will be very small or even zero for idea cases. The larger the difference, the lesser the similarity, so both absolute difference and similarity are inverse proportional.

D. Similarity measurement

Finally, to compute the similarity score between two encodings after calculating the absolute difference, an output layer with a sigmoid unit has been used. If both input images are of the same class then it produces 1, while 0 for

TABLE I. CONFUSION MATRIX

Predicted Diseases \ Actual Diseases	Powdery Mildew	Tan Spot	Sharp Eyespot	Leaf Blotch	Leaf Rust	Stem Rust	Black Chaff	Bacterial Streak	Wheat Blast	Phoma Spot	Stagonospora Nodorum
Powdery Mildew	47		1	1	1			1			1
Tan Spot	1	46				2				2	
Sharp Eyespot		1	48				1		1	2	1
Leaf Blotch		2		45		1					
Leaf Rust	1				47		1	2			1
Stem Rust						44					1
Black Chaff			1				45				
Bacterial Streak	1			2			1	47			
Wheat Blast				1	1	1			49		
Phoma Spot					1	1				46	2
Stagonospora Nodorum		1		1			2				44

III. EXPERIMENT AND RESULTS

We have experimented to demonstrate the performance of the proposed one-shot network. We have fine-tuned the MobileNetv3 model which is used as a feature extractor in the recognition pipeline. We have presented the results of mobilenetv3 fine-tuning as well as a one-shot network for disease recognition.

A. Training setup

Pre-trained Mobilenetv3 is fine-tuned on the 1080 Ti Nvidia GPU enabled machine having 16 GB RAM. We have used RMSprop optimizer having 0.8 momentum, the initial learning rate is 0.1 having 0.01 decay rate after every 5 epochs, with 16 batch size. A dropout of 0.3 is also used to avoid overfitting and create robustness. We also have used batch normalization in each convolutional layer with 0.7 average decay.

B. Dataset

First, we have used then pretrained the MobileNetv3 network, and the whole network is fine-tuned on the Plantvillage dataset [10] which has around 20k labeled image of 15 categories, while the last two dense layers are fine-tuned on around 1450 images of CGIAR Computer Vision for Crop Disease Kaggle dataset and 235 images google images. The last two dense layers are fine-tuned on the images of 11 categories, which have also been used to train our one-shot recognition network. For training the one-shot network, we have 40 images of each 11 classes.

C. MobileNetv3 performance

Mobilenetv3 is fine-tuned for 3500 epochs, and has achieved 97.78% training accuracy, while 96.3% validation accuracy. We have split the dataset into 80-20 ratio where 80 % is kept for training, while 20% for the validation of the network. As the model has achieved very high training accuracy which means that model is not underfitted. On the other hand, the model has obtained very good accuracy on the validation image which also not faced large overfitting. Fig. 6 is showing the accuracy graph, while Fig. 7 is showing the validation accuracy on 3500 epochs.

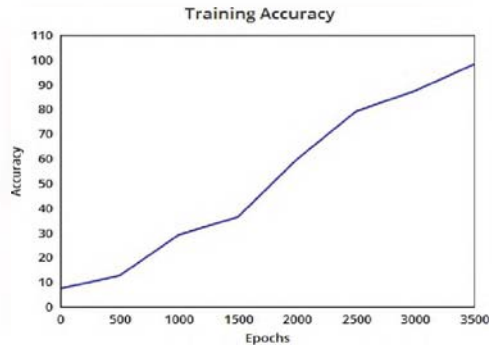

Fig. 5: Mobilenetv3 training accuracy graph.

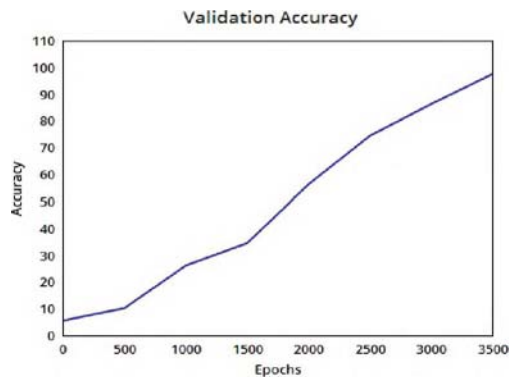

Fig. 6: Mobilenetv3 validation accuracy graph

D. One-shot network performance

We have trained our one-shot network on the 40 images of each class, and test on 50 images of each of 11 diseases. For validation, we have adopted the N-way one-shot learning because for each input pair, our model determines a similarity score between 1 to 0. We have validated the model against 4-way and 8-way one-shot learning. Our model gives more wrong predictions of 8-way one-shot learning, than 4-way validation. Tab. 1. Shows the aggregated confusion matrix of the one-shot disease recognition network. We have achieved 92.34% accuracy, 84.01% precision, and 85% recall. On the Stagonospora nodorum and Stem rust diseases, our recognition model performed slightly poor, while performance was highest on the wheat blast. Both precision and recall are lower which is not good because both false positive and false negative cases can miss guide the framer.

Performance comparison with other approaches who have worked on wheat diseases is given Tab. 2. Although VGG [9] has achieved higher accuracy, it does not accommodate different types of disease which were not present during training because it was based on standard deep learning based CNN model.

TABLE II. PERFORMANCE COMPARISON

Model	*Dataset*	**Accuracy**	*Continual learning*
Few-Shot [8]	PlantVillege	91.4%	Yes
VGG [9]	Wheat Disease Database	93%	No
statistical inference [11]	Wheat 2014	80%	No
Our	CGIAR, Google images	92.2%	Yes

IV. CONCLUSION

Our study illustrates a one-shot learning network which not only can recognize the plant disease accurately, but also can accommodate new disease types, were absent during training. MobileNetv3 extract features which showed above 96% accuracy, and our model has achieved around 92% accuracy. This illustrates that models with few examples can be developed which also gives higher accuracy.

ACKNOWLEDGMENT

Authors pay their thanks to Intelligent Criminology Lab, National Center for Artificial Intelligence, Al-Khawarizimi Institute of Computer Science, UET Lahore for providing research platform, technical support, and financial assistance.

REFERENCES

[1] Lu J, Hu J, Zhao G, Mei F, Zhang C. An in-field automatic wheat disease diagnosis system. Computers and electronics in agriculture. 2017 Nov 1;142:369-79. [2] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).

[2] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).

[3] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. InProceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 4510-4520).

[4] Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for oneshot image recognition. InICML deep learning workshop 2015 Jul 10 (Vol. 2).

[5] Howard A, Sandler M, Chu G, Chen LC, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, Le QV. Searching for mobilenetv3. InProceedings of the IEEE/CVF International Conference on Computer Vision 2019 (pp. 1314-1324).

[6] Sardogan M, Tuncer A, Ozen Y. Plant leaf disease detection and classification based on CNN with LVQ algorithm. In2018 3rd International Conference on Computer Science and Engineering (UBMK) 2018 Sep 20 (pp. 382-385). IEEE.

[7] Khamparia A, Saini G, Gupta D, Khanna A, Tiwari S, de Albuquerque VH. Seasonal crops disease prediction and classification using deep convolutional encoder network. Circuits, Systems, and Signal Processing. 2020 Feb;39(2):818-36

[8] Argüeso D, Picon A, Irusta U, Medela A, San-Emeterio MG, Bereciartua A, Alvarez-Gila A. Few-Shot Learning approach for plant disease classification using images taken in the field. Computers and Electronics in Agriculture. 2020 Aug 1;175:105542.

[9] Khan MZ, Harous S, Hassan SU, Khan MU, Iqbal R, Mumtaz S. Deep unified model for face recognition based on convolution neural network and edge computing. IEEE Access. 2019 May23;7:72622-33

[10] Hughes D, Salathé M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060. 2015 Nov 25.

[11] Johannes A, Picon A, Alvarez-Gila A, Echazarra J, RodriguezVaamonde S, Navajas AD, Ortiz-Barredo A. Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. Computers and electronics in agriculture. 2017 Jun 1;138:200-9

[12] Younis H, Bhatti MH, Azeem M. Classification of skin cancer dermoscopy images using transfer learning. In2019 15th International Conference on Emerging Technologies (ICET) 2019 Dec 2 (pp. 1-4). IEEE.