Plant Disease Detection using Deep Learning

Article *in* International Journal of Recent Technology and Engineering · April 2020

DOI: 10.35940/ijrte.A2139.059120

5 authors, including:

Murk Chohan
Begum Nusrat Bhutto Women University Sukkur
7 PUBLICATIONS **60** CITATIONS
SEE PROFILE

Adil Khan
Sukkur Institute of Business Administration
6 PUBLICATIONS **60** CITATIONS
SEE PROFILE

Saif Hassan Katper
Sukkur Institute of Business Administration
6 PUBLICATIONS **83** CITATIONS
SEE PROFILE

Muhammad Saleem Mahar
Sukkur Institute of Business Administration
4 PUBLICATIONS **60** CITATIONS
SEE PROFILE

Plant Disease Detection using Deep Learning

Murk Chohan, Adil Khan, Rozina Chohan, Saif Hassan Katpar, Muhammad Saleem Mahar

Abstract: ***[Context] Plants play an essential role in climate change, agriculture industry and a country's economy. Thereby taking care of plants is very crucial. Just like humans, plants are effected by several disease caused by bacteria, fungi and virus. Identification of these disease timely and curing them is essential to prevent whole plant from destruction. [Objective] This paper proposes a deep learning based model named plant disease detector. The model is able to detect several diseases from plants using pictures of their leaves. [Methodology] Plant disease detection model is developed using neural network. First of all augmentation is applied on dataset to increase the sample size. Later Convolution Neural Network (CNN) is used with multiple convolution and pooling layers. PlantVillage dataset is used to train the model. After training the model, it is tested properly to validate the results. [Results] We have performed different experiments using this model. 15% of data from PlantVillage data is used for testing purpose that contains images of healthy as well as diseased plants. Proposed model has achieved 98.3% testing accuracy. [Conclusion] This study is focused on deep learning model to detect disease in plant leave. But, in future model can be integrated with drone or any other system to live detect diseases from plants and report the diseased plants location to people so that they can be cured accordingly.***

Keywords: Plant Disease, Convolution Neural Network (CNN), Deep Learning, Agriculture, and PlantVillage

I. INTRODUCTION

Plants play a vital role in economy and climate change. Since climate change has become a global issue also addressed in UN general assembly 2019, various countries including Pakistan are on mission to plant more and more trees and plants to keep climate balance. Many studies have proved that extinction of plants due to industry use have caused damage to ozone layer and thus resulting global warming. The rate of climate change forecast for the future is 10-100 times faster than the rate of DE glacial warming [1]. These plants also pay a major role in food industry as well. Balance of global food production is also a major issue [2]. Apart from this in health care also plants play a vital role [3]. Overall plants are very essential for human survival therefore it's also a worldwide concern to take care of them.

Just like human health plants health can also be affected by several diseases. In economic terms, annual losses in food, fiber and ornamental production systems caused by plant pests and diseases are estimated in the hundreds of billions of dollars [4]. These diseases are caused by fungi or fungal like orgasms. However, other serious diseases of food and feed crops are caused by viral and bacterial organisms [5]. Some of the diseases may be of spreadable nature, means they may spread from one plant to other hence needed to be identified and taken care timely. It's very challenging job to detect disease in plants in very early stages [3].

Some of the common symptoms of disease in plants disease are Leaf rust (common leaf rust in corn), Stem rust (wheat stem rust), Sclerotinia (white mold), Powdery mildew, Birds-eye spot on berries (anthracnose), Damping off of seedlings (phytophthora), Leaf spot (septoria brown spot), Chlorosis (yellowing of leaves). These diseases can be identified by physical condition of plants leaves. The experts can identify either the plant is defected or not by looking at leaves stems or fruit. This approach requires having lots human resources for this particular job. In this era of technology and automation it is not very efficient approach, it would be much better if we have an automated system which detects disease in plants automatically. There are many researches already done to fill this purpose most of them utilize traditional machine learning approaches [6]. The purpose of this study is to create such automated system for detecting diseases in plants by using deep learning technique. Deep learning is subset of machine learning. The advantage of deep learning over machine learning is that one does not need to worry about domain expertise as no feature engineering is required in this, unlike traditional machine learning approaches [7]. Our system just like other previous researches utilizes images of plants leaves to detect disease in plants. Plant disease detector is computer vision based automated plant disease diagnostic system which utilizes machine learning techniques to correctly identify disease and healthy plants also the type of disease. For achieving so deep learning network for images like Convolution neural network (CNN) can be utilized. CNN is used to extract features from images i.e. horizontal edges, vertical edges, RGB values etc. CNN is best deep learning neural network for visual feature extraction [8]. The CNN based network can be trained for detecting disease in plants by providing huge amount of images of healthy and sick plants and trained model in future can be used to predict the disease in plants by images of plants leaves.

II. METHODOLOGY

Deep learning is powerful machine learning approach which have mitigated the traditional machine learning headache of feature engineering. It doesn't need any domain expertise now and all credit goes to deep learning. The core of deep learning is artificial neural network (ANN). Artificial neural networks are mathematical models that replicate with their neurons and synapses interconnecting them the general principles of brain function [9].To implement neural network one of the most standard library is Tensorflow

Revised Manuscript Received on April 21, 2020.

***** Correspondence Author

Murk Chohan*, Department of Computer Science, Sukkur IBA University, Pakistan. Email: murk.mcs18@iba-suk.edu.pk

Adil Khan, Department of Computer Science, Sukkur IBA University, Pakistan. Email: adil.khan.cs13@gmail.com

Rozina Chohan, Department of Computer Science, Shah Abdul Latif University, Khairpur, Pakistan. Email: rozinachohan@yahoo.com

Saif Hassan Katpar, Department of Computer Science, Sukkur IBA University, Pakistan. Email: saifhassankatper@gmail.com

Muhammad Saleem Mahar, Department of Computer Science, Sukkur IBA University, Pakistan. Email: msaleem.mcs18@iba-suk.edu.pk

[10]. It provides all libraries related to artificial neural network. With the help of Tensorflow one can perform classification tasks on text as well as images.

A. Convolution Neural Network (CNN)

Convolution Neural Networks (CNNs) [11] are used to detect the disease in plant's leaves. CNN is an evolution of simple ANN that gives better result on images. Because images contains repeating patterns of particular thing (any image). Two important functions of CNN are convolution and pooling. Convolution is used to detect edges of patterns in an image and pooling is used to reduce the size of an image. CNN architectures that were applied on a problem are following: (a) Simple CNN, (b) VGG [12], and (iii) InceptionV3. Moreover training of these models are done using Jupyter notebook and Keras API of Tensorflow. Keras is tensor flow's high level API for building and training deep learning models.

B. Dataset Discussion

Two datasets are used to perform plant disease detection. First dataset consists of 15 classes and second one consists of 38 classes. Both databases have number of images of each plant. First dataset have total 2952 images. Final findings of this work is on PlantVillage dataset which contains 38 classes of different plants. It is also openly available on internet. Description of these classes and dataset is given in following Table- I (a) and (b).

Table- I (a): Dataset Description

Class	Plant Name	Healthy or Diseased	Disease Name	Images (Number)
C_0	Apple	Diseased	Apple_scab	2016
C_1	Apple	Diseased	Black_rot	1987
C_2	Apple	Diseased	Cedar_apple_rust	1760
C_3	Apple	Healthy	-	2008
C_4	Blueberry	Diseased	-	1816
C_5	Cherry_(including_sour)	Diseased	Powdery_mildew	1683
C_6	Cherry_(including_sour)	Healthy	-	1826
C_7	Corn_(maize)	Diseased	Cercospora_leaf_spotGray_leaf_spot	1642
C_8	Corn_(maize)	Diseased	Common_rust	1907
C_9	Corn_(maize)	Diseased	Northern_Leaf_Blight	1908
C_10	Corn_(maize)	Healthy	-	1859
C_11	Grape	Diseased	Black_rot	1888
C_12	Grape	Diseased	Esca_(Black_Measles)	1920
C_13	Grape	Diseased	Leaf_blight (Isariopsis_Leaf_Spot)	1722
C_14	Grape	Healthy	-	1692
C_15	Orange	Diseased	Haunglongbing (Citrus_greening)	2010
C_16	Peach	Diseased	Bacterial_spot	1838
C_17	Peach	Healthy	-	1728
C_18	Pepper_bell	Diseased	Bacterial_spot	1913
C_19	Pepper_bell	Healthy	-	1988
C_20	Potato	Diseased	Early_blight	1939
C_21	Potato	Diseased	Late_blight	1939
C_22	Potato	Healthy	-	1824
C_23	Raspberry	Healthy	-	1781
C_24	Soybean	Healthy	-	2022
C_25	Squash	Diseased	Powdery_mildew	1736
C_26	Strawberry	Diseased	Leaf_scorch	1774
C_27	Strawberry	Healthy	-	1824
C_28	Tomato	Diseased	Bacterial_spot	1702

By using this table you can come to know number of images in each class. Each class contains approximately 2000 images. Fourteen different plants are available in this dataset. For every plant healthy as well as diseased images of leaves are available. Most of the images belongs to Tomato and Apple plants. Least images are from Raspberry, Soybean, and Squash class. Below image show some images of different leaves which are available in dataset.

Table- I (b): Dataset Description

Class	Plant Name	Healthy or Diseased	Disease Name	Images (Number)
C_29	Tomato	Diseased	Early_blight	1920
C_30	Tomato	Diseased	Late_blight	1851
C_31	Tomato	Diseased	Leaf_Mold	1882
C_32	Tomato	Diseased	Septoria_leaf_spot	1745
C_33	Tomato	Diseased	Spider_mites Two-spotted_spider_mite	1741
C_34	Tomato	Diseased	Target_Spot	1827
C_35	Tomato	Diseased	Tomato_Yellow_Leaf_Curl_Virus	1961
C_36	Tomato	Diseased	Tomato_mosaic_virus	1790
C_37	Tomato	Healthy	-	1926
Total				**70295**

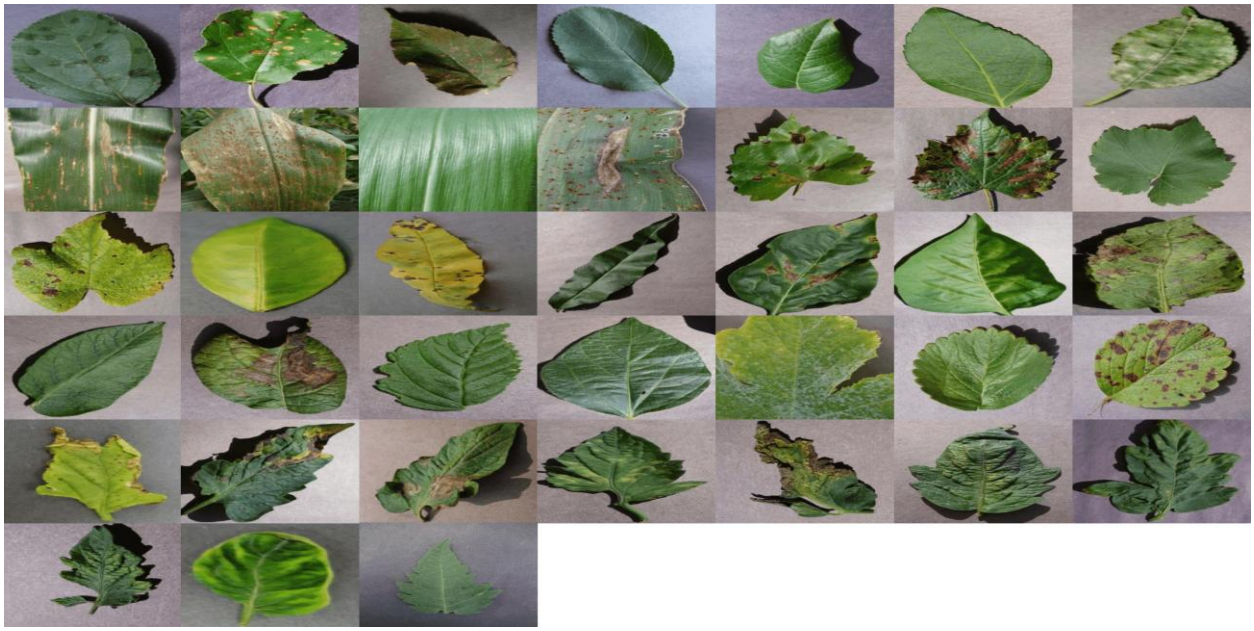

Fig. 1: Sample of Images from PlantVillage Dataset

Dataset is divided into two parts one for training and other for Testing. Splitting of dataset is 80/20 ratio randomly. 80% for the training dataset and rest 20% for testing dataset. Training dataset consists 56,236 image and testing consists 14,059 images. Training of model is done using 56,236 images and 14,056 images were kept unseen by model so that accuracy of model can be checked.

C. Model Description

First some preprocessing is applied on dataset in form of augmentation to increase size of dataset in order to achieve better accuracy. Then images size are reduced by 256x256 pixels. After that a convolution neural network based model will be created with multiple pooling and convolution layers and a dense layer for prediction. Five convolution layers with 3x3 filter are used and five MaxPooling2D layers with 2x2 filter. Batch Normalization is also used in this model. Batch normalization is used to scale data on particular scale but the difference is that it not just does it on input layer but it also do it at other hidden layers. At last model is trained on PlantVillage dataset.

Table- II: CNN Training Parameters

Parameter	Value
Epochs	25
Batch Size	32
Learning Rate	1e-1
Activation in middle layers	Relu
Activation in Final layer	Softmax

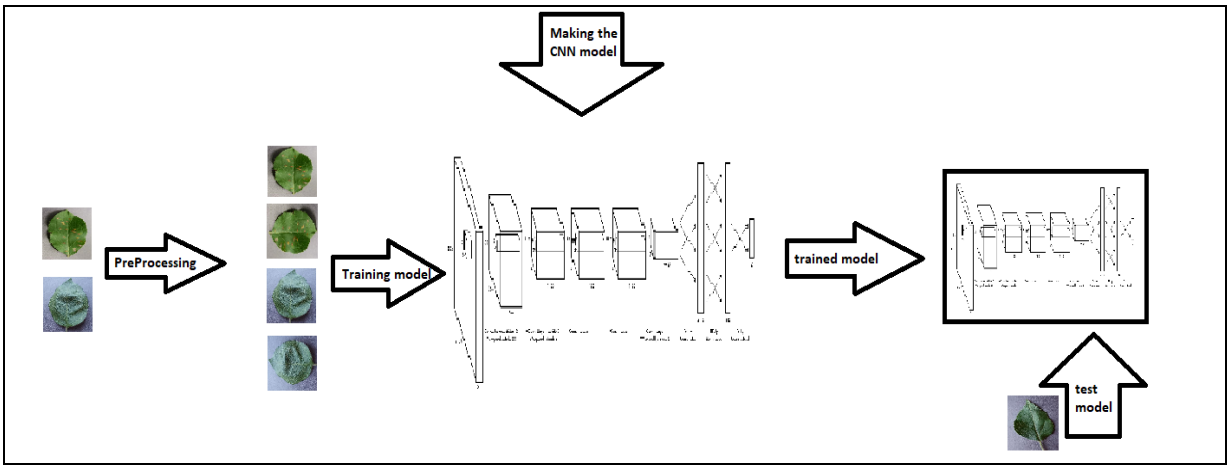

Fig. 2: Applied Methodology

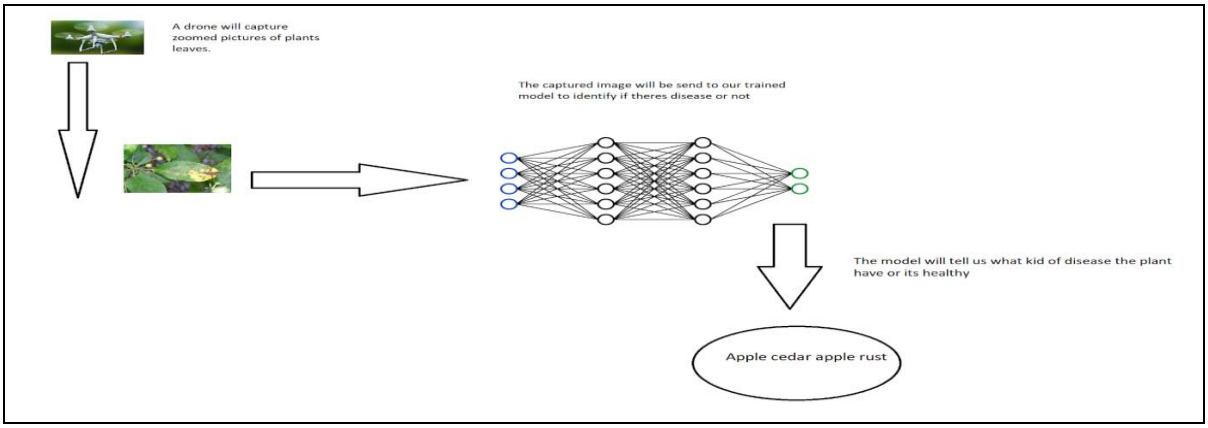

Fig. 3: Testing of an Image

III. RESULT AND DISCUSSION

This study shows the importance of plant disease detection in these days. This model were developed using Deep Learning in python. 20% (14,059) images from PlantVillage dataset were used to test the accuracy of this model. These images are from 38 different classes. 20% of each class randomly selected for testing. Some real time images were also used. Those images were captured from local environment. They do not belong to any class which are present in dataset. But model give us more than 95% accuracy on those images as well by telling either leaf is healthy of unhealthy. Total 100 images were used and 96 were classified correctly. Some images were captures at night with the help of flash light and some images have dirt upon it so that they were misclassified. Some of the images which we captured from local environment.

Unhealthy

Unhealthy

Fig. 4: Locally Captured Images

Testing dataset gives accuracy more than 98%. It means 1379 images from 14,059 images were classified correctly by model. Below is the Training and Validation accuracy graph generated by our model on testing dataset.

Unhealthy

Healthy

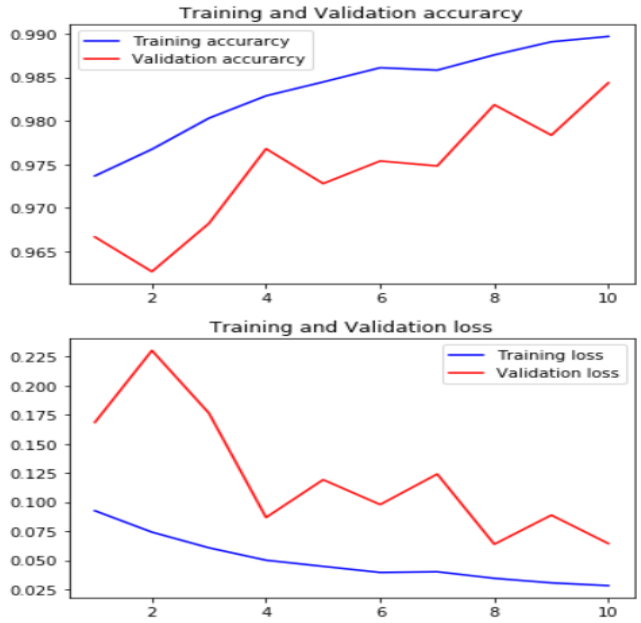

Fig. 5: Training and Validation Accuracy on Testing Dataset

Classes like Corn_(maize) healthy, Tomato Tomato_mosaic_virus, Strawberry healthy, and Corn_(maize) Common_rust give approximately 100% accuracy. Just 1 to 5 images misclassified from them. Apple Cedar_apple_rust,Cherry_(including_sour) healthy, Grape Black_rot, and Raspberry healthy classes gives less accuracy than other classes. Below is the table of image classes which gives better accuracy while testing on our model.

Table- III: Image Classes with Better Accuracy

Class	Total	Correct Classified	Mis-classified
Corn_(maize)___healthy	186	186	0
Tomato___Tomato_mosaic_virus	179	176	3
Strawberry___healthy	182	175	7
Corn_(maize)___Common_rust_	191	188	3
Corn_(maize)___Northern_Leaf_Blight	191	175	16

Table- IV: Final Model Performance

Model	Dataset for Training	Dataset for Testing	Training Accuracy	Testing Accuracy
CNN	PlantVillage (80%)	PlantVillage (20%)	99%	98%+
CNN	PlantVillage (80%)	Actual Environment (100 Images)	99%	95%+

IV. CONCLUSION AND FUTURE WORK

This study has utilized deep learning capabilities to achieve automatic plant disease detection system. This system is based on a simple classification mechanism which exploits the feature extraction functionalities of CNN. For prediction finally, the model utilizes the fully connected layers. The research was carried out using the publically accessible collection of 70295 images, and 100 images from experimental conditions and actual environment. The system has achieved an overall 98% testing accuracy on publically accessible dataset, and performed well on images of Sukkur IBA University plants. It is concluded from accuracy that CNN is highly suitable for automatic detection and diagnosis of plants. This system can be integrated into mini-drones to live detection of diseases from plants in cultivated areas. Though this system is trained on Plant Village dataset with only 38 classes it could tell if the plant has a disease or not as somehow symptoms are same in all kinds of plants. In addition, more actual environment images can be added to the dataset to improve the accuracy on real-condition images of leaves and classify more plant types as well as disease types. In the future, this system can also adopt 3 layer approach where the first layer detects if there's any plant in an image or not, second layer tells the plant type and the third layer tells if there is any disease or not and what type of disease is there if any.

REFERENCES

1. Huntley, B. (1991). How plants respond to climate change: migration rates, individualism and the consequences for plant communities. *Annals of Botany*, 15-22.
2. Fang, Y., & Ramasamy, R. P. (2015). Current and prospective methods for plant disease detection. *Biosensors*, *5*(3), 537-561.
3. Mustafa, M. S., Husin, Z., Tan, W. K., Mavi, M. F., & Farook, R. S. M. Development of automated hybrid intelligent system for herbs plant classification and early herbs plant disease detection. *Neural Computing and Applications*, 1-23.
4. Williams, S. D. (2017, February 1). Plants Get Sick Too! Retrieved from https://ohioline.osu.edu/factsheet/plpath-gen-1
5. Isleib, J., & Michigan State University. (2018, October 2). Signs and symptoms of plant disease: Is if fungal, viral or bacterial? Retrieved from https://www.canr.msu.edu/news/signs_and_symptoms_of_plant_disease_is_it_fungal_viral_or_bacterial
6. Shruthi, U., Nagaveni, V., & Raghavendra, B. K. (2019, March). A Review on Machine Learning Classification Techniques for Plant Disease Detection. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)* (pp. 281-284). IEEE.
7. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.
8. Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 806-813).
9. Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*, 311-318.
10. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)* (pp. 265-283).
11. LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), 1995
12. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

AUTHORS PROFILE

Murk Chohan Completed her BE Computer System Engineering from Quaid e Awam University of Engineering & Technology Nawabshah in 207 and received her MS Computer Science degree specialized in Data Knowledge Engineering in 2019 from Sukkur IBA University. He has been researching in the field of Deep Learning, Software Engineering, and Cloud Computing. Currently she is a faculty member of Computer Science department at IBA-IET Khairpur. Pakistan.

Adil Khan is a faculty member of Computer Science department at IBA Community College Ubauro, Pakistan. He completed his BS Computer Science in 2017 and MS Computer Science specialized in Data Knowledge Engineering from Sukkur IBA University in 2019, both with distinction (Gold Medalist). He has been researching in the field of Deep Learning, Software Engineering, and Cloud Computing.

Rozina Chohan is a faculty member of Computer Science Department at Shah Abdul Latif University, Khairpur, Pakistan. She has earned Doctor of Philosophy (Ph.D.) in Information Security Identity Theft from UCLan University, United Kingdom. She obtained her MS degree from University of Northampton, United Kingdom. She has been mainly researching in the field of Cyber Security, Information Security, and Deep Learning.

Saif Hassan Katper received the B.S degree in Computer Science from Sukkur IBA University, Pakistan in 2016 and M.S degree in Computer Science from Mohammad Ali Jinnah University, Karachi, Pakistan (Gold Medalist in MS). He is currently teaching at Sukkur IBA University. He has teaching experience of more than 4 years. His specialization areas include Deep Learning, Computer Vision. Mr. Saif's awards include PM ICT (R&D) Fund Scholarship for Four Years B.S Program, Gold Medalist in Masters, online certifications and many appreciation awards and certificates for conducting workshops/seminars.

Muhammad Saleem obtained BS Computer Science degree from University of Sindh, Jamshoro. Currently he is enrolled in MS Computer Science in Sukkur IBA University. He is faculty member of Computer Science Department at IBA Community College, Jacobabad, Pakistan with over 6 years of experience.