AN INNOVATIVE METHOD TO ANALYZE THE PREDICTION RATE AND ACCURACY FOR IDENTIFICATION OF PLANT LEAF DISEASE WITH CONVOLUTIONAL NEURAL NETWORK OVER K-NEAREST NEIGHBOR ALGORITHM

L. Lakshmi Narendra M[1], K. Malathi[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The primary goal of this study paper is to identify plant leaf disease to discover the greatest accuracy using K-Nearest Neighbors (KNN) and an Innovative Convolutional Neural Network (CNN).
Methods and Materials: The data set in this present work utilizes the publicly available Kaggle data set for plant leaf disease detection. The sample size of classification of leaf disease detection with improved accuracy rate was sample 80 (Group 1=40 and Group 2 =40) and calculation is carried out utilizing G-power of 0.8 with alpha and beta qualities are 0.05, 0.2 with a confidence interval at 95%. Accuracy is performed with the dataset from the Kaggle. The two groups are K-Nearest Neighbors (N=10) and Convolutional Neural Network algorithms (N=10).
Results: An Innovative Convolution Neural Network (CNN) is preferred for the Identification of Plant Leaf Disease. The accuracy is analyzed based on two algorithms, in which CNN reported accuracy of 93.68% and KNN reported accuracy of 81.31% for plant leaf disease detection. The two algorithms CNN and KNN are statistically satisfied with the independent sample T-Test (α=.001) value ($p<0.05$) i.e. p=0.0321
Conclusion: Identification of Plant Leaf Disease significantly seems to be better in CNN than KNN.

Keywords: Innovative Conventional Neural Network (CNN), K-Nearest Neighbor (KNN), Deep Learning, Detection, Image processing.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, 602105.
[2*]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, 602105.

1. Introduction

The goal of the research paper is to detect disease in plants, which is crucial. Nowadays, there is a minimum application-based system yet to detect the disease in plants. The previous approach for detecting plant diseases is straightforward with human eye monitoring by experts in which case identification and detection of plant disease occurred (Jasim and AL-Tuwaijari 2020). This system uses image processing techniques for identification of plant leaf disease. The agricultural landmass is quite simply being a food source supply in today's world. Agriculture productivity is critically important to the Indian economy. (Boulent et al. 2019). Crop leaf disease is a substantial danger to food security,but their rapid identification remains difficult in several plant elements due to a lack of required infrastructure. To beat the foremost drawback and increase the economic rate, people have a tendency to use strategies of plant disease detection (Mohanty, Hughes, and Salathé 2016). Plant disease detection is one of the numerous agricultural applications required to increase production. According to the latest research, WMSN using camera capacity has been made to capture the morbid leaves in the farm and utilize it to obtain the solutions (Aasha Nandhini et al. 2017). Plant leaf disease identification is carried out by researchers in order to enhance business. Totally IEEE published 20 linked articles, while Google Scholar published 6 associated articles like ResearchGate, Sciencedirect (Ngugi, Abelwahab, and Abo-Zahhad 2021) .

Identification of plant disease exploitation innovative Convolution Neural Network that tells the way to realize the assorted diseases that have an effect on plant leaves. During this system, the accuracy of various diseases is 99.7%. Conventional Neural Network (CNN) model leads to scan of the photographs and will significantly influence the economical identification of the diseases, should have the capability of detecting disease in real-time agricultural systems with an accuracy of 98.45% (Hassan et al. 2021) . Innovative Convolution network to detect plant diseases as depicted by images of healthy plant leaves and a plant infected with a disease leaves with an of 96.34% (Shrestha et al. 2020). Image processing techniques can be used to detect plant diseases, Image acquisition, image pre-processing, image segmentation, feature extraction, and classification are all steps in disease detection of results 91.27% (Agarwal et al. 2020) .

Our institution is passionate about high quality evidence based research and has excelled in various domains (Vickram et al. 2022; Bharathiraja et al. 2022; Kale et al. 2022; Sumathy et al. 2022; Thanigaivel et al. 2022; Ram et al. 2022; Jothi et al. 2022; Anupong et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Palanisamy et al. 2022). The main problem with this existing method is with a large data set.There are a few diseases that are detected,however this is insufficient for farmers or those involved in the agricultural sector.The Increase the number of diseases to make it more reliable for them, as well as the number of image datasets to achieve the best results.TheThe total number of photos is divided into two sections, one for training and one for testing, In general 70 percent the majority of the photos will be in the training section, while the rest will be in the test section. The proposed model would make a distinction among healthy and diseased leaves. Farmers can easily strengthen their growth momentum so that diseases do not spread across the civilized state.The two algorithms CNN and KNN analyzed among them and KNN has very little accuracy than CNN for detection of diseases. It is shown in a very low percentage while analyzing the disease image in the dataset. The study aims to improve the accuracy of the detection of disease, improve the correctness percentage of the recognized plant disease, and reduce the loss of data while training and testing the dataset.

2. Methods and Material

The proposed work's study setting is completed in the OOAD Laboratory, SIMATS, Saveetha School of Engineering, Department of Computer Science and Engineering, Chennai. sample size had been calculated by using clincalc.com by keeping G power (Kane, Phar, and BCPS) the calculation is performed utilizing G-power and the analysis's minimum power is set at 0.8,with alpha and the maximum allowable mistake is fixed beta qualities as 0.50 with threshold value as 0.05% and Confidence Interval is 95%. Mean and standard deviation has been calculated based on the previous literature for size calculation. The two groups are used, namely K-Nearest Neighbor (N=10) as an existing model as a group. 1 and Convolutional Neural Network (N=10) as a Proposed model as a group. 2.

Data Preparation

The Convolutional Neural Network is to find Plant Leaf disease images that have been stored in the kaggle dataset, to train and test through the dataset. The dataset contains 1000 data in the form of images taken as samples of various disease leaves from kaggle. There are 750 trained images and 250 tested images (Caldeira, Santiago,

and Teruel 2021). The sample image of the leaf present in the kaggle dataset has been shown in fig. 1.

K-Nearest Neighbor Algorithm

K-nearest neighbors (KNN) classifiers can classify Alternaria alternata, anthracnose, microorganism blight, leaf spot, and canker of various plant species as examples of plant diseases. (Asif, Rahman, and Hena 2020). It contains a variety of steps in image acquisition, the image pre-processing, the segmentation, the options extraction,and the classifier victimization KNN (Sharma 2021). K-Nearest Neighbor is the supervised Deep learning algorithm. Fig. 2 represents the working flow of the KNN, The algorithmic rule will be wont to solve each classification and regression drawback statement. the amount of nearest neighbors to a replacement unknown variable that should be foreseen or classified.

KNN algorithm steps

1. Give the input as an image with pixel 28*28.
2. Take input and crop the image to present in the code
3. Pre-process the image and find whether it is leaf disease or not.
4. If yes, then it will go to the tensor flow function and exit.
5. Once the image is in the tensor flow function it will pre-process the image.
6. After the preprocessing is done the images are well processed and produce the output image with the text format.

Convolution Neural Network (CNN)

Deep learning method in Conventional Neural Network (CNN) is the deep neural network with different activation functions and which is formed as a three-layer function that will produce the results. It is a deeper neural network there will be multiple sections which are the input layer, the hidden layer, and the output layer. (Boulent et al. 2019). Deep learning techniques, and in particular CNN, have junction rectifiers to important progress in the image process. CNN features can feed to a hidden layer followed by associate degree output classification layer exploitation soft-max and ReLU. The functioning of the Conventional Neural Network(CNN) has been delineated in Fig. 3. Soft-max is the performance that may manufacture the prices of the specific value that planned the value to find the accuracy. The Fig. 3 flow has been mentioned in the below steps.

CNN algorithm steps

1. Downloading the dataset to load
2. Initialize the variables to train and test the data.
3. Define a model. fit () function to describe the components which are to be accessed for running the code to get accuracy.
4. Define the Categorical () function to categorize the data.
5. Print the model. fit () function with the required epochs and find the accuracy.

For comparing both the models, the dataset has been trained with five different sample sizes. the accuracy values are recorded.

The system configuration is used for the algorithm to run in a 64 - bit Operating System, 4GB RAM PC, and using Windows 10, Google Colab, and Microsoft Office for software specification.

To estimate the training model's performance, the information has been separated for testing and training to verify the dataset load and reshape the information arrays to reason the photographs. Normalize the element values of grayscale pictures All the layers are operated through the ReLU activation to the specific cross-entropy to search out the loss function. The model is evaluated with the fit() operation that has the metrics to validate the accuracy of the information.

Statistical Analysis

SPSS is a statistical software package that is used for data analysis of the K-Nearest Neighbor (KNN) algorithm and the Innovative Convolutional Neural Network(CNN) algorithm. The independent sample t-test was used to calculate the mean, standard deviation, and standard error mean statistical significance between groups, and then the comparison of the two groups with the SPSS software.

3. Results

Conventional Neural Network (CNN) algorithm forms the layers with all the images of each number, whenever it runs at different times due to the initialization of sample size (N=10). The layers formed due to the iterations, the accuracy value changes with the duration of running time and produces the accuracy of the period which is shown in Table 1. Conventional Neural Network (CNN) has better accuracy the KNN due to the activation functions and metrics, the KNN algorithm has not compatible with the advanced activation functions that square measure solely restricted to the adam, adaleta, and degrade that takes longer, and therefore the functions don't seem to be taking the total information to research the similar disease within the dataset whereas the Conventional Neural Network (CNN) takes the information and forms layers with every leaf are

individual basis and eventually offers the result. Regarding the activation functions, the accuracy has been modified and has established that CNN is better than KNN.

Table. 1 represents the data collection from the N=20 samples of the dataset for CTC with the size of 28*28 pixels to gain accuracy (%) and CNN to gain accuracy (%) is calculated based on Eqn. 1.

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (1)$$

Where :

TP = True Positive
TN - True Negative
FP - False Positive
FN - False Negative

Loss: A scalar worth that endeavor to limit during the preparation of the model. The lower the misfortune, the nearer the expectations are to the genuine names.

The IBM SPSS version 21 statistical software is used for the study. Shape and size are independent variables, and size is the dependent variable accuracy (%). Present work-study Identification of Plant Leaf Disease.

The datasets are prepared in SPSS using ware N=10 as the number of samples for, K-Nearest Neighbors and Convolution Neural Network. The group ID is provided as a grouping variable, and accuracy is provided as a testing variable. The group ID is given as 1 for KNN and group ID. 2 for CNN. Group Statistics is applied for the Statistical Package for the Social Sciences (SPSS) dataset and shown in Table. 2. By performing the statistical analysis group statistics represent the comparison of the accuracy of Identification of Plant Leaf Disease of KNN and CNN. The CNN algorithm had the highest accuracy (92.67). KNN had the lowest accuracy (89.07) in Table. 2.

Table. 3. represents the Independent Sample T-Test is being used for the sample collections by fixing the level of significance as 0.005 with a confidence interval of 95 %. After applying the SPSS calculation, KNN has accepted a statistically significant value ($p<0.05$). From Fig. 4 it was represented by a simple bar Mean of Accuracy KNN error range (0.82 - 0.86) and CNN error range (0.92 - 0.94).

4. Discussion

The overall results that are shown in some variations in the accuracy values due to the advancements of the activation functions which proved that the Innovative Convolutional Neural Network with an accuracy of 93.68% is better than the K- Nearest Neighbors with an accuracy of 81.31% in identification of plant leaf disease. There is a statistically significant difference in Plant Leaf Disease using image accuracy of two algorithms having the significant accuracy value of 0.001 ($p<0.005$ T-Test for Independent Samples).

Accuracy for the Identification of Plant plant disease with totally different optimizers used as activation functions to optimize the performance with the accuracy of 89.99%. Illness-affected pictures are taken once training the model to hold out varied steps like pre-processing, segmentation, feature extraction, feature fusion to mix all options, and classification of illness. For the classification task here used progressive Deep learning algorithms. A big action to the planned model is exploitation Restricted Boltzmann Machine deep learning for feature extraction to the illness image with the best accuracy of 97.4%. The Conventional Neural Network (CNN) classification algorithmic program is employed to extract the options, and may simply get the correct category with the accuracy of 98.59%. However, the planned model, the CNN algorithmic program is employed for extracting the feature as similar and in addition to searching out the structure with all the information within the dataset of every leaf. The terms feature extraction and account the structure square measure similarly (Misra and Wu 2020).

Currently, out there datasets don't contain pictures gathered and tagged from real-life things. Another limitation is that the proposed strategies cannot currently observe multiple diseases in one image or multiple occurrences of a similar disease in one image. Therefore, careful thought ought to tend once selecting among these three networks on the condition that every design offers some blessings and suffers some limitations (Arsenovic et al. 2019).

In the future, can able to improve this classification for any development in Deep Convolutional Neural Network and also the applications of some massive complicated blatant knowledge. They improve this technique to acknowledge the assorted morbid leaf pictures. Systems area unit to be developed nonetheless to investigate massive areas of pictures.

5. Conclusion

In this research, Identification of Plant Leaf Disease performed using the image dataset seems to be better accuracy (93.67%) using an innovative Convolution Neural Network than K-Nearest Neighbors (81.07%). The clarity of Identification of Plant Leaf Disease is found with good accuracy.

DECLARATIONS

Conflict of Interests

This manuscript contains no conflicts of interest.

Authors Contributions

Author LLN was involved in conceptualization, data collection, data analysis, manuscript writing. Author KM was involved in conceptualization, guidance, and critical review of the manuscript.

Acknowledgments

The authors would like to thank Saveetha School of Engineering and Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the infrastructure required to complete this work successfully.

Funding

Thanks to the following organizations for their financial assistance, which allowed us to complete the study.

1.Chipontime Technologies Pvt.Ltd, Bangalore
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Aasha Nandhini, S., R. Hemalatha, S. Radha, and K. Indumathi. 2017. "Web Enabled Plant Disease Detection System for Agricultural Applications Using WMSN." Wireless Personal Communications 102 (2): 725–40.

Agarwal, Mohit, Abhishek Singh, Siddhartha Arjaria, Amit Sinha, and Suneet Gupta. 2020. "ToLeD: Tomato Leaf Disease Detection Using Convolution Neural Network." Procedia Computer Science. https://doi.org/10.1016/j.procs.2020.03.225.

Anupong, Wongchai, Lin Yi-Chia, Mukta Jagdish, Ravi Kumar, P. D. Selvam, R. Saravanakumar, and Dharmesh Dhabliya. 2022. "Hybrid Distributed Energy Sources Providing Climate Security to the Agriculture Environment and Enhancing the Yield." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102142.

Arsenovic, Marko, Mirjana Karanovic, Srdjan Sladojevic, Andras Anderla, and Darko Stefanovic. 2019. "Solving Current Limitations of Deep Learning Based Approaches for Plant Disease Detection." Symmetry. https://doi.org/10.3390/sym11070939.

Asif, Md Khalid Rayhan, Md Asfaqur Rahman, and Most Hasna Hena. 2020. "CNN Based Disease Detection Approach on Potato Leaves." 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). https://doi.org/10.1109/iciss49785.2020.9316021.

Bharathiraja, B., J. Jayamuthunagai, R. Sreejith, J. Iyyappan, and R. Praveenkumar. 2022. "Techno Economic Analysis of Malic Acid Production Using Crude Glycerol Derived from Waste Cooking Oil." Bioresource Technology 351 (May): 126956.

Boulent, Justine, Samuel Foucher, Jérôme Théau, and Pierre-Luc St-Charles. 2019. "Convolutional Neural Networks for the Automatic Identification of Plant Diseases." Frontiers in Plant Science 0. https://doi.org/10.3389/fpls.2019.00941.

Caldeira, Rafael Faria, Wesley Esdras Santiago, and Barbara Teruel. 2021. "Identification of Cotton Leaf Lesions Using Deep Learning Techniques." Sensors 21 (9). https://doi.org/10.3390/s21093169.

Hassan, Sk Mahmudul, Arnab Kumar Maji, Michał Jasiński, Zbigniew Leonowicz, and Elżbieta Jasińska. 2021. "Identification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach." Electronics. https://doi.org/10.3390/electronics10121388.

Jasim, Marwan Adnan, and Jamal Mustafa AL-Tuwaijari. 2020. "Plant Leaf Diseases Detection and Classification Using Image Processing and Deep Learning Techniques." 2020 International Conference on Computer Science and Software Engineering (CSASE). https://doi.org/10.1109/csase48920.2020.9142097.

Jothi, K. Jeeva, K. Jeeva Jothi, S. Balachandran, K. Mohanraj, N. Prakash, A. Subhasri, P. Santhana Gopala Krishnan, and K. Palanivelu. 2022. "Fabrications of Hybrid Polyurethane-Pd Doped ZrO2 Smart Carriers for Self-Healing High Corrosion Protective Coatings." Environmental Research. https://doi.org/10.1016/j.envres.2022.113095.

Kale, Vaibhav Namdev, J. Rajesh, T. Maiyalagan, Chang Woo Lee, and R. M. Gnanamuthu. 2022. "Fabrication of Ni–Mg–Ag Alloy Electrodeposited Material on the Aluminium Surface Using Anodizing Technique and Their Enhanced Corrosion Resistance for Engineering Application." Materials Chemistry and Physics. https://doi.org/10.1016/j.matchemphys.2022.125900.

Misra, Siddharth, and Yaokun Wu. 2020. "Machine Learning Assisted Segmentation of Scanning Electron Microscopy Images of Organic-Rich Shales with Feature Extraction and Feature Ranking." Machine Learning for Subsurface Characterization. https://doi.org/10.1016/b978-0-12-817736-5.00010-7.

Mohanty, Sharada P., David P. Hughes, and Marcel Salathé. 2016. "Using Deep Learning for Image-Based Plant Disease Detection." Frontiers in Plant Science 0. https://doi.org/10.3389/fpls.2016.01419.

Ngugi, Lawrence C., Moataz Abelwahab, and Mohammed Abo-Zahhad. 2021. "Recent Advances in Image Processing Techniques for Automated Leaf Pest and Disease Recognition – A Review." Information Processing in Agriculture. https://doi.org/10.1016/j.inpa.2020.04.004.

Palanisamy, Rajkumar, Diwakar Karuppiah, Subadevi Rengapillai, Mozaffar Abdollahifar, Gnanamuthu Ramasamy, Fu-Ming Wang, Wei-Ren Liu, Kumar Ponnuchamy, Joongpyo Shim, and Sivakumar Marimuthu. 2022. "A Reign of Bio-Mass Derived Carbon with the Synergy of Energy Storage and Biomedical Applications." Journal of Energy Storage. https://doi.org/10.1016/j.est.2022.104422.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102102.

Sharma, Shivam. 2021. "KNN - The Distance Based Machine Learning Algorithm." May 15, 2021. https://www.analyticsvidhya.com/blog/2021/05/knn-the-distance-based-machine-learning-algorithm/.

Shrestha, Garima, Deepsikha, Majolica Das, and Naiwrita Dey. 2020. "Plant Disease Detection Using CNN." 2020 IEEE Applied Signal Processing Conference (ASPCON). https://doi.org/10.1109/aspcon49795.2020.9276722.

Sumathy, B., Anand Kumar, D. Sungeetha, Arshad Hashmi, Ankur Saxena, Piyush Kumar Shukla, and Stephen Jeswinde Nuagah. 2022. "Machine Learning Technique to Detect and Classify Mental Illness on Social Media Using Lexicon-Based Recommender System." Computational Intelligence and Neuroscience 2022 (February): 5906797.

Thanigaivel, Sundaram, Sundaram Vickram, Nibedita Dey, Govindarajan Gulothungan, Ramasamy Subbaiya, Muthusamy Govarthanan, Natchimuthu Karmegam, and Woong Kim. 2022. "The Urge of Algal Biomass-Based Fuels for Environmental Sustainability against a Steady Tide of Biofuel Conflict Analysis: Is Third-Generation Algal Biorefinery a Boon?" Fuel. https://doi.org/10.1016/j.fuel.2022.123494.

Vickram, Sundaram, Karunakaran Rohini, Krishnan Anbarasu, Nibedita Dey, Palanivelu Jeyanthi, Sundaram Thanigaivel, Praveen Kumar Issac, and Jesu Arockiaraj. 2022. "Semenogelin, a Coagulum Macromolecule Monitoring Factor Involved in the First Step of Fertilization: A Prospective Review." International Journal of Biological Macromolecules 209 (Pt A): 951–62.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." Fuel. https://doi.org/10.1016/j.fuel.2022.123814.

Dey, N., Kumar, G., Vickram, A. S., Mohan, M., Singhania, R. R., Patel, A. K., ... & Ponnusamy, V. K. (2022). Nanotechnology-assisted production of value-added biopotent energy-yielding products from lignocellulosic biomass refinery–a review. Bioresource Technology, 344, 126171.

Tables and Figures

Table 1. Data collection from the N=10 samples of the dataset for KNN with the size of 28 * 28 pixels to gain accuracy (%) and CNN to gain accuracy(%).

Samples (N)	**K-Nearest Neighbors**	**Convolution Neural Network**
	Accuracy(%)	**Accuracy(%)**
1	82.01	93.06
2	80.02	93.79
3	86.02	92.80
4	74.35	93.61
5	76.80	94 .42
6	79.50	97.33
7	75.12	93.34
8	86.09	93.84
9	86.10	94.09
10	86.09	94.19

Table 2. Comparison of the accuracy of Identification of Plant Leaf Disease of KNN and CNN.CNN algorithm had the highest accuracy (93.68) . KNN had the lowest accuracy (81.31).

		N	**Mean**	**Std.Deviation**	**Std. Error Mean**
Accuracy	KNN	10	81.3110	4.73714	1.49801
	CNN	10	93.6850	1.4256	.48135

Table 3. The Independent Sample T-Test is used for the sample collections by fixing the level of significance as 0.05 with a confidence interval as 95 %. After applying the SPSS calculation, CNN has accepted a statistically significant value(P<0.05).

	Levene's Test for Equality of Variances		**T-Test for Equality of Means**						
	f	sig	t	df	Sig. (2-tiled)	Mean Difference	Std. Error Difference	95% Confidence Interval	
								Lower	Upper
ACCURACY Equal variances assumed	17.486	0.001	-7.90	18	000	-12.37200	1.56438	-15.6586	-9.085
Equal variance not assumed			-7.90	10.617	000	-12.37200	1.5648	-15.8304	-8.913

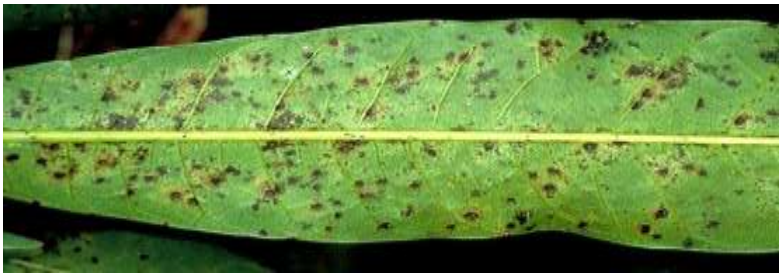

Fig. 1. Identification of Plant Leaf Disease

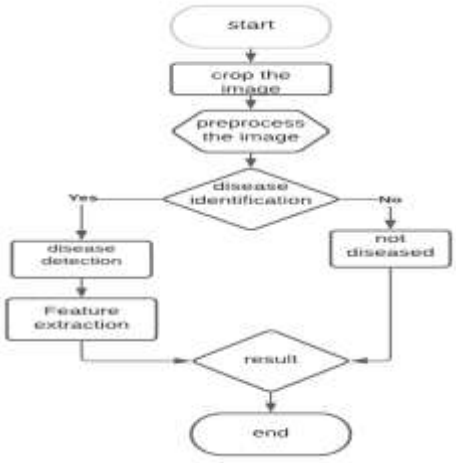

Fig.2. Flow chart of K-Nearest Neighbor

FLOWCHART

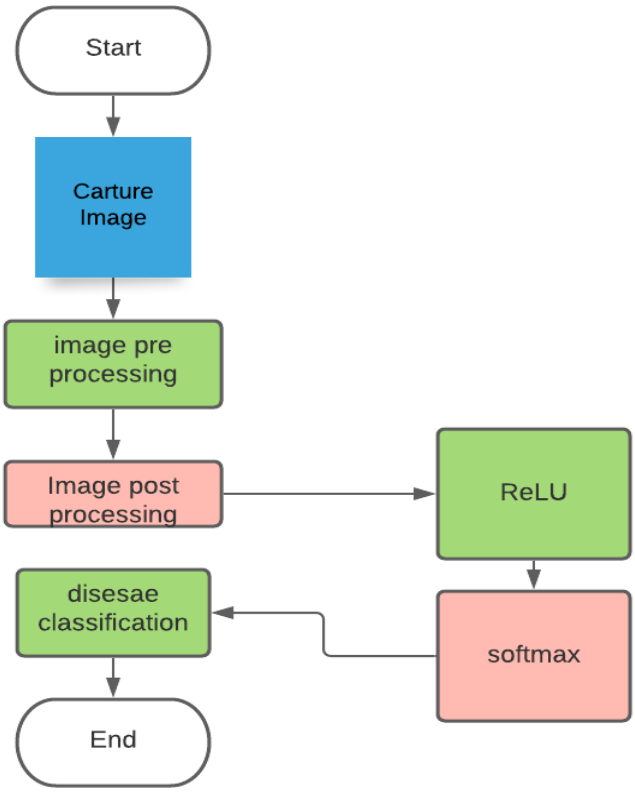

Fig.3. Flow chart of Convolution Neural Network

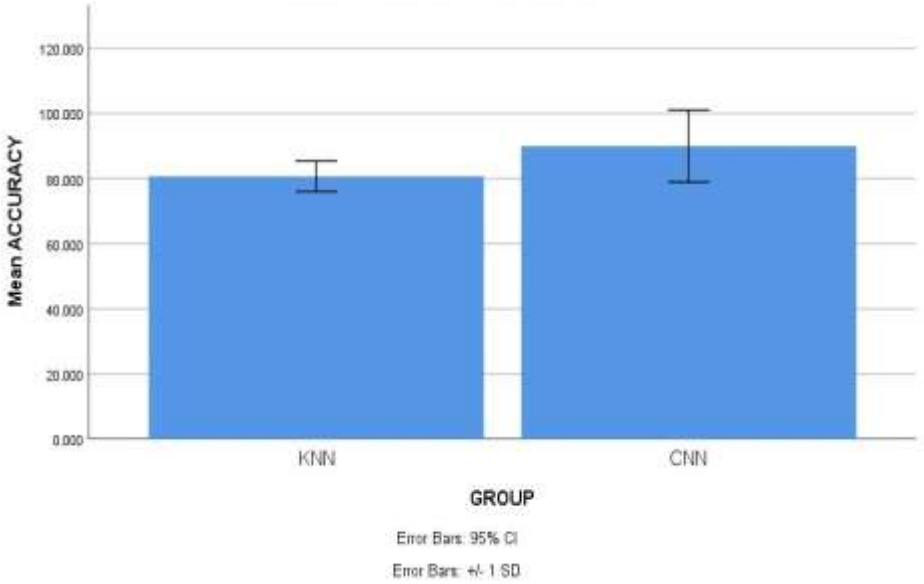

Fig. 4. Simple Bar Mean of Accuracy KNN Error Range (0.82 - 0.86) and CNN Error Range (0.92 - 0.94) with Mean accuracy of detection ± 1 SD.X Axis: KNN vs CNN Y-Axis: Mean accuracy of the detection ± 1 SD.