ISSN: 0974-5823 Vol. 7 (Special Issue, Jan.-Mar. 2022)

International Journal of Mechanical Engineering

Automatic Plant Leaf Disease Detection Using Convolutional Neural Network In Comparison With Support Vector Machine Classification

G. Manohar Reddy[1], Radhika Baskar[2]

[1]Research Scholar, Department of Electronics and Communication Engineering,
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105.

[2]Project Guide, Corresponding Author, Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105.

ABSTRACT

Aim: To perform leaf disease detection using Convolutional Neural Network (CNN) algorithm and comparing its accuracy with Support Vector Machine (SVM) algorithm. **Materials and Methods:** In this research work, the plant leaf disease detection has been carried out using machine learning algorithms such as SVM (N=10) and CNN (N=10) and the accuracy was determined for the same. Two groups are statistically analysed with the sample size 20 for both the group, with a pretest power of 80%. **Result:** The performance of the CNN algorithm is analyzed and it is observed that the accuracy is 97.09% for CNN and 91.8% for SVM algorithm. From the statistical analysis, it is observed that there is a significant difference between the two algorithms as $p < 0.05$. **Conclusion:** The result shows that CNN algorithm's accuracy was better than SVM algorithm for leaf disease detection.

Keywords: Machine Learning, CNN algorithm, SVM algorithm, Novel Classifier, Leaf disease, Accuracy.

INTRODUCTION

Leaf disease detection is important to keep crops healthy and sustained forever. By using machine learning techniques plant leaf disease can be detected. Efficient leaf disease prediction has been carried out using machine learning techniques such as K-Nearest Neighbour (KNN), CNN, a novel classifier and SVM [1]. In detecting plant diseases, the usual method of person analysis by the naked eye is a huge task. The model developed offers an accurate solution to the current issue. It takes less time, cost effective, and less manpower by these techniques. If it's once trained, the classifier can also be deployed into application, which is very easy to use. SVM is a supervised learning algorithm and it goes well on classification tasks or issues. The main application of this system is prediction of leaf disease can be done easily with real time experience [2]. [3] demonstrated that this system can be used in all areas for prediction and recommendation. The leaf disease detection work [3] proposed a survey and stated that its a reliable, safe and accurate system which was detected at every stage and used to avoid losses to the economy. As per the statista, the numbers have gradually risen every day for the last decade. Applications of leaf disease detection include precision agriculture, finding the amount of pesticides, and analysis of fertilizers used for various crops [4]. With these statements, it can be believed that applications are crucial and important in moulding the future farming.

Approximately 50 research articles have been published on leaf disease identification concept. The usage of CNN, novel classifier, in plant leaf disease identification has achieved best results in previous years. Due to present technology emergence of most best results, the layered network has been favourable among researchers [5]. In the paper, [6] Mrunalini represented the technique to identify the different diseases using CNN technique which has 95% accuracy. This approach gives feature set extraction as a color co-occurrence technique. A work proposed [7] for proposed CNN, SVM and NB and analysed by comparing those, found that CNN has achieved the best result with accuracy of 94% in classification of plant disease. Machine learning techniques such as KNN, CNN, SVM, a novel classifier can be utilized for effective analysis of plant leaf disease detection.Previously our team has a rich experience in working on various research projects across multiple disciplines[8]–[18]

From literature it is observed that there is a lack of plant disease detection using predictive algorithms. Studies claim that prediction of leaf disease accuracy was in the range of 94%. The aim of this research is to be carried out to get precise accuracy by implementing the CNN algorithm and comparing its features with SVM classification.

MATERIALS AND METHODS

This research is carried out in Image Processing Laboratory, Department of Electronics and Communication Engineering, Saveetha School of Engineering, SIMATS, Chennai. CNN, a novel classifier, is considered as the experimental group and SVM as the control group The sample dataset contains nearly 2500 images of 39 classes consisting of nearly 26 different crops plants. And 80% of the images are taken for training and 20% for testing [19] . For each algorithm 10 sample iteration and G power was calculated as actual power of 80% and alpha value set as 0.005.

Classification by CNN (Sample Preparation Group 1): This part states and explains the steps in [20]. Totally, twenty images were tested ranging from 100 x 100 to 255 x 255 [6]. With this process, the remaining layers are released. To get the proper tuning process, a plot which displays learning rate vs loss was taken, and analysed. Model Optimisation: To enhance the model's performance, additional settings of augmentation are to be added, for that operations included such as change in brightness, and warp. Next, the last two layers are trained at specified default training rate. Prediction: Finally for the interpretation, visualizations were made based on validation and test datasets. Additionally, the prediction function was created to predict the accurate disease.

SVM classification (Sample Preparation Group 2): Data Collection: The snapshots are taken from the datasets of plant Village dataset available from kaggle which are used to train the system, the sample pictures are in the form of PNG or JPG. All the pattern pictures are in the form of RGB. The picture set consists of healthy pictures and also diseases such as bacterial spot, black rot, mosaic virus, northern blight, rust etc. Preprocessing: Secondly, to create a directory for specific tremendous and poor pictures if those cannot exist in the given dataset into a variable, a characteristic is created to load all the folders consisting of pictures into arrays. So, preprocessing techniques are used to reduce or eliminate the past noise and to suppress the unwanted noise which is in images which takes place due to many reasons such as cam settings, light variants. Image Segmentation: This is a technique of dividing an image into the range of pixels up to their depth levels. And assigning labels to every pixel in that image such as those pixels with identical labels will give some characteristics [19]. Feature Extraction: The process of extracting the related records from all the pictures and transferring the information into a group of elements with their labels was taken as character extraction. Here, features were extracted based upon the color, shape, size, text. This is called Feature Extraction.SVM classification: The features generated are included into linear SVM which is a supervised computing learning algorithm that can be used for classification challenges [21]. These classifiers can assign the label to the photos and it gives which category it belongs to, from where the classifier is defined primarily based on feature.

For that testing and training has to be done and trainable parameters are calculated. Then the evaluation of the proposed model by a classifier and accuracy and loss are found. The CNN, a novel classifier, normally consists of nodes, and an arrow shows that it represents a connection from the output of each node. Each batch size is taken as 32 and the model has been trained for 25 epochs. At the initial stage, gaining knowledge of the rate set to 0.01. And the early stopping has been additionally used to monitor the loss and give the training as if it increases. Here all the 26 crops plants are checked by using CNN novel classifier and SVM classifier.

Statistical analysis

An analysis done for comparing both the CNN and SVM using the SPSS tool [22]. Two independent variables are taken and variable 1 is CNN algorithm and variable 2 is SVM algorithm. In that SPSS tool, the accuracy and loss of both the CNN and SVM algorithms has been taken and compared. Analysis is done and accuracy is plotted in a bar graph for comparison graphically. Finally, the descriptive statistics are applied for data in SPSS.

RESULTS

To get the proper tuning process, a plot which displays learning rate vs loss was taken, and analysed as shown in Fig. 1. From this system, a proper learning was selected, and the system model was run. With the achieved results recorded, the model file was again recreated to the additional four image sizes. All the processes remain constant in each trial in addition to the learning rate. With this, the fine tuning process, a plot of learning rate and loss is analysed. From this, a proper learning is taken. The result of this phase proves that it is easy to achieve an accuracy greater than 90% for the image sizes 155 x 155 to 255 x 255. From Table 1 it is observed that the accuracy is increased even as the image sizes increase. And From the SPSS tool, mean accuracy, loss and standard deviation are analysed as shown in Table 2, CNN has better accuracy than the SVM classifier and standard deviation also varies slightly and shows that CNN, a novel classifier, is more reliable for better results. An Independent sample t-test is done for CNN and SVM using the SPSS tool where the significant value is less than 0.05, and CNN has lesser significance error than SVM as shown in Table 3. The bar graph analysis done for the mean accuracy and loss of CNN (3.70) and SVM (4.51) using the SPSS tool which is shown in Fig. 2, describes CNN as having better prediction accuracy than SVM. Finally the plant leaf disease is predicted using the predict disease function created by taking the path from the dataset as shown in Fig. 3. The result can be for all 10 different crop plants by using CNN it gives 97.9% and by using SVM it gives 91.80%. The accuracy has been calculated using the Equation (1),

$$\text{Accuracy} = (TP+FP)/(TP+FP+TN+FN) \tag{1}$$

Where TP - True Positive, FP- False Positive, TN- True Positive, FN- False Negative.

DISCUSSION

From the analysis above, it is proved that CNN and SVM machines are different from each other. And also it is observed that CNN algorithm seems to have achieved more accuracy than SVM. CNN accuracy is significantly better than SVM accuracy with p-value less than 0.05.

Based on the previous research findings, it is observed that the CNN classification performance [23] can be improved in terms of accuracy, sensitivity and specificity if complexity-based features are considered (Accuracy of 97%). CNN achieves higher performance as it uses structural risk minimization principle [23]. The CNN classification performance can be improved with appropriate kernels [24]. The multiple function kernel provides better accuracy (89.24%) compared with CNN (81.61%) [25]. For the CNN classifier, the overall accuracy of the system is 95% compared to SVM (88%) [26]. Tomato and corn crops are checked by using CNN and SVM classifiers. The result for tomato crop by using SVM gives 60-70% and by CNN gives 80-85%. For corn, using CNN it gives 70-75% and by using SVM it gives 55-65%. [27]. Based on the similar and opposing findings, the CNN appears to be significantly better when compared to the SVM algorithm [28].

The achieved accuracy would depend on many factors such as stages of disease, types of disease, and background data. With the minimum dataset the proposed machine learning algorithm proved to be accurate in detection and classification of plant leaf disease using image processing. High dimensional data is also one of the limitations, this work can be further extended to the identification with improved accuracy using different deep learning techniques. This may certainly help all the users to avoid or get rid of such diseases in coming years, and these may also increase the crop yield.

CONCLUSION

In this research, a pre-trained CNN was properly tuned. Based on the performance analysis, CNN model provides the most accurate results compared to SVM based models. The accuracy of CNN algorithm (97.9%) based leaf disease detection appears to be better than SVM algorithm (91.80%).

DECLARATIONS

Conflict of interests

No conflict of interests in this manuscript.

Authors Contributions

Author GMR was involved in data collection, data analysis, manuscript writing. Author RB was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Opthought, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

[1] G. Prem, M. Hema, L. Basava, and A. Mathur, "Plant Disease Prediction using Machine Learning Algorithms," *International Journal of Computer Applications*, vol. 182, no. 25. pp. 1–7, 2018. doi: 10.5120/ijca2018918049.

[2] B. S. Prajwalgowda, M. A. Nisarga, M. Rachana, S. Shashank, and S. R. B. S, "Paddy Crop Disease Detection using Machine Learning," *International Journal of Engineering Research & Technology*, vol. 8, no. 13, Aug. 2020, Accessed: Mar. 22, 2021. [Online]. Available: https://www.ijert.org/paddy-crop-disease-detection-using-machine-learning

[3] S. Ashok, G. Kishore, V. Rajesh, S. Suchitra, S. G. Gino Sophia, and B. Pavithra, "Tomato Leaf Disease Detection Using Deep Learning Techniques," *2020 5th International Conference on Communication and Electronics Systems (ICCES)*. 2020. doi: 10.1109/icces48766.2020.9137986.

[4] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine Learning in Agriculture: A Review," *Sensors*, vol. 18, no. 8. p. 2674, 2018. doi: 10.3390/s18082674.

[5] Y. Toda and F. Okura, "How Convolutional Neural Networks Diagnose Plant Disease," *Plant Phenomics*,

vol. 2019. pp. 1–14, 2019. doi: 10.1155/2019/9237136.

[6] M. Vejmelka, P. Musilek, M. Paluš, and E. Pelikán, “K-MEANS CLUSTERING FOR PROBLEMS WITH PERIODIC ATTRIBUTES,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 04. pp. 721–743, 2009. doi: 10.1142/s0218001409007338.

[7] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, “A comparative study of fine-tuning deep learning models for plant disease identification,” *Comput. Electron. Agric.*, vol. 161, Mar. 2018, doi: 10.1016/j.compag.2018.03.032.

[8] D. Ezhilarasan, T. Lakshmi, M. Subha, V. Deepak Nallasamy, and S. Raghunandhakumar, “The ambiguous role of sirtuins in head and neck squamous cell carcinoma,” *Oral Dis.*, Feb. 2021, doi: 10.1111/odi.13798.

[9] R. Balachandar *et al.*, “Enriched pressmud vermicompost production with green manure plants using Eudrilus eugeniae,” *Bioresour. Technol.*, vol. 299, p. 122578, Mar. 2020.

[10] S. Muthukrishnan, H. Krishnaswamy, S. Thanikodi, D. Sundaresan, and V. Venkatraman, “Support vector machine for modelling and simulation of heat exchangers,” *Therm. Sci.*, vol. 24, no. 1 Part B, pp. 499–503, 2020.

[11] A. Kavarthapu and K. Gurumoorthy, “Linking chronic periodontitis and oral cancer: A review,” *Oral Oncol.*, p. 105375, Jun. 2021.

[12] S. C. Sarode, S. Gondivkar, G. S. Sarode, A. Gadbail, and M. Yuwanati, “Hybrid oral potentially malignant disorder: A neglected fact in oral submucous fibrosis,” *Oral Oncol.*, p. 105390, Jun. 2021.

[13] Hannah R, P. Ramani, WM Tilakaratne, G. Sukumaran, A. Ramasubramanian, and R. P. Krishnan, “Author response for ‘Critical appraisal of different triggering pathways for the pathobiology of pemphigus vulgaris—A review.’” Wiley, May 07, 2021. doi: 10.1111/odi.13937/v2/response1.

[14] D. Sekar, D. Nallaswamy, and G. Lakshmanan, “Decoding the functional role of long noncoding RNAs (lncRNAs) in hypertension progression,” *Hypertension research: official journal of the Japanese Society of Hypertension*, vol. 43, no. 7. pp. 724–725, Jul. 2020.

[15] P. Appavu, V. Ramanan M, J. Jayaraman, and H. Venu, “NOx emission reduction techniques in biodiesel-fuelled CI engine: a review,” *Australian Journal of Mechanical Engineering*, vol. 19, no. 2, pp. 210–220, Mar. 2021.

[16] S. Menon, H. Agarwal, S. Rajeshkumar, P. Jacquline Rosy, and V. K. Shanmugam, “Investigating the Antimicrobial Activities of the Biosynthesized Selenium Nanoparticles and Its Statistical Analysis,” *Bionanoscience*, vol. 10, no. 1, pp. 122–135, Mar. 2020.

[17] R. Gopalakrishnan, V. M. Sounthararajan, A. Mohan, and M. Tholkapiyan, “The strength and durability of fly ash and quarry dust light weight foam concrete,” *Materials Today: Proceedings*, vol. 22, pp. 1117–1124, Jan. 2020.

[18] V. R. Arun Prakash, J. F. Xavier, G. Ramesh, T. Maridurai, K. S. Kumar, and R. B. S. Raj, “Mechanical, thermal and fatigue behaviour of surface-treated novel Caryota urens fibre–reinforced epoxy composite,” *Biomass Conversion and Biorefinery*, Aug. 2020, doi: 10.1007/s13399-020-00938-0.

[19] A. John., D. Praveen Dominic, M. Adimoolam, and N. M. Balamurugan, “Share Market Data Prediction Strategies using Deep Learning Algorithm,” *Recent Advances in Computer Science and Communications*, vol. 13. 2019. doi: 10.2174/2666255813666191209093139.

[20] G. Shrestha, Deepsikha, M. Das, and N. Dey, “Plant Disease Detection Using CNN,” *2020 IEEE Applied Signal Processing Conference (ASPCON)*. 2020. doi: 10.1109/aspcon49795.2020.9276722.

[21] D. Das, M. Singh, S. S. Mohanty, and S. Chakravarty, “Leaf disease detection using support vector machine,” presented at the 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, Jul. 2020. doi: 10.1109/iccsp48568.2020.9182128.

[22] S. Landau, *A Handbook of Statistical Analysis Using SPSS*. Chapman & Hall/CRC, 2017.

[23] S. Chandaka, A. Chatterjee, and S. Munshi, “Cross-correlation aided support vector machine classifier for

classification of EEG signals," *Expert Systems with Applications*, vol. 36, no. 2. pp. 1329–1336, 2009. doi: 10.1016/j.eswa.2007.11.017.

[24] Y. Ma, X. Ding, Q. She, Z. Luo, T. Potter, and Y. Zhang, "Classification of Motor Imagery EEG Signals with Support Vector Machines and Particle Swarm Optimization," *Comput. Math. Methods Med.*, vol. 2016, p. 4941235, May 2016.

[25] Y. Bai and M. Tang, "Robust visual tracking via augmented kernel SVM," *Image and Vision Computing*, vol. 32, no. 8. pp. 465–475, 2014. doi: 10.1016/j.imavis.2014.04.008.

[26] M. H. Selamat, M. Hakeem Selamat, and H. M. Rais, "Image face recognition using Hybrid Multiclass SVM (HM-SVM)," *2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*. 2015. doi: 10.1109/ic3ina.2015.7377765.

[27] O. Asmae, R. Abdelhadi, C. Bouchaib, S. Sara, and K. Tajeddine, "Parkinson's Disease Identification using KNN and ANN Algorithms based on Voice Disorder," *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*. 2020. doi: 10.1109/iraset48871.2020.9092228.

[28] K. L. Perry, "An anomalous detection of Tomato yellow leaf curl virus in tomato in New York State," *Plant Disease*. 2021. doi: 10.1094/pdis-02-21-0356-pdn.

Tables and figures

Table 1. Training of images (4 epochs, max_LR=slice(1e-05,1e-04). As the image size increases, accuracy and in turn the time also increases.

Test	Image size	Train loss	Valid loss	Accuracy	Time (hrs)
1	155	0.1650	0.122	0.9555	2.83
2	195	0.1587	0.114	0.9567	3.54
3	224	0.1777	0.125	0.9753	4.25
4	255	0.1608	0.134	0.9846	5.43

Table 2. Descriptive Statistics of Accuracy and Loss for Plant leaf disease Detection using CNN and SVM. CNN got the highest accuracy (97.9%) and the lowest loss appeared (3.70%). The statistics of accuracy and loss using CNN and SVM by using the SPSS tool as shown.

	Groups	N	Mean	Std.Deviation	Std.Error Mean
Accuracy	CNN	10	97.90	0.731	0.231
	SVM	10	91.80	0.789	0.249
Loss	CNN	10	3.70	1.494	0.473
	SVM	10	4.51	2.670	0.844

Table 3. An independent sample t test of Accuracy and loss for detecting plant leaf disease using CNN and SVM. Statistically significant difference in two algorithms (p<0.05). CNN has appeared to have significantly lesser loss than the SVM algorithm.

		Lavene's test for equality of variances		T-test for Equality of Means					95% confidence interval of the difference	
		F	Sig	t	df	sig(2 tailed)	Mean diff	Std.error	Lower	Upper
Accuracy	Equal Variances assumed	0.279	0.044	15.196	18	0.000	5.168	.340	4.453	5.883
	Equal Variances not assumed			15.196	17.897	0.000	5.168	.340	4.453	5.883
Loss	Equal Variances assumed	1.743	0.203	.842	18	0.411	0.815	0.967	-1.218	2.848
	Equal Variances not assumed			.842	14.136	0.414	0.815	0.967	-1.258	2.888

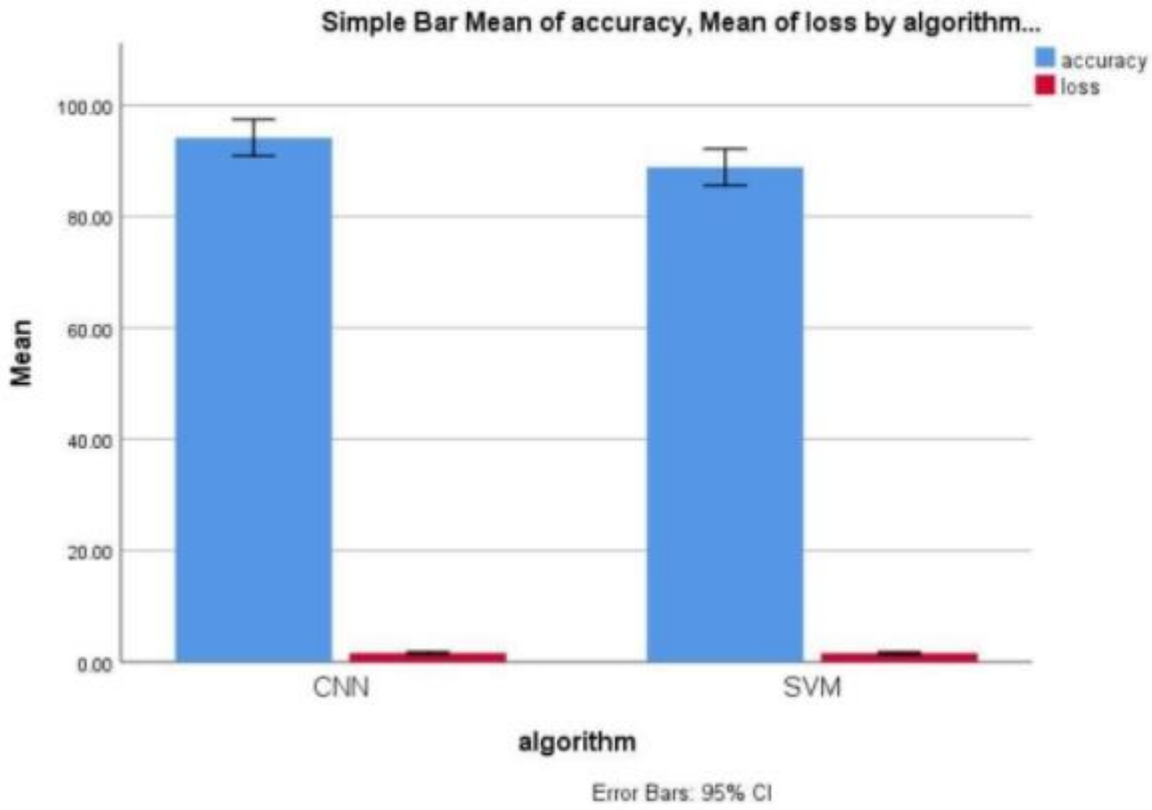

Fig. 1. Comparison of CNN algorithm and SVM classifier done in SPSS tool in terms of mean accuracy and loss. The accuracy of CNN is better than SVM and the standard deviation of CNN is slightly better and reliable than SVM. X axis: CNN vs SVM classifier, Y axis: Mean accuracy and loss. Error bars were represented as ± 1 SD.

Fig. 2. The line graph shows a) the training accuracy and validation accuracy b) training loss and validation loss of all 25 epochs.

Fig. 3. Predicting the disease of the leaf using predicted disease function. CNN has best results when compared to SVM technique.