Title page 1

Comparison of convolutional neural networks (CNNs) and support vector machines (SVMs) to maximise accuracy in identifying the tea leaf disease

V.Mahitha[1], K.Anbazhagan[2]

V.Mahitha[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. Pincode:602105.
vardhireddymahitha1440.sse@saveetha.com

K.Anbazhagan[2]
Research Guide, Corresponding Author,
Department of Computer Science and Engineering,
Saveetha School of Engineering,
Saveetha Institute of Medical And Technical Sciences,
Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

Keywords: agricultural productivity,improved nutrition, novel convolutional neural networks (CNNs), novel support vector machines (SVMs), deep learning,quality leaf

ABSTRACT:
Aim: The main goal of the study is to identify the tea leaf disease as accurately as possible by

comparing convolutional neural networks (CNNs) with support vector machines (SVMs). **Materials and Methods:** Convolutional neural networks (CNNs) (N = 20) and support vector machines (SVMs) (N = 20) are used in the detection of tea leaf illness. In this study,the final sample size calculation is done with a G-power pretest of 80 and a CI of 95%. The data is divided into training, validation, and test sets after dataset identification. Following the creation of a CNN architecture specifically designed for image classification, it is assembled using a suitable loss function and optimizer, trained on the training set, and then validated on a different validation set. SVM is a powerful machine learning algorithm that excels at classifying and categorising data points and is widely employed in diverse fields, including disease diagnosis in tea plantations. **Results:** identification of tea leaf disease detection by convolutional neural networks is 94%, and support vector machines is 76%. There's a huge distinction between the Convolutional Neural Network Technique (CNN) and support vector machines, with $p = 0.011$ ($p<0.05$). **Conclusion: A** convolutional neural network seems more accurate than supporting vector machines in identifying tea leaf disease.

Keywords:Agricultural productivity,IImproved nutrition, novel convolutional neural networks (CNNs), novel support vector machines (SVM), deep learning,quality leaf

INTRODUCTION

Tea, one of the most popular drinks in the world, is important to many countries' economies and cultures. However, a number of diseases that affect tea leaves can have a significant negative influence on the quantity and quality of tea crops and reduce agricultural productivity (Preedy(Preedy 2012). In order to maintain quality leaf tea output and guarantee the financial sustainability of tea estates, early detection and efficient control of these illnesses are essential. Technological and data analytics breakthroughs in recent years have transformed disease detection techniques, providing novel ways to track and identify illnesses of the tea leaf (Bao(Bao et al. 2022). The purpose of this introduction is to discuss the significance of disease detection in the tea industry and improved nutrition in tea leaves, emphasising the difficulties producers encounter and the opportunities presented by contemporary methods to overcome these difficulties ((Bao et al. 2022; Dai et al. 2020). Through the utilisation of cutting-edge technologies and interdisciplinary teamwork, scientists are working to create dependable, economical, and efficient ways to protect tea plants from illnesses, protect tea growers' livelihoods, and satisfy the increasing demand for premium tea products worldwide (Wang(Wang et al. 2021).

There are around 456 papers uploaded to IEEE resources, and 63 papers have been uploaded to Google Updated Scholar for the past 4 years. (Chatterjee et al. 2021)2021) have introduced a new model called a novel convolutional neural network (CNN) model for tea leaf disease detection with an approximate success rate of around 90%. (Venkata Krishna and Obaidat 2020) used support vector machines for disease detection and improvised models to give more accurate results than existing algorithms. (Dash et al. 2020) used the support vector machine algorithm to find the hidden patterns in the dataset to classify the data more efficiently. The maximum accuracy achieved was nearly 90%.(Wang et al. 2021))a pre-owned fuzzy method to identify tea leaf diseases. It is the combination of novel convolutional neural network algorithms with support vector machines, and the maximum accuracy achieved is around 94%.

All of the deep learning and machine learning models that have been around for a while perform less accurately in terms of tea leaf disease identification. Given this, the objective of the current paper is to detect activities with substantially better and more accurate outcomes utilising support vector machines and the Novel Convolutional Neural Network Algorithm. In comparison to previous models, these methods significantly improve the recognition of tea leaf illnesses by modifying the models and using a larger dataset with more parameters and various results (Preedy 2012). By utilising an enhanced, new convolutional neural network technique, the goal is to raise the accuracy rate of support vector machines for the diagnosis of tea leaf illness.

MATERIALS AND METHODS

The proposed project's research environment is accomplished in the Computer Vision Department, Saveetha School of Engineering, SIMATS, Chennai. The project has two groups, the first of which is the Convolutional Neural Technique (CNN) and the second of which is the support vector machines (SVMs). Using prior study results, the sample size was estimated in the Sample Size Calculator (clincalc.com) with a level of 0.07 and a G capacity of 35% with a certainty range of 99.0%.

The current dataset that is being followed is the train dataset, gathered from Kaggle.com. The database is composed of 885 files belonging to 8 classes. Using 177 files for validation. They contain data on diseased tea leaves like 'algal', 'anthrone', 'bird eye', 'brown blight', 'grey light', 'healthy','red spot', 'white spot', and healthy tea leaves. For training, 80% of the dataset was used, and 20% for testing.

The Windows 11 OS was used as the evaluation platform. Six gigabytes of RAM and a Ryzen i5 processor comprised the hardware setup. 64-bit system sorting was employed. The Python programming language was used to implement the code. Here, the healthy leaf is the independent variable, and the dependent variables are the algae, anthrone, and bird eye. To increase accuracy, the analysis is conducted using both independent and dependent variables.

Convolutional Neural Networks:

The noteworthy achievements of convolutional neural networks (CNNs) in excelling at image recognition tasks underscore their proficiency in automatically acquiring hierarchical features from images. This inherent capability positions CNNs as well-suited for discerning patterns linked to diverse diseases. Within the context of this study, we will harness the power of

CNNs by training them on a meticulously curated dataset comprising labelled images of tea leaves. This training process will empower the model to discern and internalise the distinctive visual characteristics distinguishing between healthy and diseased tea leaves, thereby enhancing its accuracy in disease identification.

ALGORITHM:

Step 1:Collect a dataset of labelled tea leaf images, with annotations indicating whether each image contains a healthy leaf or is affected by a specific disease.

Step 2: Split the dataset into training, validation, and test sets.

Step 3:Apply data augmentation techniques (rotation, flipping, and scaling) to artificially increase the diversity of the training dataset. This helps improve the model's generalisation.

Step 4:Design the CNN architecture for tea leaf disease detection. A common architecture might include convolutional layers for feature extraction and pooling layers for down-sampling.

Step 5:Train the CNN using the prepared dataset.

Step 6:Evaluate the trained model on the test set to assess its generalisation to unseen data.

Step7:validate the model

Step8:get the accuracy

Support Vector Machines

SVMs are effective in classifying data into distinct categories by finding the optimal hyperplane that separates different classes with the maximum margin. In our approach, SVMs will complement CNNs by refining the classification results obtained from the neural network. The combination of CNN and SVM aims to enhance the robustness and reliability of disease detection.

ALGORITHM:

Step 1:Collect a dataset of tea leaf images, labelled with the corresponding disease classes (e.g., healthy, diseased).

Step 2:Preprocess the images, which may include resizing, normalisation, and feature extraction.

Step 3:Extract relevant features from the preprocessed images. This step transforms the images into feature vectors that represent the characteristics of the tea leaf

Step 4:Split the dataset into training and testing sets.

Step 5:Train the SVM model using the training dataset.

Step 6:Use the extracted features and corresponding labels to fit the SVM model.

Step 7:Evaluate the trained SVM model on the testing dataset.

Step 8:Use the trained model to predict the disease class of unseen tea leaf images in the testing dataset.

Step 9: Analyse the results and visualise the classification outcomes

Step 10:Investigate misclassifications and refine the model if necessary.

STATISTICAL ANALYSIS

The statistical analysis is carried out using the IBM SPSS 26.0.1 software for both the proposed model and the compared model. The healthy tea leaf disease images in this dataset are independent variables. 'Algal', 'anthrone', 'bird eye', and 'brown blight' are dependent variables. The analysis done in this study is an independent sample T test analysis. This analysis is done on both models. After analysing the mean accuracy, the standard deviation and standard error are noted.

RESULT

After comparing and analysing both models, it is observed that the convolutional neural network model has significantly higher performance than the support vector machines in identifying tea leaf disease.

In Table 1, the accuracy values of convolutional neural networks and support vector machines were noted. Each algorithm contains 10 attributes.

In Table 2, the group statistical analysis is performed for both models on all iteration variables. The proposed novel convolutional neural networks have outperformed the smoothed moving average model. The mean accuracy of the novel convolutional neural network model observed is 94%, and the novel support vector machine model is 76%. The standard deviation observed for the novel convolutional neural networks is 4.9711, and for the support vector machine model, it is 4.9349. The mean standard error observed in novel convolutional neural network is 1.1116 and in novel support vector machine (SVM) model, it is 1.1035

In Table 3, the independent sample T test is performed on both the proposed model and the compared model. The significance values shown on the CNN and SVM are equal variances of 011 and 010, respectively, for the test performed. The convolutional neural network is better because of the significant difference in variance. The mean difference is equal for both the convolutional neural networks and support vector machines. At a 95% confidence interval of the difference, both the lower interval and upper interval values are the same.

Figure 1 shows the graphical representation of the accuracy difference between the convolutional neural networks and the support vector machines. It can be seen that convolutional neural networks has shown a good difference in accuracy compared to support vector machines

DISCUSSION

According to the research study, the convolutional model with an accuracy of 94% has shown better performance compared to the support vector machine model, which has an accuracy of 76 in comparison for forecasting product demand. There is no in-between difference on a sample T-test with a value of $p = 0.754$ ($p > 0.05$).

Convolutional neural networks, support vector machines, and other algorithms are used in the analysis (Zhou and Chen 2018). Convolutional neural networks (CNNs) are outperforming support vector machines (ANNs) in a comparison that compares the two (Balas, Solanki, and Kumar 2023). The accuracy of the CNN algorithm is 94%, while that of the SVM algorithm is 74.6%. (Nakamatsu et al. 2020). Additionally, a comparison between the two algorithms reveals that the convolutional neural networks (CNNs) algorithm is outperforming the support vector machines (SVMs). Additionally, it has produced findings that are comparable to our own (Kim 2017). When combined with supervised pattern recognition and advanced analytics algorithms, convolutional neural network algorithms diagnose tea leaf disease more accurately and with better performance than any other technique. (Bindhu, Chen, and Tavares 2020). Furthermore, we can draw the conclusion that the convolutional neural network algorithm performs better than the support vector machine method and appears to be more exact based on the preceding talks and facts.

There are some limitations with the novel convolutional neural network that consists of clusters of a large number of CNN classifiers, which takes more time to get executed compared to other machine learning and advanced analytics algorithms for tea leaf disease detection. In future work, this model will be improved with enhanced qualities, less execution duration, and more exact outcomes. This might have a better future, as the number of actions has been increasing every day.

CONCLUSION

In this study, the convolutional neural network model has been compared to the support vector machine model for tea leaf disease detection. The convolutional neural network model had better performance with an accuracy difference of 18% than the compared model.

DECLARATIONS

Conflict of interests

There are no conflicts of interest in this manuscript.

Author Contributions

Author VM was involved in a literature study, data collection, data analysis, and manuscript writing. The author, AK, is involved in data verification, validation, and review of the manuscript.

Acknowledgement

The authors would like to express their gratitude towards Saveetha School of Engineering and Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for providing the necessary infrastructure to carry out research studies.

Funding: We thank the following organisations for providing financial support that enabled us to complete the study.

1. Techno soft labs Pvt Ltd
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

References:

Balas, Valentina E., Vijender Kumar Solanki, and Raghvendra Kumar. 2023. *Recent Advances in Internet of Things and Machine Learning: Real-World Applications*. Springer.

Bao, Wenxia, Tao Fan, Gensheng Hu, Dong Liang, and Haidong Li. 2022. "Detection and Identification of Tea Leaf Diseases Based on AX-RetinaNet." *Scientific Reports* 12 (1): 2183.

Bindhu, V., Joy Chen, and João Manuel R. Tavares. 2020. *International Conference on Communication, Computing and Electronics Systems: Proceedings of ICCCES 2019*. Springer Nature.

Chatterjee, Jyotir Moy, Abhishek Kumar, Pramod Singh Rathore, and Vishal Jain. 2021. *Internet of Things and Machine Learning in Agriculture: Technological Impacts and Challenges*. Walter de Gruyter GmbH & Co KG.

Dai, Weidong, Zhengyan Hu, Dongchao Xie, Junfeng Tan, and Zhi Lin. 2020. "A Novel Spatial-Resolution Targeted Metabolomics Method in a Single Leaf of the Tea Plant (Camellia Sinensis)." *Food Chemistry* 311 (May): 126007.

Dash, Subhransu Sekhar, C. Lakshmi, Swagatam Das, and Bijaya Ketan Panigrahi. 2020. *Artificial Intelligence and Evolutionary Computations in Engineering Systems*. Springer Nature.

Kim, Phil. 2017. *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence*. Apress.

Nakamatsu, Kazumi, Roumen Kountchev, Ari Aharari, Nashwa El-Bendary, and Bin Hu. 2020. *New Developments of IT, IoT and ICT Applied to Agriculture: Proceedings of ICAIT 2019*. Springer Nature.

Preedy, Victor R. 2012. *Tea in Health and Disease Prevention*. Academic Press.

Venkata Krishna, P., and Mohammad S. Obaidat. 2020. *Emerging Research in Data Engineering Systems and Computer Communications: Proceedings of CCODE 2019*. Springer Nature.

Wang, Yuting, Zhongwei Yu, Jun Jiang, Yaxin Li, and Shui Yu. 2021. "Mulberry Leaf Attenuates Atherosclerotic Lesions in Patients with Coronary Heart Disease Possibly via 1-Deoxynojirimycin: A Placebo-Controlled, Double-Blind Clinical Trial." *Journal of Food Biochemistry* 45 (1): e13573.

Zhou, Jianlong, and Fang Chen. 2018. *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent*. Springer.

TABLES AND FIGURES

Table 1. Accuracy values of cnn and svm

S. No	CNN	SVM
1	94.00	76.00
2	93.00	75.00
3	91.60	73.40
4	92.00	72.90
5	90.80	68.60
6	89.80	66.90
7	88.70	74.90
`8	87.00	75.00
9	92.90	69.90
10	45.00	41.00

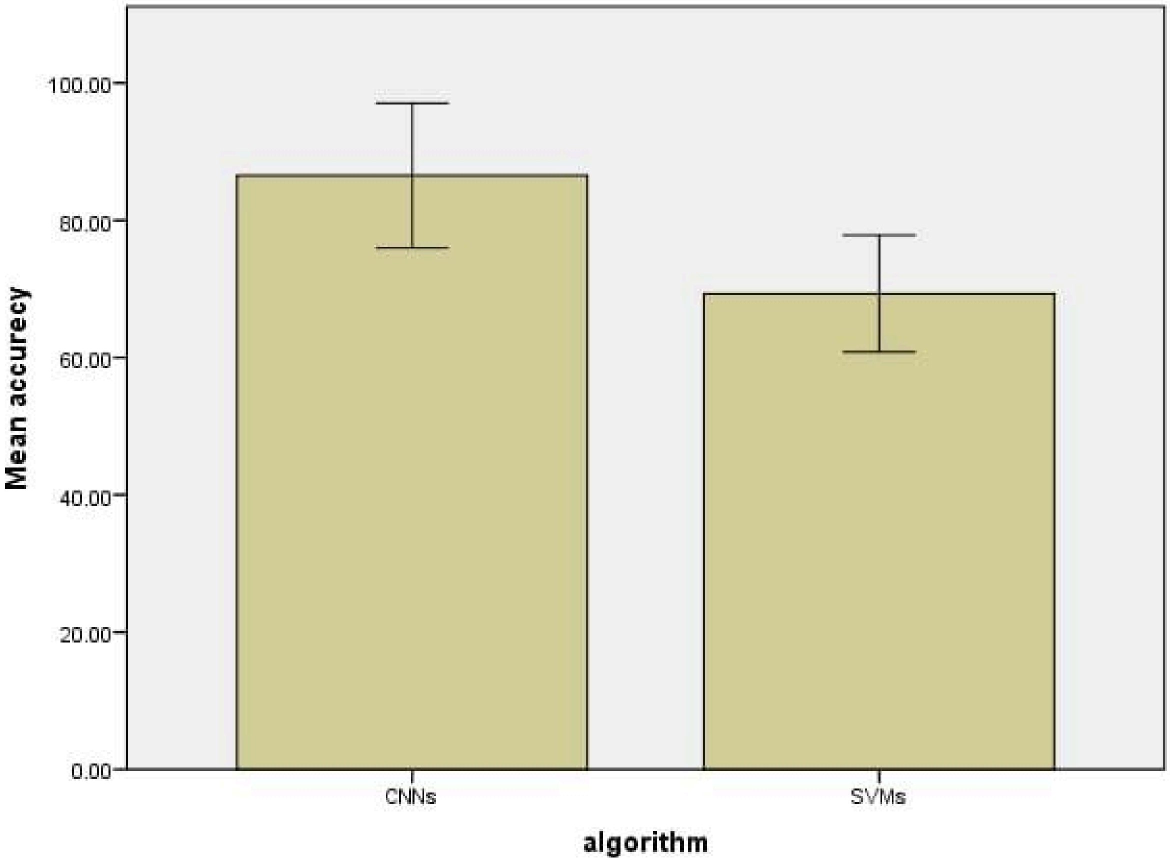
Table 2. Group Statistics Results (The mean of the Novel Convolutional Neural Network Algorithm is 86.4800, which is higher than the mean of the support vector machines, which is

69.3000; the average variance for the Novel CNN technique is 4.65, while the support vector machine is 3.68.).

	GROUPS	N	Mean	Std. Deviation	Std. Error Mean
ACCURACY	CNNs	10	86.4800	14.72879	4.65765
	SVMs	10	69.3000	11.05746	3.68582

Table 3: Independent Sample T-Test is applied with the sample collections by fixing the level of significance as ($p = 0.01$, $p<0.05$) with a confidence interval of 95% after applying the algorithm.

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	f	sig	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.078	.784	2.848	17	.011	17.18000	6.03264	4.45224	29.90776
Equal variances not assumed			2.892	16.515	.010	17.18000	5.93961	4.62044	29.73956



Error Bars: 95% CI
Eroor Bars:+/-2 SD