International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 1 Jan. 2026, pp: 170-176 www.ijaem.net ISSN: 2395-5252

Advanced Computer Vision Techniques for Tomato Plant Diseases Detection and Analysis

Sani Ibrahim
Department of Computer Science, School of Technology, Kano State Polytechnic

Date of Submission: 05-01-2026 Date of Acceptance: 16-01-2026

ABSTRACT
Tomato leaf diseases significantly reduce global crop yield and threaten food security, especially in regions where agriculture is the primary livelihood. Traditional disease identification methods rely heavily on expert inspection, which is time-consuming, subjective, and inaccessible to many farmers. This study presents an enhanced deep learning framework for automated detection and classification of tomato leaf diseases using a merged dataset sourced from Kaggle and PlantVillage, comprising 25,849 images across eleven disease classes and healthy leaves. A fine-tuned VGG19 convolutional neural network architecture was employed and improved through L1 regularization and targeted feature extraction to minimize overfitting and enhance generalization. The model achieved a validation accuracy of ***99.6%****, outperforming several state-of-the-art approaches reported in the literature. Experimental results across multiple train–test splits demonstrate consistent performance, highlighting the robustness of the proposed model in varying environmental conditions. The framework is designed for eventual integration into user-friendly agricultural systems, enabling real-time, non-expert-friendly disease diagnosis. This research contributes a diverse hybrid dataset, an optimized CNN architecture, and a pathway toward scalable precision agriculture tools.*

Keywords: *Tomato leaf disease, Deep learning, Computer vision, VGG19, CNN, Agriculture, Image classification*

I. Introduction

Scarcity of the food and current global inflation are among the world's best problem facing people today. Because of the rapid growth in the world's population, global warming that distorts the environment, and other problems within the environment which most of the food industries find it difficult to control the demand chain and supply chain (Silva & Brown, 2023). However, using existing and new agricultural trends to put more effort into harvesting more food is the primary concern to change the above case (Silva & Brown, 2023). Agricultural biological variety is preliminary to providing food and raw materials to humans. The microorganisms such as bacteria, nematodes, and fungi; the soil on which plants are planted; global warming that results in temperature extremes; the weather that results in changing the amount of moisture and humidity; and other elements that repeatedly destroy a plant, it can develop a disease. Many diseases affecting plants might tend to influence the growth function, and structures of plants and crops, which automatically affect the living, that are dependent on them. Many farmers still use manual ways to identify plant diseases, since it is challenging to do so early on and hurts productivity. To address the issues, many deep learning (DL), image processing, and machine learning (ML) techniques are being developed, by which the detection of disease in a plant is performed by images of plant leaves (Van Teeffelen, 2023).

Tomato is one of the most common fruits used by people all over the world (Manje et al., 2022). Many farmers in Nigeria harvest tomatoes and grow these fruits on a commercial scale and sell them which are their primary income. The demand for Tomato fruit globally is for all seasons (Sharma et al., 2022). However, it is very paramount for farmers to be well knowledgeable about the fruits they harvest and to be self-sufficient, and it is important for them to know the way to properly tackle the chemicals and pesticides in their farms. Because, this used to happen each season before the tomatoes to be harvested, Tomatoes were destroyed due to the attack of insects and diseases which led to poor crop harvest. There is little difference in the quantity of crops, quality of crops, price of tomatoes, and demand in the tomato market globally (Sharma et al., 2022). Insect attacks, soil, bad weather, the environment bad insecticides as well as polluted water are the major causes of Tomato diseases (Sharma et al., 2022). The agricultural sector plays a critical role in sustaining global food production and addressing food security challenges. However, plant diseases remain a significant threat to crop yield and quality, with potentially devastating economic consequences. Tomato plants

International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 1 Jan. 2026, pp: 170-176 www.ijaem.net ISSN: 2395-5252

(Solanumlycopersicum), a staple in diets worldwide, are particularly susceptible to a range of diseases caused by various pathogens, including bacteria, fungi, and viruses. Timely and accurate detection of these diseases is crucial to mitigate their impact in an agricultural environment.

This research tends to detect leaf diseases of tomatoes using up-to-date technology

"Image Processing". If any of the plants had been affected by any diseases farmers had to wait for days for disease confirmation. But computer science and well-built technologies have made many things possible. Farmers can easily learn about pesticides, seeds, and even diseases through a phone call to the local agriculture consultant. However, it might not be possible for the consultant to give any suggestions without observing the plant. And that is a very lengthy process that consumes more time. Image processing made that task so easy that now only by snapping pictures of affected plants and inputting those images into the disease-detecting system farmers can learn about which disease attack their plant.

2.0 Related Work

Balafaset al.(2023) in their work they performed an extensive computational studyon five state-sof-the-art object detection algorithms, which they used Plant- Doc dataset to detect the diseases on the leaves, and eighteen state-of-the-art classification algorithms on PlantDoc dataset, to predict whether or not there is a disease in a leaf and which one it is. Based on their finding for plant diseases detection YOLOv5 is able to achieve a high accuracy on both aspect of detection small or larger spot on leaves images which is the indication that shows the plant was affected by the diseases. ResNet50 achieved a total accuracy of 61.01% and was trained on about 18 minutes, while MobileNetv2 achieved a total accuracy of 59.74% and was trained on about 16 minutes. However, they revealed that with regard to classification problem, the networks MobileNetv2 and ResNet50 had the most optimal trade-off on accuracy and training time.

Ballester and Araujo(2018) in their study they chose to use dataset from sketches with two well known CNN model GoogleNet and AlexNet. They concluded that the test networks are not usefulto perform sketch classification and their classifications are different from the classifications of images by humans.

In a research conducted by Bhatkoti(2019) they display that Bayesian method and other machine learning methods have failed to produce desired results in diagnosing

Alzheimer's Disease. They explored that deep learning framework k-Sparse Autoencoder which was modifed which improves classification accuracy and speeds.

They validated mKSA using a benchmark dataset of Alzheimer's Disease which was compared with other models which shows that the modifed techniques was superior than the other techniques.A review paper conducted by (Chin et al., 2023)they used drones to collect the dataset they selected a 38 primary studies.

Cınar and Arslan(2021) revealed that CNN architecture when mergered feature vector in the last pooling layer of both CNN architectures, and the resulting feature vector is classified by the Support Vector Machine (SVM) will resulted in coming with CNN architecture with high accuracy.

In comparative study was conducted by Hatem et al.(2023) were they compare different CNN transfer learning models, they chose six different metrics in assessing the models, accuracy, precision, specificity, recall, F-score and time in assessing the performance GoogleNet, AlexNet, VGG16, and ResNet50 each have an accuracy of 98.57%, 98.81%, 99.05%, and 99.36% based on their findings they concluded that RestNet50 has highest performance based on the metrics they chose.

Infet al.(2022) conducted a predictive analytic model for research were support viector machine (SVM) algorithm were employed To overcome the problem with missing values, they used the linear interpolation. To deal with the imbalanced dataset, they get rid of over-sampling method, since it proved to be more effective for small dimension datasets. The problem with their research is the dataset were too small, they limited their research on one diseases that is late blight.

In a research work conducted by Jung et al.(2023) they focus on comparitive analysis of five different CNN models to assessed the performance of each models based on the accuracy, precision, recall, and F1-score and the performance were calculated and evaluated using confusion matrix. Their research shows that EfficientNet has the highest performance. But the dataset size were too small and the accuracy of the model sometimes depends on the size of the dataset, the algorithm used for their model were not well define.

LailatulMufidah(2021) used fuzzy neural network was used for diseases classification in their study. Learning of the system has been performed using a cross-validation approach by applying a gradient descent algorithm. To conduct the learning, the mean of the respective attributes was used to replace all missing values in the CKD dataset. After learning, the designed FNN system is applied for the detection of diseases. The model accuracy was obtained as 99.75%, sensitivity- 100%, specificity-

International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 1 Jan. 2026, pp: 170-176 www.ijaem.net ISSN: 2395-5252

99.34%, precision was 99.6%, F1 score- 99.8%. in the research they used 16 different rules and 400 number of dataset.

In a research paper by Suhamanet al.(2023)in their research they used CNN model with less complexity their aim is to develop a model that can be use for older farmers. However, the performance of their model in terms of accuracy is 89%.

In study conducted by Theses (2021) used to compare and assess the performance of five different pretrained deep learning convolutional neural networks for classification of crops diseases for apples, grapes, and tomatoes. Based on their findings they concluded that VGG16 is the best model in classification while the worst performance is ResNet50.

Tunku and Rahman(2020) in their research work they developed a framework that aimed at classifying the crops according to their class and detect the spot of the leaf that is affected. To achieve their aim they combine the single shot MultiBox and MobileNet to detect a multiple object.

Wallelign(2020)in their research they used Deep Learning Convolutional Neural Network to design model, they evaluated two model one for classification and one grades. They revealed that dataset shift and small imbalanced dataset are the main issues that hinders model performance and reliability of a model. The drawback of their research work is they used small numeber of images for the training of the model.

In research paper conducted by Xu (2023) in the research various approach were analysed such as Naïve Bayse, Stochastic Gradient Decent, Decision Tree, Support Vector Machine and Neural Network. The study revealed that Naive Bayes model has been the model with the highest accuracy (90.6%), highest precision (96.4%).

III. Methodology

3.1 Dataset Collection and Composition

A hybrid dataset totaling **25,849** images was constructed by merging two publicly available sources: Kaggle and PlantVillage. The dataset covers **11 disease classes** and healthy leaves. Images vary in resolution, lighting, backgrounds, and disease severity.

Classes include:

Bacterial Spot, Early Blight, Late Blight, Leaf Mold, Powdery Mildew, Septoria Leaf Spot, Spider Mites, Target Spot, Tomato Mosaic Virus, Tomato Yellow Leaf Curl Virus, and Healthy.

3.2 Image Preprocessing

Images were standardized to **256 × 256 × 3** resolution before training. Data augmentation included rotation, zooming, flipping, shifting, and brightness variation to increase dataset diversity and reduce overfitting.

3.3 Model Architecture

The VGG19 CNN architecture was selected for its balance between depth and computational feasibility. Modifications included:

- Application of **L1 regularization** to encourage sparsity
- Additional dropout layers to reduce overfitting
- Fine-tuning of fully connected layers for eleven-class classification
- Softmax classifier for probability mapping

3.4 Training Strategy

Models were trained and evaluated under various split ratios (60–40, 70–30, 80–20) to assess robustness. The Adam optimizer was used with a learning rate schedule to accelerate convergence.

3.5 Evaluation Metrics

Performance was assessed using:

i. Accuracy
ii. Precision
iii. Recall
iv. Loss
v. Confusion matrix

These metrics provided insight into both overall performance and per-class predictive capacity.

IV. RESULTS AND DISCUSSION

4.3.3 Result based on 80% Training and 20% Testing + Regularization Technique

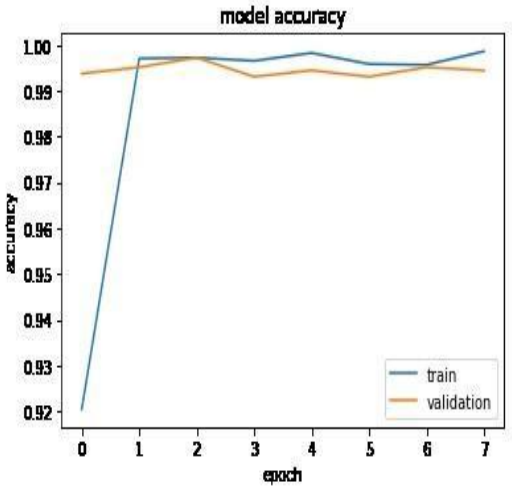

Fig. 4.1 Model Accuracy on 80-20 splitting basis

International Journal of Advances in Engineering and Management (IJAEM)
Volume 8, Issue 1 Jan. 2026, pp: 170-176 www.ijaem.net ISSN: 2395-5252

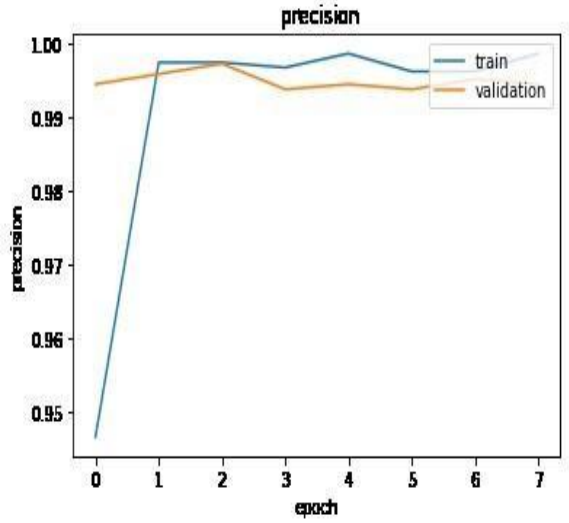

Fig. 4.2 Model Precision on 80-20 splitting basis

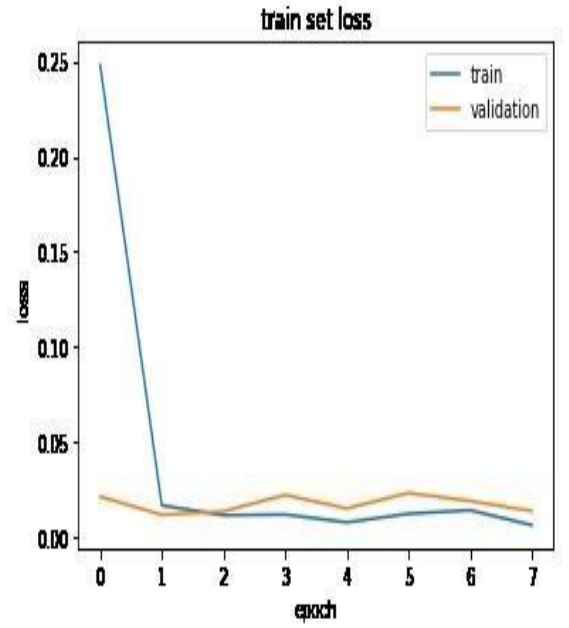

Fig. 4.3 Model Loss on 80-20 splitting basis

	precision	recall	f1-score	support
0	1.00	1.00	1.00	210
1	0.99	0.99	0.99	189
2	1.00	1.00	1.00	210
3	1.00	0.99	0.99	202
4	1.00	1.00	1.00	205
5	1.00	0.99	1.00	216
6	0.99	1.00	1.00	192
7	0.98	0.99	0.99	184
8	1.00	1.00	1.00	192
accuracy			1.00	1800
macro avg	1.00	1.00	1.00	1800
weighted avg	1.00	1.00	1.00	1800

Fig. 4.4 Model Weight on 80-20 splitting basis

Table 4.4 Evaluation summary on 80-20 splitting basis

S/N	Evaluation Metrics	Result
1	Test Accuracy	0.996%
2	Test Precision	0.994%
3	Test Recall	0.996%
4	Test Loss	0.015%

4.1Performance Metrics Analysis of Experiment

1. Test Accuracy: 99.6%

i. Interpretation: The model achieved a test accuracy of 99.6% as shown in table 4.4, indicating that it correctly classified 99.6% of the validation dataset. This high accuracy reflects the model's ability to generalize well to unseen data, suggesting that it has effectively learned the underlying patterns associated with tomato diseases.

ii. Contextual Comparison: In the context of tomato disease detection, this accuracy is significantly higher than many models reported in the literature. For instance, previous studies have achieved accuracy rates ranging from 67% to 95%, indicating that the current model demonstrates superior performance.

2. Test Precision: 99.4%

i. Interpretation: The precision of 99.4% indicates that when the model predicts a disease, it is correct 99.4% of the time. This metric is crucial in agricultural applications, as high precision minimizes the likelihood of false positives, thereby reducing unnecessary treatments and resource expenditures.

ii. Implications: A precision of 99.4% suggests that the model is highly reliable in its positive predictions, making it a valuable tool for farmers and agricultural practitioners seeking to manage disease outbreaks effectively.

3. Test Recall: 99.6%

i.Interpretation: The recall of 99.6% signifies that the model successfully identified 99.6% of the actual diseased instances in the validation dataset. This metric is particularly important in scenarios where failing to detect diseased plants can lead to significant crop losses.

4. Test Loss: 0.015% ii.Interpretation: The test loss of 0.015% reflects the model's performance in terms of how well it predicts the validation data. A low loss value indicates a good fit to the data.

iii.Analysis: The very low loss value suggests that the model has effectively learned the patterns in the training data and is performing well on the validation set, indicating strong model training and optimization.

4.5 Discussion of Findings

The evaluation metrics present a highly favorable view of the model's performance. The accuracy of 99.6%, precision of 99.4%, and recall of 99.6% indicate that the model is not only reliable in its predictions but also effective in identifying diseased plants. The low test loss further reinforces the model's capability to generalize well to unseen data.

4.1 Performance Across Experiments

The enhanced VGG19 achieved:

- **99.6% validation accuracy**
- Strong precision and recall values across all classes
- Consistently low loss values across split ratios

This performance exceeds those of many comparable studies reviewed.

4.2 Analysis of Feature Extraction and Regularization

L1 regularization effectively reduced overfitting by penalizing less important weights, which improved generalization. Enhanced feature extraction allowed the model to detect subtle disease patterns, resulting in fewer false positives and false negatives.

4.3 Comparison with Existing Studies

Compared with studies using AlexNet, GoogleNet, Xception, and EfficientNet, the proposed model demonstrates competitive or superior accuracy, especially given the larger and more diverse dataset used. Many earlier studies relied on small datasets (often fewer than 10,000 images) or fewer disease categories, limiting applicability.

The integration of deep feature extraction and regularization distinguishes this model as a more scalable and generalizable solution.

V. CONCLUSION

This research presents a robust and high-accuracy CNN-based framework for detecting tomato leaf diseases. By merging multiple datasets, optimizing VGG19 architecture, and applying effective regularization, the system achieved 99.6% accuracy and demonstrated strong generalization across diverse image conditions.

The framework contributes:

1. A large, diverse hybrid dataset covering eleven disease classes.
2. An optimized VGG19 model with L1 regularization for improved performance.
3. A foundation for deployment into real-time agricultural diagnostic systems.

Future work includes integrating the model into mobile platforms for farmers, expanding the dataset with field-captured images, and exploring lightweight architectures for edge deployment.

REFERENCES

[1]. Ahmed, M. R., Haque, R., Rahman, S. M. A., Afridi, S., Abir, M. F. F., Hossain, M. F., Khan, M. S. H., & Uddin, M. M. (2024). Towards Automated Detection of Tomato Leaf Diseases. *Proceedings - 6th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2024*, *May*, 387–392. https://doi.org/10.1109/ICEEICT62016.2024.10534559

[2]. Ai, I. (2023). *A Smartphone-Based Detection System for Tomato Leaf Disease Using EfficientNetV2B2 and Its Explainability with Artificial*.

[3]. Almas, B., Sathesh, K., &Rajasekaran, S. (2020). *A Deep Analysis of Google Net and AlexNet for Lung Cancer Detection*. *December 2019*. https://doi.org/10.35940/ijeat.B3226.129219

[4]. Alpyssov, A., Uzakkyzy, N., Talgatbek, A., Moldasheva, R., Bekmagambetova, G., Yessekeyeva, M., Kenzhaliev, D., &Yerzhan, A. (2023). *ASSESSMENT OF PLANT DISEASE DETECTION BY DEEP*. 41–48. https://doi.org/10.15587/1729-4061.2023.274483

[5]. Arafat, I. S., Aswath, S., &Nishath, S. M. H. (2024). *Automatic Early Detection of Tomato Leaf Disease using IoT and Deep Learning*. 1–17.

[6]. Arivazhagan, M. . (2023). Hemp Disease Detection and Classification using Machine Learning. *Interantional Journal of Scientific Research in Engineering and Management*, *07*(11), 1–11. https://doi.org/10.55041/ijsrem27412

[7]. Balafas, V., Karantoumanis, E., Louta, M., &Ploskas, N. (2023). Machine Learning and Deep Learning for Plant Disease Classification and Detection. *IEEE Access*, *11*, 114352–114377. https://doi.org/10.1109/ACCESS.2023.3324722

[8]. Ballester, P., & Araujo, R. M. (2018). *On the performance of GoogLeNet and AlexNet applied to sketches On the Performance of GoogLeNet and AlexNet Applied to Sketches*. *February 2016*.

[9]. Bay, H., Tuytelaars, T., & Van Gool, L. (2006). *SURF: Speeded Up Robust Features.* ECCV.

[10]. Bayram, H. Y., &Alatas, B. (2022). *Traitement du Signal*. *December*. https://doi.org/10.18280/ts.390537

[11]. Bhatkoti, P. (2019). *Early Diagnosis of Alzheimer's Disease: A Multiclass Deep Learning Framework with Modified k-Sparse Autoencoder Classification IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF INFORMATION TECHNOLOGY IN SCHOOL OF COMPUTING AND MATHEMATICS*. *April*.

[12]. ChandrakantPatkar, U., Bharadwaj Shrives, S., Patil, U. S., JanardanPatankar, A., Jain, N., Kumari, M., &Chandhoke, A. (2023). International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Object Detection using Machine Learning and Deep Learning. *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE*, *2024*(1s), 466. www.ijisae.org

[13]. Chin, R., Catal, C., &Kassahun, A. (2023). *Plant disease detection using drones in precision agriculture*.

[14]. Çınar, A., &Arslan, S. (2021). and neutrophils on white blood cells using hybrid

[15]. Alexnet - GoogleNet -SVM. *SN Applied Sciences*, *3*(4), 1–11. https://doi.org/10.1007/s42452-021-04485-9

[16]. Foody, G. M., &Mathur, A. (2006). *The use of small training sets containing mixed pixels for accurate hard image classification.* Remote Sensing of Environment.

[17]. Hatem, A. S., Altememe, M. S., &Fadhel, M. A. (2023). *Identifying corn leaves diseases by extensive use of transfer learning : a comparative study*. *November 2022*, 1030–1038. https://doi.org/10.11591/ijeecs.v29.i2.pp1030-1038

[18]. Hendrickx, I., & Van Bosch, A. Den. (2005). Hybrid algorithms with instance-based classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3720 LNAI*(October 2005), 158–169. https://doi.org/10.1007/11564096_19

[19]. Inf, S. S. D., Maridina, F., Supervisor, M., &Fenu, P. G. (2022). *PhD DEGREE Mathematics and Computer Science TITLE OF THE PhD THESIS Predictive Analytics Models and Tools for Decision Support of Stakeholders in Digital Agriculture Scientific and Disciplinary Sector(s)* (Issue February). https://www.francescamalloci.com

[20]. Jung, M., Song, J. S., Shin, A. Y., Choi, B., Go, S., Kwon, S. Y., Park, J., Park, S. G., & Kim, Y. M. (2023). Construction of deep learning - based disease detection model in plants. *Scientific Reports*, 1–14. https://doi.org/10.1038/s41598-02334549-2

[21]. Karim, M., &Ul, M. (2020). *Rice Disease and Pest Detection Using Deep Learning*. http://dspace.ewubd.edu/handle/123456789/3557%0Ahttp://dspace.ewubd.edu:8080/bitstream/handle/123456789/3557/Md._Minhaz_Ul_Karim.pdf?sequence=1&isAllowed=y

[22]. Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). *ImageNet Classification with Deep Convolutional Neural Networks.*Advances in Neural Information Processing Systems.

[23]. LailatulMufidah, K. T. (2021). *No CHRONIC KIDNEY DISEASE DETECTION USING FUZZY NEURAL NETWORKTitle*. *7*(3), 6.

[24]. Laptev, N., Laptev, V., Gerget, O., &Kolpashchikov, D. (2021). *Integrating Traditional Machine Learning and Neural Networks for Image Processing.*GraphiCon 2021.

[25]. Likas, A., Vlassis, N., &Verbeek, J. J. (2003). *The global k-means clustering algorithm.* Pattern Recognition.

[26]. Lu, J., & Tan, L. (2021). *Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification*. 1–18.

[27]. Manakitsa, N., Maraslidis, G. S., &Moysis, L. (2024). *A Review of Machine Learning and Deep Learning for Object Detection , Semantic Segmentation , and Human Action Recognition in Machine and Robotic Vision*.

[28]. Manje, S., Pawar, T., Waseem, K. M., Asst, S. S., Student, U. G., &Engg, C. (2022). *Tomato Leaves Diseases Detection Using Image Processing and Deep Learning*. *01*, 8–11.

[29]. Michalski, B., &Wójcik, M. P. (2022). *Comparison of LeNet-5 ,AlexNet and GoogLeNet models in handwriting recognition PorównaniemodeliLeNet -5 , AlexNetiGoogLeNet w rozpoznawaniupismaręcznego*. *23*(April), 145–151.

[30]. Muhammad, N. A., Nasir, A. A., Ibrahim, Z., &Sabri, N. (2018). *Evaluation of CNN ,Alexnet and GoogleNet for Fruit Recognition*. *12*(2), 468–475. https://doi.org/10.11591/ijeecs.v12.i2.pp468-475

[31]. Müller, E. J., Palesi, F., Hou, K. Y., Tan, J., Close, T., Gandini Wheeler-Kingschott, C. M., D'Angelo, E., Calamante, F., & Shine, J. M. (2023). Parallel processing relies on a distributed, low-dimensional cortico-cerebellar architecture. *Network Neuroscience*, *7*(2), 844–863. https://doi.org/10.1162/netn_a_00308

[32]. Opadeji, V., Mahmudur Rahman, M., & Professor, A. (2021). *Title of Thesis: ACCURATE SICKLE CELL DETECTION USING DEEP TRANSFER LEARNING BASED FEATURE EXTRACTION, CLASSIFICATION, AND RETRIEVAL OF BLOOD SMEAR IMAGES*. *May*.