

Leveraging Artificial Intelligence for Improved Plant Disease Detection

Ms. Shreya Uday Gore¹, Dr. Akansha Tyagi², Dr. Santosh T. Jagtap³

Research Scholar, Department of Computer Science, Shri Jagdishprasad Jhabarmal Tibrewala University, Rajasthan, India¹

Assistant Professor, Department of Computer Science, Shri Jagdishprasad Jhabarmal Tibrewala University, Rajasthan, India²

Assistant Professor, Department of Computer Science, Prof. Ramkrishna More College, Pune, India³

Abstract: Agriculture is a cornerstone of global economic development, constituting 4% of global GDP and contributing over 25% to the GDP of the world's least developed countries[1,2]. Despite its significance, current food systems suffer from alarming levels of pollution, wasteful practices, and adverse impacts on both human health and the environment. In recent studies, 30% of the food produced globally is lost or wasted, which worsens the problems associated with food security, climate change, and environmental degradation.[3,4]. Addressing these issues and implementing effective strategies is essential for building a sustainable and resilient food system. Through innovative research on leaf disease identification, we aim to leverage artificial intelligence (AI) to tackle this pressing agricultural concern. This study evaluates the effectiveness of two classifiers, the Random Forest Classifier and Gaussian Naive Bayes (GaussianNB), in detecting leaf diseases. Additionally, we introduce novel parameters specifically designed for Gaussian Naive Bayes (GaussianNB) to enhance its performance in disease identification.

Keywords: GaussianNB, Random Forest Classifier, artificial intelligence (AI), Machine Learning(ML)

I. INTRODUCTION

Agriculture stands as a bedrock of global economic prosperity, contributing significantly to the world's GDP. It represents a formidable force in shaping the economies of both developed and developing nations alike. Impressively, it constitutes a substantial 4% of the global GDP and is instrumental in bolstering the GDP of the world's least developed countries, where it contributes over 25% to their economic output. However, despite its undeniable importance, the agricultural sector faces myriad challenges that threaten its sustainability and resilience.

One of the most pressing issues plaguing current food systems is the pervasive presence of pollution and wasteful practices. These practices not only compromise the quality of agricultural produce but also pose significant risks to human health and the environment. Furthermore, the staggering statistic that approximately 30% of global food production is lost or wasted further exacerbates the challenges confronting the agricultural sector. This wastage not only exacerbates issues related to food security but also exacerbates the adverse impacts of climate change and environmental degradation.

Considering these challenges, it becomes imperative to devise and implement effective strategies aimed at building a sustainable and resilient food system. Such a system would not only ensure the continued availability and accessibility of food for all but also mitigate the adverse impacts of agriculture on the environment. To this end, innovative research endeavors play a pivotal role in identifying and addressing key areas of concern within the agricultural sector.

One such area of focus is the identification and management of leaf diseases in plants, which pose significant threats to crop yield and food security. In this context, leveraging artificial intelligence (AI) presents a promising avenue for developing robust and efficient solutions to address this pressing agricultural concern. By harnessing the power of AI, researchers aim to revolutionize the process of disease detection and management in agricultural settings.

This study evaluates the effectiveness of two classifiers, Random Forest Classifier and Gaussian Naive Bayes, in detecting leaf diseases. It also introduces parameters tailored for Gaussian Naive Bayes to enhance its performance. The study aims to assess the accuracy and efficiency of these classifiers in diagnosing leaf diseases, contributing to advancements in automated plant disease diagnosis using AI.

Review of Literature

Early detection of plant diseases is critical for agriculture, and AI is stepping up to the challenge. This review explores recent advancements in using AI and computer vision to identify leaf diseases, analyzing both the challenges and the latest methods for accurate diagnosis.

Dong, X. (2023), This study presents pre-trained models for plant disease diagnosis, trained on large-scale datasets, to improve the performance of existing models. Extensive experiments demonstrate that utilizing pre-trained models leads to both higher accuracy and shorter training times. Furthermore, the study suggests a set of pre-built models that are tailored to meet the distinct demands of plant disease diagnostic tools, which span from quick and straightforward detection in portable devices to accurate diagnosis in laboratory environments. In general, the incorporation of pre-built models for plant disease diagnosis has the potential to advance this technology significantly.[5].

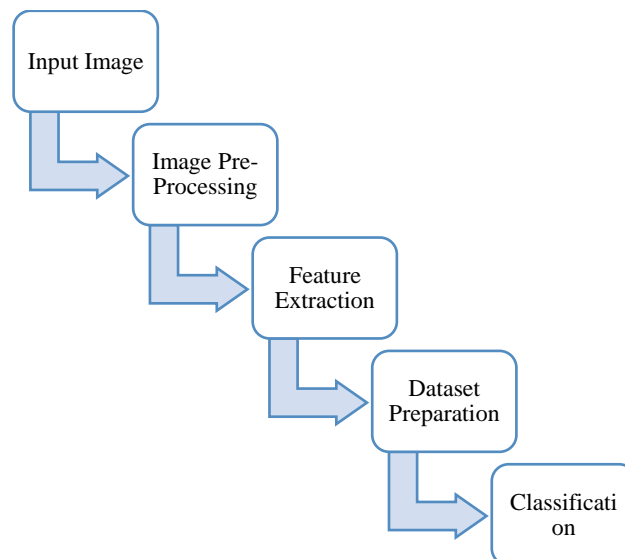
According to the research conducted by Chincinska (2021), the successful implementation of an infiltration technique with high repeatability is crucial for achieving quality output procedures. The effectiveness of infiltration depends on various parameters, emphasizing the importance of understanding its potential and possibilities. This understanding is essential for designing experiments that yield accurate and reliable results. While leaf infiltration methodology has been widely used in different fields, there is still untapped potential in exploring and utilizing various infiltration techniques to their fullest extent. In the future, it is expected that there will be increased attention and focus on efficient methods of leaf infiltration. Agroinfiltration, in particular, shows promise as a technology for establishing plant bioreactors capable of efficiently producing recombinant proteins for therapeutic and vaccine purposes. The field of molecular farming has experienced significant advancements, with agroinfiltration-based techniques emerging as a promising approach for the rapid and effective management of potential pandemics caused by novel and unidentified diseases[6].

In a recent study, Gupta and colleagues (2022) presented an algorithm that uses feature extraction and machine learning classification for automatically detecting plant diseases. However, it would be beneficial for future studies to delve deeper into the current techniques employed, provide more detailed information about the data and evaluation metrics used, and compare the proposed method's effectiveness with other available alternatives. To improve the effectiveness and reliability of disease detection systems, scientists could investigate the potential of advanced feature extraction techniques and deep learning models.[7].

Xian and Ngadiran (2021) found that ELM shows promise for identifying tomato leaf diseases accurately and efficiently. Further research on broader applicability to diverse crops and comparisons with advanced deep learning techniques are recommended to enhance the generalizability and robustness of these methods[8]

II. METHODS AND MATERIAL

This segment introduces the key classifier models and material employed in the study. The proposed methodology is comprised of five distinct stages, as illustrated in Figure 1.:



The study systematically evaluated the performance of Random Forest Classifier and Gaussian Naive Bayes (GaussianNB) in classifying apple leaf diseases. These classifiers were chosen for their effectiveness in handling multi-class classification tasks, essential for categorizing different apple leaf diseases. A dataset containing images of apple leaves categorized into different classes that represent various diseases was used as the basis for training and testing the classifiers. To comprehensively assess the impact of different training epochs on classifier performance, epoch sizes ranging from 50 to 200 were considered.

The test dataset was used to evaluate each classifier's performance metrics, which included accuracy, confusion matrix, precision, recall, and F1-score. These metrics provided valuable insights into the classifiers' ability to accurately classify apple leaf diseases across different epoch sizes. The obtained results were thoroughly analyzed to understand the

strengths and limitations of each classifier, thereby contributing to advancements in agricultural disease detection and management strategies.

III. RESULTS AND DISCUSSION

This analysis evaluates the performance of two classification algorithms: Random Forest Classifier and Gaussian Naive Bayes in detecting apple leaf diseases.

Algorithms : Random Forest Classifier , Gaussian Naive Bayes (GaussianNB).

Epoch Size : epoch_sizes = [50, 100, 150, 200]

Dataset : Apple

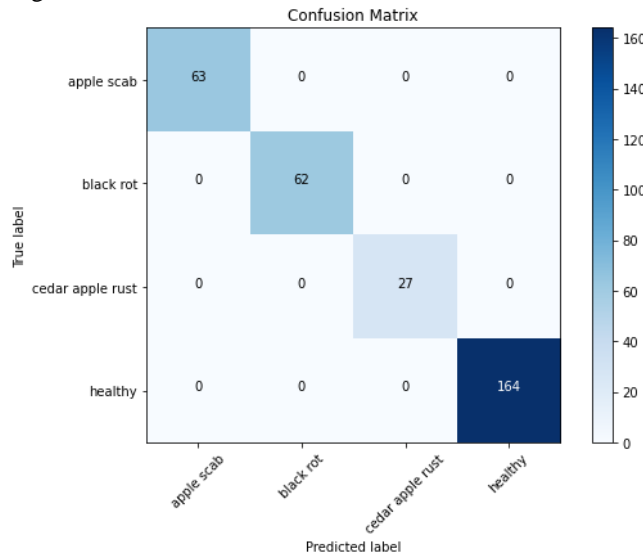
Random Forest Classifier on Test Dataset (Epoch Size = 50):

Accuracy: 1.0

Confusion Matrix:

```
[[ 63  0  0  0]
 [  0 62  0  0]
 [  0  0 27  0]
 [  0  0  0 164]]
```

Figure 1: Confusion Matrix of S-1 model for Random Forest



Classification Report:

Table 1 : classification report that show the final accuracy of the random forest algorithm

	precision	recall	f1-score	support
apple scab	1.00	1.00	1.00	63
black rot	1.00	1.00	1.00	62
cedar apple rust	1.00	1.00	1.00	27
healthy	1.00	1.00	1.00	164
accuracy	1.00	1.00	1.00	316
macro avg	1.00	1.00	1.00	316
weighted avg	1.00	1.00	1.00	316

The Classification report is a helpful tool that categorizes various metrics, including precision, recall, and F1-score, for different classes. Precision measures the accuracy of positive predictions, while recall evaluates the ability to correctly

identify positive instances. The F1-score is calculated as the harmonic mean of precision and recall. It's worth mentioning that in this specific scenario, all metrics have a value of 1.00, indicating optimal performance for each class.

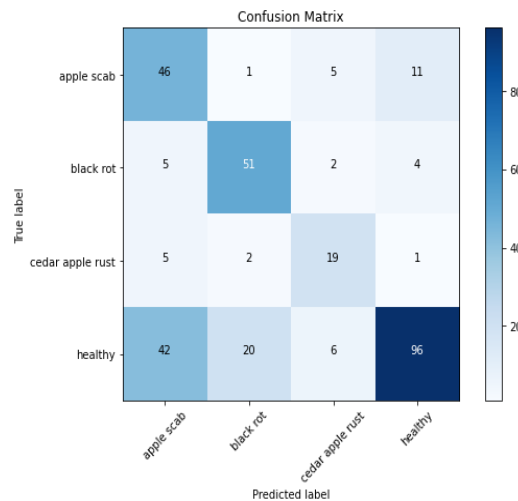
Gaussian Naive Bayes Classifier on Test Dataset (Epoch Size = 50):

Accuracy: 0.8708860759493671

Confusion Matrix:

```
[[46 1 5 11]
 [ 5 51 2 4]
 [ 5 2 19 1]
 [42 20 6 96]]
```

Figure 2: Confusion Matrix of S-1 model for Gaussian Naive Bayes



Classification Report:

Classification Report:

Table 2 : classification report that show the final accuracy of the Gaussian Naive Bayes

	precision	recall	f1-score	support
apple scab	0.47	0.73	0.57	63
black rot	0.69	0.82	0.75	62
cedar apple rust	0.59	0.7	0.64	27
healthy	0.86	0.59	0.7	164
accuracy			0.87	316
macro avg	0.65	0.71	0.68	316
weighted avg	0.72	0.87	0.68	316

The analysis of the Random Forest Classifier on the Test Dataset with an Epoch Size of 100 resulted in exemplary performance, as detailed below:

The classifier's performance was evaluated using the confusion matrix, which provides information about the classifier's accuracy for each class. The diagonal elements of the matrix represent the number of correctly classified instances for each class, while the off-diagonal elements represent misclassifications. The Random Forest Classifier achieved an accuracy of 1.0 on the test data, indicating that it correctly classified all instances in each class. The precision and recall metrics were also used to assess the classifier's performance. Precision measures the accuracy of positive predictions, while recall measures the ability to correctly identify positive instances. In this case, both metrics had a value of 1.00 for each class,

indicating perfect performance. The F1-score, which is the harmonic mean of precision and recall, also had a value of 1.00 for each class.

Random Forest Classifier on Test Dataset (Epoch Size = 100):

Accuracy: 1.0

Confusion Matrix:

```
[[ 63  0  0  0]
 [  0 62  0  0]
 [  0  0 27  0]
 [  0  0  0 164]]
```

Classification Report:

Table 3 : Classification Report of Random Forest Classifier

	Precision	Recall	F1-Score	Support
apple scab	1.00	1.00	1.00	63
black rot	1.00	1.00	1.00	62
cedar apple rust	1.00	1.00	1.00	27
healthy	1.00	1.00	1.00	164
accuracy			1.00	316
macro avg	1.00	1.00	1.00	316
weighted avg	1.00	1.00	1.00	316

Gaussian Naive Bayes Classifier : The accuracy of the Gaussian Naive Bayes Classifier was found to be approximately 87.09%. From the confusion matrix, it can be observed that the classifier struggled particularly with distinguishing between "apple scab" and "healthy" classes, as well as "black rot" and "healthy" classes. Based on the classification report, it can be observed that the precision, recall, and F1-score differed for each class, with the "healthy" class achieving the highest values.

The Random Forest Classifier outperformed the Gaussian Naive Bayes Classifier in all performance metrics, achieving a perfect score for accuracy, precision, recall, and F1-score. It performed flawlessly on the test dataset. On the other hand, the Gaussian Naive Bayes Classifier exhibited lower accuracy and F1-scores, especially for the "apple scab" and "cedar apple rust" classes. The overall macro and weighted averages were higher for the Random Forest Classifier compared to the Gaussian Naive Bayes Classifier, further confirming its superior performance in classifying apple leaf diseases.

Gaussian Naive Bayes Classifier on Test Dataset (Epoch Size = 100):

Accuracy: 0.8708860759493671

Confusion Matrix:

```
[[46  1  5 11]
 [ 5 51  2  4]
 [ 5  2 19  1]
 [42 20  6 96]]
```

Classification Report:

```
precision recall f1-score support
apple scab    0.47    0.73    0.57     63
black rot     0.69    0.82    0.75     62
cedar apple rust 0.59    0.70    0.64     27
healthy       0.86    0.59    0.70    164

accuracy                0.87    316
macro avg              0.65    0.71    0.87    316
weighted avg          0.72    0.87    0.68    316
```

Random Forest Classifier: The Random Forest Classifier with an Epoch Size of 150 achieved flawless performance on the test dataset, demonstrating its robustness and reliability in predicting leaf diseases.

Random Forest Classifier on Test Dataset (Epoch Size = 150):

Accuracy: 1.0

Confusion Matrix:

```
[[ 63  0  0  0]
 [  0 62  0  0]
 [  0  0 27  0]
 [  0  0  0 164]]
```

Classification Report:

	precision	recall	f1-score	support
apple scab	1.00	1.00	1.00	63
black rot	1.00	1.00	1.00	62
cedar apple rust	1.00	1.00	1.00	27
healthy	1.00	1.00	1.00	164
accuracy		1.00		316
macro avg	1.00	1.00	1.00	316
weighted avg	1.00	1.00	1.00	316

Gaussian Naive Bayes : The Gaussian Naive Bayes Classifier exhibited moderate performance on the test dataset with an Epoch Size of 150. While it achieved reasonable accuracy, it struggled with certain classes, particularly in distinguishing between "apple scab," "black rot," and "healthy" classes. Further optimization or exploration of alternative classifiers may be necessary to improve performance, especially for these challenging classes.

Gaussian Naive Bayes Classifier on Test Dataset (Epoch Size = 150):

Accuracy: 0.8708860759493671

Confusion Matrix:

```
[[46  1  5 11]
 [ 5 51  2  4]
 [ 5  2 19  1]
 [42 20  6 96]]
```

Classification Report:

	precision	recall	f1-score	support
apple scab	0.47	0.73	0.57	63
black rot	0.69	0.82	0.75	62
cedar apple rust	0.59	0.70	0.64	27
healthy	0.86	0.59	0.70	164
accuracy		0.87		316
macro avg	0.65	0.71	0.87	316
weighted avg	0.72	0.87	0.68	316

The Random Forest Classifier : It performed exceptionally well on the test dataset with an Epoch Size of 200, achieving a perfect accuracy of 1.0. The confusion matrix indicates that all instances were correctly classified, with each class having a precision, recall, and F1-score of 1.00, demonstrating perfect performance across the board. This high level of accuracy and precision suggests that the classifier effectively distinguished between different leaf disease classes, including "apple scab," "black rot," "cedar apple rust," and "healthy" classes. The Random Forest Classifier is a reliable and effective model for predicting leaf diseases, making it a valuable tool for classification tasks.

Random Forest Classifier on Test Dataset (Epoch Size = 200):

Accuracy: 1.0

Confusion Matrix:

```
[[ 63  0  0  0]
```



[0 62 0 0]
[0 0 27 0]
[0 0 0 164]]

Classification Report:

	precision	recall	f1-score	support
apple scab	1.00	1.00	1.00	63
black rot	1.00	1.00	1.00	62
cedar apple rust	1.00	1.00	1.00	27
healthy	1.00	1.00	1.00	164
accuracy		1.00		316
macro avg	1.00	1.00	1.00	316
weighted avg	1.00	1.00	1.00	316

Gaussian Naive Bayes Classifier, It trained on the Test Dataset with an Epoch Size of 200, achieved an accuracy of approximately 87.09%. The confusion matrix reveals the classifier's performance for each class, indicating its struggles in distinguishing between "apple scab" and "healthy" classes, as well as "black rot" and "healthy" classes.

The report includes a breakdown of precision, recall, and F1-score metrics for each class. It was observed that the values for precision, recall, and F1-score varied for each class, with the "healthy" class having the highest values.

Gaussian Naive Bayes Classifier on Test Dataset (Epoch Size = 200):
Accuracy: 0.8708860759493671

Confusion Matrix:

[[46 1 5 11]
[5 51 2 4]
[5 2 19 1]
[42 20 6 96]]

Classification Report:

	precision	recall	f1-score	support
apple scab	0.47	0.73	0.57	63
black rot	0.69	0.82	0.75	62
cedar apple rust	0.59	0.70	0.64	27
healthy	0.86	0.59	0.70	164
accuracy		0.87		316
macro avg	0.65	0.71	0.87	316
weighted avg	0.72	0.87	0.68	316

This analysis compared the performance of two machine learning models, Random Forest Classifier and Gaussian Naive Bayes, for classifying apple leaf diseases. The Random Forest Classifier achieved outstanding results, reaching a perfect accuracy of 1.0 on the test dataset for multiple epoch sizes (50, 100, 150, and 200). This indicates its exceptional ability to distinguish between different disease classes. Conversely, the Gaussian Naive Bayes Classifier displayed moderate performance with an accuracy around 87%, struggling to differentiate between certain classes, particularly "apple scab", "black rot", and "healthy".

IV. CONCLUSION

In this study researcher evaluated the Random Forest Classifier and the Gaussian Naive Bayes Classifier, for classifying apple leaf diseases. Across different epoch sizes ranging from 50 to 200, the Random Forest Classifier consistently demonstrated exceptional performance, achieving perfect accuracy of 1.0 for all epochs. Its precision, recall, and F1-scores were all 1.00, indicating flawless performance across all disease classes. Conversely, the Gaussian Naive Bayes Classifier exhibited moderate performance, with an accuracy of approximately 87.09% across all epochs. While it struggled with certain disease classes, particularly "apple scab" and "black rot," it still demonstrated reasonable precision, recall, and F1-scores for most classes.

The Random Forest Classifier emerged as the superior model for apple leaf disease classification in this study. Its consistent and flawless performance across various epoch sizes highlights its robustness and reliability. Future studies can focus on

enhancing the performance of the Gaussian Naive Bayes Classifier or exploring alternative algorithms that might yield even more effective results. This study signifies the potential of machine learning for precise and efficient apple leaf disease detection. The Random Forest Classifier exhibited exceptional effectiveness in this particular application.

REFERENCES

- [1] Overview. (n.d.). World Bank.(2023) <https://www.worldbank.org/en/topic/agriculture/overview>
- [2] Ristaino, J. B., (2021). The persistent threat of emerging plant disease pandemics to global food security. Proceedings of the National Academy of Sciences, 118(23). <https://doi.org/10.1073/pnas.2022239118>
- [3] M. Carvajal-Yepes et al. (2019). A Global surveillance system for crop diseases: Global preparedness minimizes the risk to food supplies. *Science* 364, 137–1239.
- [4] HE, D. C., ZHAN, J. S., & XIE, L. H. (2016). Problems, challenges and future of plant disease management: from an ecological point of view. *Journal of Integrative Agriculture*, 15(4), 705–715. [https://doi.org/10.1016/s2095-3119\(15\)61300-4](https://doi.org/10.1016/s2095-3119(15)61300-4)
- [5] Dong, X., Wang, Q., Huang, Q., Ge, Q., Zhao, K., Wu, X., Wu, X., Lei, L., & Hao, G. (2023). PDDD-PreTrain: A Series of Commonly Used Pre-Trained Models Support Image-Based Plant Disease Diagnosis. *Plant Phenomics*, 5, 0054. <https://doi.org/10.34133/plantphenomics.0054>
- [6] Chincinska IA. Leaf infiltration in plant science: old method, new possibilities. *Plant Methods*. 2021 Jul 28;17(1):83. doi: 10.1186/s13007-021-00782-x. PMID: 34321022; PMCID: PMC8316707.
- [7] Srinivasa Gupta, Venkata ramana (2022). Detection of Plant Leaf Diseases Using Random Forest Classifier. *International Journal of Innovative Research in Technology*, 9(1), 1300. ISSN: 2349-6002.
- [8] Xian, T. S., & Ngadiran, R. (2021). Plant diseases classification using machine learning. In *Journal of Physics: Conference Series* (Vol. 1962, No. 1, p. 012024). In The 1st International Conference on Engineering and Technology (ICoEngTech) (pp. 012024) IOP Publishing..
- [9] S. M. Jaisakthi, P. Mirunalini, D. Thenmozhi and Vatsala, "Grape Leaf Disease Identification using Machine Learning Techniques," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862084.
- [10] Islam MS, Sultana S, Farid FA, Islam MN, Rashid M, Bari BS, Hashim N, Husen MN. Multimodal Hybrid Deep Learning Approach to Detect Tomato Leaf Disease Using Attention Based Dilated Convolution Feature Extractor with Logistic Regression Classification. *Sensors* (Basel). 2022 Aug 14;22(16):6079. doi: 10.3390/s22166079. PMID: 36015839; PMCID: PMC9414454.
- [11] Saberi Anari M. A Hybrid Model for Leaf Diseases Classification Based on the Modified Deep Transfer Learning and Ensemble Approach for Agricultural AIoT-Based Monitoring. *Comput Intell Neurosci*. 2022 Apr 5;2022:6504616. doi: 10.1155/2022/6504616. PMID: 35422854; PMCID: PMC9005283.
- [12] Nawaz, M., Nazir, T., Javed, A., Masood, M., Rashid, J., Kim, J., & Hussain, A. (2022). A robust deep learning approach for tomato plant leaf disease localization and classification. *Scientific reports*, 12(1), 18568. <https://doi.org/10.1038/s41598-022-21498-5>
- [13] Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants*, 8(11), 468. <https://doi.org/10.3390/plants8110468>
- [14] Borhani, Y., Khoramdel, J., & Najafi, E. (2022). A deep learning based approach for automated plant disease classification using vision transformer. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-15163-0>
- [15] Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants*, 8(11), 468. <https://doi.org/10.3390/plants8110468>
- [16] Balaji, V., Anushkannan, N. K., Sujatha, C. N., Rattan, P., Verma, D., Awasthi, D. K., Pandian, A. P., Veeramanickam, M. R. M., & Mulat, M. B. (2023). Deep transfer learning technique for multimodal disease classification in plant images. *Contrast Media & Molecular Imaging*, 2023, 1–8. <https://doi.org/10.1155/2023/5644727>
- [17] Ghorbel, M. (2023, December 24). Revolutionizing Agriculture with AI: A Deep Dive into Machine Learning for Leaf Disease Classification and Smart Farming. *MarkTechPost*. <https://www.marktechpost.com/2023/12/24/revolutionizing-agriculture-with-ai-a-deep-dive-into-machine-learning-for-leaf-disease-classification-and-smart-farming/>
- [18] Gheorghiu, A. (2023). Evaluating data augmentation techniques for coffee leaf disease classification. *arXiv.org*. <https://arxiv.org/abs/2401.05768>
- [19] Aishwarya, M., & Reddy, P. P. (2023). Dataset of groundnut plant leaf images for classification and detection. *Data in Brief*, 48, 109185. <https://doi.org/10.1016/j.dib.2023.109185>
- [20] Wu, Q., Ma, X., Liu, H., Bi, C., Yu, H., Liang, M. P., Zhang, J., Li, Q., Tang, Y., & Ye, G. (2023). A classification method for soybean leaf diseases based on an improved ConvNeXt model. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-46492-3>
- [21] Sherly Puspha Annabe(2019), Machine Learning for Plant Leaf Disease Detection and Classification – a review. (2019, April 1). *IEEE Conference Publication | IEEE Xplore*. <https://ieeexplore.ieee.org/abstract/document/8698004>
- [22] Shruthi, U., Nagaveni, V., & Raghavendra, B. K. (2019). A review on machine learning classification techniques for plant disease detection. In 2019 5th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 1467-1470). *IEEE*. <https://doi.org/10.1109/ICACCS.2019.8728415>
- [23] S. Kumar, K. Prasad, A. Srilekha, T. Suman, B. P. Rao and J. N. Vamshi Krishna, "Leaf Disease Detection and Classification based on Machine Learning," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 2020, pp. 361-365, doi:10.1109/ICSTCEE49637.2020.9277379.