IECAG 2025 Conference

The 5th International Electronic Conference on Agronomy

15–18 December 2025 | Online

MDPI

White Star Line SLU

VitiNet – An open-set framework for OOD-robust grape leaf disease classification

Bruna Condé[1] and Raúl Moreno[1]

[1]White Star Line SLU, Seville, Spain

INTRODUCTION & AIM

Grape leaf diseases are a major threat to viticulture, causing significant yield and quality losses worldwide[1,2]. Still, disease detection in vineyards is still being performed mostly manually, leading to misidentification, ineffective treatments, and unnecessary chemical inputs[2]. Ultimately, it results in harming both fruit quality and productivity as well as the environment[2].

Deep learning has emerged as a promising solution, enabling automated image-based classification of plant diseases[2,4]. Convolutional neural networks (CNN), including architectures such as ResNet, have demonstrated satisfactory performance in controlled datasets[3,4]. Nevertheless, the models' increasing computational demands hinder development and deployment in mobile or resource-constrained vineyard contexts[5]. Additionally, they are vulnerable to non-grape leaves, vineyard tool and complex backgrounds. Those are commonly classified as out-of-distribution (OOD) inputs[6]. Conventional classifiers trained only on curated images may give overconfident but incorrect predictions in theses scenarios, limiting their real-word applicability[6].

Therefore, the aim of this project was to: i) develop a fast, accurate, non-invasive, and accessible tool for grape leaf disease detection in vineyard conditions; ii) implement an open-set deep learning framework that integrates public datasets and OOD-awareness to improve robustness and generalisation; iii) establish a lightweight CNN backbone (e.g. ResNet18) to serve as scaffolding for future domain adaptation with real-world field data.

METHOD

Computing resources
EuroHPC MareNostrum5 (BSC, Spain)
PyTorch

Datasets
- Plant Village (PV)
- Niphad Grape Leaf Disease (NGLD)
- COCO
- private images
- Final: ~50.000 images

OOD assembly
- private images
- Non-grape leaf PV images & background class

Preprocessing
- Duplicates removal
- Overlay composites (PV/NGLD masks on COCO backgrounds),
- Heavy augmentation (albumentations)
- Stratified train/validation/test splits

Training
ResNet18

Pre-Training
- Plant Village (PV)
- Full training

Stage 1
- PV + NGLD
- freeze backbone
- train head
- unfreeze backbone
- full training

Stage 2
- PV + NGLD + OOD
- freeze backbone
- train head
- freeze backbone (except last block)
- train head + 4th layer

Optimisation/Regularization
- AdamW optimzer
- Early stopping

Evaluation
Validation set
Independent test set

Performance metrics
- F1-score
- Precision-recall
- Accuracy
- Confusion matrix

Interpretability
- Gradient-weighted Class Activation Mapping (Grad-CAM)

Prediction
- Confidence score distribution
- Temperature scaling (T-certainty)

Deployment
- Classification threshold

RESULTS & DISCUSSION

The final model achieved an overall test accuracy of 98.2% and a weighted F1-score of 0.982. The model showed robust OOD rejection, with the engineered 'other' class reaching 99.6% precision. Nevertheless, the primary limitation was the Bacterial_rot class, which showed the lowest F1-score (0.807). Analysis of the confusion matrix (Fig. 1) revealed this weakness was originated from a two-way confusion with the OOD class. The largest source of misclassifications originates from the 'OOD class being misclassified as Bacterial_rot, where 9 counts were observed (Fig. 1). Grad-CAM analysis showed the model attended to brown, necrotic like textures and edges similar to veins present in the OOD images, resulting in a high confidence false positive for Bacterial_rot (Fig. 2). On the other hand, the model it activates on the symptomatic region, but the global appearance was judged as atypical, leading to a false negative (Fig 3).

Analysis of prediction confidence showed the need for a safety mechanism. While the lowest confidence for a correct prediction was 0.44, the highest confidence for an incorrect prediction was extremely high at 0.99. Further analysis uncovered the confidence threshold of 0.96 was required to achieve the target 95% recall for disease classes (Fig. 3).

Ultimately, the model weaknesses were mitigated by selecting a threshold where ambiguous predictions are flagged for human review, to make sure only reliable results are automated. Consequently, two thresholds were suggested for deployment: T-disease = 0.96 and T-other = 0.99.

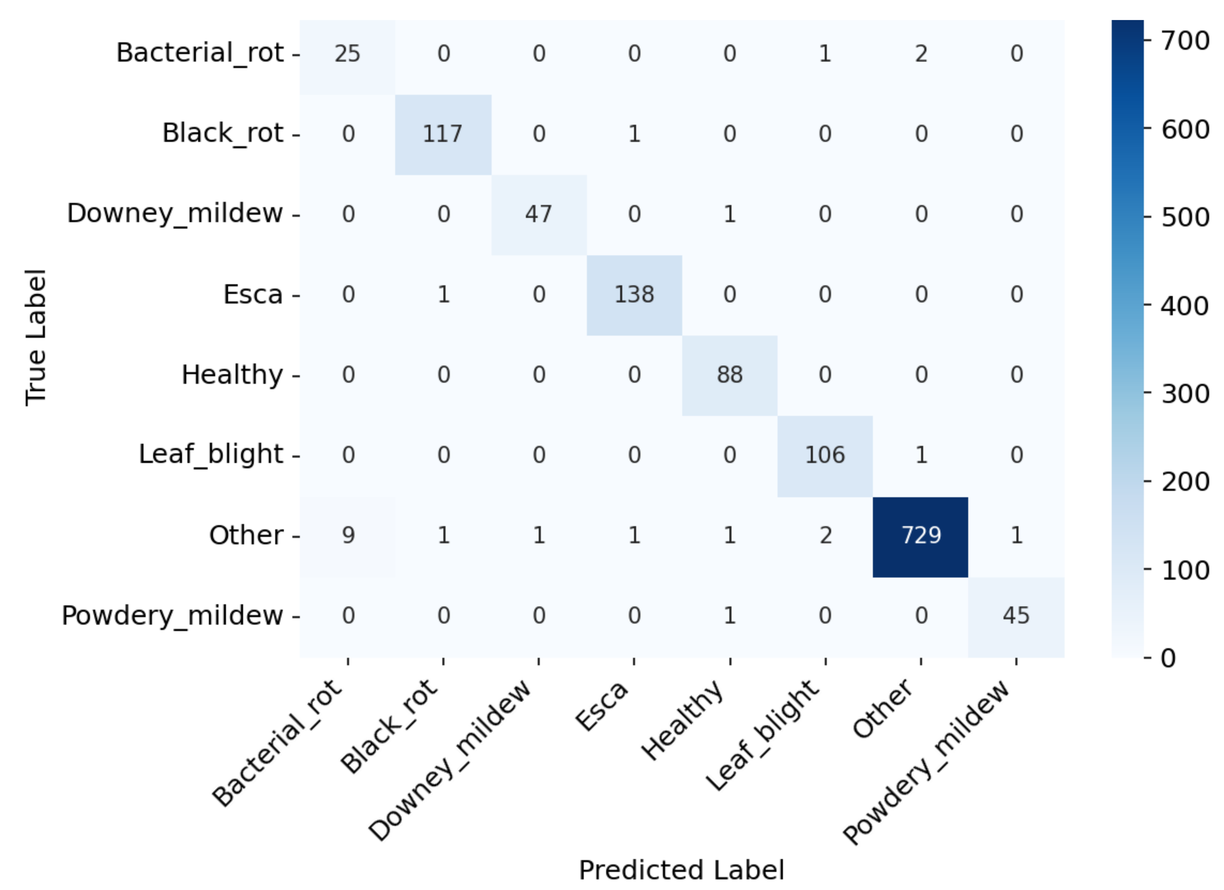

Figure 1. Confusion matrix of the final model on the test set. The matrix shows the counts of true labels (y-axis) versus the predicted label (x-axis). The values represent the total count of image; lighter colour means lower counts, while darker colours means higher counts.

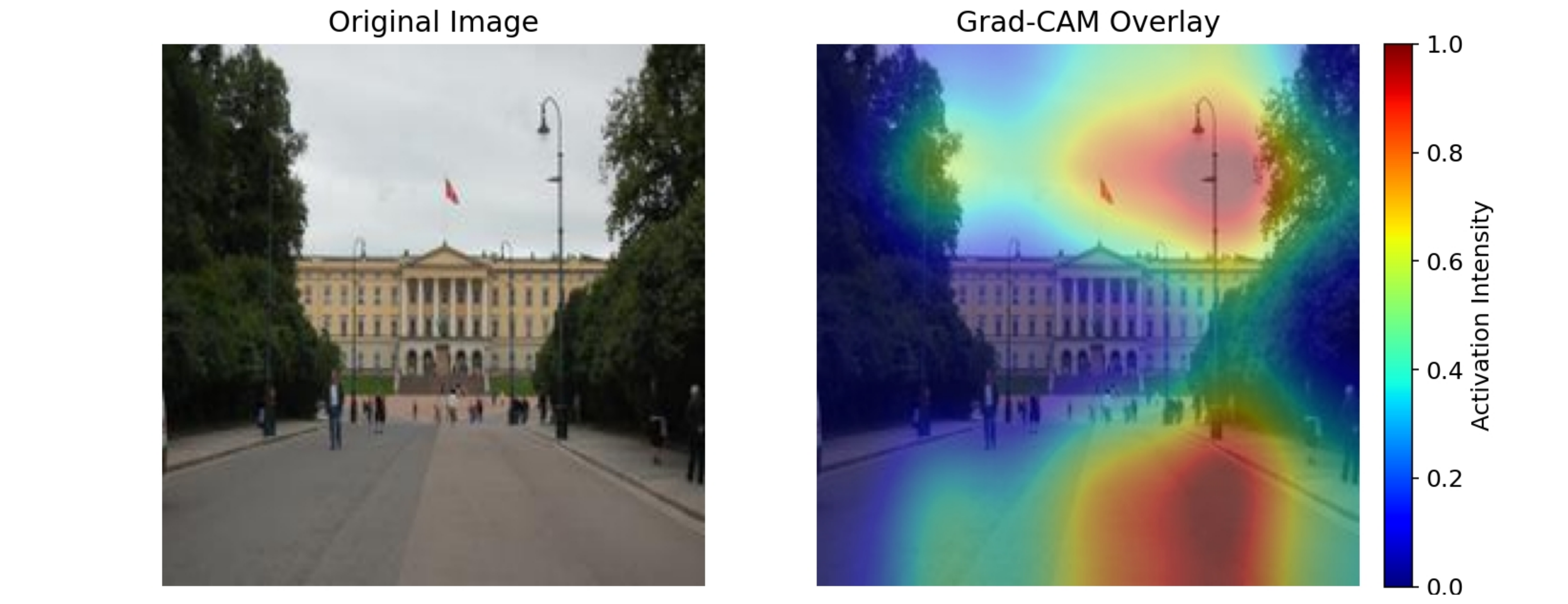

Figure 2. (a) Original image of 'Other'; **(b)** Corresponding Grad-CAM visualizaton shows the model activation intensity. True: Other | Predicted: Bacterial_rot (confidence: 0.96)

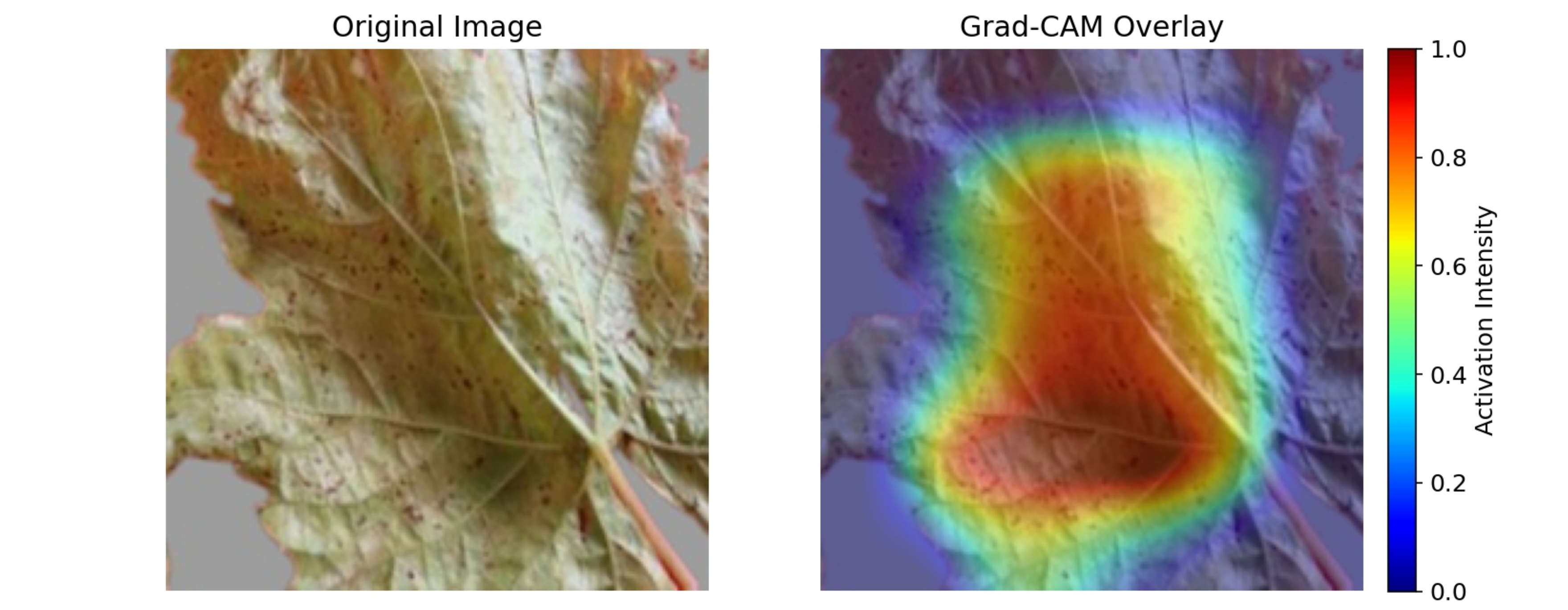

Figure 3. (a) Original image of Bacterial_rot; **(b)** Corresponding Grad-CAM visualization shows the model activation intensity. True: Bacterial_rot | Predicted: Other (confidence: 0.96)

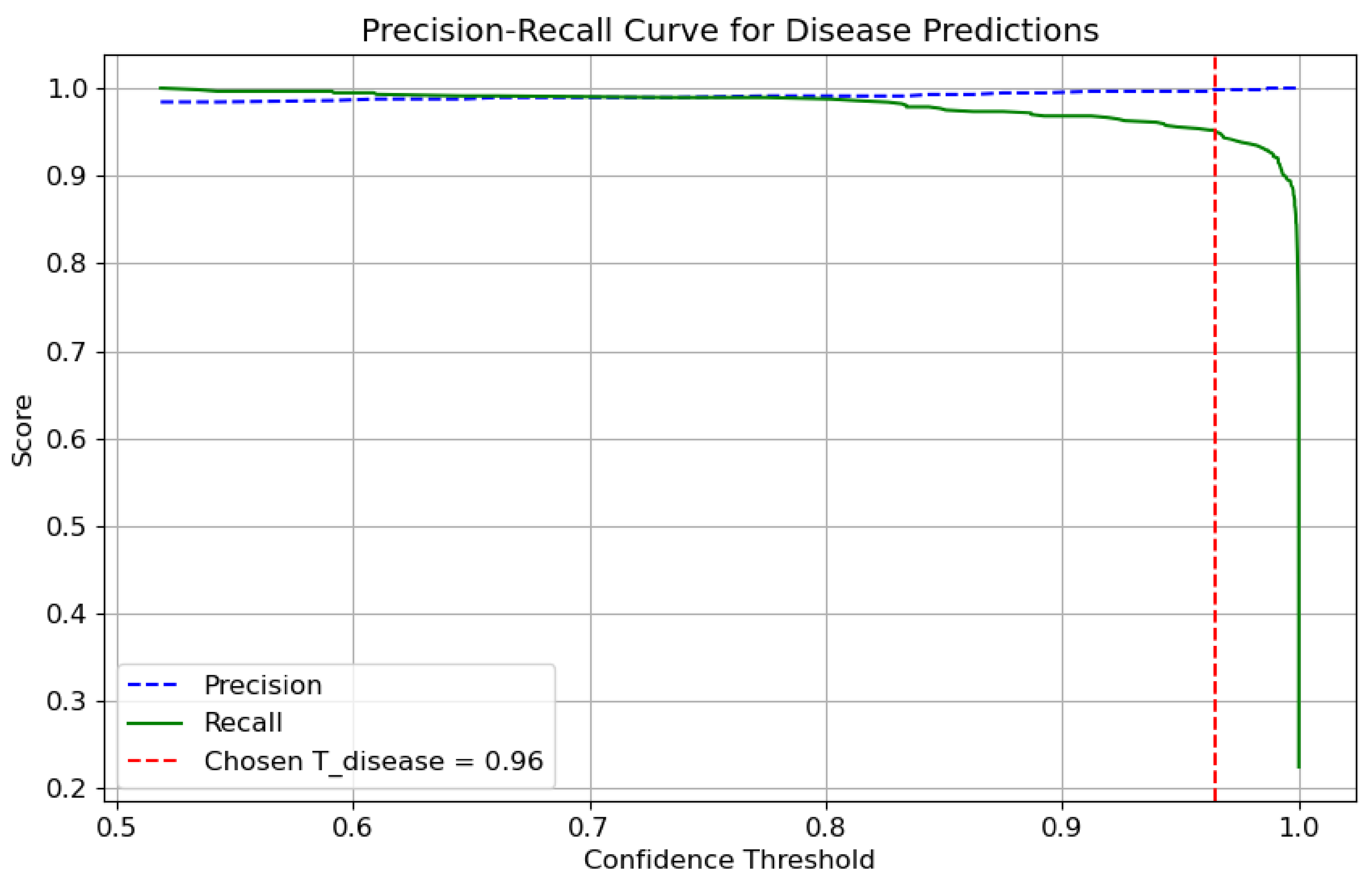

Figure 4. **Precision-Recall curve for disease class predictions on the test set.** The vertical red line marks the threshold T_disease = 0.96, where recall is approximately 0.95.

CONCLUSIONS / FUTURE WORK

This project successfully developed VitiNet, a research grape leaf disease classifier that is fast, accurate, and accessible to grape growers. The combination of staged fine-tuning, integration of a OOD class and risk mitigating thresholding protocol for deployments resulted in a successful model Further improvements should focus on: i) collecting more diverse field images of Bacterial_rot; ii) explore other techniques to address class imbalance; and iii) continuously improve the OOD class with real world vineyard data.

REFERENCES

[1]Gessler, C., Pertot, I., & Perazzolli, M. (2011). Plasmopara viticola: a review of knowledge on host-pathogen interaction, epidemiology and disease control. Plant Pathology, 60(1), 3-15.

[2]Mahlein, A. K. (2016). Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping. Plant disease, 100(2), 241-251.

[3]He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[4]Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in plant science, 7, 1419.

[5]Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017).

[6]Hendrycks, D., & Gimpel, K. (2017). A baseline for detecting misclassified and out-of-distribution examples in neural networks. In International Conference on Learning Representations.