44. Small data deep learning methodology for in-field disease detection: Case study of potato crops

D. Herrera-Poyatos[1], J. Domínguez-Rull[1], R. Montes[1], I. Hernández[2], I. Barrio[2], J. Tardaguila[2], F. Herrera[1] and A. Herrera-Poyatos[1]
[1]*Andalusian Institute of Data Science and Computational Intelligence (DaSCI), Spain*
[2]*Televitis Research Group, University of La Rioja, 26006, Logroño, Spain*

Abstract

Early detection of diseases in crops is essential to prevent harvest losses and improve the quality of the final product. In this context, the combination of machine learning and proximity sensors is emerging as a technique capable of achieving this detection efficiently and effectively. This approach has been applied to potato crops to detect late blight (*Phytophthora infestans*). However, most of these AI models found in the specialised literature have been developed using leaf-by-leaf images taken in the lab, which does not represent field conditions and limits their applicability. This study presents the first machine learning model capable of detecting mild symptoms of late blight in potato crops through the analysis of high-resolution RGB images captured directly in the field, overcoming the limitations of other publications in the literature and presenting real-world applicability.

Keywords: computer vision, disease detection, deep learning, small data

Introduction

Agricultural production plays a significant role in the global economy and food supply. However, it faces serious challenges such as climate change, along with diseases and pests that can devastate entire crops, directly affecting both quality and quantity of the fruit and vegetables produced (Liu and Wang, 2021; Thakur *et al*., 2022). Among these diseases, potato late blight (*Phytophthora infestans*) is a fungus that manifests as dark spots on the leaves of potato plants. Under humid conditions, a plant affected by late blight can deteriorate rapidly. Additionally, the tubers of infected plants rot quickly when stored for consumption or replanting, making treatment of potato late blight critical for agricultural producers. It is important to note that climate change has intensified the occurrence and pressure of late blight on crops due to changes in temperature and humidity across many agricultural regions (Gullino *et al*., 2021). Consequently, the use of pesticides to prevent the appearance of late blight has increased considerably, leading to collateral damage such as contamination of food products, as well as significant CO_2 emissions from increased pesticide production and a notable rise in pesticide-related costs for farmers (Heinrich-Böll-Stiftung, 2022). For this reason, the European Union has developed legislation to reduce pesticide use by 50% by 2030 (European Commission Food Safety, 2020).

In this context, the primary tool to minimise pesticide use is the early detection of diseases and pests, so that pesticide treatments can be applied with varying intensities across different areas of the land according to the potential incidence of disease in those zones (Cucak *et al*., 2021). Conventional methods of visual inspection by farmers to detect diseases and pests, though prevalent, are fraught with challenges such as the labour intensity involved and the subjectivity and potential for human error. Moreover, due to the small size of symptoms, early stages of an infestation can easily go unnoticed, and not all plants are affected simultaneously. Therefore, there is a need to develop automatic and precise detection methods.

In an effort to improve the detection and management of agricultural diseases, non-invasive technologies have been explored for crop protection. Proximal sensing and, in particular, RGB image analysis of crops using machine learning has emerged as a vital tool, providing an economical and straightforward alternative for visual disease identification. However, this tool encounters challenges such as interference from external conditions and the high complexity of the images to be analysed. For this reason, previous scientific studies on potato late blight have focused on applying machine learning techniques to synthetic datasets, where images have been taken in a laboratory setting, leaf by leaf (Mahum *et al.*, 2023; Oppenheim *et al.*, 2019; Rashid *et al.*, 2021; Van De Vijver, 2020), limiting their applicability.

This study introduces the In-field Small Data Disease Detection Deep Learning methodology, abbreviated ISD^{4}L that, using high-resolution RGB images acquired in the field, allows for the detection and localisation of diseases and pests in potato crops. This proposal stands out for the combination of several state-of-the-art computer vision techniques to address the complexities of images taken under field conditions, representing a significant advancement over the state of the art. Namely, the ISD^{4}L methodology employs the concept of patching to exploit high-resolution, incorporates a novel data augmentation scheme that allows machine learning algorithms to perform under the small data paradigm, and utilises convolutional neural networks with focal loss that prioritise more complicated examples during training. The obtained model correctly classifies 21 out of 22 high-resolution images of the TelevitisPotatoDiseases data set.

Materials and methods

The dataset TelevitisPotatoDiseases

This work introduces the dataset TelevitisPotatoDiseases, a database of 22 images of potato plants taken in the field, 9 of which show symptoms of late blight (*Phytophthora infestans*).

These images are high-resolution (4000×6000 pixels), providing enough detail to visualise early symptoms of late blight. The images were taken from two experimental commercial potato plots located in northern Spain (Basque Country and La Rioja) during the 2022 season, using a Sony Alpha 7-II (Sony, Tokyo, Japan) mirrorless RGB digital camera equipped with a Zeiss 24/70 mm lens with optical stabilization. Figure 1 shows an image of this dataset, showcasing that some symptoms are small and difficult to detect with the human eye. For each infected plant, the dataset includes a segmentation map of the plant and symptoms, where symptoms of late blight are highlighted in red. The information regarding symptoms contained in this map is key for the development of machine learning models.

(a) (b)

Figure 1. Example of an instance in the TelevitisPotatoDiseases dataset. The images were taken with the camera pointing towards the ground, from a height of around one meter. (a) Potato plant with symptoms of late blight. (b) Segmentation of the plant and symptoms.

The ISD^{4}L methodology: In-field Small Data Disease Detection with Deep Learning

It is well known that traditional CNNs may struggle to detect fine-grained patterns when the original images are massively down-scaled to the network input size. Thus, when high-resolution images are available, splitting these images into patches (smaller subimages) and training models on these patches has become a standard technique in classification based on fine-grained patterns, see, for instance, (Herrera-Poyatos, 2024). This approach has been shown to be promising in the case of disease detection; see (Hernández, 2024). More precisely, in (Herrera-Poyatos, 2024; Hernández, 2024), high-resolution images in the training set are split into patches to generate the new training set of patches, where further data augmentation may be applied if necessary. Then, at prediction time, new high-resolution images are split into patches, potentially with a sliding window method, and prediction is obtained by aggregating the evaluations of the trained model on these new patches. In the case of disease detection, this aggregation is extremely simple, the high-resolution image presents symptoms of a disease if one of its patches does. One can highlight two issues with this approach:

1. Limited image augmentation. Once high-resolution images have been split into patches to generate the training set of patches, current methods further increase the size of the training set mainly by performing symmetries and rotations on each patch. However, if one performs rotations with an arbitrary angle θ, the rotated patch may not fit into a square anymore, and thus, the remaining pixels must be filled with blanks/reflections, thus losing information and compromising performance of the trained models. Another approach is restricting rotations to the angles π/2, π, 3π/2, which limits the amount of data augmentation.
2. Amplification of false positives. High accuracy of a model trained on patches may not translate to high accuracy of predictions for the high-resolution images. Indeed, if a high-resolution image does not present symptoms of a disease, a false positive on a patch translates to a false positive for the full image prediction. Since each high-resolution image is split into multiple patches at the prediction stage, the probability of a false positive occurring is significantly amplified.

The methodology introduced in this work, namely the ISD^{4}L methodology, presents a solution for both problems. There are two key ingredients in this proposal. The first one is a novel data augmentation scheme that is able to extract a high number of patches from each high-resolution image without much redundancy and loss of information. The second one is the application of focal loss as loss function of the CNN architectures which forces the model to prioritise on the most complicated examples of the training set. Indeed, false positives tend to arise in those patches that present a complex mix of leaves, trunks and background.

This section is organised as follows. First, we describe the patch generation algorithm of the ISD^{4}L methodology, then we delve into the training phase, introducing the focal loss, and finally we explain the prediction phase of the ISD^{4}L methodology.

Data augmentation phase: random patch generation

As mentioned at the beginning of this section, the main idea to leverage the high resolution of the images in classification based on fine-grained patterns is to extract patches (smaller subimages), on which CNNs are trained. The ISD^{4}L methodology extracts random patches from high-resolution images with different characteristics: size, rotation angle and position. This maximises the amount of information extracted from each image when generating the training set of patches. Concretely, for each image in the training set, ρ patches are extracted. The parameter ρ could be optimised, and has been set to ρ=200 in the experiments presented here. Each patch is extracted as follows:

1. Random Rotation. The high-definition image is rotated by an angle θ, where θ is a random number between −π and π (for clarification, here π corresponds to 180 degrees).
2. Random Zoom. The size of the patch is randomly selected. To do this, the ISD^{4}L methodology samples a number t between $0.15l$ and $0.25l$, where l is the number of pixels on the shortest side of the image. In the TelevitisPotatoDiseases data set, l=4000. The extracted patch will have size t times t.

3. Random patch extraction. The rotated high-resolution image is cropped to the largest rectangle of pixels that does not include blanks. Then, a random square, with t pixels per side, is extracted from the rotated high-resolution image.

Patches are labelled as healthy or affected by a disease as follows. If the patch is extracted from a high-resolution image that shows no symptoms (it has been labelled as healthy), then this will be the label for the patch. If the patch has been extracted from a high-resolution image that presents symptoms, then the same transformations used to extract the patch are performed to the segmentation mask of the image, extracting the same patch for this segmentation mask. If the segmentation map for the patch contains highlighted symptoms, then the patch is labelled according to the corresponding disease. Otherwise the patch is labelled as healthy.

The use of rotations and zoom ensures that the resulting model is not affected by the angle at which the image is taken or the distance from the sensor to the plant. Moreover, it increases the variety of the patches generated. Due to the randomness of this process, ρ could be significantly large if needed, although this would increase the training time.

Training phase: introducing focal loss

After data preprocessing, a CNN is trained on the patches extracted. Before training occurs patches in the training set must be scaled to the input size of the neural network. Indeed, note that the patches generated may have distinct size due to the random zoom operation. Note that the impact of downscaling patches is minor compared to the impact of downscaling the full high-resolution image. Here the input size of the utilised CNN model is 380x380, which was chosen to maximise the information provided to the neural network while maintaining a moderate size for the network layers to avoid overfitting and reduce computation time.

The CNN architecture used in this work is the EfficientNetV2 deep learning model (Tan and Le, 2021), which has shown excellent results in similar problems (Hernández, 2024). The authors of EfficientNetV2 propose seven different architectures depending on the input image size. In this work, the EfficientNetV2B3 model is used, where the input images are 380x380 pixels, to maximize the information available to the neural network while avoiding overfitting. Weights of the model are initialised using transfer learning to the weights of EfficientNetV2B3 that performed best on ImageNet. The code used in this work uses the libraries TensorFlow and Keras. We modified the last three layers of EfficientNetV2B3, replacing them with a pooling layer, a dense layer with ReLU activation, and, in the final layer, a single neuron with a sigmoid activation function, which outputs a confidence score (between 0 and 1) estimating if the patch shows symptoms of late blight.

Traditionally, the CrossEntropy (CE) loss function is used in binary classification due to its desirable mathematical properties. However, when hard-to-classify examples are the minority, these difficult examples do not sufficiently affect the loss function, and the model ends up making significant errors on them. This behaviour is exacerbated in our problem, where a single false positive with a large error is enough to classify a high-resolution image without symptoms as a plant affected by late blight.

To avoid false positives with large errors, we propose using the focal loss function with parameters α=0.5 and γ=2 (Johnson, 2019). This loss function is a modification of CE that is used in problems with class imbalance. The parameter $\alpha \in (0, 1/2]$ is a weight applied to the majority class, while the minority class is weighted by $1 - \alpha$. In this case, the default value 0.5 is used, as extreme class imbalance is not an issue. More important is the parameter γ, which adjusts the weight assigned to an instance based on the error made and, as a consequence, examples with lower error contribute even less to the loss function. Let I be an instance in the dataset, c_I {0,1} the true class of instance I, and p_I the probability, according to the model, that I belongs to class 1. The focal loss function with α=0.5 is given by

$$\mathrm{Focal}_I\,(p) = -c_I \log(p_I)(1 - p_I)^{\gamma} - (1 - c_I) \log(1 - p_I) p_I{}^{\gamma} \quad (1)$$

Prediction phase
This section explains how the ISD^{4}L methodology applies a deep learning model, trained on patches, to detect late blight in a high-resolution image. One initial idea is to divide the high-resolution image into patches using a grid and evaluate the model on them. This algorithm has the following problem: pixels on the edge of a patch tend to have less relevance in the final prediction. To solve this problem, the sliding window method is applied. Suppose the high-resolution image has $n \times m$ pixels (in this case n=4000 and m=6000), and that we want to extract patches of size $t \times t$, where t is an even number that is a divisor of n. In the sliding window method, for each pair of positive integers $i \in [0, 2(n/t-1)]$ and $j \in [0, 2(\lfloor m/t \rfloor-1)]$, the method extracts the subimage of size t whose top-left vertex has coordinates ($it/2$, $jt/2$). In our model, we choose $t=n/5=800$. In total, we apply the deep learning model to $(2n/t - 1)(2\lfloor m/t \rfloor - 1)=117$ patches of the original image.
Recall that, for each subimage I, the CNN returns the predicted confident score that the subimage shows symptoms of late blight. Let $\hat{p_I}$ denote the output of the network. The probability that the original image presents symptoms of late blight, denoted by P, is at least the probability that a particular subimage shows symptoms. Therefore, if the predictions were correct, we could take $max_I\ \hat{p_I}$ as a lower bound for P. To avoid the effect of a single false positive from the CNN radically affecting the evaluation of the original image, ISD^{4}L includes a threshold for the final classification. That is, the model decides that a high-resolution image presents symptoms of late blight if and only if $max_I\ \hat{p_I} \geq 0.8$. This approach reduces the impact of local failures on the final model.

Experimental framework, results and analysis

The experiment was conducted using the EfficientNetV2B3 model, which has proven effective in image classification tasks due to its optimized architecture in terms of size and computational efficiency. The model was configured with an input size of 380x380 pixels, trained for 100 epochs for each fold with a batch size of 32 on a Nvidia A100-SXM4 GPU with 40GB of memory. The original dataset TelevitisPotatoDiseases consists of 22 high-resolution images of potato plants. Recall that the parameter ρ=200 indicates the number of patches extracted of each high-resolution image. In order to obtain statistical significant results, given the size of the data set, the proposal is evaluated using Leave-One-Out validation. Here we briefly discuss this validation approach in this experimentation, given the fact that one has to differentiate between high-resolution images and patches. First, ρ patches are extracted from each high-resolution image using the ISD^{4}L methodology. Then, for each high-resolution image I, a model is trained on the patches of the rest of the high-resolution images (21ρ=4200 patches in total), and the resulting model is evaluated on the patches of I, obtaining both an accuracy of prediction at patch level and a prediction for the high-resolution image (which is Late Blight if one of the patches has this as a prediction with probability at least 0.8, and healthy otherwise). This iteration is called a fold. Table 1 shows the accuracy of the trained CNN on each one of the folds of the leave-one-out validation. Note that most folds present an extremely high accuracy, except for a couple of folds, where accuracy drops due to the fact that there is a significant difference between the image where the model is evaluated and the images in the training set. We hypothesize that this issue would be resolved with a larger training set that represents wider in-field complexities.
Table 2 shows the prediction of the leave-one-out validation for each high-resolution image. Recall that a threshold of 0.8 has been applied here, that is, a high-resolution image is classified as *Late Blight* if and only if one of the patches has been classified as *Late Blight* by the CNN with probability at least 0.8. This criterion minimises the incidence of false positives in the overall prediction. The results indicate a significant number of correctly classified images, demonstrating the model's reliability and effectiveness in distinguishing between healthy and infected plants. Indeed, the model correctly classifies 21 out of the 22 images, only making one mistake classifying a healthy plant as *Late Blight* due to committing a false positive on a patch, whose confident score exceeded the 0.8 threshold.

Table 1. Accuracy of prediction (at patch level) for each fold of the leave-one-out validation.

Fold	Accuracy
1	1
2	0.984
3	0.9127
4	0.992
5	1
6	0.770
7	0.992
8	0.992
9	0.976
10	1
11	1
12	1
13	0.984
14	1
15	0.992
16	1
17	0.944
18	1
19	0.992
20	0.96
21	0.873
22	1

The average accuracy among all folds is 0.9711.

Table 2. Count of images with and without pest and the correctness of the prediction among all folds of Leave-One-Out Cross Validation training with focal loss.

Image class	Correct prediction	Incorrect prediction
Late blight	9	0
Healthy	12	1

The results obtained from the leave-one-out cross-validation yield a 0.9545 accuracy in classifying high-resolution images of potato plants based on the presence or absence of *late blight* symptoms.

Conclusions

This work introduces the ISD^{4}L methodology, addressing the development of deep learning models to detect potato plants infected with late blight in high-resolution RGB images taken directly in the field. This methodology includes data augmentation scheme based on patches that permits the training of models using only a few high-resolution images in the context of small data. Moreover, it exploits the concept of focal loss function to focus learning on the hardest instances. The developed model following the ISD^{4}L methodology correctly detects all cases of late blight in leave-one-out validation, indicating a high level of accuracy and effectiveness in detecting the disease under field conditions, and only makes a false positive error on images of healthy plants. Moreover, it presents an accuracy of detection of late blight at patch level of 97.11%. These promising results reinforce the potential use of machine learning for the early detection and localisation of diseases and pests in agriculture, which will enable better treatment and reduction of their impact on crops. Future work to make this technology applicable in real-world scenarios should focus on the development of a large dataset of high-resolution images taken under field conditions that represents all possible complexities of the problem, including multiple diseases and pests that may affect potato crops.

Acknowledgements

This work has been developed as part of the project NoPest (Novel Pesticides for a Sustainable Agriculture), which received funding from the European Union Horizon 2020 FET Open program under Grant agreement ID 828940.

References

Cucak, M., Andrade-Moral, R., Fealy, R., Lambkin, K., & Kildea, S. (2021). Opportunities for improved potato late blight management in the Republic of Ireland: Field evaluation of the modified Irish rules crop disease risk prediction model. *Phytopathology*, 111(8), 1349–1360.

European Commission Food Safety. Farm to fork targets. (2020). Available online at https://food.ec.europa.eu/plants/pesticides/sustainable-use-pesticides/farm-fork-targets-progress_en (accessed January 2025).

Gullino, M.L., Albajes, R., Al-Jory, I., Angelotti, F., Chakraborty, S., Garrett, K.A., Hurley, B.P., Juroszek, P., Makkouk, K., Pan, X., & Stephenson, T. (2021). Scientific review of the impact of climate change on plant pests. FAO, Rome, on behalf of the IPPC Secretariat.

Hernández, I., Gutiérrez, S., Barrio, I., Íñiguez, R., & Tardáguila, J. (2024). In-field disease symptom detection and localisation using explainable deep learning: use case for downy mildew in grapevine. Computers and Electronics in Agriculture, 226, 109478.

Herrera-Poyatos, D., Herrera Poyatos, A., Montes Soldado, R., De Palacios, P., Esteban, L., García Iruela, A., García Fernández, F., & Herrera, F. (2024). Deep learning methodology for the identification of wood species using high-resolution macroscopic images. Non-peer reviewed preprint at arXiv: 2406.11772. https://doi.org/10.48550/arXiv.2406.11772

Heinrich-Böll-Stiftung (2022). Pesticide atlas: Facts and figures about toxic chemicals in agriculture. Heinrich-Böll-Stiftung, Berlin. Available online at https://eu.boell.org/en/PesticideAtlas (accessed January 2025).

Johnson, J.M., & Khoshgoftaar, T. (2019). Survey on deep learning with class imbalance. Journal of Big Data, 6(1), 1–54.

Liu, J., & Wang, X. (2021).Plant diseases and pests detection based on deep learning: a review. Plant Methods, 17, 1–18.

Mahum, R., Munir, H., Mughal, Z., Awais, M., Khan, F.S., Saqlain, M., Mahamad, S., & Tlili, I. (2023). A novel framework for potato leaf disease detection using an efficient deep learning model. Human and Ecological Risk Assessment, 29(2), 303–326.

Oppenheim, D., Shani, G., Erlich, O., & Tsror, L. (2019). Using deep learning for image-based potato tuber disease detection. Phytopathology, 109(6), 1083–1087.

Rashid, J., Khan, I., Ali, G., Almotiri, S.H., AlGhamdi, M.A., & Masood, K. (2021). Multi-level deep learning model for potato leaf disease recognition. Electronics, 10(17), 2064.

Tan, M., & Le, Q. (2021). EfficientNetV2: Smaller Models and Faster Training. In Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 10096–10106.

Thakur, P.S., Khanna, P., Sheorey, T., & Ojha, A. (2022). Trends in vision-based machine learning techniques for plant disease identification: A systematic review. Expert Systems with Applications, 208, 118117.

Van De Vijver, R., Koen, M., Heungens K., Somers, B., Nuyttens, D., Borra-Serrano, I., Lootens, Roldán-Ruiz, I., Vangeyte, J., & Saeys, W. (2020). In-field detection of alternaria solani in potato crops using hyperspectral imaging. Computers and Electronics in Agriculture, 168, 105106.