Stage - Internship M2

Explainable AI for Enhanced Visual Interpretability in image-based applications

- **Summary**

This project funded by O3T project (Région Occitanie) aims to develop an ensemble approach to explainable AI for image-based applications. We will consider the case of plant disease detection, combining techniques like CAM, Grad-CAM, and Saliency Maps. By improving model interpretability, the approach aims to provide clearer, more accurate explanations of AI predictions. In the case of plant disease detection, this will help farmers and agronomists better understand diagnoses, enhancing trust in AI systems and improving plant health management. The methods developed aim to be generic enough to be applied in other application contexts.

- **Description**

Deep learning has emerged as a powerful tool for automatic classification e.g. for detection of plant diseases through image analysis. Models like convolutional neural networks (CNNs) have demonstrated remarkable performance in classifying images or image parts by learning complex patterns from images. However, despite their accuracy, these models are often criticized for being "black boxes," offering little insight into how they arrive at their decisions. This lack of interpretability restricts their practical deployment, especially in domains where trust in AI systems is paramount.

Explainable AI (XAI) addresses this challenge by providing insights into the decision-making processes of AI models. Techniques such as Class Activation Mapping (CAM), Gradient-weighted CAM (Grad-CAM), and Saliency Maps create visual explanations, highlighting the regions in an image that are most influential in a model prediction. For example, in the case of plant disease detection, these methods can show which parts of a leaf image led the model to classify it as "infected with a fungal disease". However, each XAI method has limitations. For instance, CAM Requires architectural modifications to the model, which may not always be feasible; Grad-CAM Relies on gradient information, which can sometimes lead to noisy or overly generalized explanations; Saliency Maps Can produce explanations that are challenging to interpret visually.

Relying on a single XAI method often results in explanations that may miss critical aspects or fail to resonate with end-users, such as farmers or agronomists. Drawing inspiration from ensemble learning—a technique in machine learning that combines the strengths of multiple models to improve overall performance—this project proposes creating an ensemble of XAI methods. By integrating outputs from CAM, Grad-CAM, and Saliency Maps, the goal is to deliver richer, more accurate, and user-friendly explanations.

The project focuses on developing an ensemble approach to explainable AI techniques—such as CAM, Grad-CAM, and Saliency Maps—to improve the interpretability and reliability of deep learning models used in plant disease detection. The aim is to analyze how combining multiple explanation methods can provide a richer and more comprehensive understanding of model predictions, aiding in the diagnosis and management of plant health.

The developed approaches will be validated on state-of-the-art plant disease detection datasets, e.g., **plant_leaves**, **PlantVillage**, and **plantae_k**, as well as on a custom dataset we constructed.

- **Objectives:**

To advance this project, we aim to recruit an intern who will contribute to achieving the following objectives:

1. **Literature Review:** Analyze existing XAI techniques (e.g., CAM, Grad-CAM, Saliency Maps) and their applications in plant disease detection.
2. **Design an Ensemble XAI Framework:**
    - Combine multiple XAI techniques to generate unified and richer explanations of model predictions.
    - Define metrics to evaluate the quality and usability of the explanations in agricultural applications.
3. **Implementation and validation:**
    - Apply the framework to benchmark datasets of plant disease images (e.g., **plant_leaves**, **PlantVillage**, and **plantae_k**) and compare its performance with individual XAI methods.
    - Work on validating the framework on our wheat disease dataset, ensuring the severity scores generated align closely with those of human operators.
    - Investigate whether finer representations of ensemble XAI are improving model insights.
    - Assess whether a better understanding of decision-making processes can enhance model generalization—a critical challenge in real-world use cases.

References:

- Dwivedi, R., et al. (2024). An efficient ensemble explainable AI (XAI) approach for morphed face detection. *Pattern Recognition Letters*, *184*, 197-204.
- Selvaraju, R. R., et al. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE int. conference on computer vision* (pp. 618-626).

- Garouani, M., et al. (2023). Unlocking the Black Box: Towards Interactive Explainable Automated Machine Learning. In *Intelligent Data Engineering and Automated Learning – IDEAL 2023.*
- *Menya, E., Interdonato, R., Owuor, D., & Roche, M. (2024). Explainable epidemiological thematic features for event-based disease surveillance. Expert Systems with Applications, 250, 123894.*

- **Lab & Supervision**

~~Lab : In~~stitut de Recherche en Informatique de Toulouse (IRIT), Toulouse, France

Supervisors :

- Moncef GAROUANI ; Josiane Mothe IRIT Toulouse

- **Mathieu ROCHE CIRAD Montpellier**
- **Jordan BERNIGAUD-SAMATAN INRAE Montpellier**

Applications to be send to : moncef.garouani@irit.fr ; josiane.mothe@irit.fr

Applications consist of: CV including skills related to the project, transcripts.