

Développement de méthodes de *deep learning* sur des images hyperspectrales de grains de blé, pour la classification de variétés.

Contexte

L'agroécologie repose sur le développement de nouvelles diversités génétiques assurant des services écosystémiques (au-delà de la stabilité des rendements), avec des intrants limités (engrais et pesticides), tout en s'adaptant à la variabilité climatique et aux événements extrêmes dans le contexte du changement climatique en cours. Le mélange de plusieurs variétés au champ permet d'atténuer les effets de ces aléas, en offrant de multiples avantages : augmenter et stabiliser les rendements, réduire la pression des pathogènes et adventices, améliorer la qualité des sols.

Afin d'améliorer notre compréhension des interactions entre différentes variétés d'un mélange, de nombreuses expérimentations sont nécessaires pour évaluer les performances de différents mélanges de génotypes. Pour cela, il est essentiel de pouvoir mesurer individuellement chaque composante à l'aide de méthodes non destructives. Ainsi, la reconnaissance de variétés par analyse d'images de grains a été étudiée chez plusieurs espèces avec succès, dans le but, soit d'estimer la composition d'un mélange (e.g. pour évaluer la contribution de chaque génotype au rendement total), soit de la coupler avec un système de tri (e.g. pour pouvoir constituer un semi dans les proportions souhaitées). Cependant, les approches fondées sur l'imagerie RVB reposent principalement sur des caractéristiques morphologiques et colorimétriques, qui peuvent s'avérer insuffisantes pour discriminer des variétés morphologiquement proches.

L'originalité de ce stage réside dans l'exploitation de l'imagerie hyperspectrale (IHS) pour dépasser ces limites. L'IHS permet d'aller plus loin que l'imagerie RVB, car elle fournit une information spectrale liée aux propriétés biochimiques et structurales des tissus végétaux en plus des informations morphologiques. Cette technique, qui fournit une image basée sur la mesure de centaines de longueurs d'onde, montre un fort potentiel pour obtenir des modèles de classification, plus précis qu'avec l'imagerie RVB qui n'exploitent que trois longueurs d'onde.

Objectifs du stage

L'ambition de ce stage est de pouvoir déterminer automatiquement la variété de blé dur d'une graine sur la base de son image hyperspectrale. Pour cela, l'idée est d'ajuster un modèle de reconnaissance de variétés de blé dur qui sera entraîné sur un large panel représentatif de la diversité mondiale de cette espèce. Au cours du stage, l'étudiant.e développera donc des approches de classification basées sur des modèles de fondation disponibles et l'apprentissage auto-supervisé, en adaptant des stratégies de pré-entraînement issues de l'imagerie multi- et hyperspectrale afin d'obtenir des représentations spectrales robustes. Ces représentations seront ensuite exploitées pour estimer les proportions de chaque variété de blé dur dans un mélange.

Ce stage s'intègre dans le projet PEPR Past2eco (<https://www.pepr-agroeconum.fr/les-projets-finances/ressources-genetiques/laureat-aap/past2eco>) dont un des objectifs est la discrimination de variétés de blé. Ce projet, démarré en mars 2026 rassemble des chercheurs en génétique des populations, des spécialistes du phénotypage et des chercheurs en analyse d'images. L'étudiant.e sera pleinement intégré dans ce projet.

Le travail initié dans ce stage se poursuivra par une thèse déjà financée à partir de septembre 2027.

L'étudiant·e se verra confier les missions suivantes :

- Effectuer une étude bibliographique des méthodes de deep learning adaptées aux images hyperspectrales, et des travaux sur la discrimination d'espèces/variétés basée sur les images de grains.
- Réaliser des acquisitions d'images hyperspectrales de grains à partir d'échantillons du projet Past2eco présents dans les banques de graines de l'UMR AGAP Institut (~ 100 images de grain acquises en 10 min).
- Mettre en forme les images acquises (extraction des régions d'intérêt, prétraitements), et stocker les données dans un format optimisé pour l'entrée des modèles d'apprentissage.
- Proposer, en accord avec ces encadrants, les méthodes de classification les plus adaptées aux IHS, en intégrant l'information spatiale et spectrale, puis les mettre en œuvre sur les données acquises.
- Définir et appliquer une stratégie de validation des méthodes, afin d'optimiser les paramètres des modèles, et de garantir la robustesse des prédictions.

Formation :

Étudiant·e en dernière année d'École d'Ingénieur ou Master 2, dans le domaine des mathématiques, la science des données, l'intelligence artificielle, ou écologie/agroécologie avec une forte composante IA .

Lieux du stage et encadrement :

Equipe Ge2pop, Campus Lavalette à Montpellier, UMR AGAP, bâtiment ARCAD: 10 rue Arthur Young, Montpellier (<https://umr-agap.cirad.fr/recherches/equipes-de-recherche/genomique-evolutive-et-gestion-des-populations/contexte>).

Stage supervisé par :

Martin Ecarnot (INRAE), Vincent Ranwez (Institut Agro Montpellier), Gustavo Andrade-Miranda (IMT Mines Alès), Benjamin Bourel (INRIA de Montpellier).

Profil et compétences requises :

L'étudiant·e recruté·e devra avoir un bon niveau en programmation (Python), une connaissance approfondie des méthodes de machine learning, une bonne connaissance en « computer vision ». Il·elle saura travailler en collaboration et être à l'écoute des conseils de ses encadrants. Il·elle devra faire preuve de capacité d'écoute et de synthèse, d'initiative et de créativité. L'expérience d'exécution de scripts sur serveurs distants et utilisation du terminal sous Linux sera un plus. Enfin, bien que non indispensable, un ou plusieurs des éléments suivants seront appréciés : intérêt marqué pour l'interdisciplinarité , connaissance en botanique ou agroécologie.

Dates :

6 mois à partir de février 2027

Rémunération :

Gratification entre 660 € /mois

Contacts :

martin.ecarnot@inrae.fr ; gustavo.andrade-miranda@mines-ales.fr ; vincent.ranwez@supagro.fr ; benjamin.bourel@inria.fr