

# Le machine learning à l'usage de non-spécialistes

*Exemple de la plateforme EcoTaxa*



# EcoTaxa aujourd'hui

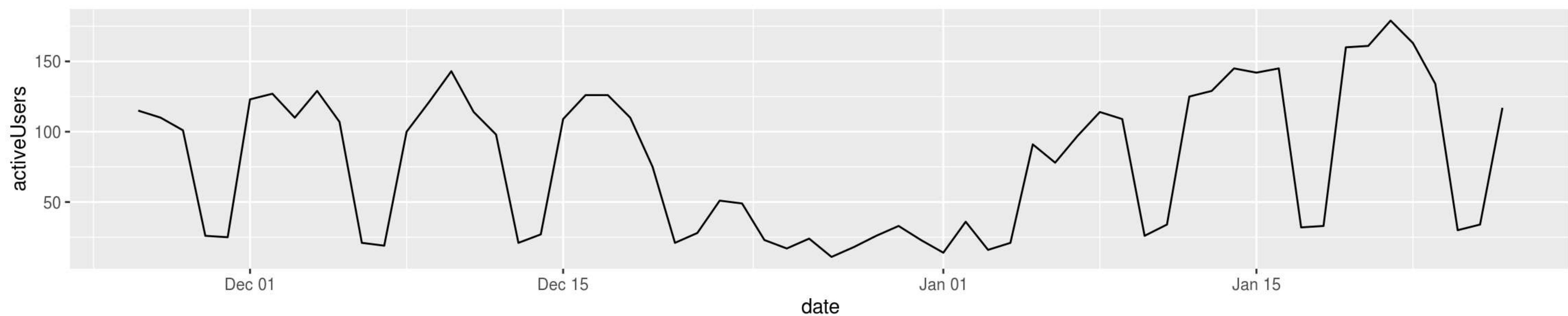
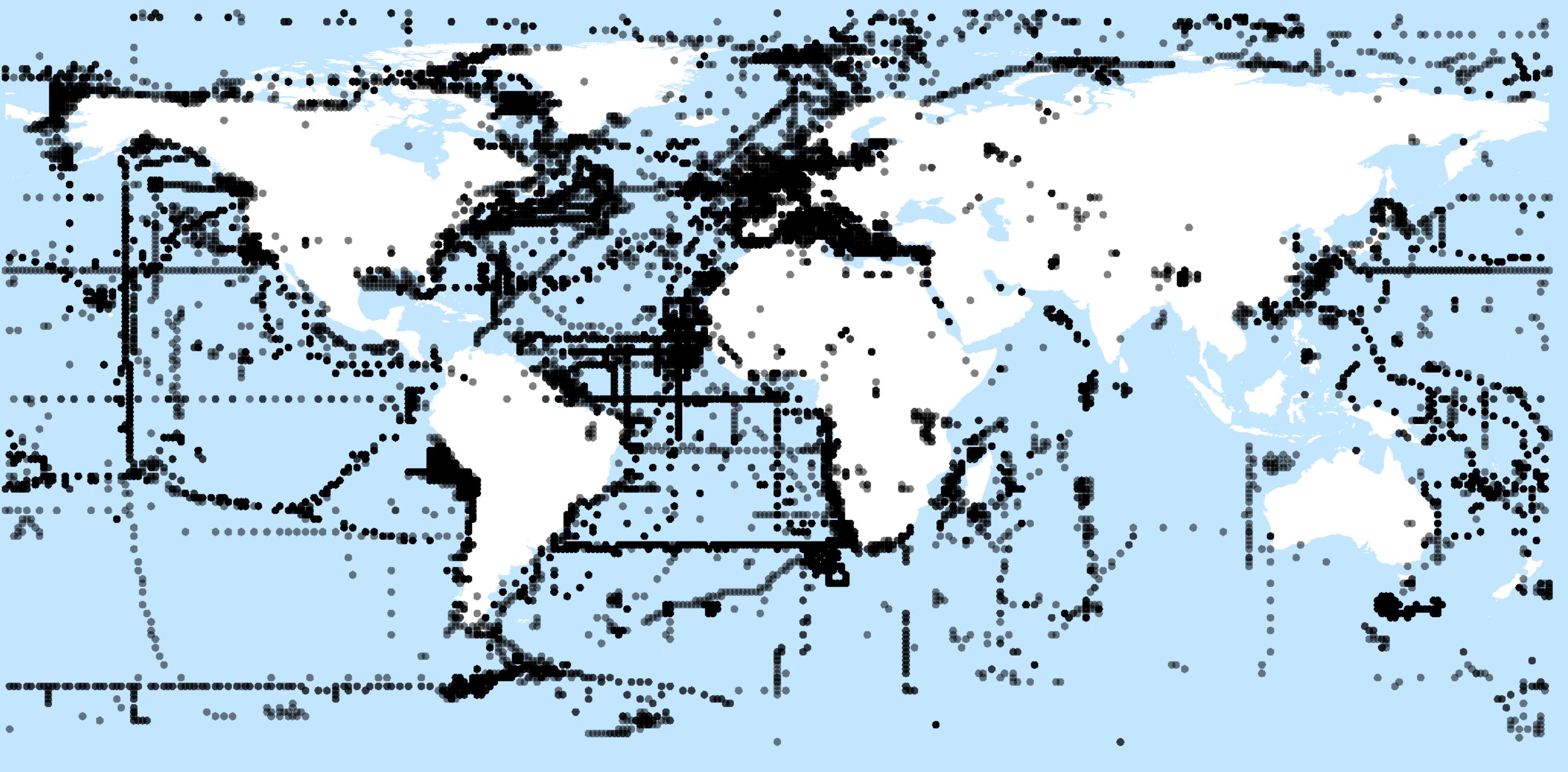
750M images, 250M validées par un opérateur humain

+15M par mois

5,000 to 10,000 validations/opérateur/j

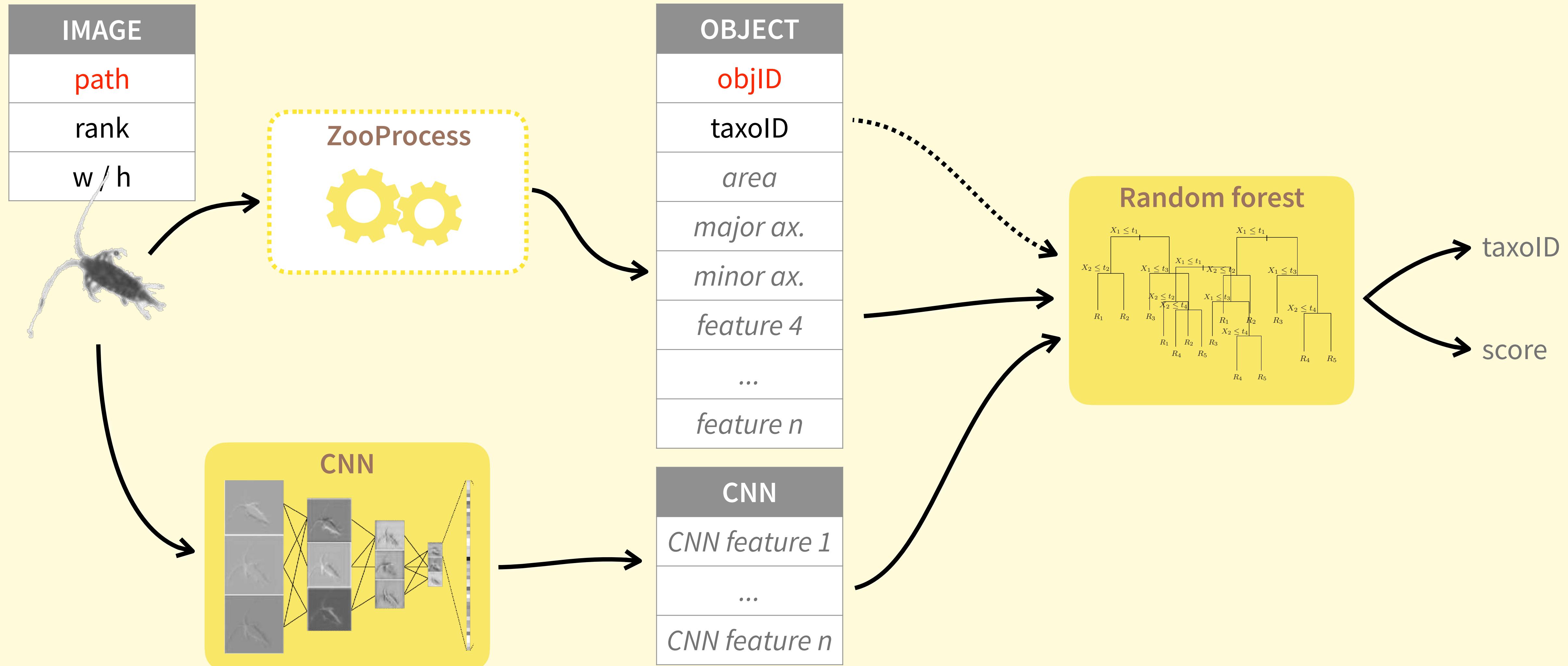
>4500 users from ~900 organisations

**~125 utilisateurs concurrents**

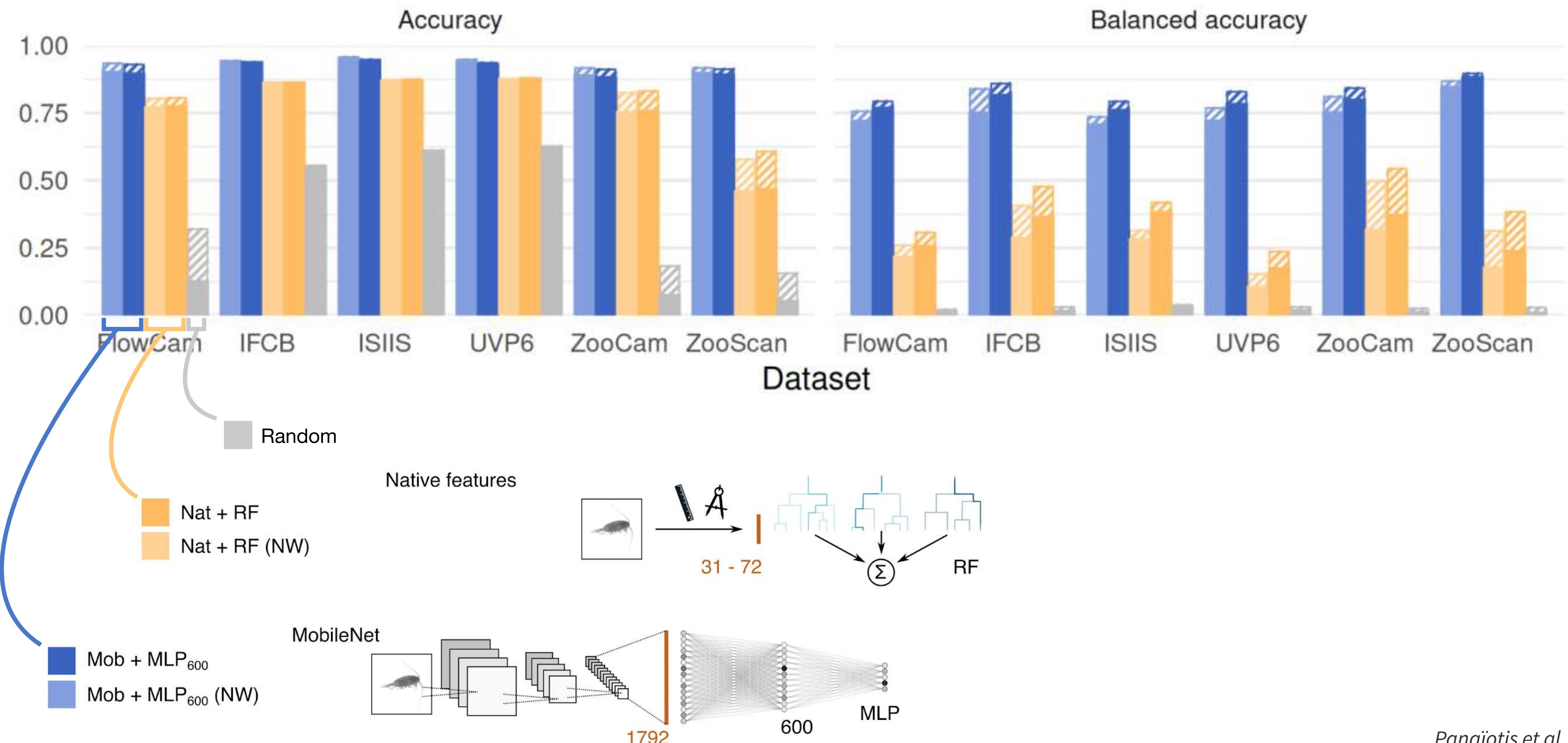




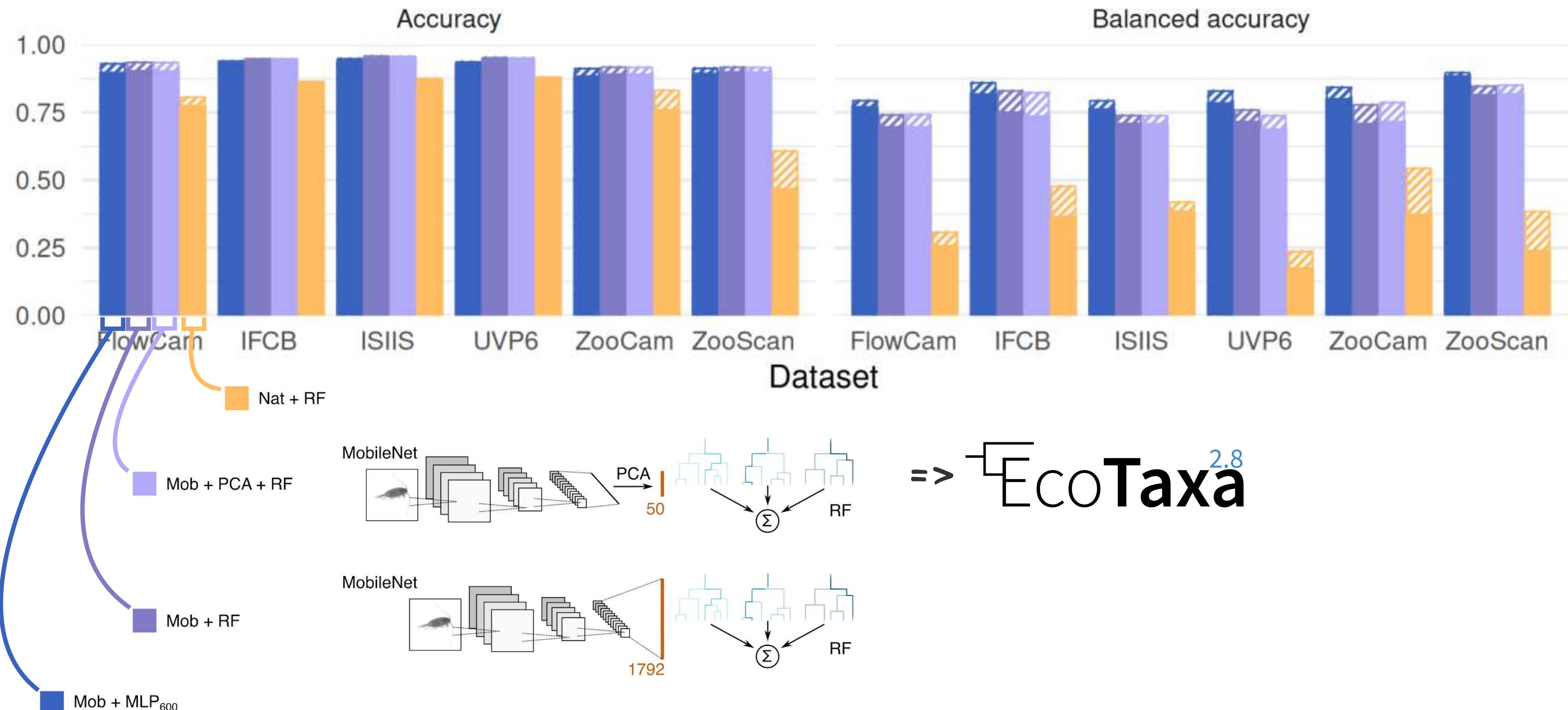
# L'architecture de machine learning



# Classique vs deep: le deep est meilleur *surtout pour les classes rares*



# Pourquoi? Ce sont les descripteurs, pas le classifieur



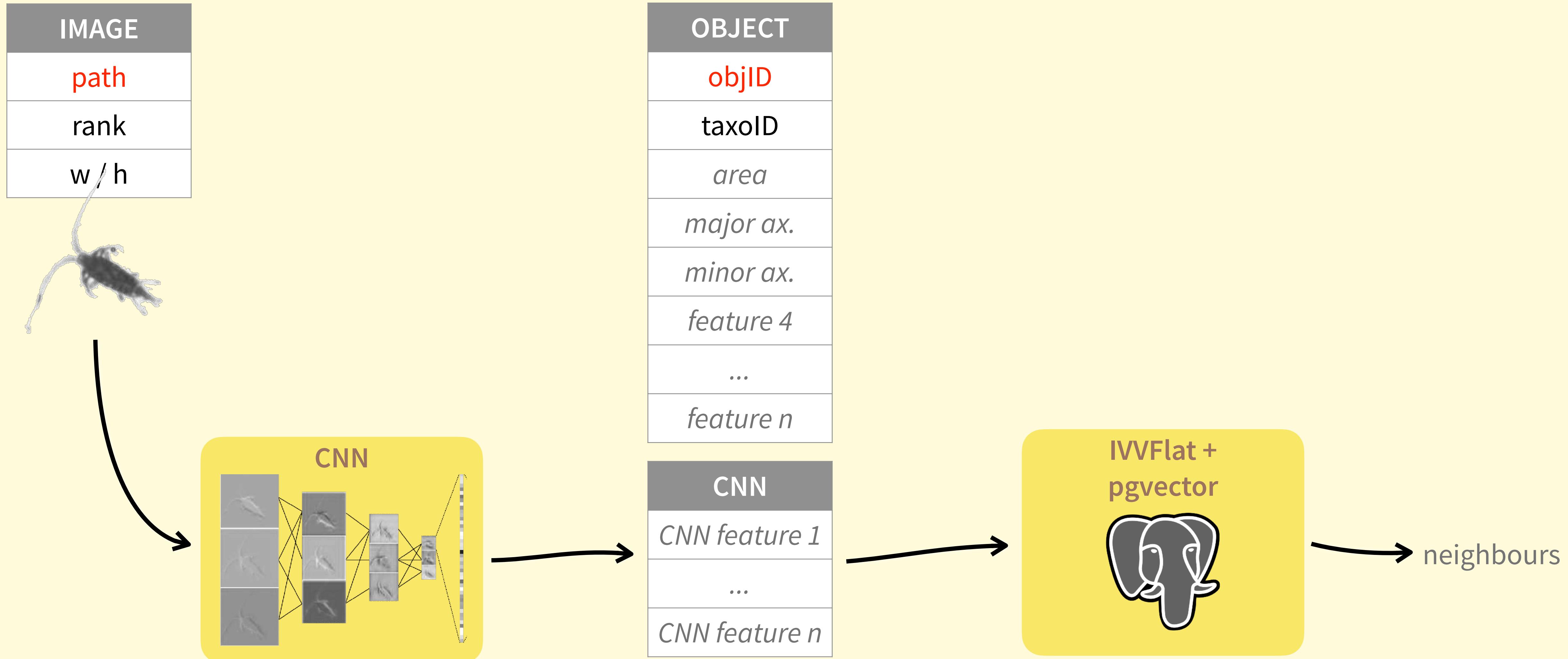
# Descripteurs natifs + PCA(descripteurs deep fixes) + RF ... Pourquoi?

1. **Simplicité (d'implémentation):** il suffisait d'ajouter des descripteurs deep à une infrastructure de classification déjà disponible 😊
2. **Rapidité:** entraîner un RF en parallèle est plus rapide qu'entraîner un CNN complet; fine tuner uniquement la partie dense d'un CNN (i.e. classifieur) n'a pas vraiment d'intérêt
3. **Robustesse:** facile de combiner n'importe quels descripteurs natifs et des descripteurs deep (pas sensible à la distribution)
4. **Stabilité:** RF a montré une grande stabilité de performance dans diverses configurations
5. **Calibration:** RF calcule, de façon native, des probabilités d'être dans chaque groupe. Celles-ci sont plus fiables et mieux distribuées que celles d'un MLP

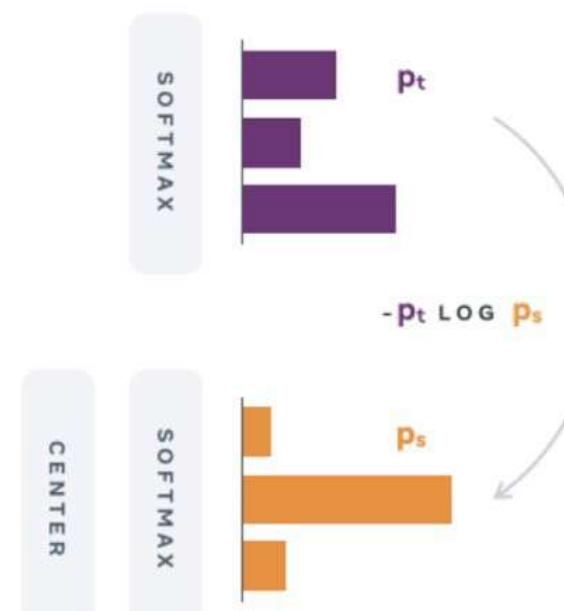
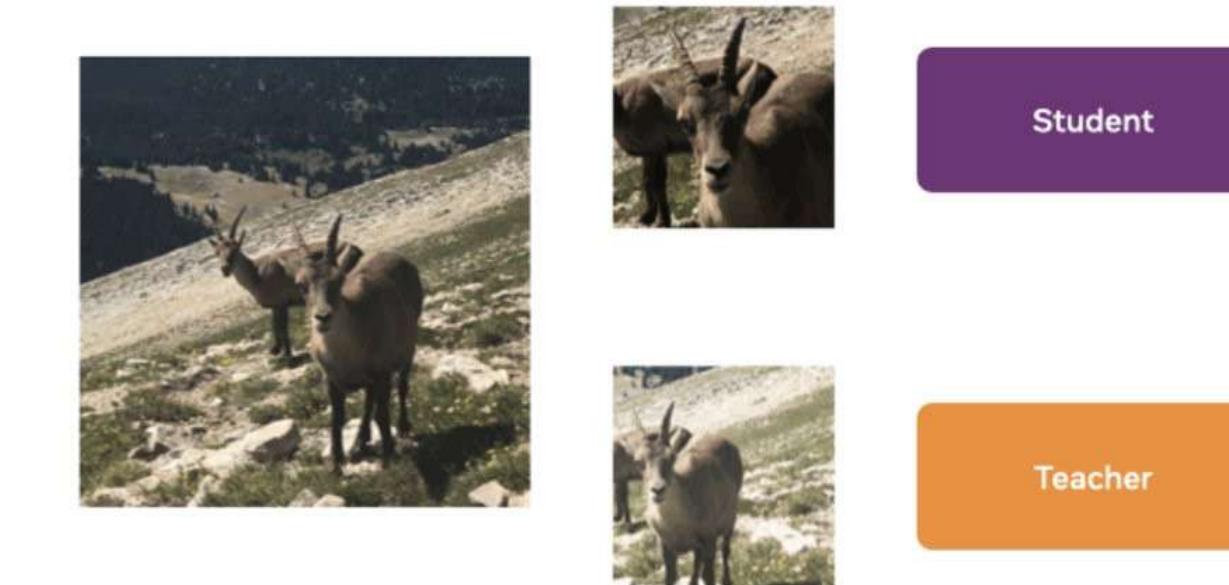
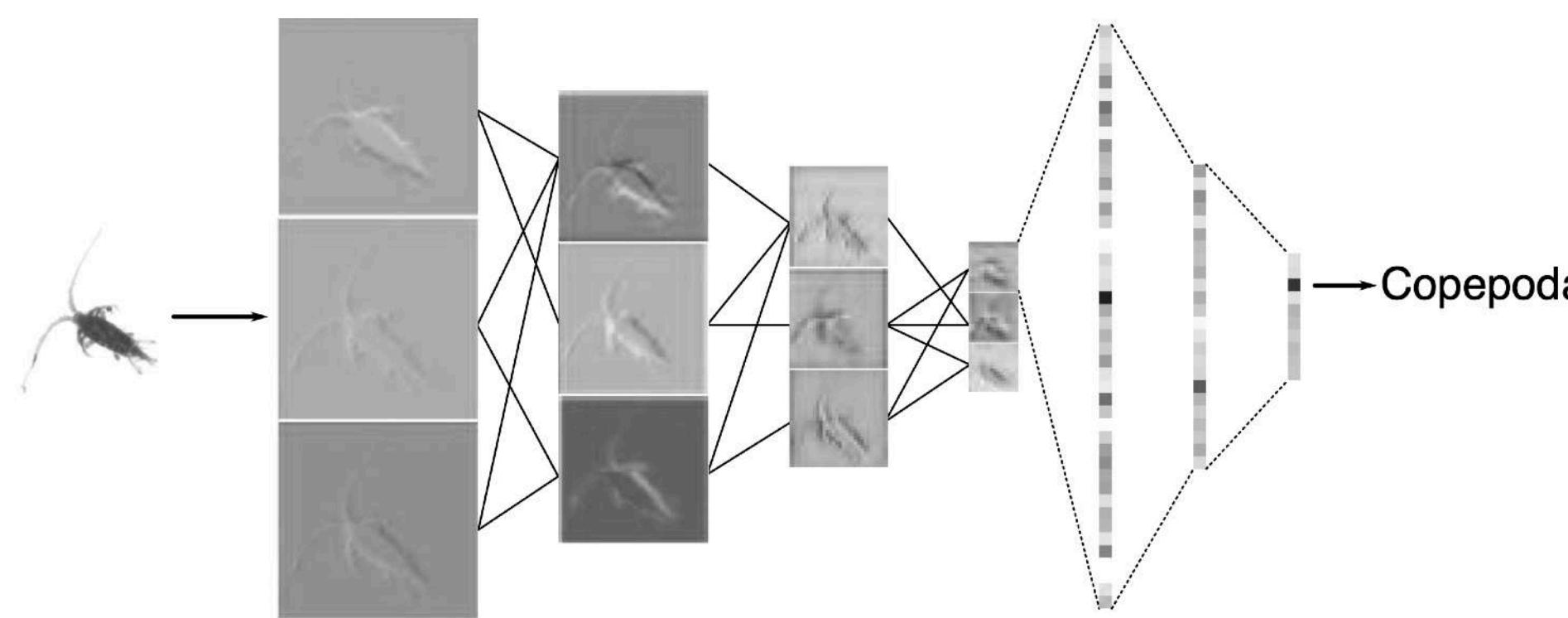
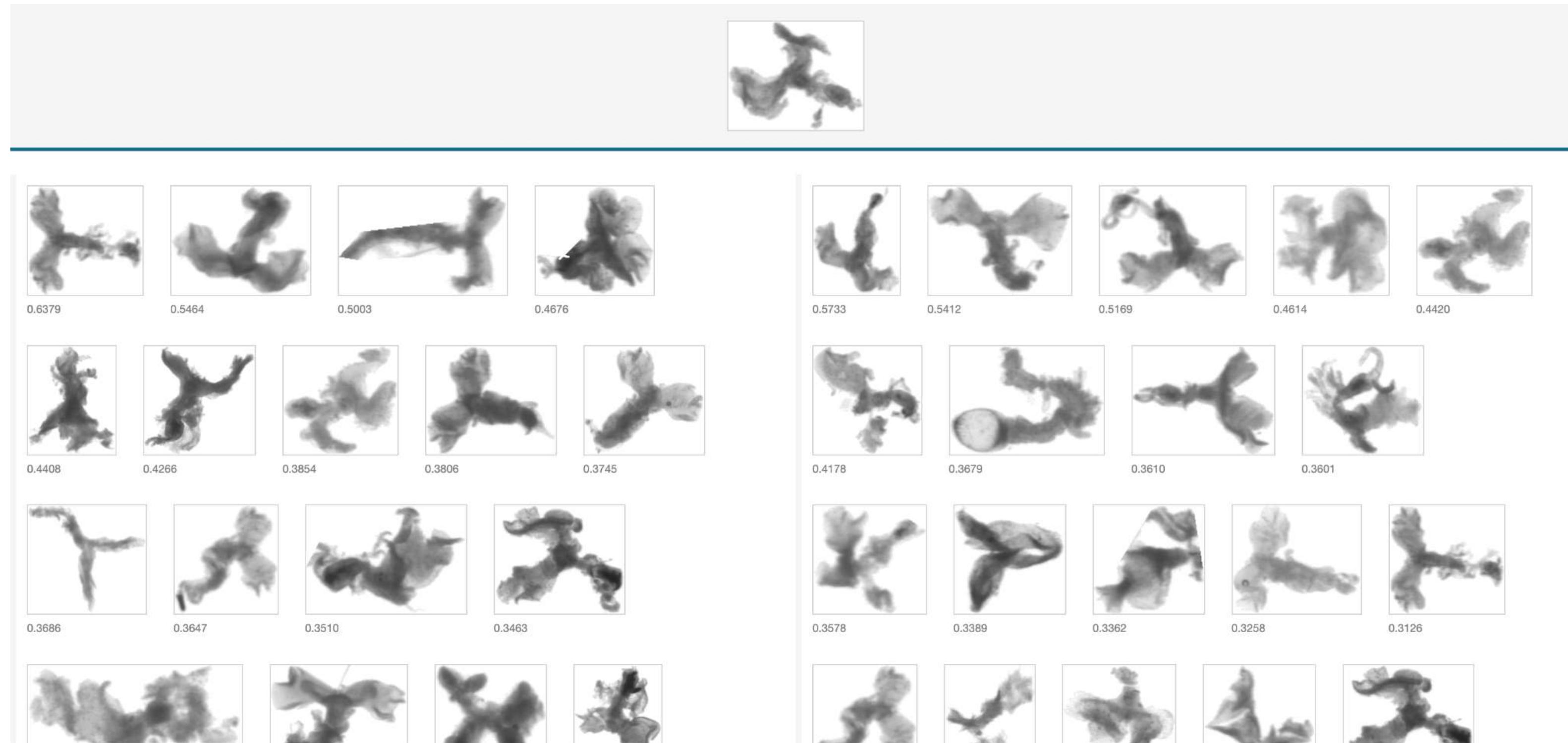




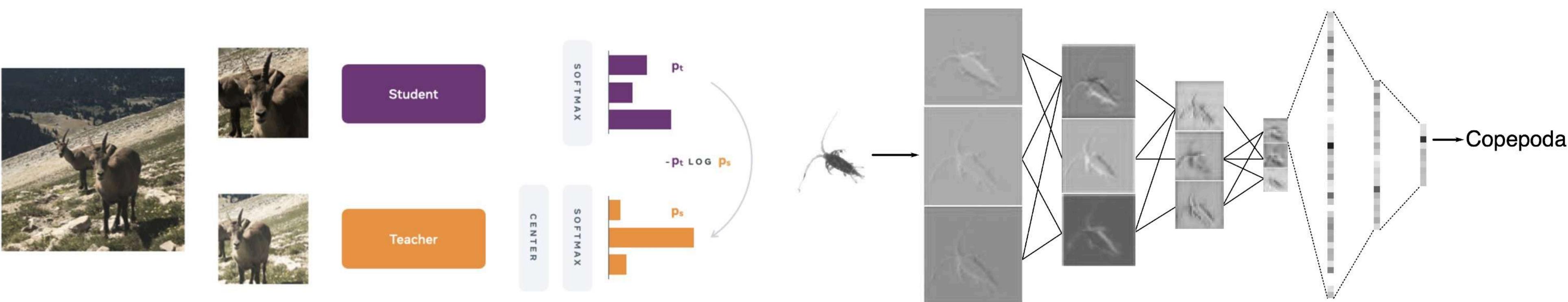
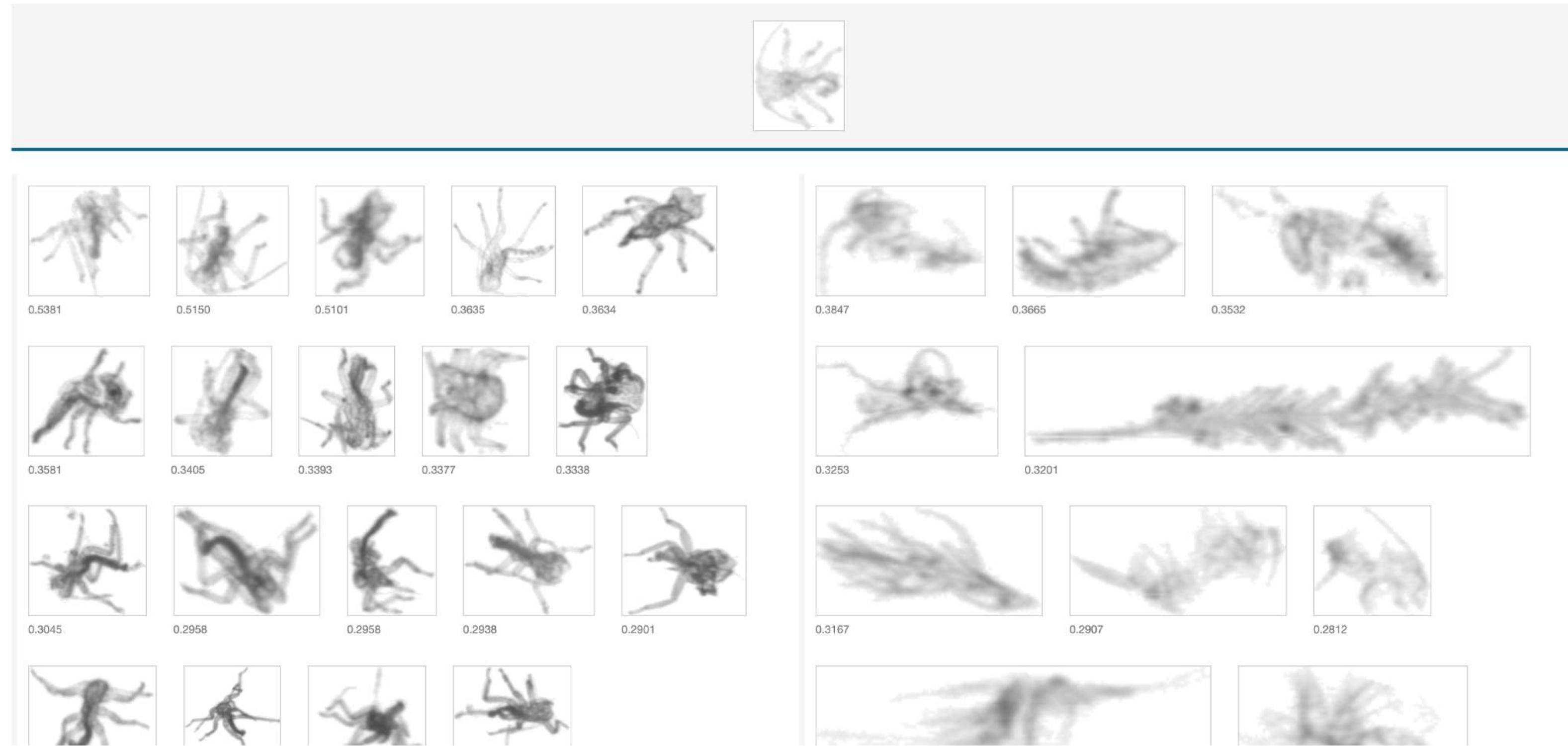
# Une autre utilisation de l'infrastructure de machine learning



# Quels descripteurs de deep learning?



# Quels descripteurs de deep learning?



A wide-angle photograph of a calm ocean under a clear blue sky. The horizon is visible in the distance, and the water is a deep, uniform blue.

Merci !