



5BIP012 (6 ECTS)

Intelligence artificielle et apprentissage profond pour la biologie

1^{er} semestre 2025

Apprenez à construire et utiliser des réseaux de neurones profonds pour l'analyse de données biologiques.

Objectifs pédagogiques

Poussées par les avancées en intelligence artificielle, les techniques d'apprentissage profond offrent un potentiel exceptionnel pour l'analyse, l'interprétation et la modélisation des données biologiques.

Cette unité d'enseignement vise à former les étudiants à la conception et à l'utilisation de réseaux de neurones pour développer des modèles appliqués à l'analyse de données transcriptomiques. Un accent particulier sera mis sur la bibliothèque Keras (disponible sous R et Python) pour concevoir différentes architectures de réseaux de neurones et tester leur capacité à analyser, classer et interpréter les données d'expression génique.

Les étudiants acquerront des compétences clés en programmation et en modélisation, tout en approfondissant leur compréhension des mécanismes moléculaires sous-jacents et des principes qui régissent les réseaux de neurones. Ces acquis leur permettront de proposer des approches innovantes pour l'analyse de données biologiques complexes.

Programme pédagogique

Cette unité d'enseignement comprend des cours et des projets d'analyse de données menés par des groupes d'étudiants.

Le programme pédagogique couvre les aspects suivants :

- Introduction à l'apprentissage profond et à ses applications dans les sciences du vivant.
- Exploration du logiciel ML Studio pour expérimenter des modèles d'apprentissage.
- Concepts fondamentaux de la conception des réseaux de neurones.
- Utilisation de la bibliothèque Keras pour concevoir et entraîner des réseaux neuronaux.
- Préparation, encodage et normalisation des données pour l'apprentissage automatique.
- Étude de la topologie des réseaux neuronaux, y compris les perceptrons multicouches (MLP).
- Présentation des réseaux récurrents (RNN) et convolutifs (CNN).
- Évaluation des performances des modèles à l'aide des matrices de confusion et courbes ROC.
- Introduction aux autoencodeurs et aux techniques de réduction de dimensionnalité.
- Applications de l'intelligence artificielle générative en biologie.
- Comparaison des architectures de réseaux et optimisation des hyperparamètres.
- Régularisation pour éviter le surapprentissage.
- Prédiction d'états physiopathologiques à partir des données transcriptomiques.
- Réalisation de projets sur des cas d'étude biologiques concrets.
- Utilisation de Git pour la gestion collaborative et le versionnement des projets.
- Communication des résultats et rédaction de rapports d'analyse.

Organisation

Cette unité d'enseignement de 6 ECTS se déroule au premier semestre de l'année universitaire.

Les modalités de contrôle des connaissances sont les suivantes :

- Un rapport d'analyse sous forme d'article scientifique (60 % de la note finale)
- Une soutenance orale pour présenter et discuter les résultats obtenus (40 % de la note finale)

Le projet porte sur l'analyse d'une sous-partie de l'atlas de données TCGA, en mobilisant des réseaux de neurones pour identifier des groupes de patients, prédire leur statut clinique et évaluer la qualité des prédictions. Ce jeu de données comprend le profil transcriptomique de plus de 20 000 échantillons de cancers primaires et de tissus normaux appariés, couvrant 33 types de cancer chez l'Homme.