

# IMMERSION ET MAITRISE DES CONCEPTS FONDAMENTAUX

Sorelle Perelle NGUIAKAM

August-September 2025

# INTRODUCTION

Les maladies auto-immunes (MAI) regroupent des pathologies chroniques où le système immunitaire, en rupture de tolérance immunitaire, attaque ses propres cellules et tissus. Cette dérégulation, impliquant l'activation anormale des lymphocytes T et B et la production d'auto-anticorps, entraîne une inflammation persistante et des lésions tissulaires pouvant toucher un organe unique ou plusieurs systèmes. Leur complexité biologique et clinique se traduit par une hétérogénéité des symptômes, une évolution imprévisible, l'absence de biomarqueur universel et un risque de chevauchement entre affections, compliquant le diagnostic et la prise en charge. L'intégration et l'analyse des vastes données médicales issues de sources multiples (dossiers électroniques, imagerie, données omiques) constituent un défi majeur. Dans ce contexte, l'Intelligence Artificielle (IA) et le Machine Learning (ML) offrent des solutions prometteuses grâce à leur capacité à traiter des données multimodales et à identifier des schémas complexes. Ils ouvrent la voie à un diagnostic plus précoce, à la prédiction de l'évolution de la maladie, à la découverte ou au repositionnement de médicaments, et à la personnalisation des traitements. Ce stage de Master 1 a pour objectif d'acquérir les bases théoriques sur les MAI et les approches IA/ML, de réaliser une revue de littérature approfondie et de conduire une analyse critique de l'état de l'art afin de préparer de futures recherches appliquées dans ce domaine.

# 1 Bases Médicales des Maladies Auto-Immunes (MAI)

Les MAI sont des pathologies chroniques résultant d'un dérèglement du système immunitaire, caractérisé par une rupture de la tolérance immunitaire. Au lieu de protéger l'organisme contre les agents pathogènes, le système immunitaire attaque ses propres cellules et tissus.

Sur le plan immunologique, cette dérégulation se manifeste par :

- rupture de la tolérance immunitaire, où les cellules immunitaires considèrent les éléments du "soi" comme étrangers ;
- activation anormale des lymphocytes T et B, qui amplifie la réaction auto-immune ;
- production d'auto-anticorps dirigés contre des protéines, acides nucléiques ou récepteurs membranaires, provoquant des lésions directes ou médiées par l'activation du complément et la formation de complexes immuns.

Ces anomalies entraînent une inflammation chronique et des dommages tissulaires pouvant être :

- localisés à un seul organe (MAI spécifiques d'organe) comme le diabète de type 1, la thyroïdite de Hashimoto ou le vitiligo ;
- étendus à plusieurs systèmes simultanément (MAI systémiques) comme le lupus érythémateux systémique ou la polyarthrite rhumatoïde.

La complexité des MAI résulte de plusieurs facteurs :

- **Hétérogénéité clinique** : les symptômes varient considérablement d'un patient à l'autre, même pour une même maladie ;
- **Syndromes de chevauchement** : un patient peut présenter simultanément les caractéristiques de plusieurs MAI ;
- **Évolution imprévisible** : alternance de phases de poussées et de rémissions, rendant le cours difficile à anticiper ;
- **Absence de biomarqueur universel** : aucun test unique ne permet un diagnostic fiable pour toutes les MAI, nécessitant une approche combinant données cliniques, biologiques et d'imagerie.

Les causes des MAI sont multifactorielles, impliquant une interaction complexe entre :

- facteurs génétiques : certaines variations, notamment dans les gènes HLA, augmentent la susceptibilité ;
- facteurs environnementaux : infections, tabagisme, exposition à certaines substances chimiques, alimentation et composition du microbiote ;
- facteurs épigénétiques : modifications de l'expression des gènes (méthylation de l'ADN, modifications d'histones) influencées par l'environnement ;
- facteurs hormonaux : une prévalence plus élevée chez les femmes suggère un rôle modulateur des hormones sexuelles.

Cette combinaison d'éléments explique la diversité des tableaux cliniques et la difficulté d'établir un diagnostic précoce et précis. Elle justifie également l'intérêt croissant pour des approches d'analyse de données complexes comme l'Intelligence Artificielle (IA) afin de mieux comprendre et gérer ces maladies.

## 2 Fondamentaux de l'Intelligence Artificielle (IA) et du Machine Learning (ML)

L'Intelligence Artificielle (IA) est un domaine de l'informatique visant à concevoir des systèmes capables de reproduire des fonctions cognitives humaines, telles que l'apprentissage, la reconnaissance de motifs, le raisonnement et la prise de décision. En médecine, l'IA permet d'analyser rapidement de vastes ensembles de données complexes et hétérogènes pour identifier des corrélations, prédire des évolutions cliniques et proposer des stratégies thérapeutiques personnalisées.

Le Machine Learning (ML), une branche de l'IA, se concentre sur la capacité des algorithmes à apprendre à partir des données sans être explicitement programmés. On distingue principalement trois types d'apprentissage :

- **Apprentissage supervisé** : l'algorithme apprend à partir de données étiquetées pour prédire un résultat précis (exemple : diagnostiquer une maladie à partir d'images médicales) ;
- **Apprentissage non supervisé** : l'algorithme regroupe les données selon leurs similarités, sans étiquettes préalables (exemple : identification de nouveaux sous-groupes de patients) ;
- **Apprentissage par renforcement** : le système améliore sa stratégie en recevant un retour positif ou négatif après chaque action.

L'application du ML en biomédecine suit un processus structuré, souvent appelé pipeline :

1. **Collecte des données** : dossiers médicaux électroniques (DME), imagerie, données "omiques" (génomiques, transcriptomiques, protéomiques), données de capteurs biométriques ;
2. **Prétraitement et nettoyage** : harmonisation des formats, traitement des valeurs manquantes, suppression des doublons, correction des erreurs ;
3. **Sélection des variables pertinentes** : élimination du bruit et des caractéristiques non informatives ;
4. **Entraînement et validation des modèles** : utilisation de techniques comme la validation croisée pour éviter le surapprentissage ;
5. **Évaluation des performances** : mesure via précision, rappel, F-mesure, AUC (aire sous la courbe ROC) ;
6. **Interprétation et intégration clinique** : présentation des résultats de manière compréhensible pour les professionnels de santé.

### Algorithmes clés

Parmi les algorithmes les plus utilisés en santé, on retrouve :

- **Forêts Aléatoires (Random Forest)** : robustes face aux données bruitées et performantes pour la classification de grands ensembles de variables ;
- **Machines à Vecteurs de Support (SVM)** : adaptées aux problèmes complexes et non linéaires ;
- **Réseaux de Neurones et Deep Learning** : puissants pour l'analyse d'images médicales et de données massives, mais souvent considérés comme des « boîtes noires » ;

- **XGBoost** : performant pour les données structurées et très utilisé en compétition de data science ;
- **Naïve Bayes** : simple, rapide et efficace, notamment avec des données bruitées ;
- **Régressions linéaires ou logistiques** : utiles pour leur interprétabilité et leur capacité à modéliser les relations entre variables ;
- **Réseaux de Neurones Graphiques (GNN)** : adaptés à l'analyse de données en forme de graphes, comme les réseaux d'interactions protéine–protéine.

La combinaison de ces méthodes avec des bases de données médicales de plus en plus riches ouvre la voie à une médecine prédictive et personnalisée.

### 3 Premières Intersections entre l'IA et les MAI

L'application de l'IA et du ML aux MAI ouvre des perspectives prometteuses pour améliorer le diagnostic, le pronostic et la prise en charge thérapeutique grâce à leur capacité d'analyse de grandes quantités de données hétérogènes.

#### Domaines d'application

Les domaines d'application majeurs incluent :

- **Diagnostic précoce** : identification de signatures biologiques et cliniques spécifiques permettant de détecter une maladie avant l'apparition des symptômes majeurs ;
- **Pronostic et suivi de la maladie** : prédiction de l'évolution, des poussées et des rémissions à partir de données longitudinales ;
- **Découverte et repositionnement de médicaments** : identification de nouvelles cibles thérapeutiques et réutilisation de molécules existantes pour d'autres indications ;
- **Aide à la décision clinique** : systèmes intelligents fournissant aux médecins des recommandations personnalisées en fonction du profil du patient.

#### Types de données utilisées

Plusieurs types de données sont exploitées dans ces approches :

- **Données cliniques** : antécédents, symptômes, examens physiques ;
- **Données biologiques** : analyses de sang, auto-anticorps, marqueurs inflammatoires ;
- **Imagerie médicale** : IRM, échographie, microscopie, vidéocapsule ;
- **Données « omiques »** : génomiques, transcriptomiques, protéomiques, métabolomiques ;
- **Données issues de capteurs connectés** : suivi en temps réel de paramètres physiologiques.

#### Avantages et défis

Les avantages de l'IA dans le domaine des MAI incluent :

- **Analyse multidimensionnelle** : intégration simultanée de données cliniques, biologiques et d'imagerie ;
- **Précision accrue** : amélioration des taux de sensibilité et de spécificité des outils diagnostiques ;
- **Personnalisation des soins** : adaptation des stratégies thérapeutiques à chaque patient ;
- **Gain de temps** : accélération du processus diagnostique et réduction du délai de mise en place du traitement.

L'IA présente un potentiel dans le domaine des maladies auto-immunes, mais elle se heurte à de nombreux problèmes et défis

- **Validation insuffisante des modèles** : Il est crucial de renforcer la validation croisée et les tests indépendants pour améliorer la précision et la fiabilité des modèles.
- **Difficultés d'intégration des données** : L'intégration de données issues de sources multiples (DME, données génomiques, bases de données publiques, etc.) est complexe et nécessite une standardisation.
- **Puissance de calcul** : Les modèles d'IA complexes, notamment le Deep Learning, nécessitent d'importantes ressources informatiques, ce qui peut poser des défis.
- **Manque de modèles de réponse immunitaire "sains"** : Il y a un besoin de modèles complets pour différencier les états normaux des états pathologiques.
- **Biais algorithmiques et équité des soins** : Le risque de reproduire ou d'amplifier les inégalités de soins est une préoccupation majeure. Ce point est d'autant plus pertinent que des disparités existent dans l'application de l'IA en médecine, notamment entre les pays à revenus élevés et ceux à revenus faibles ou intermédiaires, où les modèles formés dans une région peuvent être sous-optimaux s'ils sont appliqués dans une autre en raison du "décalage des ensembles de données".
- **Confidentialité et sécurité des données** : La protection des données des patients est un enjeu crucial.
- **Qualité et standardisation des données** : nécessité de bases de données fiables, bien annotées et harmonisées ;
- **Interprétabilité des modèles** : importance de comprendre comment l'algorithme aboutit à une prédiction, en particulier en contexte médical ;
- **Intégration dans la pratique clinique** : adaptation des outils aux contraintes et besoins du terrain.

Ainsi, l'IA et le ML constituent des outils puissants pour la recherche et la clinique dans le domaine des maladies auto-immunes, mais leur efficacité repose sur la qualité des données, la transparence des méthodes et une collaboration étroite entre médecins, biologistes et data scientists.

## Conclusion

Le présent rapport a permis de réaliser avec succès l'Étape 1 de notre mission, axée sur l'Immersion et la Maîtrise des Concepts Fondamentaux des maladies auto-immunes (MAI) et de l'Intelligence Artificielle/Machine Learning (IA/ML). Nous avons établi une base théorique solide en décrivant la complexité des MAI – marquées par une rupture de tolérance immunitaire, une grande hétérogénéité clinique et des mécanismes multifactoriels impliquant des composantes génétiques, environnementales, épigénétiques et hormonales. Parallèlement, nous avons exploré les principes fondamentaux de l'IA et du ML, en détaillant les types d'apprentissage, le pipeline méthodologique et les algorithmes clés utilisés en biomédecine, tels que les Forêts Aléatoires, les Machines à Vecteurs de Support, les Réseaux de Neurones/Deep Learning, XGBoost, Naïve Bayes, les Régressions et les Réseaux Neuronaux Graphiques. Cette phase initiale a également mis en lumière les premières intersections prometteuses entre l'IA et les MAI, notamment pour le diagnostic précoce, le pronostic, la découverte de médicaments et l'aide à la décision clinique, tout en identifiant les types de données exploitées et les avantages potentiels de ces technologies. Cependant, il est impératif de souligner les défis persistants qui nécessitent des avancées significatives. Ces défis incluent la nécessité d'une validation rigoureuse des modèles, la résolution des difficultés d'intégration et de standardisation des données, la gestion de la puissance de calcul, l'amélioration de l'interprétabilité des modèles, la prise en compte des biais algorithmiques pour garantir l'équité des soins, l'absence de modèles complets de réponse immunitaire "saine" et la protection de la confidentialité des données. Les perspectives futures s'orientent vers des collaborations interdisciplinaires, le développement des jumeaux numériques, l'échange de données à grande échelle et une intégration éthique de l'IA dans la pratique clinique. Ces connaissances fondamentales et cette analyse critique de l'état de l'art nous préparent idéalement pour l'Étape 2 de la mission. Cette prochaine phase nous permettra d'explorer de manière approfondie les méthodologies spécifiques et les applications concrètes du Machine Learning dans les maladies auto-immunes, en s'appuyant sur les concepts et les défis identifiés ici. Nous serons ainsi en mesure d'analyser comment l'IA contribue à l'identification des patients, à l'évaluation des facteurs de risque et du pronostic, au diagnostic, à la classification des sous-types, à la surveillance et à l'aide à la décision, ainsi qu'à la découverte de médicaments.