

École Nationale Supérieure d'Informatique et d'Analyse des Systèmes
Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur

AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT

Madame Nissrine HATIBI

Soutiendra publiquement sa thèse de Doctorat en Informatique

Le 13 juin 2026 à 10h30 au Grand Amphi à l'ENSIAS de Rabat

Intitulé de la thèse

PREDICTIVE MODELING AND EXPLAINABILITY IN ARTIFICIAL INTELLIGENCE FOR PANDEMIC PREVENTION AND THERAPEUTIC DISCOVERY

Président :

Pr. Rachid OULAD HAJ THAMI, PES, ENSIAS, Université Mohammed V de Rabat

Directrice de thèse :

Pr. Mounia ABIK, PES, ENSIAS, Université Mohammed V de Rabat

Rapporteurs :

Pr. Karim BAINA, PES, ESIN, Université Internationale de Rabat

Pr. Zahra MUNGLOO-DILMOHAMUD, Associate Professor, FoICDT, University of Mauritius

Pr. Mohammed RZIZA, PES, FSR, Université Mohammed V, Rabat

Examineurs :

Pr. Brahim BENAJI, PES, ENSAM, Université Mohammed V, Rabat

Pr. Youness TABII, PES, ENSIAS, Université Mohammed V, Rabat

Pr. Abdelalim SADIQ, PES, FS, Université Ibn Tofail, Kénitra



Résumé:

Le deep learning présente un potentiel majeur en biologie computationnelle et en médecine, mais l'opacité des modèles limite encore son adoption dans des contextes à fort enjeu. Concevoir des systèmes d'IA fiables pour des tâches biologiques complexes nécessite des architectures capables de traiter des données hétérogènes tout en produisant des explications claires et biologiquement fondées. Cette exigence est particulièrement critique pour des applications telles que la surveillance précoce des épidémies et la conception rationnelle de médicaments, qui demandent à la fois précision prédictive et explicabilité.

Cette thèse propose un cadre computationnel unifié visant à transformer des prédictions opaques en découvertes fiables et fondées sur des résultats. Le cadre s'articule autour de deux pipelines complémentaires. Le pipeline de prévention se concentre sur l'évaluation du risque viral et démontre que des représentations de données optimales révèlent des signaux évolutifs souvent ignorés par les approches classiques. Il exploite également l'incertitude des prédictions comme indicateur des transitions biologiques à haut risque. Appliqué à l'Influenza A, ce pipeline identifie les souches pandémiques connues et extrait des biomarqueurs corrélés au potentiel zoonotique. Le pipeline thérapeutique, quant à lui, vise à accélérer l'identification de candidats thérapeutiques grâce à la fusion multimodale de données. Descripteurs physicochimiques, fingerprints moléculaires et encodages structuraux en graphes sont combinés pour construire des modèles performants sur six récepteurs distincts. Les techniques d'IA explicable permettent de relier les prédictions aux déterminants structuraux, produisant des cartes d'attribution au niveau des résidus cohérentes avec la biophysique. Appliqué à un récepteur d'Influenza A, le pipeline génère une liste priorisée de candidats inhibiteurs pour validation expérimentale, complétant ainsi la boucle prédiction-expérimentation.

Ces contributions, basées sur l'incertitude des modèles, des représentations biologiques et la fusion multimodale, peuvent être appliquées à tout domaine traitant des données complexes et hétérogènes. Elles définissent une norme rigoureuse pour une IA biomédicale fiable, délivrant des prédictions précises et explicables, généralisables aux systèmes viraux et à d'autres champs scientifiques.

Mots-clés:

Intelligence Artificielle Explicable - Biologie Computationnelle - Apprentissage Profond pour les Applications Biomédicales - Biomarqueurs de Transition Zoonotique - Découverte de Médicaments Basée sur la Structure - Données Biologiques Multi-Représentation - Architectures de Fusion Multimodale

Abstract:

Deep learning holds significant promise in computational biology and medicine, yet the opacity of neural models limits their adoption in high-stakes decision-making. Building reliable AI systems for complex biological tasks requires architectures capable of handling heterogeneous data while also producing clear and consistent explanations. Addressing this challenge is essential for applications such as early outbreak surveillance and rational therapeutic design, which demand both predictive accuracy and explainability.

This thesis introduces a unified computational framework designed to shift from opaque prediction toward trustworthy and insight-driven discovery. The framework is organized around two complementary pipelines. The prevention pipeline focuses on viral threat assessment and demonstrates how the choice of an optimal data representation can highlight patterns consistent with evolutionary signals that traditional approaches overlook. It also introduces a novel reinterpretation of prediction uncertainty, transforming ambiguous outputs into informative markers of high-risk biological state transitions. Applied to Influenza A, this pipeline successfully identifies known pandemic strains and extracts potential codon-level biomarkers linked to zoonotic potential. The therapeutics pipeline aims to accelerate computational drug candidate identification through multimodal data fusion.

Physicochemical descriptors, molecular fingerprints, and structural graph encodings are combined to build predictive models that perform consistently across six receptor targets. Integrated explainable AI techniques map learned patterns to meaningful structural determinants, generating residue-level attribution profiles that are aligned with established biophysical knowledge. When extended to an Influenza A receptor, the pipeline produces a prioritized shortlist of candidate inhibitors suitable for experimental evaluation, thereby closing the loop between prediction and validation.

This work introduces advances grounded in model uncertainty, biologically meaningful representations, and early-stage multimodal fusion. These contributions generalize to any domain involving complex and heterogeneous data. Through structured explainability, the framework establishes a rigorous standard for trustworthy biomedical AI, delivering accurate and explainable predictions that are generalizable across viral systems and related scientific domains.

Keywords:

Explainable Artificial Intelligence - Computational Biology - Deep Learning for Biomedical Applications - Zoonotic Transition Biomarkers - Structure-Based Drug Discovery - Multi-Representation Biological Data - Multi-Modal Fusion Architectures