

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

AITOUHANNI IMANE

Soutiendra publiquement sa thèse de Doctorat en Informatique

Le 19 Décembre 2025 à 15h00 au Grand Amphi à l'ENSIAS de Rabat

Intitulé de la thèse

**CROSS-DOMAIN AI CORRECTION FRAMEWORKS: BOOSTING
MODEL RELIABILITY IN HEALTHCARE AND ENVIRONMENTAL
SCIENCES**

Président :

Pr. Ali IDRI, PES, ENSIAS, Université Mohammed V de Rabat

Directeur de thèse :

Pr. Amine BERQIA, PES, ENSIAS, Université Mohammed V de Rabat

Rapporteurs :

Pr. Mohamed ET-TOLBA, PES, INPT

Pr. Yousef FARHAOUI, PES, FST ERRACHIDIA

Pr. Zhour MADINI, PES, ENSA KENITRA

Examineur(s) :

Pr. Hanane EL BAKKALI, PES, ENSIAS, Université Mohammed V de Rabat

Résumé:

À une époque où les modèles d'apprentissage automatique sont de plus en plus utilisés dans des domaines critiques tels que la santé, le développement pharmaceutique et la protection de l'environnement, la nécessité de systèmes prédictifs fiables et interprétables n'a jamais été aussi pressante. Ce mémoire aborde le défi persistant de l'incertitude des prédictions en proposant des méthodologies de correction modulaires capables d'améliorer la fiabilité des modèles sans nécessiter leur réentraînement.

Trois tâches prédictives majeures — la classification du cancer du sein, la prédiction du risque de diabète et l'estimation de la solubilité aqueuse (LogS) pour la durabilité environnementale ont été sélectionnées pour valider les cadres méthodologiques proposés. Dans les contextes de classification, les stratégies de correction basées sur des seuils de confiance probabilistes ont permis d'améliorer significativement la précision et de réduire les faux positifs, renforçant ainsi la confiance clinique et diagnostique. Pour les tâches de régression, la correction basée sur les résidus via des modèles de régression Ridge secondaires a affiné de manière sélective les prédictions difficiles, réduisant les MAE et RMSE sans compromettre l'efficacité computationnelle.

Une analyse comparative entre les cas d'usage a démontré la généralisation, l'évolutivité et la pertinence pratique des approches de correction. Au-delà des gains de performance empirique, ce travail souligne l'importance de la quantification de l'incertitude, de l'explicabilité et du développement d'une intelligence artificielle durable. Des recommandations pour une application plus large ainsi que des pistes de recherche futures incluant des systèmes de correction adaptatifs et l'intégration dans des architectures d'apprentissage profond ont également été proposées.

En définitive, ce travail contribue à faire progresser le domaine de l'apprentissage automatique fiable en offrant des solutions légères, évolutives et adaptables aux différents domaines pour améliorer la qualité des prédictions dans des environnements à enjeux élevés.

Mots-clés:

Apprentissage automatique, Fiabilité des prédictions, Méthodologies de correction, Classification du cancer du sein, Prédiction du risque de diabète, Solubilité aqueuse (LogS), Durabilité environnemental.

Abstract:

In an era where machine learning models are increasingly deployed in critical domains such as healthcare, pharmaceutical development, and environmental protection, the need for reliable and interpretable predictive systems has never been greater. This thesis addresses the persistent challenge of prediction uncertainty by proposing modular correction methodologies capable of enhancing model reliability without retraining.

Three major predictive tasks—breast cancer classification, diabetes risk prediction, and aqueous solubility (LogS) estimation for environmental sustainability—were selected to validate the proposed frameworks. In classification contexts, correction strategies based on probabilistic confidence thresholds significantly improved precision and reduced false positive rates, thereby enhancing clinical and diagnostic trust. In

regression settings, residual-based correction through secondary Ridge Regression models selectively refined difficult predictions, reducing MAE and RMSE without compromising computational efficiency.

A comparative analysis across uses cases demonstrated the generalizability, scalability, and practical relevance of the correction approaches. Beyond empirical performance gains, the thesis emphasized the importance of uncertainty quantification and sustainable AI development. Recommendations for broader application and future research directions including adaptive correction systems and integration with deep learning architectures—were also outlined.

Ultimately, this work advances the state of trustworthy machine learning by providing lightweight, scalable, and domain-adaptable solutions for enhancing prediction quality in high stakes environments.

Keywords:

Machine Learning, Prediction Reliability, Correction Methodologies, Breast Cancer Classification, Diabetes Risk Prediction, Aqueous Solubility (LogS), Environmental Sustainability.