

TITRE DU PROJET (limité chacun à 255 caractères)

**Traduction des modèles d'intelligence artificielle en scores cliniques interprétables
pour la prédiction des événements cardiovasculaires**

***Translating Artificial Intelligence Models into Interpretable Clinical Risk Scores for Predicting
Cardiovascular Events***

1) Renseignements administratifs sur la direction de thèse¹ (1 page maximum) :

Directeur de thèse HDR :

Nom : **Chopard**

Prénom : **Romain**

Section CNU : Cardiologie (51-02).

Grade : **PUPH**

HDR : Date de soutenance Juin 2019 Discipline : Cardiologie

Coordonnées (adresse, courriel, téléphone) : service de cardiologie, CHU de BESANCON, 3 boulevard
Alexandre Fleming, 25000, BESANCON. rchopard@chu-besancon.fr ; chopardromain@icloud.com ;
03 81 66 81 66 poste 69462

Unité d'appartenance (intitulé, label, n°, directeur) : **SINERGIES, EA 4662, Frédéric Auber**

2) Descriptif du projet de thèse (devra inclure les rubriques suivantes) :

Résumé en français et anglais (limité chacun à 1800 caractères)

Les systèmes de score clinique sont largement utilisés en médecine pour prédire le risque d'événements et soutenir la prise de décision. Ces modèles de stratification du risque sont généralement des systèmes additifs, parcimonieux, reposant sur des points entiers, conçus pour une utilisation simple et rapide au lit du patient [1]. Toutefois, leur développement s'appuie traditionnellement sur des approches fondées sur la régression logistique, avec arrondi des coefficients et discrétisation heuristique des variables continues. Or, ces choix méthodologiques peuvent ne pas capturer des relations complexes, non linéaires ou interactives entre les prédicteurs et les événements, et ainsi conduire à une perte d'information prédictive et à une diminution des performances globales [2].

¹ ATTENTION : selon l'article 16 de l'arrêté du 25 mai 2016, le total d'encadrants ne peut pas dépasser 2, sauf si l'un des encadrants appartient au monde socio-économique, qui peut venir en sus, ou en cas de co-tutelle; Le décompte des co-encadrements se fera au prorata du nombre d'encadrants : 1 pour 1 encadrant, ½ pour deux encadrants.

L'essor de l'intelligence artificielle (IA) et des méthodes d'apprentissage automatique (Machine Learning, ML) permet désormais d'intégrer un plus grand nombre de données et une plus grande diversité de variables pronostiques, tout en améliorant la précision prédictive grâce à la modélisation d'interactions d'ordre supérieur et de relations non linéaires entre les variables explicatives [3, 4]. Ces approches offrent des performances discriminantes souvent supérieures, généralement évaluées par des métriques telles que l'aire sous la courbe ROC (AUC), mais leur complexité limite leur interprétabilité et leur utilisation directe en pratique clinique [5].

Le cadre méthodologique Unified Automatic Clinical Scoring (UNI-ACS) a été proposé afin de combler l'écart entre des modèles de ML performants et des scores cliniques directement utilisables en routine [5].

Dans le cadre de ce projet de thèse d'université, nous proposons :

1. Élaborer un modèle de représentation des connaissances cliniques pour la prédiction des événements cardiovasculaires aigus en intégrant les méthodes d'IA agentiques ;
2. Intégrer notre approche de modélisation dans le cadre la méthode UNI-ACS afin de traduire des algorithmes d'IA/ML classiques (p. ex., XGBoost, Random Forest) en scores cliniques et biologiques interprétables et utilisables au lit du patient ;
3. Comparer ses performances à celles d'autres approches de traduction (p. ex., AutoScore), en termes de discrimination, calibration et préservation du classement relatif du risque ainsi que les modèles de survie de type DeepSurv et DeepHit ;
4. Intégrer l'ensemble des points précédents en développant un système intelligent d'assistance clinique pour l'aide à la décision lors des événements cardiovasculaires aigus.

(english version)

Clinical scoring systems are widely used in medicine to predict the risk of clinical events and support decision-making. These risk-stratification models are generally sparse, additive systems based on integer point values, designed for simple and rapid bedside use [1]. However, their development has traditionally relied on logistic regression-based approaches, with coefficient rounding and heuristic discretization of continuous variables. Such methodological choices may fail to capture complex, non-linear, or interactive relationships between predictors and outcomes, potentially leading to a loss of predictive information and reduced overall performance [2].

The rapid expansion of artificial intelligence (AI) and machine-learning (ML) methods now enables the integration of a larger number and greater diversity of prognostic variables, while improving predictive accuracy through the modeling of higher-order interactions and non-linear relationships among predictors [3, 4]. Although these approaches often demonstrate superior discriminative performance—commonly assessed using metrics such as the area under the receiver-operating characteristic curve (AUC)—their complexity limits interpretability and direct implementation in routine clinical practice [5].

The Unified Automatic Clinical Scoring (UNI-ACS) framework has been proposed to bridge the gap between high-performing ML models and clinically usable risk scores [6].

As part of this university thesis project, we propose:

1. Developing a clinical knowledge representation model to predict acute cardiovascular events by integrating agent-based AI methods ;
2. Integrate our modelling approach into the UNI-ACS method in order to translate classic AI/ML algorithms (e.g. XGBoost, Random Forest) into clinical and biological scores that are interpretable and usable at the patient's bedside ;
3. Compare its performance with that of other translation approaches (e.g., AutoScore) in terms of discrimination, calibration, and preservation of relative risk ranking, as well as with the DeepSurv and DeepHit survival models ;
4. Integrate all of the above by developing an intelligent clinical assistance system to aid decision-making during acute cardiovascular events.

Préciser le domaine de compétence dans la liste ci-dessous :

2 choix possibles maximum –les intitulés sont ceux d’ABG),

Ex : vous ne pouvez pas sélectionner uniquement informatique lorsqu’il est indiqué informatique, électronique :

- **Santé, médecine humaine, vétérinaire**
- **Informatique, électronique**

Indiquez quelques mots clés concernant votre sujet) :

1. Stratification pronostique
2. Traduction des modèles
3. Score clinique interpretable
4. Intelligence artificielle
5. Systems multi-agents
6. LLM

- Financement du projet – partie Recherche (montants acquis, type de contrat) :

- Dotation récurrente du Laboratoire SINERGIES : environ 6000 €.
- Utilisation d’une partie du budget APICHU (fonctionnement) à hauteur de 8000 € pour l’achat d’un ordinateur portable nécessaire au démarrage de la thèse et pour le temps de calcul au Méso centre BFC.
- En 2026 dépôt d’un projet région : Inter GS INTHERAPI/EIPHI.

- Connaissances et compétences requises

1. Connaissances en intelligence artificielle et science des données

Le candidat doit maîtriser les fondements de l'apprentissage automatique appliqué aux données de santé:

- Concepts de **machine learning supervisé et non supervisé**
- Algorithmes classiques : régression logistique, forêts aléatoires, gradient boosting, réseaux de neurones
- Méthodes d'**évaluation des performances des modèles** (AUC-ROC, sensibilité, spécificité, précision, calibration)
- Techniques d'**explicabilité des modèles** (Explainable AI : SHAP, LIME, importance des variables)
- Prétraitement et structuration des données : nettoyage, gestion des données manquantes, normalisation, feature engineering
- Programmation scientifique, idéalement en **Python ou R** (bibliothèques telles que scikit-learn, pandas, TensorFlow/PyTorch).

2. Connaissances en biostatistiques

Une solide base en statistiques appliquées à la recherche biomédicale est requise :

- Régression statistique et modèles multivariés
- Analyse de survie
- Validation interne et externe de modèles prédictifs
- Méthodes de calibration et de discrimination des scores de risque
- Analyse de cohortes cliniques et interprétation des résultats statistiques.

3. Connaissances de base en cardiologie et épidémiologie

Le candidat doit comprendre le contexte clinique et les enjeux de la prédiction cardiovasculaire :

- Principaux **événements cardiovasculaires majeurs** (infarctus du myocarde, accident vasculaire cérébral, insuffisance cardiaque, mortalité cardiovasculaire)
- Facteurs de risque cardiovasculaire (hypertension, diabète, dyslipidémie, tabagisme, âge, etc.)
- Principaux **scores de risque cardiovasculaire existants** (Framingham, SCORE2, GRACE)
- Notions d'épidémiologie clinique et d'analyse de bases de données médicales.

4. Compétences méthodologiques en recherche

Le candidat devra également maîtriser :

- Les méthodes d'analyse de **bases de données cliniques ou hospitalières**
- Les principes de **reproductibilité scientifique et de gestion des données de recherche**
- Les aspects réglementaires liés aux données de santé (RGPD, anonymisation)
- La rédaction d'articles scientifiques et la communication des résultats.

5. Compétences spécifiques liées au sujet de thèse

Un intérêt particulier pour l'**interprétabilité des modèles d'IA en médecine** est attendu, notamment :

- Traduction de modèles complexes en **scores cliniques simples et utilisables par les médecins**
- Sélection de variables pertinentes pour la décision clinique
- Construction et validation de **scores de risque interprétables** ou de nomogrammes
- Comparaison des performances avec les scores cliniques existants.

Le candidat idéal doit être titulaire d'un **Master 2 ou diplôme d'ingénieur** en :

- intelligence artificielle ou data science appliquée à la santé
- biostatistiques ou bioinformatique
- génie biomédical