



الجمهورية الجزائرية الديمقراطية الشعبية
The People's Democratic Republic of Algeria
وزارة التعليم العالي والبحث العلمي
The Ministry of Higher Education and Scientific Research
المدرسة العليا للإعلام الآلي - 8 ماي 1945 - سيدي بلعباس
The Higher School of Computer Science -08 May 1945- Sidi Bel Abbas

UNIVERSITÉ
TÉLUQ

MASTER THESIS

To obtain the diploma of **Master**
Field: **Computer Science**
Specialty: **Artificial Intelligence and Data Science**

Theme

**From Supervised to Self-Supervised Learning for Heterogeneous
Physiological Signals:**
Foundations and a Framework for Irregular Multi-Source Representation Learning

Presented By
Yacine DAIT DEHANE

Dr. KHALDI Belkacem	President	Dr. NAIB Rabiaa	Examiner
Dr. Chaib Souleyman	Supervisor	Pr. BENABDESLEM Khalid	Co-supervisor
Pr. Henni Khadidja	Co-supervisor	Dr. Hakimi Yacine	Co-supervisor

Defended 23 September 2026 · Academic year 2025/2026

This research work was supported by



**Digital Research
Alliance of Canada**

Abstract

Continuous physiological monitoring provides large amounts of heterogeneous and partially observed data, yet learning from such signals remains challenging due to irregular sampling, missing modalities, and substantial variation across cohorts. Existing supervised approaches commonly address these issues through resampling and imputation, which may introduce information that is not present in the original observations. This thesis investigates alternative learning strategies that preserve the structure of the observed physiological data while improving cross-cohort generalization.

We first establish a common evaluation framework for supervised hypoglycemia detection using continuous glucose monitoring (CGM). The observed cross-cohort performance degradation highlights the limitations of models trained and evaluated under cohort-specific settings and motivates the use of self-supervised representation learning. We then study different pretraining objectives using CGM and evaluate their representations within and across cohorts. Building on these findings, we investigate the contribution of additional physiological channels to CGM-based downstream tasks and identify the configurations that provide the most consistent representations.

These studies motivate the design of GlucoAnchor, a multimodal foundation model for heterogeneous and partially observed physiological time series. The proposed framework avoids treating missing observations as values that must be reconstructed and instead learns directly from recorded measurements. By combining self-supervised representation learning with a unified evaluation protocol, this work provides a principled framework for learning generalizable physiological representations across cohorts and signal modalities. The results establish the basis for extending foundation-model approaches to continuous physiological monitoring beyond CGM-specific applications.

Keywords : Self-Supervised Learning, Foundation Models, Representation Learning, Irregularly Sampled Time Series, Partially Observed Data, Multi-Channel Physiological Signals, Continuous Glucose Monitoring

Résumé

La surveillance physiologique continue produit d'importants volumes de données hétérogènes et partiellement observées, dont l'exploitation reste difficile en raison de l'échantillonnage irrégulier, de l'absence de certaines modalités et des différences entre cohortes. Les approches supervisées traitent généralement ces contraintes par rééchantillonnage et imputation, au risque d'introduire des informations absentes des observations originales. Ce mémoire étudie des stratégies alternatives fondées sur l'apprentissage auto-supervisé afin de préserver la structure des données observées et d'améliorer la généralisation inter-cohortes.

Nous établissons d'abord un cadre d'évaluation commun pour la détection de l'hypoglycémie à partir de la mesure continue du glucose (CGM). La dégradation observée des performances entre cohortes met en évidence les limites de modèles évalués dans des configurations spécifiques et motive l'étude de l'apprentissage de représentations auto-supervisé. Différents objectifs de pré-entraînement sont ensuite étudiés à partir du CGM et comparés au sein et entre les cohortes. L'analyse est étendue à d'autres canaux physiologiques afin d'évaluer leur contribution aux tâches en aval fondées sur le CGM et d'identifier les configurations produisant les représentations les plus robustes.

Ces résultats motivent le développement de GlucoAnchor, un modèle fondation multimodal pour les séries temporelles physiologiques hétérogènes et partiellement observées. Le cadre proposé évite de considérer les observations manquantes comme des valeurs devant être reconstruites et apprend directement à partir des mesures effectivement enregistrées. En combinant apprentissage auto-supervisé et protocole d'évaluation unifié, ce travail propose un cadre pour l'apprentissage de représentations physiologiques généralisables à différentes cohortes et modalités, au-delà des seules applications fondées sur le CGM.

Mots-clés : Apprentissage Auto-Supervisé, Modèles Fondamentaux, Apprentissage de Représentations, Séries Temporelles Irrégulières, Données Partiellement Observées, Signaux Physiologiques Multi-Canaux, Mesure Continue du Glucose

ملخص

تُنتج المراقبة الفسيولوجية المستمرة كميات كبيرة من البيانات المتباينة والمرصودة جزئياً، غير أن الاستفادة منها في تعلم التمثيلات تظل تحدياً بسبب عدم انتظام أخذ العينات، واختلاف جداول القياس بين القنوات، وغياب بعض القنوات لدى بعض المرضى. وتعالج المناهج الخاضعة للإشراف هذه التحديات عادةً من خلال إعادة أخذ العينات وتعويض القيم المفقودة، مما قد يؤدي إلى إدخال معلومات غير موجودة في القياسات الأصلية. تبحث هذه الأطروحة في استراتيجيات بديلة قائمة على التعلم الذاتي الإشراف، تهدف إلى الحفاظ على بنية الملاحظات الفعلية وتحسين قابلية تعميم التمثيلات بين المجموعات.

نُقيم أولاً نماذج التعلم الخاضع للإشراف لكشف نقص سكر الدم باستخدام المراقبة المستمرة للغلوكوز (نحج) ضمن إطار تقييم موحد. ويكشف الانخفاض الملحوظ في الأداء بين المجموعات عن محدودية التعميم خارج المجموعة المستخدمة في التطوير، مما يحفز دراسة التعلم الذاتي الإشراف للتمثيلات. وتُقارن بعد ذلك أهداف مختلفة للتدريب المسبق باستخدام نحج، داخل المجموعات وفيما بينها، ضمن بروتوكول تقييم متسق. كما يُوسَّع التحليل ليشمل قنوات فسيولوجية إضافية بهدف تحديد مساهمتها في المهام اللاحقة القائمة على نحج واختيار التكوينات الأكثر ملاءمة للتعلم متعدد الوسائط.

استناداً إلى هذه النتائج، تُقدّم جُلّنا نُحج بوصفه نموذج أساس متعدد الوسائط للإشارات الفسيولوجية الزمنية المتباينة والمرصودة جزئياً. ويقوم الإطار المقترح على مبدأ عدم اعتبار القياسات المفقودة قيمة ينبغي إعادة بنائها، بل التعلم مباشرةً من الملاحظات المسجلة فعلياً. ومن خلال الجمع بين التعلم الذاتي الإشراف، والتقييم عبر المجموعات، والتعلم متعدد القنوات، يضع هذا العمل أساساً لتعلم تمثيلات فسيولوجية قابلة للتعميم عبر المجموعات والوسائط، مع إمكانية توسيعها إلى مهام لاحقة تتجاوز نطاق نحج.

الكلمات المفتاحية: التعلم الذاتي الإشراف؛ نماذج الأساس؛ تعلم التمثيلات؛ السلاسل الزمنية غير المنتظمة؛ البيانات المرصودة جزئياً؛ الإشارات الفسيولوجية متعددة القنوات؛ المراقبة المستمرة للغلوكوز
