الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي - 8 ماي 1945 - سيدي بلعباس
École Supérieure en Informatique - 08 Mai 1945 - Sidi Bel Abbès

Dissertation

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

Theme

Multimodal Machine Learning for Cardiomyopathy Analysis

Presented By : Benyamina Yacine Lazreg

Submitted : 09/09/2025
Supervised by :

Dr. Nassima Dif	President
Dr. Benabdeslem Khalid	Supervisor
Dr. Philippe Chevalier	Co-supervisor
Dr. Chaib Souleymene	Co-supervisor
Dr. Taouli Amina	Examinator

Academic year : 2024/2025

Abstract

Cardiomyopathy, a disease of the heart muscle, presents a significant diagnostic challenge due to its complex and heterogeneous nature. This dissertation explores the application of multimodal machine learning for the analysis and classification of cardiomyopathy by integrating diverse data sources: tabular clinical records and genetic variants, time-series electrocardiogram (ECG) waveforms. The primary objective is to bridge the gap that exists in the current literature, which neglect feature selection and the use of genetic data and often focuses on single modalities.

This work introduces a novel supervised multimodal feature selection method and evaluates its efficacy against existing unsupervised techniques within a deep learning multimodal architecture. We analyse and process different data types and conduct a series of experiments to compare baseline model performance with soft and hard feature selection strategies.

Our results demonstrate that the proposed supervised feature selection method significantly improves model stability and generalization by aligning the selection process directly with the classification objective, outperforming the existing unsupervised approach. The final refined model, shows substantial improvement over the baseline, achieving strong performance across various metrics.

Ultimately, this dissertation highlights the potential of multimodal deep learning and targeted feature selection to address the complexities of cardiomyopathy analysis, offering a promising direction for enhancing clinical decision-support systems.

Keywords— Cardiomyopathy, Multimodal Machine Learning, Feature Selection, Deep Learning

Résumé

La cardiomyopathie, une maladie du muscle cardiaque, constitue un défi diagnostique majeur en raison de sa nature complexe et hétérogène. Ce mémoire explore l'application de l'apprentissage automatique multimodal pour l'analyse et la classification de la cardiomyopathie en intégrant diverses sources de données : dossiers cliniques tabulaires et variants génétiques, ainsi que les signaux électrocardiographiques (ECG) temporels. L'objectif principal est de combler une lacune existante dans la littérature, qui néglige souvent la sélection de caractéristiques et l'utilisation des données génétiques, et qui se limite généralement à une seule modalité.

Ce travail introduit une nouvelle méthode supervisée de sélection multimodale de caractéristiques et évalue son efficacité par rapport aux approches non supervisées existantes dans une architecture multimodale en apprentissage profond. Nous analysons et traitons différents types de données et menons une série d'expériences comparant la performance des modèles de base avec des stratégies de sélection de caractéristiques « douces » et « dures ».

Nos résultats démontrent que la méthode proposée de sélection supervisée améliore significativement la stabilité et la généralisation du modèle en alignant le processus de sélection directement avec l'objectif de classification, surpassant les approches non supervisées existantes. Le modèle final raffiné montre une amélioration substantielle par rapport au modèle de référence, atteignant de solides performances sur divers indicateurs.

En définitive, ce mémoire met en évidence le potentiel de l'apprentissage profond multimodal et de la sélection ciblée de caractéristiques pour relever la complexité de l'analyse de la cardiomyopathie, ouvrant une voie prometteuse pour renforcer les systèmes d'aide à la décision clinique.

Mots-clés— Cardiomyopathie, Apprentissage automatique multimodal, Sélection de caractéristiques, Apprentissage profond

الملخص

تُعَدّ اعتلالات عضلة القلب من الأمراض المعقدة والمتنوعة التي تطرح تحديات كبيرة على مستوى التشخيص. يتناول هذا البحث تطبيق التعلم الآلي متعدد الوسائط لتحليل وتصنيف اعتلال عضلة القلب من خلال دمج مصادر بيانات مختلفة: السجلات السريرية الجدولية والمتغيرات الجينية، إضافةً إلى إشارات تخطيط القلب الكهربائي (ECG) الزمنية. الهدف الأساسي هو سد الفجوة الموجودة في الأدبيات الحالية التي غالبًا ما تُهمل اختيار الخصائص واستخدام البيانات الجينية وتركّز على وسائط منفردة.

يقدّم هذا العمل منهجية جديدة لاختيار الخصائص بشكل إشرافي متعدد الوسائط، ويقيّم فعاليتها مقارنة بالأساليب غير الإشرافية القائمة ضمن إطار تعلم عميق متعدد الوسائط. قمنا بتحليل ومعالجة أنواع مختلفة من البيانات وإجراء سلسلة من التجارب لمقارنة أداء النماذج الأساسية مع استراتيجيات اختيار الخصائص «المرنة» و«الصارمة».

أظهرت النتائج أن الطريقة المقترَحة للاختيار الإشرافي للخصائص تُحسّن بشكل ملحوظ من استقرار النماذج وقدرتها على التعميم من خلال مواءمة عملية الاختيار مباشرة مع هدف التصنيف، متفوقةً على الأساليب غير الإشرافية الحالية. كما أظهر النموذج المُحسَّن تحسناً جوهرياً مقارنة بالنموذج المرجعي، محققًا أداءً قوياً عبر عدة مقاييس.

في النهاية، يسلط هذا البحث الضوء على الإمكانات الكبيرة للتعلم العميق متعدد الوسائط واختيار الخصائص الموجّه في معالجة تعقيدات اعتلال عضلة القلب، مما يفتح آفاقًا واعدة لتعزيز أنظمة دعم القرار السريري.

الكلمات المفتاحية--- اعتلال عضلة القلب، التعلم الآلي متعدد الوسائط، دمج المعلومات، الرعاية الصحية الذكية