

الجمهورية الشعبية الديمقراطية الجزائرية
وزارة التعليم العالي والبحث العلمي
المدرسة العليا للإعلام الآلي 08 ماي 1945 بسيدي بلعباس

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès



THESIS

To obtain the diploma of **Engineering Degree**
Field: **Computer Science**
Major: **Artificial Intelligence and Data Science**

Topic

**An End-to-End Deep Learning Framework for
Automatic Cardiomegaly Diagnosis Using Anatomical
Biomarkers and Multimodal Data**

Presented by:
Mariem Boukennouche

Submission Date: **September, 2026**
In front of the jury composed of:

Pr. ELOUALI Nadia
Dr. Samir AWAD
Dr. SAIDI Imène

President
Supervisor
Examiner

Academic Year : 2025/2026

Abstract

Cardiothoracic ratio (CTR) and cardiopulmonary area ratio (CPAR) remain the standard digital biomarkers for automated cardiomegaly assessment, yet both reduce the heart and lung segmentation mask to scalar measurements, discarding shape information such as curvature, symmetry, and orientation. This work extends the multimodal fusion framework of Duvieusart et al. (2022) by reimplementing its pipeline on the MIMIC-CXR/MIMIC-IV cohort and introducing an extended set of interpretable geometric descriptors, computed from the same segmentation masks as CTR and CPAR, as additional digital biomarkers. Across a gradient-boosted fusion model and a deep multimodal network, adding these descriptors consistently improves classification performance (AUC up to 0.8959, over 15 fold/seed combinations), with two of the three new descriptors ranking among the most informative features overall. An anatomy-guided attention mechanism, supervised against the heart segmentation mask, is further proposed to correct a shortcut-learning issue identified via Grad-CAM, yielding the best-performing multimodal variant (test AUC 0.8706). The resulting pipeline is packaged into a deployable demonstration application.

Keywords: cardiomegaly, chest radiograph, multimodal deep learning, digital biomarkers, cardiothoracic ratio, image segmentation, interpretability

الملخص

تُعد نسبة القلب إلى الصدر (CTR) ونسبة مساحة القلب إلى الرئتين (CPAR) المقياسين المعيارين لتقييم تضخم القلب آلياً، غير أن كليهما يحتزل قناع تجزئة القلب والرئتين إلى قياس عددي واحد، متجاهلاً معلومات الشكل كالانحناء والتماثل والاتجاه. يوسّع هذا العمل إطار الدمج متعدد الوسائط الذي اقترحه Duvieusart وآخرون (2022) من خلال إعادة تنفيذ نظامه على مجموعة بيانات MIMIC-CXR/MIMIC-IV، وإدخال مجموعة موسّعة من الواصفات الهندسية القابلة للتفسير، محسوبة من أقنعة التجزئة نفسها المستخدمة لحساب CTR وCPAR، كمقاييس حيوية رقمية إضافية. وعبر نموذج دمج قائم على التعزيز التدريجي وشبكة عصبية عميقة متعددة الوسائط، أدت إضافة هذه الواصفات إلى تحسين ثابت في أداء التصنيف (AUC بلغ 8959.0، عبر 15 تركيبة من الطيات والبذور العشوائية)، مع تصدر اثنين من الواصفات الثلاثة الجديدة قائمة أكثر الخصائص إفادة. كما اقترحت آلية انتباه موجهة تشریحياً، تخضع للإشراف بقناع تجزئة القلب، لمعالجة مشكلة تعلّم مختصر كشفها تحليل Grad-CAM، فأُسفرت عن أفضل نسخة متعددة الوسائط (AUC اختبار 8706.0). وقد جرى تعبئة الخطط الإجرائي الناتج ضمن تطبيق توضيحي قابل للنشر.

الكلمات المفتاحية: تضخم القلب، الأشعة السينية للصدر، التعلم العميق متعدد الوسائط، المؤشرات الحيوية الرقمية، النسبة القلبية الصدرية، تجزئة الصورة، قابلية التفسير

Résumé

Le rapport cardiothoracique (RCT) et le rapport de surface cardiopulmonaire (RSCP) restent les biomarqueurs numériques standards pour l'évaluation automatisée de la cardiomégalie, mais tous deux réduisent le masque de segmentation cœur-poumons à des mesures scalaires, en écartant les informations de forme telles que la courbure, la symétrie et l'orientation. Ce travail étend le cadre de fusion multimodale de Duvieusart et al. (2022) en réimplémentant son pipeline sur la cohorte MIMIC-CXR/MIMIC-IV et en introduisant un ensemble étendu de descripteurs géométriques interprétables, calculés à partir des mêmes masques de segmentation que le RCT et le RSCP, comme biomarqueurs numériques supplémentaires. Sur un modèle de fusion par gradient boosting et un réseau multimodal profond, l'ajout de ces descripteurs améliore de façon constante les performances de classification (AUC jusqu'à 0,8959, sur 15 combinaisons de folds/seeds), deux des trois nouveaux descripteurs figurant parmi les caractéristiques les plus informatives. Un mécanisme d'attention guidé par l'anatomie, supervisé par le masque de segmentation du cœur, est en outre proposé pour corriger un problème d'apprentissage de raccourci identifié par Grad-CAM, produisant la meilleure variante multimodale (AUC test 0,8706). Le pipeline résultant est packagé dans une application de démonstration déployable.

Mots clés : cardiomégalie, radiographie thoracique, apprentissage profond multimodal, biomarqueurs numériques, rapport cardio-thoracique, segmentation d'image, interprétabilité