

الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المدرسة العليا للإعلام الآلي - 08 ماي 1945 - بسيدي بلعباس
École Supérieure en Informatique
- 08 Mai 1945 - Sidi Bel Abbès



END-OF-STUDIES THESIS

To obtain the diploma of Engineer (Ingénieur d'État)
Field: Computer Science
Specialty: Artificial Intelligence and Data Science (AI&DS)

Theme

**TaViT: A Treatment-Aware Temporal Vision
Transformer for Longitudinal Glioma Trajectory
Modeling and Visualization**

Presented by:

BOUFABA Mohamed

Defended on: **July 2026**

[President Name]

Prof. Belkacem CHIKHAOU

Prof. Belkacem KHALDI

[Examiner Name]

[Guest Name]

In front of the jury composed of:

President

Supervisor

Supervisor

Examiner

Invited

Academic Year: 2025 / 2026

Abstract

This engineering report implements and validates TaViT, a treatment-aware temporal Vision Transformer that estimates the volume trajectory of glioma sub-regions from a sequence of magnetic resonance imaging studies. The pipeline has three stages. A fine-tuned nnU-Net backbone segments the whole tumour, the tumour core, and the enhancing tumour from four-channel MRI and reaches a mean Dice score of 0.877 across the three sub-regions. The bottleneck of this segmentation network supplies a compact per-scan embedding. The choice of backbone was settled by experiment: an 18-test battery compared a convolutional encoder and a transformer encoder as embedding sources, and the convolutional one was selected on evidence headed by an effective rank of 165.7 against 7.6. A temporal Transformer then reads the embeddings as a sequence, conditioned on a per-scan treatment token and a patient-level molecular vector, with a time-aware attention bias that respects the irregular spacing of clinical scans. A conditional diffusion model completes the pipeline by turning the predicted trajectory into a smooth video of tumour evolution that a clinician can inspect.

Trained on the MU-Glioma-Post cohort of 203 patients and 596 post-operative scans under three-fold cross-validation, the model estimates sub-region volumes with a coefficient of determination of 0.918 for the whole tumour, 0.953 for the tumour core, and 0.952 for the enhancing tumour, with a mean absolute error of 15.5, 4.7, and 3.7 millilitres. The fit is stable across folds, and the model recovers the direction of change, growth or shrinkage, in 83.3% of consecutive scan pairs. Applied without adaptation to the independent multi-site BraTS-PTG cohort, the model’s accuracy collapses, but calibration on only ten local patients recovers a whole-tumour coefficient of determination of 0.820, about 89% of the in-distribution value, which points to a practical way to onboard a new site. Monte-Carlo dropout adds a per-prediction uncertainty estimate, and attention maps make the model’s temporal focus inspectable.

Keywords: glioma; longitudinal MRI; Vision Transformer; diffusion model; treatment-aware modelling; explainable AI; nnU-Net; external validation.

Résumé

Ce rapport de projet de fin d'études met en œuvre et valide TaViT, un Vision Transformer temporel tenant compte du traitement, qui estime la trajectoire volumique des sous-régions d'un gliome à partir d'une séquence d'examens d'imagerie par résonance magnétique. Le pipeline comporte trois étapes. Un réseau nnU-Net affiné segmente la tumeur entière, le noyau tumoral et la tumeur rehaussée à partir d'une IRM à quatre canaux, et atteint un score de Dice moyen de 0,877 sur les trois sous-régions. Le goulot d'étranglement de ce réseau de segmentation fournit un plongement compact par examen. Le choix du réseau a été tranché par l'expérience : une batterie de dix-huit tests a comparé un encodeur convolutif et un encodeur à transformeur comme sources de plongement, et le convolutif a été retenu sur la base d'un rang effectif de 165,7 contre 7,6. Un Transformer temporel lit ensuite les plongements comme une séquence, conditionné par un jeton de traitement par examen et par un vecteur moléculaire au niveau du patient, avec un biais d'attention sensible au temps qui respecte l'espacement irrégulier des examens cliniques. Un modèle de diffusion conditionnel complète le pipeline en transformant la trajectoire prédite en une vidéo fluide de l'évolution tumorale que le clinicien peut inspecter.

Entraîné sur la cohorte MU-Glioma-Post de 203 patients et 596 examens postopératoires en validation croisée à trois plis, le modèle estime les volumes des sous-régions avec un coefficient de détermination de 0,918 pour la tumeur entière, 0,953 pour le noyau tumoral et 0,952 pour la tumeur rehaussée, et une erreur absolue moyenne de 15,5, 4,7 et 3,7 millilitres. L'ajustement reste stable d'un pli à l'autre, et le modèle retrouve le sens de l'évolution, croissance ou régression, dans 83,3% des paires d'examens consécutifs. Appliqué sans adaptation à la cohorte multicentrique indépendante BraTS-PTG, le modèle voit sa précision s'effondrer, mais un calibrage sur seulement dix patients locaux permet de retrouver un coefficient de détermination de 0,820 pour la tumeur entière, soit environ 89% de la valeur en distribution, ce qui ouvre une voie pratique pour intégrer un nouveau site. Le dropout de Monte-Carlo ajoute une estimation d'incertitude par prédiction, et les cartes d'attention rendent inspectable la focalisation temporelle du modèle.

Mots-clés : gliome ; IRM longitudinale ; Vision Transformer ; modèle de diffusion ; modélisation tenant compte du traitement ; IA explicable ; nnU-Net ; validation externe.

ملخص

ينفذ هذا التقرير الهندسي ويتحقق من صحة نظام TaViT، وهو محوّل رؤية زمني وإعّج بالعلاج يقدر المسار الحجمي للمناطق الفرعية للورم الدبقي انطلاقاً من سلسلة فحوصات بالرنين المغناطيسي. يتألف خطّ المعالجة من ثلاث مراحل: أولاً، شبكة nnU-Net مضبوطة تجزئ الورم الكامل والنواة الورمية والورم المعزّز من تصوير رباعي القنوات، محقّقة معدّل Dice يبلغ 0.877 على المناطق الثلاث. ثمّ يوفّر عنق الزجاجة في شبكة التجزئة تمثيلاً مدججاً لكلّ فحص. وقد حُسم اختيار الشبكة الأساسية تجريبياً: قارنت مجموعة من ثمانية عشر اختباراً بين مُرمّز التفاني ومُرمّز محوّل كمصادر للتمثيل، واختير المُرمّز الالتفاني بناءً على رتبة فعّالة بلغت 165.7 مقابل 7.6. بعد ذلك يقرأ محوّل زمني التمثيلات كسلسلة، مشروطاً برمز العلاج لكلّ فحص ومتّجه جزئيّ على مستوى المريض، مع انخياز انتباه حسّاس للزمن يراعي التباعد غير المنتظم للفحوصات السريرية. ويُكمل نموذج انتشار مشروط خطّ المعالجة بتحويل المسار المتوقّع إلى فيديو سلس لتطوّر الورم يمكن للطبيب الاطلاع عليه.

دُرّب النموذج على مجموعة بيانات MU-Glioma-Post المكوّنة من 203 مرضى و596 فصاً بعد الجراحة باستخدام التحقق المتقاطع ثلاثي الطيّات، وحقق معامل تحديد 0.918 للورم الكامل و0.953 للنواة الورمية و0.952 للورم المعزّز، مع خطأ مطلق متوسّط 15.5 و4.7 و3.7 ملليتر. يبقى الأداء مستقرّاً عبر الطيّات، ويستعيد النموذج اتّجاه التغيّر، نموّاً أو تراجعاً، في 83.3% من أزواج الفحوصات المتتالية. وعند تطبيقه دون تكييف على مجموعة بيانات BraTS-PTG المستقلّة متعدّدة المراكز ينهار الأداء، لكنّ معايرة على عشرة مرضى محليّين فقط تستعيد معامل تحديد 0.820 للورم الكامل، أي نحو 89% من القيمة الأصليّة، ممّا يشير إلى طريقة عمليّة لضمّ مواقع جديدة. ويضيف إسقاط Monte-Carlo تقديراً لعدم اليقين لكلّ تنبؤ، فيما تجعل خرائط الانتباه تركيز النموذج الزمني قابلاً للفحص.

الكلمات المفتاحية: الورم الدبقي؛ التصوير بالرنين المغناطيسي الطولي؛ محوّل الرؤية؛ نموذج الانتشار؛ النمذجة الواعية بالعلاج؛ الذكاء الاصطناعي القابل للتفسير؛ nnU-Net؛ التحقق الخارجي.