Supplementary Material for METIS 2026: Bridging Clinical and Computational Communities for the Next Generation of Medical Imaging AI

Alin Navas[1,*], Mehdi Astaraki[8,‡], Fatima Ez-Zahraa Bazay[11,‡], Francesco Chiumento[10,‡], Sanket Kachole[2,‡], Jongwoon Lee[9,‡], Ezekiel Ayodeji Oladejo[13,‡], António S Soares[12,‡], Jeffrey D Rudie[17,18], Evan Calabrese[14], Mariam S Aboian[19,20], Aon Safdar[1,*], Mohamed Saadeldin[1,*], Nuala A. Healy[3,*], Gilberto Ochoa-Ruiz[5,*], Sharib Ali[6,*], Bulent Yilmaz[7,*], Bartlomiej W. Papież[4,*], Spyridon Bakas[2,15,16,†,*], Kathleen M. Curran[1,†,*]

[1]University College Dublin, Dublin, Ireland
[2]Division of Computational Pathology, Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA
[3]Department of Radiology, Royal College of Surgeons Ireland and Department of Radiology, Beaumont Hospital, Dublin, Ireland
[4]University of Oxford, Oxford, UK
[5]Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, Monterrey, Mexico
[6]University of Leeds, Leeds, UK
[7]Gulf University for Science and Technology, Kuwait
[8]Department of Medical Radiation Physics, Stockholm University & Karolinska Institutet, Stockholm, Sweden
[9]University of Rochester, Rochester, NY, USA
[10]SFI Centre for Research Training in Machine Learning (ML-Labs), Dublin City University, Dublin, Ireland
[11]LRIT, Faculty of Sciences in Rabat, Mohammed V University, Rabat, Morocco
[12]Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal, and Instituto Português Oncologia, Lisbon, Portugal
[13]Slum and Rural Health Initiative Network (SRHIN), Ibadan, Nigeria
[14]Duke University Medical Center, Durham, USA
[15]Departments of Biostatistics and Health Data Science; Radiology and Imaging Sciences; Neurological Surgery, Indiana University School of Medicine, Indianapolis, IN, USA
[16]Department of Computer Science, Luddy School of Informatics, Computing, and Engineering, Indianapolis, IN, USA
[17]Department of Radiology, University of California San Diego, San Diego, CA, USA
[18]Department of Radiology, Scripps Clinic Medical Group, San Diego, CA, USA
[19]Department of Radiology, Children's Hospital on Philadelphia (CHOP), Philadelphia, PA, USA
[20]Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[‡]Accepted proposal lead
* People involved in the METIS organizing committee
[†]Equal senior corresponding authors: spbakas@iu.edu; kathleen.curran@ucd.ie

1. Accepted Proposal #1

1.1. Title : An Integrated Computational Framework for Guideline-Compliant, Uncertainty-Guided Clinical Target Volume Delineation in Glioblastoma Radiotherapy

Table 1. Author and affiliation of accepted proposal 1

Author	Affiliation
Mehdi Astaraki	Karolinska Institute

1.2. Clinical problem : Glioblastoma is the most aggressive primary adult brain tumor, with annual incidence of 3.53/100,000 individuals and median overall survival of 15-18 months[1]. Standard of care is surgery followed by temozolomide and radiotherapy (RT)[2]. RT planning requires defining the visible Gross Tumor Volume (GTV) and applying an expansion margin to construct the Clinical Target Volume (CTV) covering subclinical microscopic spread. Automated GTV segmentation is largely an addressed computational problem as indicated by the MICCAI Brain Tumor Segmentation (BraTS) challenge[3], where state-of-the-art deep learning (DL) models match expert neuroradiologists' performance. However, automated definition of the CTV and the deployment of such DL models remains an unsolved, high-priority clinical bottleneck. Unlike visible GTV, microscopic infiltration lacks distinct contrast on MRI. Clinicians must estimate anisotropic cellular migration along white matter tracts while manually enforcing complex, non-uniform anatomical barrier rules (e.g., stopping expansion at the skull or falx cerebri). Therefore, manual CTV delineation is labor-intensive, unstandardized, and subject to inter-observer variability. This proposal addresses this unmet need by introducing a computational framework automating uncertainty guided CTV generation, while adhering to the 2023-ESTRO-EANO consensus recommendations[4]. Serving radiation oncologists, neuroradiologists, and medical physicists, success will be measured by comparison with manually defined CTV and reduced inter-observer variance.

1.3. How AI could address this clinical need : This project proposes three distinct, synergistic AI solutions for automated, and interactive CTV segmentations. First, a deterministic guideline-compliant pipeline operationalizes 2023 ESTRO-EANO consensus recommendations by extracting GTV masks from multi-parametric MRIs and executing anatomically truncated 3D margin expansions using whole-brain parcellation maps. Second, a direct DL model trained across an international multi-center consortium via a privacy-preserving Federated Learning (FL) framework [5] learns diverse multi institutional delineation styles, generating probabilistic CTV prediction maps and epistemic uncertainty maps. Third, an interactive uncertainty-guided human-in-the-loop workstation fuses these complementary outputs, highlighting spatial discordance zones, where

deterministic rules disagree with data-driven probabilistic predictions as interactive uncertainty overlays. Incorporating a promptable, click-conditioned 3D segmentation architecture, the workstation enables radiation oncologists to rapidly refine target contours in real time with minimal positive or negative clicks within uncertainty regions, achieving an optimal balance between algorithmic automation, physical safety, and complete clinician autonomy.

1.4. Data available for model development : Model development and multi-center training utilize established open-access benchmarks and retrospective clinical imaging cohorts. Primary GTV segmentation architectures will be trained on the open-access BraTS benchmark dataset comprising over 3,800 annotated MRIs[6][3], incorporating channel-dropout augmentations ensuring high segmentation accuracy despite missing MRI sequences. Initial clinical evaluation of the deterministic expansion pipeline employs a retrospective clinical cohort of over 300 adult GBM cases aggregated across three university hospitals in Europe and North America, containing baseline planning MRIs co-registered with clinical CT scans and corresponding expert-contoured DICOM-RT objects(RTSTRUCT). Furthermore, the global FL consortium currently encompasses three clinical centers contributing retrospective cohorts of at least 100 annotated GBM RT planning cases each, which we hope METIS will help to expand this FL consortium. Data challenges, including scanner heterogeneity, variable slice thicknesses, and missing modalities, will be systematically mitigated using standardized spatial normalization, face/skull-stripping preprocessing pipelines, and class-imbalance loss functions tailored for target segmentation.

1.5. How the proposed model could be externally validated : Validation of the proposed framework follows a two-tier evaluation strategy incorporating computational stability analysis and multi-center clinical benchmarking. First, a large-scale computational stability analysis will execute the deterministic pipeline across primary GTV predictions from over 100 top-performing BraTS challenge algorithms, quantifying the spatial error-dampening capabilities of rule-based anatomical truncations [7]. Second, multi-center clinical validation across hold-out testing sets from participating institutes will benchmark all three AI solutions against reference manual contours using volumetric segmentation, and dosimetric coverage metrics. Regulatory governance, data privacy, and strict GDPR compliance are guaranteed by performing local automated DICOM de-identification and face/skull-stripping algorithms prior to model processing. Patient imaging datasets will remain strictly secure behind local institutional firewalls; the FL infrastructure transmits only encrypted model parameters and secure gradient updates, ensuring exceptional overall feasibility for conducting a compliant, multi institutional international validation study without cross-border transfer of individual image files.

1.6. How the AI solution would fit into routine clinical practice : The core deliverable of this project is a modular, clinical-grade package incorporating all three CTV solutions, deployed on research PACS and research versions of RT Planning Systems (e.g., RayStation) within clinical centers. Integrated into rou-

tine pre-treatment workflows, the pipeline automatically processes the MRI volume, generates the guideline-compliant CTV, and maps spatial uncertainty zones directly into image viewer windows. Radiation oncologists can review automated contours, inspect highlighted uncertainty regions, and apply click-guided interactive corrections prior to final dosimetric planning. Potential adoption barriers, such as inter-institutional protocol variations and clinician skepticism toward fully automated delineation, are mitigated by the framework's modular design. Solution 1 provides parameter toggles allowing clinicians to adapt expansion margins and edema inclusions to institutional standards, while Solution 3 preserves complete physician autonomy by placing the clinician in an interactive decision-support loop. Clinical impact will be rigorously evaluated across four primary axes: inter-observer consistency, workflow efficiency, dosimetric quality, and pattern-of-failure classification. Success will be demonstrated by a statistically significant reduction in contouring variability among clinicians, reduced manual editing time per patient, verified preservation of 95% dose coverage to subclinical targets with improved sparing of adjacent organs at risk, and enhanced spatial classification of future recurrences during post treatment follow-up.

Reference

[1] K. Bender et al., "What is the role of the subventricular zone in radiotherapy of glioblastoma patients?," Radiotherapy and Oncology, vol. 158, pp. 138–145, May 2021, doi: 10.1016/j.radonc.2021.02.017.

[2] R. Stupp et al., "Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma," New England Journal of Medicine, vol. 352, no. 10, pp. 987–996, Mar. 2005, doi: 10.1056/NEJMOA043330;PAGE:STRING:ARTICLE/CHAPTER.

[3] U. Baid et al., "The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification," Jul. 2021, Accessed: Mar. 28, 2024. [Online]. Available: https://arxiv.org/abs/2107.02314v2

[4] M. Niyazi et al., "ESTRO-EANO guideline on target delineation and radiotherapy details for glioblastoma," Radiotherapy and Oncology, vol. 184, p. 109663, Jul. 2023, doi: 10.1016/j.radonc.2023.109663.

[5] S. Pati et al., "Federated learning enables big data for rare cancer boundary detection," Nature Communications 2022 13:1, vol. 13, no. 1, pp. 1–17, Dec. 2022, doi: 10.1038/s41467-022-33407-5.

[6] M. C. de Verdier et al., "The 2024 Brain Tumor Segmentation (BraTS) Challenge: Glioma Segmentation on Post-treatment MRI," Eugenio Iglesias, vol. 33, May 2024, Accessed: Jul. 28, 2024. [Online]. Available: https://arxiv.org/abs/2405.18368v1

[7] S. Bakas et al., Eds., Segmentation, Classification, and Synthesis for Brain Tumors and Traumatic Brain Injuries, vol. 16376. in Lecture Notes in Computer Science, vol. 16376. Cham: Springer Nature Switzerland, 2026. doi: 10.1007/978-3-032-16365-3.

2. Accepted Proposal #2

2.1. Title : Domain-Generalizable EEG Biomarkers for Neurodevelopmental Disorder Screening

Table 2. Author and affiliation of accepted proposal 2

Author	Affiliation
Fatima Ez-Zahraa Bazay	Mohammed V University,Morocco

2.2. Clinical problem : Diagnosis of neurodevelopmental disorders such as Autism Spectrum Disorder (ASD) and Attention-Deficit/Hyperactivity Disorder (ADHD) still relies primarily on behavioral observation and parent-report instruments in pediatric neurology and child psychiatry clinics. These evaluations are subjective, require scarce specialist availability, and frequently delay diagnosis by one to three years past first parental concern, narrowing the window for effective early intervention. ASD alone affects an estimated 1–2% of children worldwide, and ADHD a further 5–7%, placing substantial pressure on already limited specialist capacity. Clinicians currently lack an objective, repeatable physiological measure that can be used alongside behavioral assessment to support earlier and more consistent diagnostic decisions. Rigorous evaluation of EEG-based classifiers shows that reported accuracy can drop sharply on fully held-out data, underscoring how far current biomarkers are from clinical reliability [3]. Success would be measured by a reduction in time to diagnosis, and by demonstrating that an extracted biomarker remains reliable when applied to patients, devices, and sites different from those on which it was developed, rather than only performing well on a single local cohort.

2.3. How AI could address this clinical need : The proposed approach extracts EEG functional-connectivity biomarkers using models trained with domain generaliza tion techniques, namely adversarial and contrastive site-invariance training, so that performance does not collapse when applied to a new site, device, or patient population. Inputs are resting-state (or task-based, to be agreed with the clinical partner) high-density EEG recordings; outputs are an interpretable connectivity-based biomarker profile and an associated risk score, accompanied by an explainable AI component highlighting which network features drove the prediction. AI is appropriate here because relevant spatiotemporal connectivity patterns are not readily assessable by visual EEG inspection. Compared with current practice, the expected benefit is an objective, low-cost, better-tolerated, and repeatable measure that can complement behavioral evaluation, in contrast to resting-state fMRI, which is expensive and poorly tolerated by young or sensory-sensitive children.

2.4. Data available for model development : The Healthy Brain Network (HBN) EEG dataset provides openly accessible high-density pediatric EEG (ages

5–21, over 2,600 participants) via OpenNeuro and NEMAR [1], built on the broader HBN biobank of transdiagnostic pediatric phenotypic and neuroimaging data [2], and will serve as an open-access base for methodological development and preliminary cross-site benchmarking. Diagnostic phenotypic data, including clinician-confirmed diagnoses, are accessible separately through a data use agreement with the data custodian. A non-public retrospective cohort from the clinical partner's site, described only in general terms at this stage, would be required for genuine external validation. Anticipated challenges include class imbalance between affected and typically developing participants, heterogeneity in acquisition devices and montages across sites, and the initial absence of finalized annotations for any partner-site data, which would need to be defined jointly once a collaboration is established.

2.5. How the proposed model could be externally validated : External validation would rely on EEG cohorts from one or more pediatric neurodevelopmental clinics, described generically to preserve anonymity, ideally spanning more than one country or acquisition system to test genuine cross-site generalization. Key regulatory considerations include GDPR (or equivalent local data protection law), institutional ethics approval, and data-sharing agreements governing any identifiable clinical information. Potential approaches to overcome these barriers include federated or model-to-data evaluation, where models are sent to the data rather than data being centralized, together with pseudonymization and secure research environments. Given that open pediatric EEG data are already available for pretraining, a staged feasibility plan, methodological development on open data followed by a smaller-scale external validation at the partner site, is realistic within the proposed one-year mentoring timeline.

2.6. How the AI solution would fit into routine clinical practice : Intended end users are pediatric neurologists, child psychiatrists, and clinical neurophysiologists involved in the diagnostic work-up of suspected neurodevelopmental disorders. The tool would be used as an adjunct at the point of an existing clinical EEG recording, rather than requiring an additional visit, with outputs reviewed alongside standard behavioral assessment rather than replacing clinical judgment. Integration would ideally connect with existing EEG acquisition and reporting systems to minimize added workflow steps. Potential barriers to adoption include clinician trust in an unfamiliar biomarker, the need for outputs to be interpretable rather than opaque, uncertainty around reimbursement, and the practical burden of demonstrating reliability before any diagnostic use is considered. Clinical impact would be evaluated through concordance between the biomarker and existing clinical diagnosis, reduction in time to diagnosis, and structured feedback from clinicians on usability and trust, rather than classification accuracy alone.

References

[1] Shirazi, S.Y., Franco, A., Hoffmann, M.S., Esper, N.B., Truong, D., Delorme, A., Milham, M.P., Makeig, S.: HBN-EEG: The FAIR implementa-

tion of the Healthy Brain Network (HBN) electroencephalography dataset. bioRxiv (2024)

[2] Alexander, L.M., et al.: An open resource for transdiagnostic research in pediatric mental health and learning disorders. Scientific Data 4, 170181 (2017)

[3] Li, Q., Weiland, R.F., Konvalinka, I., Mansvelder, H.D., Andersen, T.S., Smit, D.J.A., Begeer, S., Linkenkaer-Hansen, K.: Intellectually able adults with autism spectrum disorder show typical resting-state EEG activity. Scientific Reports 12, 18426 (2022)

3. Accepted Proposal #3

3.1. Title : MRI-Only Amyloid-β Assessment to Prioritize Confirmatory Testing in Alzheimer's Disease

Table 3. Author and affiliation of accepted proposal 3

Author	**Affiliation**
Francesco Chiumento	Dublin City University, Ireland

3.2. Clinical problem : Amyloid-β (Aβ) confirmation is becoming increasingly important in Alzheimer's disease care pathways. In dementia assessment, patients with mild cognitive impairment (MCI) or in the early symptomatic stages increasingly need biomarker evidence of amyloid pathology, while anti-amyloid therapies such as lecanemab and donanemab require its confirmation before treatment initiation [10, 17]. In 2021, approximately 57 million people were living with dementia worldwide, and this number is projected to reach 78 million by 2030 [15, 16]. Confirmatory testing is a major bottleneck, as amyloid PET is costly, exposes patients to ionizing radiation, and does not scale to population-level screening. A health economic model found routine use at the MCI stage not cost-effective [7], while demand is projected to increase as much as 20-fold following the introduction of anti-amyloid therapies [4, 13]. Real-world data from a tertiary memory clinic show that only a limited proportion of patients meet appropriate use recommendations for these therapies [10], supporting the need for scalable strategies to prioritize candidates before confirmatory PET. Success would be measured by agreement with PET-derived Centiloid values, particularly near the positivity threshold, and by the ability to identify individuals who may benefit from confirmatory PET [2, 5, 8].

3.3. How AI could address this clinical need : The proposed task concerns MRI-only estimation of PET-derived amyloid burden. The inputs comprise one or more clinically available MRI contrasts (T1w, T2w, FLAIR, T2*). PET, APOE genotype, and cognitive scores are not required at inference, avoiding inputs often unavailable in routine workflows. Feasibility is supported by a

recently reported PET-guided knowledge distillation framework, where PET supervises training but inference uses MRI alone, achieving best Aβ-classification AUCs of 0.74 on OASIS-3 and 0.68 on ADNI [3]. Beyond binary classification, three directions are proposed: continuous Centiloid (CL) estimation; longitudinal modeling of conversion probability and annual CL accumulation rate; and region-level radiomic attribution over anatomical parcellations. Amyloid plaques are not directly visible on MRI, so burden is inferred from patterns of atrophy, white-matter alteration, and cerebrovascular change [18]. Compared with current practice, the benefit is a Centiloid-scale estimate from scans already acquired, expressed in the same units used to report amyloid PET [6].

3.4. Data available for model development : Development would use two public longitudinal cohorts accessed under data use agreements: OASIS-3 (1,379 subjects) and ADNI (over 5,100 subjects), both with subsets providing paired MRI, amyloid PET, and CL quantification [3]. Two imaging modalities are used: structural and vascular MRI (T1w, T2w, FLAIR, T2*) and paired amyloid PET acquired using PiB and florbetapir tracers. PET-derived CL values are available in the datasets and would be used in three ways: thresholded at CL>20.6 to define binary status [3], used directly as the regression target, and compared across visits in the subset with at least two PET timepoints [1]. Anticipated challenges include class imbalance across cohorts (with an Aβ-positive prevalence of approximately 22% to 50% [3]), an imbalanced distribution of CL values that may bias the regression head toward more frequent target values, residual tracer- and pipeline-related variability despite CL harmonization [11], and incomplete contrast availability across sessions.

3.5. How the proposed model could be externally validated : The next step is to assess the three directions, each using subjects from OASIS-3 and ADNI who meet the corresponding data requirements. Cross-dataset transfer of the existing binary classifier has already been tested bidirectionally [3]. External validation would extend to the AIBL dataset, which adds flutemetamol alongside the already available PiB and florbetapir tracers, allowing stratification by tracer, scanner, and baseline CL; A4/LEARN provides further preclinical cohorts. These datasets are de-identified and available under data use agreements, and since they span different countries, scanners, and tracers, a retrospective multicenter evaluation is feasible. Evaluation would be output-specific: agreement with PET-derived CL and near-threshold error for CL estimation; time-dependent discrimination and calibration for conversion risk; MAE (CL/year) for annual accumulation rate; and ROI ranking stability for interpretability. Clinical use would ultimately require prospective validation against PET, CSF, or a validated plasma assay, under local ethics approval.

3.6. How the AI solution would fit into routine clinical practice : Intended users are neuroradiologists reporting dementia MRI and memory-clinic physicians. Because MRI is already the primary modality in most dementia assessment pathways [9, 14], Centiloid estimation is the most immediate direction: it could in principle be performed on studies already available in PACS, with-

out requiring additional imaging, and the estimate delivered to the radiologist within the existing reporting workflow [12]. Several barriers remain. The three directions are proposed rather than evaluated, so their performance must be quantified; even then, they should serve as decision-support tools for prioritizing confirmatory testing, not as standalone diagnostic confirmation. Atrophy and white-matter changes also arise from aging and cerebrovascular disease [18], potentially biasing estimates when the underlying cause is not amyloid. Regulatory approval would also be required before deployment. Impact would be assessed by the proportion of confirmatory scans prioritized or deferred, time to confirmation, and agreement with measured CL. Clinician input would determine which direction to develop first and define clinically meaningful endpoints.

References

[1] Bollack, A., Collij, L.E., García, D.V., Shekari, M., Altomare, D., Payoux, P., Dubois, B., Grau-Rivera, O., Boada, M., Marquié, M., Nordberg, A., Walker, Z., Scheltens, P., Schöll, M., Wolz, R., Schott, J.M., Gismondi, R., Stephens, A., Buckley, C., Frisoni, G.B., Hanseeuw, B., Visser, P.J., Vandenberghe, R., Drzezga, A., Yaqub, M., Boellaard, R., Gispert, J.D., Markiewicz, P., Cash, D.M., Farrar, G., Barkhof, F.: Investigating reliable amyloid accumulation in Centiloids: Results from the AMYPAD Prognostic and Natural History Study. Alzheimer's & Dementia 20(5), 3429–3441 (Apr 2024). https://doi.org/10.1002/alz.13761

[2] Chattopadhyay, T., Ozarkar, S.S., Buwa, K., Joshy, N.A., Komandur, D., Naik, J., Thomopoulos, S.I., Ver Steeg, G., Ambite, J.L., Thompson, P.M.: Comparison of deep learning architectures for predicting amyloid positivity in Alzheimer's disease, mild cognitive impairment, and healthy aging, from T1-weighted brain structural MRI. Frontiers in Neuroscience 18, 1387196 (Jul 2024). https://doi.org/10.3389/fnins.2024.1387196

[3] Chiumento, F., Dietlmeier, J., Killeen, R.P., Curran, K.M., O'Connor, N.E., Liu, M.: Cross-Modal Knowledge Distillation for PET-Free Amyloid-Beta Detection from MRI. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 6164–6175 (Jun 2026)

[4] Garibotto, V., Albert, N.L., Barthel, H., van Berckel, B., Boellaard, R., Brendel, M., Cecchin, D., Ekmekcioglu, O., van de Giessen, E., Guedj, E., Lammerstma, A.A., Semah, F., Traub-Weidinger, T., Van Weehaeghe, D., Morbelli, S., for the EANM Neuroimaging Committee: The approval of a disease-modifying treatment for Alzheimer's disease: Impact and consequences for the nuclear medicine community. European Journal of Nuclear Medicine and Molecular Imaging 48(10), 3033–3036 (Sep 2021). https://doi.org/10.1007/s00259-021-05485-y

[5] Kim, D., Ottesen, J.A., Kumar, A., Ho, B.C., Bismuth, E., Young, C.B., Mormino, E., Zaharchuk, G.: Deep Learning-Based Prediction of PET Amyloid Status Using MRI. AJNR. American journal of neuroradiology 46(12), 2590–2598 (Dec 2025). https://doi.org/10.3174/ajnr.A8899

[6] Klunk, W.E., Koeppe, R.A., Price, J.C., Benzinger, T., Devous, M.D., Jagust, W., Johnson, K., Mathis, C.A., Minhas, D., Pontecorvo, M.J., Rowe, C.C., Skovronsky, D., Mintun, M.: The Centiloid Project: Standardizing Quantitative Amyloid Plaque Estimation by PET. Alzheimer's & dementia : the journal of the Alzheimer's Association 11(1), 1–15.e4 (Jan 2015). https://doi.org/10.1016/j.jalz.2014.07.003

[7] Lee, Y.S., Youn, H., Jeong, H.G., Lee, T.J., Han, J.W., Park, J.H., Kim, K.W.: Cost-effectiveness of using amyloid positron emission tomography in individuals with mild cognitive impairment. Cost Effectiveness and Resource Allocation 19(1), 50 (Aug 2021). https://doi.org/10.1186/s12962-021-00300-9

[8] Lew, C.O., Zhou, L., Mazurowski, M.A., Doraiswamy, P.M., Petrella, J.R., Alzheimer's Disease Neuroimaging Initiative: MRI-based Deep Learning Assessment of Amyloid, Tau, and Neurodegeneration Biomarker Status across the Alzheimer Disease Spectrum. Radiology 309(1), e222441 (Oct 2023). https://doi.org/10.1148/radiol.222441

[9] Raji, C.A., Benzinger, T.L.S.: The Value of Neuroimaging in Dementia Diagnosis. Continuum (Minneapolis, Minn.) 28(3), 800–821 (Jun 2022). https://doi.org/10.1212/CON.0000000000001133

[10] Rosenberg, A., Solomon, A., Bonnard, A., Daniilidou, M., Hagman, G., Hall, A., Matton, A., Öhlund-Wistbacka, U., Westman, E., Kivipelto, M.: Preparing for the implementation of anti-amyloid therapies in Europe: Assessing real-world eligibility for lecanemab and donanemab in a Swedish memory clinic. The Journal of Prevention of Alzheimer's Disease 13(3), 100476 (Jan 2026). https://doi.org/10.1016/j.tjpad.2025.100476

[11] Shekari, M., Vállez García, D., Collij, L.E., Altomare, D., Heeman, F., Pemberton, H., Roé Vellvé, N., Bullich, S., Buckley, C., Stephens, A., Farrar, G., Frisoni, G., Klunk, W.E., Barkhof, F., Gispert, J.D.: Stress testing the Centiloid: Precision and variability of PET quantification of amyloid pathology. Alzheimer's & Dementia 20(8), 5102–5113 (Jul 2024). https://doi.org/10.1002/alz.13883

[12] Tejani, A.S., Cook, T.S., Hussain, M., Sippel Schmidt, T., O'Donnell, K.P.: Integrating and Adopting AI in the Radiology Workflow: A Primer for Standards and Integrating the Healthcare Enterprise (IHE) Profiles. Radiology 311(3), e232653 (Jun 2024). https://doi.org/10.1148/radiol.232653

[13] Verger, A., Yakushev, I., Albert, N.L., van Berckel, B., Brendel, M., Cecchin, D., Fernandez, P.A., Fraioli, F., Guedj, E., Morbelli, S., Tolboom, N., Traub-Weidinger, T., Van Weehaeghe, D., Barthel, H.: FDA approval of lecanemab: The real start of widespread amyloid PET use? - the EANM Neuroimaging Committee perspective. European Journal of Nuclear Medicine and Molecular Imaging 50(6), 1553–1555 (May 2023). https://doi.org/10.1007/s00259-023-06177-5

[14] Vernooij, M.W., Pizzini, F.B., Schmidt, R., Smits, M., Yousry, T.A., Bargallo, N., Frisoni, G.B., Haller, S., Barkhof, F.: Dementia imaging in clin-

ical practice: A European-wide survey of 193 centres and conclusions by the ESNR working group. Neuroradiology 61(6), 633–642 (2019). https://doi.org/10.1007/s00234-019-02188-y

[15] World Health Organization (WHO): Global status report on the public health response to dementia. https://www.who.int/publications/i/item/9789240033245 (2021)

[16] World Health Organization (WHO): Dementia. https://www.who.int/news-room/fact-sheets/detail/dementia (2026)

[17] Yakushev, I., Verger, A., Brendel, M., Cecchin, D., Fernandez, P.A., Fraioli, F., Grimmer, T., Tolboom, N., Traub-Weidinger, T., Guedj, E., Van Weehaeghe, D.: Lecanemab approval in EU: What should we be ready for?– the EANM perspective. European Journal of Nuclear Medicine and Molecular Imaging 52(5), 1607–1610 (2025). https://doi.org/10.1007/s00259-025-07066-9

[18] Young, P.N.E., Estarellas, M., Coomans, E., Srikrishna, M., Beaumont, H., Maass, A., Venkataraman, A.V., Lissaman, R., Jiménez, D., Betts, M.J., McGlinchey, E., Berron, D., O'Connor, A., Fox, N.C., Pereira, J.B., Jagust, W., Carter, S.F., Paterson, R.W., Schöll, M.: Imaging biomarkers in neurodegeneration: Current and future practices. Alzheimer's Research & Therapy 12(1), 49 (Apr 2020). https://doi.org/10.1186/s13195-020-00612-7

4. Accepted Proposal #4

4.1. Title : Computational Pathology Breast Cancer Risk Stratification

Table 4. Author and affiliation of accepted proposal 4

Author	**Affiliation**
Sanket Kachole	Indiana University School of Medicine

4.2. Clinical problem : Breast cancer is the most frequently diagnosed malignancy in women, accounting for roughly 32% of new female cancer diagnoses and remaining the second leading cause of cancer death, with over 320,000 new invasive cases projected annually in the United States alone. Although mammographic screening has reduced mortality by approximately 20%, it carries a false-positive rate of up to 60% over ten annual screens, driving more than one million unnecessary biopsies each year and generating substantial patient anxiety, avoidable procedures, and costs exceeding $4 billion annually.

The core bottleneck is risk stratification. Established models (e.g. Gail, Tyrer–Cuzick) rely on demographic, reproductive, hormonal, and imaging features and do not use the tissue-level information already obtainable at biopsy,

despite growing evidence that the breast microenvironment (stromal remodeling, collagen architecture, epithelial change, immune infiltration) shapes cancer initiation years before radiological detectability. No deployed tool integrates tissue-based analysis into breast cancer risk assessment. Success would mean identifying high-risk women earlier while sparing low-risk women unnecessary biopsies measured by improved risk-model discrimination (AUC), higher sensitivity, and fewer false-positive recommendations.

4.3. How AI could address this clinical need : We propose a tissue-based risk-stratification model that predicts future breast cancer development from digitized biopsy specimens. Inputs: H&E- and immunohistochemistry-stained whole-slide images, from which the pipeline extracts quantitative descriptors ("pathomics") of stromal architecture including collagen fiber orientation and organization alongside epithelial morphology, cellular composition, biomarker expression, and deep-learning embeddings from pathology foundation models. Where longitudinal specimens exist, temporal (Δ) descriptors capture tissue change between timepoints. Output: a per-patient risk score stratifying women into enhanced surveillance versus routine-screening pathways, optionally fused with mammography-derived risk. AI is appropriate because these microenvironmental signatures are high-dimensional, subtle, and not reliably assessable by eye. Unlike current epidemiologic and imaging-only models, this interrogates the biological substrate of disease directly, offering earlier and more personalized stratification.

4.4. Data available for model development: Development would draw on a rare longitudinal tissue-bank resource: approximately 900 healthy donors with banked normal breast specimens, H&E glass slides and whole-slide images, associated blood-derived material, mammography for a subset, and long-term self-reported clinical follow-up, of whom 52 subsequently developed invasive breast cancer, enabling a retrospective case–control design. The primary modality is digital pathology, complemented by mammographic imaging and structured follow-up data. Preliminary work by the applicants on a paired-specimen subset has demonstrated that longitudinal collagen-architecture change is computationally measurable [1]. Ground-truth labels would be derived from documented cancer outcomes and immunohistochemistry annotations would be pathologist-guided. Anticipated challenges include marked class imbalance (few cases relative to controls), limited existing immunohistochemistry, staining and scanner heterogeneity across accrual years, and self-reported follow-up. Publicly available cohorts (e.g. TCGA-BRCA, CBIS-DDSM) offer pretraining and external reference data.

4.5. How the proposed model could be externally validated : External validation would target independent breast tissue banks and academic pathology archives with linked outcome data across multiple centers and regions, prioritizing cohorts that differ in scanner, staining protocol, and patient demographics to test generalizability. Primary tissue and outcome data are highly sensitive, validation would rely on privacy-preserving strategies rather than raw data pool-

ing: federated learning and model-to-data evaluation (moving the trained model to each site), secure research environments, and pseudonymization, all under appropriate ethics approval and data-sharing governance. Class imbalance and modest case counts would be addressed through matched case–control sampling, cross-validation with bootstrap confidence intervals, and, where useful, synthetic augmentation. A staged, multi-center validation is feasible: the modeling pipeline is established, the required tissue resources exist at several institutions, and federated frameworks for breast imaging are already in active use, providing a credible template for tissue-based extension.

4.6. How the AI solution would fit into routine clinical practice : The intended end users are pathologists and breast oncologists, with results communicated to referring clinicians and patients. The natural point of use is the moment a breast biopsy returns benign or high-risk-benign pathology: rather than defaulting to epidemiologic risk models alone, the pathologist would receive an AI-generated tissue-based risk score computed from the same digitized slide already produced for diagnosis. In practice the model would run within the digital-pathology workflow, consuming whole-slide images from the image-management system and returning a structured result into the laboratory information system and electronic health record, so that the score is available alongside the diagnostic report and can inform surveillance decisions in the breast clinic.

Barriers to adoption include the need for prospective validation before clinical use, regulatory clearance (e.g. FDA 510(k) as decision support), integration and scanner-standardization overhead, reimbursement, and clinician trust addressed through interpretable, pathologist-aligned outputs. Clinical impact would be evaluated by improved risk-model discrimination and sensitivity, a measurable reduction in false-positive biopsy recommendations, and, ultimately, prospective evidence of earlier detection and better-targeted surveillance without increasing harm.

References

[1] Anonymous: Longitudinal computational analysis of collagen fiber architecture in normal breast tissue as a prognostic biomarker for breast cancer. Accepted abstract, European Congress on Digital Pathology (ECDP) (2026)

[2] Li, H., et al.: Collagen fiber orientation disorder from H&E images is prognostic for early-stage breast cancer: clinical trial validation. npj Breast Cancer 7(1), 104 (2021)

[3] Tyrer, J., Duffy, S.W., Cuzick, J.: A breast cancer prediction model incorporating familial and personal risk factors. Statistics in Medicine 23(7), 1111–1130 (2004)

5. Accepted Proposal #5

5.1. Title : Label-Free Adaptation of IHC Grading Across Hospitals

Table 5. Author and affiliation of accepted proposal 5

Author	Affiliation
Jongwoon Lee	University of Rochester,USA

5.2. Clinical problem : Immunohistochemical (IHC) grading increasingly guides cancer treatment and biomarker development. HER2 is a clear example: differences between adjacent IHC categories can change treatment eligibility. Computational grading could improve consistency, but models often lose accuracy when moved between hospitals because staining, tissue processing, and scanning alter image appearance. Re-annotating a large local dataset at every new hospital is impractical. We therefore need a model that can adapt without target-site annotations.

5.3. How AI could address this clinical need : We propose a cell-level model that detects tumor cells and assigns an ordinal expression grade. At a new hospital, it will adapt using unlabeled local images; target-site annotations will remain withheld until evaluation. The central question is whether adaptation improves target-site performance without harming source-site performance. We will evaluate detection by precision and recall, ordinal grading by quadratic weighted kappa and mean absolute ordinal error, and case-level agreement with pathologist assessment. Metrics will be reported separately by institution rather than only as pooled averages.

5.4. Data available for model development : The current data comprise 512 $\times$ 512-pixel IHC patches acquired at 0.25 μm/pixel, with tumor-cell bounding boxes and ordinal expression grades. Dataset Tumor type Marker(s) Annotation,Hospital A Bladder cancer Nectin-4, HER2, MOC31, TROP2 Cell boxes and grades,Hospital B Bladder cancer Nectin-4 Cell boxes and grades, Public dataset Breast cancer HER2 Slide-level grade .At Hospital A, each marker has 110–140 annotated patches and approximately 1,000–1,750 graded tumor cells, plus a larger unlabeled pool from the same slides. Hospital B enables a same-organ, same-marker Nectin-4 comparison. The public HER2 set differs in tumor type and annotation level and will serve only as a secondary robustness analysis. A board-certified pathologist assigned all cell-level labels. Blinded rereading will measure intraobserver reproducibility; an independently reviewed subset will measure interobserver agreement if additional readers are available.

5.5. How the proposed model could be externally validated : Preliminary label-free adaptation recovered part of the performance lost between Hospitals A and B, but the benefit depended on the direction of transfer. Explaining and reducing this asymmetry is the main technical goal. Final validation will use independent IHC cohorts from institutions with different staining protocols and scanners. Adaptation will use only unlabeled target images; performance will then be compared with the unadapted source model using annotations withheld throughout development and adaptation.

5.6. How the AI solution would fit into routine clinical practice : The model would show the reviewing pathologist a proposed case-level grade and a per-cell overlay; unfamiliar slides would be flagged rather than assigned an unqualified score. At MÉTIS, we seek expertise in unsupervised domain adaptation and independent IHC cohorts for external evaluation. The goal is not full automation, but reliable transfer of an interpretable model without target-site labels.

6. Accepted Proposal #6

6.1. Title : Device-Agnostic Automated Visual Evaluation for Cervical Precancer Triage in HPV-Based Screen-and-Treat programs

Table 6. Authors and affiliations of accepted proposal 6

Author	Affiliation
Ezekiel Ayodeji Oladejo	Slum and Rural Health Initiative Network (SRHIN)
Elizabeth Oluyemisi Ogunseye	University of Ibadan
Dr. Uthman Babatunde	Slum and Rural Health Initiative Network (SRHIN)

6.2. Clinical problem : Cervical cancer is the leading cause of cancer death among women in much of sub-Saharan Africa, which carries close to a quarter of the global burden despite being largely preventable [1,2]. In 2022 the region saw over 100,000 new cases and more than 75,000 deaths [1]; most women present late, when five-year survival falls from above 90% to under 20%. The failure is screening: coverage sits near 4%, roughly one woman in ten is ever screened, and most programs still rely on visual inspection with acetic acid (VIA), which is subjective and poorly reproducible [3,4]. WHO now recommends a screen-triage-treat pathway built on primary HPV testing [4]; the bottleneck is triage. Once a woman is HPV-positive, a provider must decide, often the same day and without a specialist, whether she needs immediate ablative treatment or referral, and that judgment still rests on VIA. Standardizing it would let programs treat correctly at first contact, cut loss to follow-up, and move toward the 70% screening and 90% treatment targets. Success is judged against histology and by same-visit triage accuracy.

6.3. How AI could address this clinical need : We propose automated visual evaluation (AVE): a deep-learning system that reads a cervical image captured on a low-cost smartphone or handheld cerviscope after acetic acid application and returns a triage recommendation [5]. The pipeline localizes the cervix and transformation zone by detection and segmentation, applies an image-quality

gate that rejects unusable frames at the point of care, and runs an ordinal classifier separating normal, equivocal acetowhite change, precancer (CIN2+), and appearances suspicious for cancer. Each read carries a calibrated risk score and an uncertainty flag, so borderline cases are escalated rather than forced into a class. Vision suits the task because the decision is inherently visual and expert graders are scarce, and a trained model applies one standard to every woman, giving reproducible reads and an auditable trail a nurse can act on at once

6.4. Data available for model development : Development would use publicly available cervical image collections, including open cervigram archives and the large longitudinal cohort behind the original AVE algorithm, for pretraining and benchmarking [5,6]. The distinctive asset is a prospectively collected community cohort of HPV-positive women screened with smartphone-acquired images, on the order of 1500 women, with colposcopy and histology confirmation on the treated and referred subset of roughly 180 cases supplying ground-truth labels. The cohort spans several providers, devices, and outreach sites and includes a substantial share of women living with HIV. Anticipated challenges are class imbalance, since precancer is uncommon at population level; variable image quality from field conditions; inter-grader variability in labels; and heterogeneity across capture devices, the very property the model must tolerate rather than exploit.

6.5. How the proposed model could be externally validated : The central risk is portability. Published AVE models overfit to the capture device and population they were trained on and degrade elsewhere, so external validation is not a formality but the core scientific question [6]. We would validate prospectively across independent sites chosen to differ by device, operator, and geography, for example community screening programs and a separate tertiary colposcopy service, with histology as the reference standard. Domain-generalization and device-agnostic training, image quality gating, and calibration would each be tested for their effect on transported performance. Privacy is handled through ethics approval, data-sharing agreements, and pseudonymization, and cross-site learning can use federated or model-to-data evaluation so images never leave their source. Reporting would follow STARD-AI and DECIDE-AI [7]. Multi-site validation is feasible because image capture is cheap, protocols are simple, and HPV-positive triage is already a clinical step.

6.6. How the AI solution would fit into routine clinical practice : Intended users are nurses, midwives, and community health workers who already run outreach screening, not radiologists or gynecologic oncologists. The tool occupies one well-defined point in the workflow: immediately after a positive HPV result, when the provider must choose same-visit ablation or referral. It would run offline-first on an inexpensive Android device, cache results, and synchronize with the mobile screening register and referral system when connectivity allows, so it fits screen-and-treat outreach rather than assuming a hospital PACS or EHR. A borderline or low-quality read routes the woman to specialist review instead of forcing a decision. Barriers to adoption are practical and cultural:

device and lighting variability in the field, provider trust in an automated read, unreliable connectivity, and a national medical-device regulatory pathway still maturing for software of this class. We would meet these with image-quality gating, transparent uncertainty, provider training, and early regulator engagement. Impact would be judged against histology for triage accuracy, and in service terms by same-visit treatment rate, loss to follow-up, over-treatment and under treatment rates, and time from screening to treatment.

References

[1] Geographic and demographic patterns of cervical cancer in Africa using GLOBOCAN 2022. BMC Cancer, 2026.

[2] Advancing the elimination of cervical cancer across Africa: burden of HPV, screening and AI-based methods. Nature Communications, 2026.

[3] Regional and country-level trends in cervical cancer screening coverage in sub-Saharan Africa (2000 to 2020). PLOS Medicine, 2023.

[4] World Health Organization. Global Strategy to Accelerate the Elimination of Cervical Cancer (2020); WHO guideline for screening and treatment of cervical pre-cancer lesions (2021).

[5] Hu L, et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening. J Natl Cancer Inst, 2019;111(9):923-932.

[6] Desai KT, et al. The development of automated visual evaluation (AVE) for cervical cancer screening: promise and challenges in adapting deep learning for clinical testing. Int J Cancer, 2022.

[7] Reporting standards for diagnostic and early-stage clinical AI: STARD-AI and DECIDE-AI.

7. Accepted Proposal #7

7.1. Title : Predicting Post-operative Breast Morphology from 3D Surface Imaging to standardize the Mastectomy versus Breast-Conservation Decision (MATRIX)

Table 7. Author and affiliation of accepted proposal 7

Author	Affiliation
António Soares	Universidade de Lisboa, Portugal Instituto Português Oncologia, Lisboa, Portugal

7.2. Clinical problem : Breast cancer is the most commonly diagnosed cancer in women worldwide, with approximately 2.3 million new cases annually, and

most patients undergo surgery with curative intent. In breast surgical oncology, the choice between mastectomy and breast-conserving surgery with oncoplastic reshaping [1] is made in the outpatient clinic, usually by a single surgeon evaluating tumor size and location against breast morphology. The surgeon must mentally simulate how excising a given volume will alter breast shape and symmetry, then choose the technique likely to give the best oncological and aesthetic result. Volume-based criteria have been proposed by surgical societies [2,3], but their application is limited by inconsistent measurement methods and by wide variation in individual surgeon proficiency. Access to breast conservation therefore depends substantially on where a patient is treated: high-volume specialist centers offer oncoplastic techniques that lower-volume units do not. Objective, reproducible volumetric assessment would standardize this decision, shorten the oncoplastic learning curve, and allow implant size to be anticipated when mastectomy is chosen. Success would be measured by breast-conservation and re-excision rates and by patient-reported aesthetic satisfaction in non-specialist centers.

7.3. How AI could address this clinical need : We propose a generative model that predicts post-operative breast morphology from pre-operative imaging and a candidate surgical plan. Inputs are pre-operative 3D surface scans acquired standing and supine [4], MRI-derived breast and tumor volumes with tumor location coordinates, and standardized clinical measurements (sternal notch-to-nipple distance, width, projection, Regnault ptosis grade). A hybrid point-cloud/graph architecture based on PointNet++ encodes shape, tumor location, and planned technique in separate branches, and an encoder-decoder head generates the predicted post-operative surface under a loss weighted for volume and symmetry preservation. Outputs are a predicted 3D surface plus derived volume-loss and symmetry metrics for each candidate technique. AI methods are appropriate because the mapping from breast shape, tumor position, and resection volume to final morphology is high-dimensional and currently encoded only as tacit expert intuition. Relative to current practice, this converts an unrecorded mental simulation into an inspectable, comparable prediction.

7.4. Data available for model development : Development will use a prospective cohort of approximately 300 patients undergoing tumor excision with or without oncoplastic techniques at a tertiary cancer center performing around 800 breast cancer operations annually. Each patient contributes data at three time points: pre-operative (structured-light surface scans at a minimum density of 1 point/mm2, contrast-enhanced 3T MRI, clinical measurements, standardized non-identifiable photography); intra operative (oncoplastic technique and parenchymal rearrangement pattern, excised specimen weight, and specimen surface scanning at 2 points/mm2); and post-operative (repeat scans at 4-6 weeks and 6 months, patient-reported outcomes, complications). Ground truth for the generative task is the registered 6-month scan; breast and tumor volumes come from semi-automated MRI segmentation using MedSAM [5], validated against water-displacement volumetry in a subset. No public dataset pairs pre- and post-operative 3D breast surfaces with surgical technique. Antic-

ipated challenges are limited sample size, class imbalance across techniques, and registration error arising from breast deformability and interval weight change.

7.5. How the proposed model could be externally validated : The development cohort is single-center, and external validation is the principal open question this proposal brings to METIS. We envisage prospective validation in two or three European academic breast units of differing case volume, deliberately including at least one lower-volume unit, since generalization to non-specialist settings is the clinical claim being tested. Site-level ethics approval and data-sharing agreements would be required. Mitigations include restricting acquisition to the thoracic region so that no facial or otherwise identifying structures are captured, releasing derived meshes rather than raw scans, DICOM pseudonymization for MRI, and federated or model-to-data evaluation where transfer is not permitted. Acquisition uses a portable calibrated structured-light camera.

7.6. How the AI solution would fit into routine clinical practice : Intended end users are breast and oncoplastic surgeons, with secondary use by trainees and patients during consul tation. The tool would sit at the point of surgical planning, after diagnostic MRI and before the multidisciplinary team decision: the surgeon performs a brief surface scan in clinic and the platform returns predicted post-operative morphology for each candidate technique side by side, with predicted volume loss, symmetry index, and position

relative to accepted excision-volume thresholds for conservation. Delivery would be through a web-based planning interface drawing MRI from PACS and writing a summary to the EHR. Predicted integration barriers are no native handling of surface data in PACS, clinic time, scanner cost in lower-resource units, and surgeon trust in a generative prediction. Evaluation should proceed in sequential stages: geometric accuracy (mean absolute volumetric error, Hausdorff distance against 6-month scans, and comparison against the surgeon's own volume estimate), then decision impact (change in mastectomy versus conservation recommendations, and concordance between non-specialist and specialist surgeons), then patient outcomes (aesthetic satisfaction, re-excision rate) and quality of shared decision-making [6].

References

[1] Losken, A., Dugal, C.S., Styblo, T.M., Carlson, G.W.: A meta-analysis comparing breast conservation therapy alone to the oncoplastic technique. Annals of Plastic Surgery 72(2), 145–149 (2014)

[2] Chatterjee, A., Gass, J., Patel, K., et al.: A consensus definition and classification system of oncoplastic surgery developed by the American Society of Breast Surgeons. Annals of Surgical Oncology 26, 3436–3444 (2019) [3] Loibl, S., Andre, F., Bachelot, T., et al.: Early breast cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Annals of Oncology 35(2), 159–182 (2024)

[4] Stern, C.S., Plotsker, E.L., Rubenstein, R., et al.: Three-Dimensional Surface Analysis for Preoperative Prediction of Breast Volume: A Validation Study. Plastic and Reconstructive Surgery 152(6), 1153–1162 (2023)

[5] Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B., et al.: Segment anything in medical images. Nature Communications 15, 654 (2024)

[6] Josfeld, L., Keinki, C., Pammer, C., et al.: Cancer patients' perspective on shared decision-making and decision aids in oncology. Journal of Cancer Research and Clinical Oncology 147, 1725–1732 (2021)