24

Key brain region identification in obesity prediction with structural MRI and probabilistic uncertainty aware model

Walia Farzana[1], Megan A. Witherow[1], Ahmed Temtam[1], Liangsuo Ma[2], Melanie Bean[2], F. Gerry Moeller[2] and K.M. Iftekharuddin[1]
[1]Old Dominion University and [2]Virginia Commonwealth University

OBJECTIVES/GOALS: Predictive performance alone may not determine a model's clinical utility. Neurobiological changes in obesity alter brain structures, but traditional voxel-based morphometry is limited to group-level analysis. We propose a probabilistic model with uncertainty heatmaps to improve interpretability and personalized prediction. METHODS/STUDY POPULATION: The data for this study are sourced from the Human Connectome Project (HCP), with approval from the Washington University in St. Louis Institutional Review Board. We preprocessed raw T1-weighted structural MRI scans from 525 patients using an automated pipeline. The dataset is divided into training (357 cases), calibration (63 cases), and testing (105 cases). Our probabilistic model is a convolutional neural network (CNN) with dropout regularization. It generates a prediction set containing high-probability correct predictions using conformal prediction techniques, which add an uncertainty layer to the CNN. Additionally, gradient-based localization mapping is employed to identify brain regions associated with low uncertainty cases. RESULTS/ANTICIPATED RESULTS: The performance of the computational conformal model is evaluated using training and testing data with varying dropout rates from 0.1 to 0.5. The best results are achieved with a dropout rate of 0.5, yielding a fivefold cross-validated average precision of 72.19% and an F1-score of 70.66%. Additionally, the model provides probabilistic uncertainty quantification along with gradient-based localization maps that identify key brain regions, including the temporal lobe, putamen, caudate, and occipital lobe, relevant to obesity prediction. Comparisons with standard segmented brain atlases and existing literature highlight that our model's uncertainty quantification mapping offers complementary evidence linking obesity to structural brain regions. DISCUSSION/SIGNIFICANCE OF IMPACT: This research offers two significant advancements. First, it introduces a probabilistic model for predicting obesity from structural magnetic resonance imaging data, focusing on uncertainty quantification for reliable results. Second, it improves interpretability using localization maps to identify key brain regions linked to obesity.

25

The association of a documented prescription of medication for opioid use disorder (MOUD) during pregnancy with maternal outcomes

Ashley Hailer[1], Pauline Terebuh[2], Rong Xu[3], David C. Kaelber[4] and Pamela B. Davis[5]
[1]Case Western Reserve University School of Medicine, Cleveland, OH, USA; [2]Center for Artificial Intelligence in Drug Discovery, Case Western Reserve University School of Medicine, Cleveland, OH, USA; [3]Center for Artificial Intelligence in Drug Discovery, Case Western Reserve University School of Medicine, Cleveland, OH, USA; [4]Center for Clinical Informatics Research and Education, The MetroHealth System, Cleveland, OH, USA and [5]Center for Community Health Integration, Case Western Reserve University School of Medicine, Cleveland, OH, USA

OBJECTIVES/GOALS: Opioid use disorder (OUD) at delivery increased between 1999 and 2014. Clinical guidelines include medication for OUD (MOUD) for pregnant women with OUD and is associated with better fetal outcomes. Few large studies have compared prenatal MOUD outcomes to no MOUD. We evaluated the association of documented MOUD prescription during pregnancy with maternal outcomes. METHODS/STUDY POPULATION: We utilized aggregated electronic health records using the TriNetX platform to conduct a retrospective cohort study of females, aged 1249 years with a childbirth CPT code and documented opioid use via ICD-10 codes in the nine months before delivery between 2014 and 2020, comparing patients with MOUD prescription of buprenorphine or methadone during the nine months before delivery to demographically matched patients without any documented MOUD, using hazard ratios and 95% CIs for outcomes occurring one week to one or three years after childbirth. RESULTS/ANTICIPATED RESULTS: MOUD cohort (n = 6,945, 85.33% White; 82.77% Non-Hispanic or Latino) was associated with significantly higher subsequent documented MOUD prescription (HR, 9.26 [95% CI, 7.98–10.76]; 6.21 [95% CI, 5.60–6.88]) and new remission codes (HR, 2.79 [95% CI, 2.15–3.62]; 2.85 [95% CI, 2.38–3.40]) at one and three years, lower ED visits at one year (HR, 0.88 [95% CI, 0.81–0.96]), no significant association of ED visits at three years (0.95 [95% CI, 0.89–1.02]), higher outpatient visits (HR, 1.26 [95% CI, 1.20–1.32]; HR, 1.27 [95% CI, 1.21–1.33], and no significant association of inpatient visits at one and three years (HR, 0.93 [95% CI, 0.813–1.06]; 1.06 [95% CI, 0.96–1.18]) than the never-MOUD cohort (n = 4,708, 76.11% White; 75.68% non-Hispanic or Latino). DISCUSSION/SIGNIFICANCE OF IMPACT: A documented prescription for MOUD during pregnancy is associated with newly documented remission of OUD, increased outpatient visits, decreased ED visits, and additional documented MOUD prescriptions suggestive of increased access to continuity care. Efforts to increase MOUD use in pregnancy may improve maternal outcomes.

26

Evaluating AI models trained with varying amounts of expert feedback for chronic graft-versus-host disease skin assessment in photos of patients with diverse skin tones*†

Andrew McNeil[1], Kelsey Parks[1], Michael Pogharian[1], Edward W Cowen[2], Julia Lehman[3], Stephanie J Lee[4], Steven Z Pavletic[5], Benoit M Dawant[1] and Eric R Tkaczyk[1]
[1]Vanderbilt University Medical Center; [2]National Institute of Arthritis and Musculoskeletal; [3]Skin Diseases Mayo Clinic; [4]Fred Hutchinson Cancer Center and [5]Center for Cancer Research, National Cancer Institute

OBJECTIVES/GOALS: Manual skin assessment in chronic graft-versus-host disease (cGVHD) can be time consuming and inconsistent (>20% affected area) even for experts. Building on previous work we explore methods to use unmarked photos to train artificial intelligence (AI) models, aiming to improve performance by expanding and diversifying the training data without additional burden on

https://doi.org/10.1017/cts.2024.717

experts. METHODS/STUDY POPULATION: Common to many medical imaging projects, we have a small number of expert-marked patient photos (N = 36, n = 360), and many unmarked photos (N = 337, n = 25,842). Dark skin (Fitzpatrick type 4+) is underrepresented in both sets; 11% of patients in the marked set and 9% in the unmarked set. In addition, a set of 20 expert-marked photos from 20 patients were withheld from training to assess model performance, with 20% dark skin type. Our gold standard markings were manual contours around affected skin by a trained expert. Three AI training methods were tested. Our established baseline uses only the small number of marked photos (supervised method). The semi-supervised method uses a mix of marked and unmarked photos with human feedback. The self-supervised method uses only unmarked photos without any human feedback. RESULTS/ANTICIPATED RESULTS: We evaluated performance by comparing predicted skin areas with expert markings. The error was given by the absolute difference between the percentage areas marked by the AI model and expert, where lower is better. Across all test patients, the median error was 19% (interquartile range 6 – 34) for the supervised method and 10% (5 – 23) for the semi-supervised method, which incorporated unmarked photos from 83 patients. On dark skin types, the median error was 36% (18 – 62) for supervised and 28% (14 – 52) for semi-supervised, compared to a median error on light skin of 18% (5 – 26) for supervised and 7% (4 – 17) for semi-supervised. Self-supervised, using all 337 unmarked patients, is expected to further improve performance and consistency due to increased data diversity. Full results will be presented at the meeting. DISCUSSION/SIGNIFICANCE OF IMPACT: By automating skin assessment for cGVHD, AI could improve accuracy and consistency compared to manual methods. If translated to clinical use, this would ease clinical burden and scale to large patient cohorts. Future work will focus on ensuring equitable performance across all skin types, providing fair and accurate assessments for every patient.

27

Automated IRB compliance and secure data delivery in i2b2

Hakob Abajian, Venkat Vuppula Johnson Kwong, Praveen Angyan, Shakeel Usmani, Saurav Limbu Brittney Arvizu Hira Sherazi, Swaroop Samek, Maryam Abdallah, Amy Chuang, Daniella Garofalo and Neil Bahroos
University of Southern California

OBJECTIVES/GOALS: To address the manual, time-consuming processes of validating IRB compliance and ensuring the secure delivery of i2b2 data, this project automates compliance checks, streamlines Protected Health Information (PHI) access, and provides timely, secure data availability while reducing administrative burdens and non-compliance risks. METHODS/STUDY POPULATION: This project enhances the i2b2 application to automate compliance processes and facilitate secure data delivery through integration with REDCap. By linking i2b2 with the IRB system, the application performs automatic compliance checks for project requests, verifying GCP and HIPAA certifications, only allowing the release of IRB-approved PHI variables, safeguarding against unauthorized data access. Manual signatures confirm non-automated compliance processes. Once verified, the application automatically creates a REDCap project, assigns user access, and securely delivers data, ensuring compliance with HIPAA regulations. RESULTS/ANTICIPATED RESULTS: The automated system successfully streamlined IRB compliance checks and data delivery for i2b2 requests. Validation of certifications like GCP and HIPAA, now occurs automatically, significantly reducing the risk of non-compliance. Personnel access to data is limited to IRB-approved PHI, ensuring data security and adherence to institutional standards. The integration with REDCap has reduced manual processes, cutting data request processing time to approximately 30 minutes. Researchers and administrative staff experienced a notable decrease in administrative burden, with faster, more efficient access to approved data while maintaining full compliance with IRB and HIPAA regulations. DISCUSSION/SIGNIFICANCE OF IMPACT: The lessons learned can be adapted by institutions to improve compliance efficiency and reduce administrative overhead. Implementing similar automation of certification checks and data delivery, sites can enhance data security, minimize errors, and ensure faster, compliant access to research data.

28

Using AI to predict molecular subtype from histopathology slides in endometrial cancer†

Vincent Wagner[1], Jesus Gonzalez, Bosquet[2], Michael Goodheart[2], Xiaodong Wu[3] and Megan Samuelson[4]
[1]University of Iowa; [2]Division of Gynecologic Oncology, University of Iowa; [3]Department of Electrical and Computer Engineering, University of Iowa and [4]Department of Surgical Pathology, University of Iowa

OBJECTIVES/GOALS: Endometrial cancer is one of the few cancers that has both a rising incidence and mortality rate. Molecular classification is becoming more important for the management of endometrial cancer but the ability to translate this into clinical practice remains constrained. Our goal is to use AI to predict the molecular subtype from histopathology slides. METHODS/STUDY POPULATION: We utilized the open source endometrial cancer datasets from The Cancer Genome Atlas (TCGA) (N = 387) and Cancer Proteomics Transcriptomic Tumor Analysis Consortium (CPTAC) (N = 135) to develop and train a vision transformer AI model. We used a proprietary cohort of patients (N = 548) for external validation. Whole slide images (WSI) and molecular subtype data were collected. Subtypes include POLE ultramutated (POLE), microsatellite instability (MSI-H), copy-number low (CNV-L), and copy-number high (CNV-H). WSI were preprocessed, and features were extracted. Modified STAMP protocol was used in training, utilizing a pretrained foundation transformer model (Virchow2). Cross-validation of the TCGA was used for initial training, followed by testing on the CPTAC dataset and validation on our proprietary cohort. RESULTS/ANTICIPATED RESULTS: Fivefold cross-validation of the TCGA database (60% training, 20% testing, and 20% validation) developed a best overall model with a mean AUC of 0.74 (POLE 0.78, MSI-H 0.76, CNV-H 0.86, CNV-L 0.77). Overall precision 0.58, recall 0.55. CNV-H was the subtype with the most accurate prediction. CPTAC holdout testing revealed moderately high AUC (POLE 0.63, MSI-H 0.62, CNV-H 0.98, and CNV-L 0.76). Overall precision 0.54 and recall 0.58. Again, CNV-H was the most accurate

https://doi.org/10.1017/cts.2024.717