

A clinical brain MRI dataset of polymicrogyria in patients suspected of epilepsy

Llucia Coll ¹ , Abdullah Shafique ², Helene Kaas ^{1,3} , Martin Prener ^{1,3}, Lars Hageman Pinborg ^{1,3,4} , Gitte Moos Knudsen ^{1,4} , Melanie Ganz ^{1,5} 

1 Neurobiology Research Unit, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark

2 Department of Radiology, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark

3 Epilepsy Clinic, Department of Neurology, Copenhagen University Hospital – Rigshospitalet, Copenhagen, Denmark

4 Institute of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

5 Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

Abstract

Brain magnetic resonance imaging (MRI) is widely used as a diagnostic and monitoring tool for most neurological conditions. Although clinical practice produces large volumes of data, they are typically underutilised for training machine learning models that could substantially improve the feasibility, accuracy, and implementation of diagnostic tools in clinical settings. Instead, such models are usually trained on small, highly curated datasets and evaluated under controlled conditions, which limits their generalisability to the variability of real-world clinical imaging. This limitation is also reflected in publicly available datasets. And, it is particularly evident in epilepsy-related abnormalities, for which such datasets are scarce.

The PMG-BrainDrugs dataset is the first open polymicrogyria (PMG) dataset including patients across the full age range. PMG is a malformation of cortical development characterised by abnormal cortical layering. The associated abnormalities are heterogeneous and can sometimes be difficult to detect.

The dataset is comprised of 208 subjects, including 119 patients with radiologically confirmed PMG and 89 controls with negative findings. Each scanning session may include several anatomical imaging modalities (T1-, T2-, and T2*-weighted, or FLAIR sequences). For 27 subjects, 2 or 3 time points are available. The release of this dataset includes minimally processed data (with only defacing applied) as well as masks of PMG abnormalities. The dataset is available at DOI: 10.70883/GBSQ9852.

This dataset aims to bridge the gap between curated research datasets and the variability encountered in routine clinical practice. In doing so, it offers a valuable resource for developing and evaluating more robust and generalisable methods for the detection of PMG and related cortical malformations.

Keywords

polymicrogyria, MRI, BIDS, clinical acquisition, epilepsy

Article informations

<https://doi.org/10.59275/j.melba.2026-59f7>

Volume 2026, Received: 2026-06-19, Published 2026-09-21

Corresponding author: melanie.ganz@nru.dk

Special issue: MICCAI Open Data 2026 x MELBA

Guest editors: AT, MS et al.

©2026 Coll et al. License: CC-BY 4.0



1. Background

Data sharing of medical imaging data is an important aspect of current open science practices (Gorgolewski et al., 2016). In the field of neuroimaging in particular, the research community has strongly embraced data-sharing initiatives.

However, there remains a lack of datasets that provide clinical-grade data. Most publicly available neuroimaging datasets consist of data collected for specific research purposes, where image quality and acquisition protocols often differ from those used in clinical practice (Marcus et al., 2010; Petersen et al., 2010; Miller et al., 2016). As a result,

many current machine learning algorithms for detecting medical neuropathologies are trained either on research-grade data or on relatively small in-house clinical datasets. This discrepancy, often referred to as domain shift, represents a major challenge for deploying machine learning tools in real-world clinical settings. To begin addressing this limitation, we present a brain magnetic resonance imaging (MRI) dataset acquired during routine clinical practice, specifically in cases where patients were suspected of epilepsy. In particular, the dataset includes a cohort of patients diagnosed with polymicrogyria (PMG).

PMG is a common malformation of cerebral cortex development characterised by excessive folding and abnormal cortical layering (Severino et al., 2020). Epilepsy is considered the most common clinical consequence of PMG, which occurs in 65-87% of cases (Leventer et al., 2010). PMG is a frequent finding in patients with drug-resistant epilepsy and is often associated with neurodevelopmental delay, motor deficits, and cognitive impairment. Despite its clinical relevance, PMG remains challenging to study systematically due to its heterogeneous presentation, variable anatomical distribution, and limited availability of neuroimaging datasets.

MRI plays a central role in the diagnosis and characterisation of PMG, enabling non-invasive assessment of cortical morphology and lesion extent (Kolbjør et al., 2023). However, radiological presentation of PMG is often subtle and requires expert interpretation, especially in cases with mild or atypical features. Automated or semi-automated computational approaches for PMG detection and quantification are emerging, but their development is hindered by the lack of publicly available, annotated clinical datasets (Zhang et al., 2024; Rauch et al., 2025).

To date, only one PMG dataset has been made publicly available (Zhang et al., 2022). It consists of 23 pediatric PMG cases, which are shared as 2D image slices in JPEG format, instead of 3D volumes. The different slices are labelled according to the presence of abnormality, normal appearing or uncertain. However, no delineation of the abnormalities themselves is available.

To address this gap, we present and release PMG-BrainDrugs, a curated dataset of clinical brain MRI scans from patients, both pediatric and adult, screened for epilepsy within the Capital Region and Region Zealand in Denmark between 1998 and 2023. PMG-BrainDrugs includes brain MRI scans from 208 patients suspected of epilepsy, with radiology report-based annotations identifying 119 subjects with PMG and 89 no-PMG controls. For each PMG subject, at least one abnormality segmentation mask is provided. The data is structured according to the Brain Imaging Data Structure (BIDS) format (Gorgolewski et al., 2016), an open standard designed to facilitate MRI data sharing. The image data is published on the platform PublicnEUro

(<https://publicneuro.eu/>), a GDPR-compliant data-sharing site where access is granted only under a data usage agreement (DUA).

With the publication of this dataset, we aim to shed light on the characteristics of real-world, clinically acquired brain MRI data, highlighting both its strengths and limitations. The inclusion of diverse MRI modalities per case, variable numbers of scanning runs, and heterogeneous acquisition protocols reflects the variability inherent in clinical imaging. This makes PMG-BrainDrugs valuable not only for pathology-specific research but also for studies focused on MRI harmonisation, robustness testing, and domain adaptation across clinical settings. Importantly, this resource offers a unique opportunity to advance computational neuroimaging research in epilepsy and cortical malformations, particularly PMG in all age ranges. PMG-BrainDrugs can be used to train and validate machine learning models for PMG detection, support studies on cortical morphology and lesion distribution, and enable reproducible benchmarking of surface-based and volumetric analysis pipelines. By transforming heterogeneous clinical MRI scans into a structured, annotated resource, PMG-BrainDrugs makes clinical neuroimaging accessible for open science, reproducible research, and translational applications in epilepsy.

2. Summary

PMG-BrainDrugs contains brain MRI scans from 208 patients suspected of epilepsy, including 149 scanning sessions confirmed to present PMG and 92 that serve as no-PMG control cases. For each patient, at least one time point with a full clinical protocol consisting of a single or multiple sequences is provided. The available sequences per study vary in number and type, always containing at least one of the following anatomical sequences: T1-weighted (w) MPRAGE, T2-w, fluid attenuated inversion recovery (FLAIR), and T2*-w. All scans have been curated and de-identified. Figure 1 shows examples of different subjects and modalities included in the dataset. Binary PMG abnormality masks aligned with the T1-w, whenever available, if not, with T2-w or FLAIR sequences, are included for every PMG subject. Note, there are two different types of masks provided: 1) manual masks that are lesion masks with region-level accuracy and 2) pseudo-masks derived from FreeSurfer cortical parcellations, corresponding to the cortical regions containing the lesion. Both masks were based on the existing radiological reports.

Taken together, this dataset is designed for supervised machine learning tasks, including binary classification of PMG abnormalities in brain MRI. With the addition of lesion masks, it could also be used for the segmentation of cortical abnormalities. The clinical nature of this dataset, together with the availability of multiple runs and sequences,

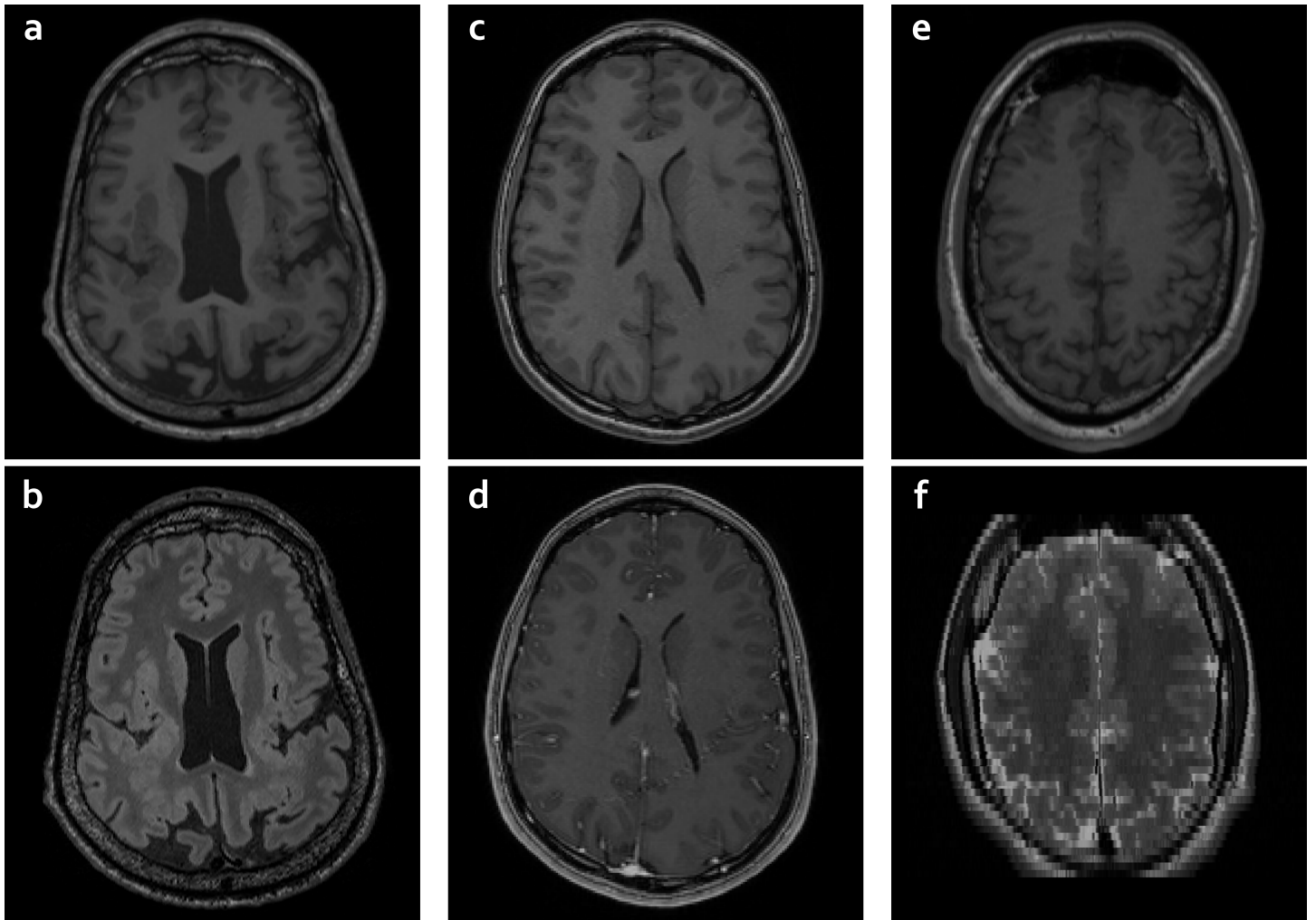


Figure 1: Example of a central axial slice taken from polymicrogyria cases from the PMG-BrainDrugs dataset. (a) sub-EP026898_ses-001_T1w, (b) sub-EP026898_ses-001_FLAIR, (c) sub-EP013035_ses-001_T1w, (d) sub-EP013035_ses-001_ce-gd-T1w (with gadolinium contrast), (e) sub-EP037840_ses-001_T1w, and (f) sub-EP037840_ses-001_T2w. Notably, the image quality differs as well as the level of abnormality.

provides real-world heterogeneity for algorithm development and benchmarking, with potential applications in clinical settings. Importantly, PMG-BrainDrugs is BIDS-compliant, which aligns with the FAIR (Findable, Accessible, Interoperable, Reusable) principles. The range of potential machine learning applications supported by this dataset, combined with the provision of minimally processed data, allows future users to apply preprocessing steps best suited to their specific use cases.

The PMG-BrainDrugs dataset is structured in BIDS format and composed of open imaging (NIfTI) and metadata (JSON) format files, requires no proprietary software, and can be viewed and analysed with any open neuroimaging software.

3. Discussion

The PMG-BrainDrugs dataset fills a critical gap in the availability of brain MRI data for epilepsy research, specifically

cortical malformations such as PMG. More importantly, it represents a step towards publicly available, real-world datasets acquired in clinical practice that reflect the diversity, quality, and acquisition heterogeneity needed to develop machine learning algorithms for feasible future clinical implementation.

Even though the dataset comprises data acquired using scanners from different vendors, it is worth noting that it is limited to a specific population and to centers located within the same regions in Denmark. This limitation raises concerns regarding geographic and demographic bias. Additionally, machine learning models trained on this dataset may generalise less well to data acquired using under-represented scanner configurations, as 3T Siemens scanners constitute the predominant scanners in the cohort.

Data access is regulated and can only be used following the conditions in the DUA that needs to be signed before accessing the data, notably no re-distribution or

re-identification, among other clauses is allowed (see DUA document for more explicit usage and terms).

Based on the same parent cohort, BrainDrugs work package 5 (WP5), we plan to release and make publicly accessible additional datasets focusing on other specific epilepsy-related focal malformations. To our knowledge, this is the first large publicly available PMG dataset. We believe it will benefit researchers in the field, as well as those interested in real-world clinical MRI scans.

4. Data Availability

4.1 Data Location

The dataset has been shared on PublicnEUro and is available at DOI: 10.70883/GBSQ9852 (Coll et al., 2026). To access the data, one must register on publicneuro.eu, where identities are verified (depending on user availability, expect a 1-day to 1-week delay). Once the identity is verified, users can request data. For this specific dataset, a DUA must be signed and re-uploaded, and approval from the data controller's legal team is required (1-2 weeks). Note that for users outside the EU or adequacy countries, institutional signing of the EU standard contractual clauses is also mandatory by law (Comission, b). After the documents are signed and approved, users receive a link to the data download via browser or command line.

4.2 Potential Use Cases

The PMG-BrainDrugs dataset offers several potential applications in machine learning, particularly in epilepsy research:

- *PMG vs no-PMG classification algorithm.* The dataset may support the development and benchmarking of supervised algorithms for the detection and classification of PMG in brain MRI against patients who are suspected of epilepsy, rather than healthy controls.
- *PMG-related cortical abnormalities segmentation.* With the inclusion of lesion annotations (masks), the dataset could facilitate the development of automated segmentation methods for cortical malformations associated with PMG.

In addition, the dataset enables broader applications in brain MRI research such as:

- *Modality synthesis.* The availability of multiple MRI acquisition modalities for some patients, together with missing modalities in others, enables the exploration of modality synthesis and image-to-image translation approaches.
- *Domain adaptation.* The heterogeneity in acquisition protocols, scanner vendors, field strengths, and image

quality may support research in domain adaptation and generalisation methods for robust clinical AI applications.

4.3 Licensing

PMG-BrainDrugs is shared under a DUA. The data presented are considered personal data and therefore fall under the EU General Data Protection Regulation (GDPR) (Comission, a). The DUA is also listed on the dataset webpage for convenience (see DUA terms). Data access is free as the data owners pay PublicnEUro for hosting of their data. Commercial usage is allowed, but needs to be approved beforehand.

4.4 Ethical Considerations

The PMG-BrainDrugs dataset was collected across different centers included in the Capital Region and Region Zealand in Denmark. The dataset spans a wide age range, including children from 1 years of age. The dataset is shared on PublicnEUro via *DTU computerome*, a Danish data center with full GDPR and NIS2 compliance. The dataset stored in PublicnEUro is version-controlled, and any updates are recorded via new versions and DOIs.

Ethics approval for data collection has been granted by the relevant ethics committees (CVK-2112460), and approval for data processing has been given by the Danish Data Protection Authorities. This data can be shared under public service as a legal basis, as recognised by the Danish Data Authority, following the GDPR of the European Union. While the risk of re-identification is low, the data presented can only be pseudonymised according to the Danish Data Authorities and is therefore considered personal data. As personal data, they fall under the EU GDPR, and are therefore shared under a DUA. Sharing under DUA for neuroscientific research is in line with the ethics approval for the project under which this data was acquired.

To ensure pseudonymisation, facial features were removed from the MRI scans using *pydeface* (Omer Faruk Gulban et al., 2026). Personal patient information contained in the raw DICOM files was removed. The accompanying JSON files contain MRI acquisition metadata, following the BIDS format standards.

A statement of intended use is described in the DUA, shared together with the dataset in PublicnEUro (see DUA terms). Fundamentally, the data must only be used for purposes of medical and cognitive neuroscience research, consistent with the purpose for which the data were collected.

5. Methods

5.1 Data collection

Data collection was performed within the BrainDrugs initiative¹. The BrainDrugs WP5 cohort is a retrospective study that includes all persons with epilepsy and suspected epilepsy in the *Sundhedsplatform*, the Danish health system platform, identified by ICD-10 codes DG40* and/or DZ033A in the Capital Region and Region Zealand in Denmark, from 1998 up to December 2023. Due to a change of electronic health system in 2016, most data is from the period 2016 to 2023.

Within the BrainDrugs WP5 cohort, the PMG-BrainDrugs inclusion criteria were the identification of PMG abnormalities mentioned in the radiology reports associated with a brain MRI. Pre-selection of PMG cases was performed by searching on pseudo-anonymised radiology report free-texts for patterns matching "polymicrogyria", "polymikrogyri", or "pmg". The initial selection based on regular expressions was manually reviewed by a medical doctor (MP) and a medical student (HK) supervised by a board-certified neurologist specialised in epilepsy (LP) to confirm whether each case represented a true PMG or not. Discrepant or complex cases were discussed until consensus was reached. Cases labeled as No-PMG do not imply the absence of epilepsy, but rather that the detected abnormality was not PMG. Additional labels such as "resection" or "other pathology present" were also recorded. An overview of the data collection flowchart is shown in Figure 2.

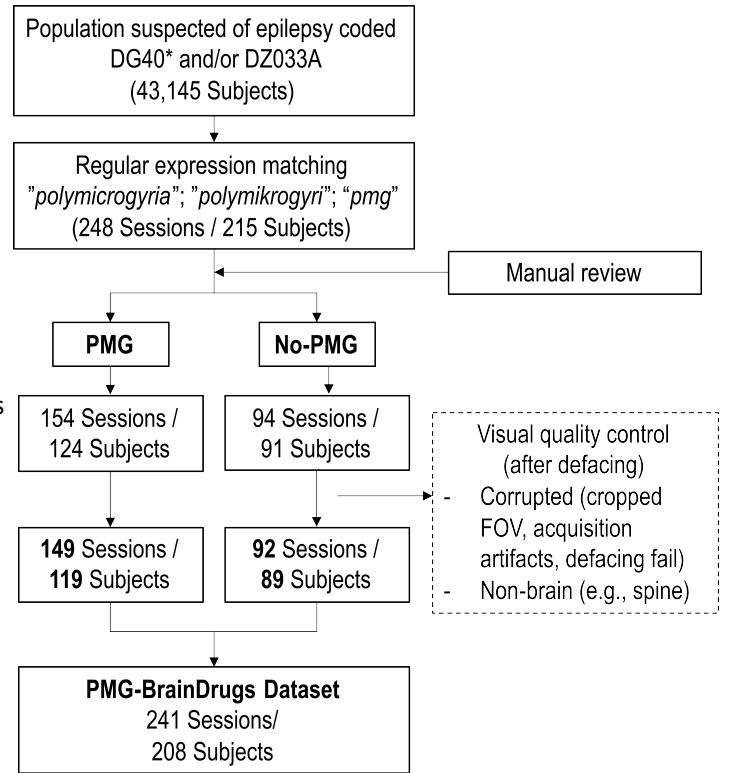


Figure 2: Flowchart illustrating the inclusion and exclusion processes of acquisition sessions and subjects included in the PMG-BrainDrugs dataset per group: polymicrogyria (PMG) and non-polymicrogyria (No-PMG).

5.2 Population

The final PMG-BrainDrugs dataset is comprised of 208 patients (106 female, 102 male), of whom 27 had multiple scanning sessions, resulting in a total of 241 sessions (see Figure 3 for the distribution of subjects with multiple time-points within the dataset). These include 149 labeled as PMG (58 female, 61 male) and 92 as No-PMG (48 female, 41 male). For PMG cases, the mean age ($\pm$ standard deviation) at the time of MRI was 23.8 ± 17.2 years, and 24.7 ± 19.8 years for No-PMG cases (see Figure 4 for the age-sex distribution). For individual details on each patient and scanning session, we refer the reader to the *participants.tsv* file shared with the dataset.

5.3 Data acquisition

MRI scans were acquired using various scanners available within the Capital Region and Region Zealand in Denmark from different manufacturers and scanner models, including Siemens (81%), Philips (15%), and GE (4%). Most scans were performed at 3 T (81%), with the remainder at 1.5 T (19%). During the study period, diverse MRI protocols were used. Most scanning sessions included T1-w (96%),

Table 1: Overview of the most common MRI acquisition protocols, scanned with a 3T Siemens Verio (Siemens Healthcare GmbH, Erlangen, Germany) at Hvidovre Hospital, Copenhagen.

	T1 MPRAGE	T2w	FLAIR	T2*
TR [ms]	1900	6180	5000	28
TE [ms]	2.23	79	395	20
TI [ms]	900	-	1800	-
Flip angle [°]	9	120	120	15
Voxel spacing [mm ³]	1x1x1	3x3x3.3	1x1x5	1.2x1.2x6

T2-w (87%), and FLAIR (92%) sequences, and a smaller percentage also had T2*-w (22%). The number of runs per sequence varied from 1 to 3. In 56% of sessions, gadolinium contrast was administered, typically corresponding to the second run of the T1-w acquisition. The most common acquisition parameters for each sequence are summarised in Table 1. An overview of the availability of sequences for each subject is included in the top-level *participants.tsv* file. Detailed sequence parameters are available for all acquisitions in the anonymised JSON files accompanying each scan, following BIDS standards.

1. <https://braindrugs.nru.dk/>

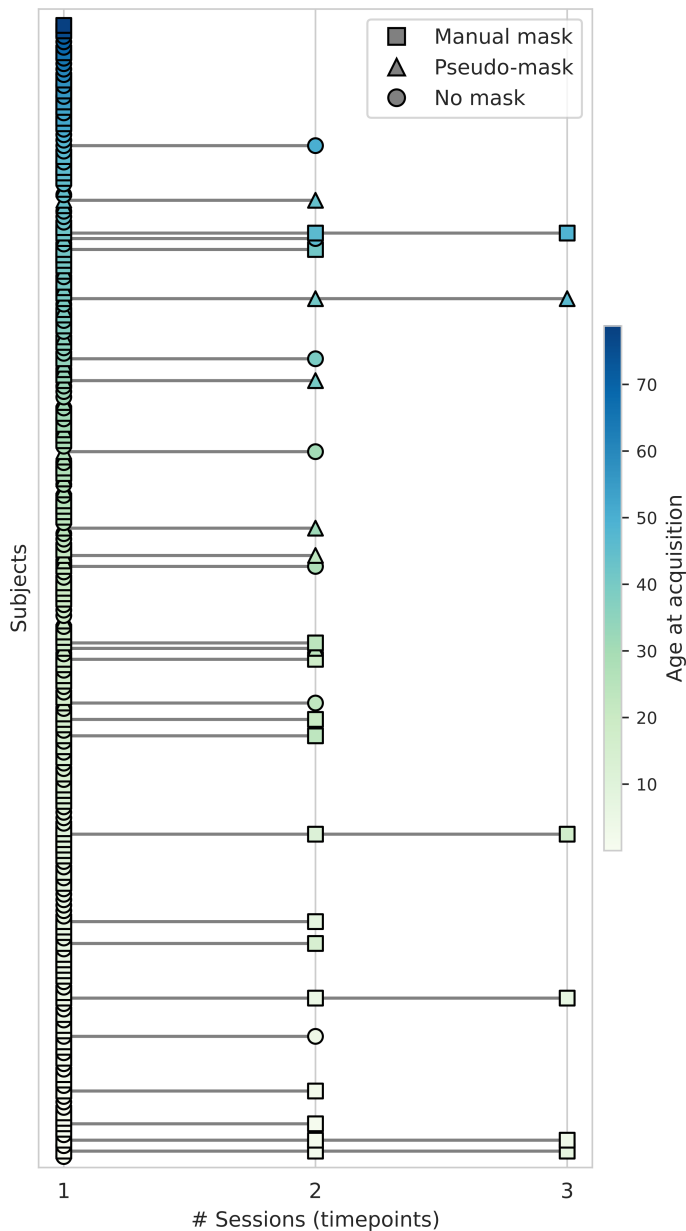


Figure 3: Overview of the subjects and their available timepoints across sessions in PMG-BrainDrugs. Each dot represents a unique subject and is coloured by the subject's age at acquisition. Individual subject identifiers are not displayed due to the large sample size. Additionally, the availability of manual or pseudo-masks is indicated by the marker shape.

5.4 Data preprocessing and organisation

MRI data were converted from DICOM to NIFTI format using *dcm2nii* (Li et al., 2016), and organised according to the BIDS standard (Gorgolewski et al., 2016).

All MRI data are provided in the recommended NIFTI format (.nii.gz) along with anonymised JSON files containing MRI acquisition parameters, ensuring all identifiable metadata is removed. The top-level BIDS directory in-

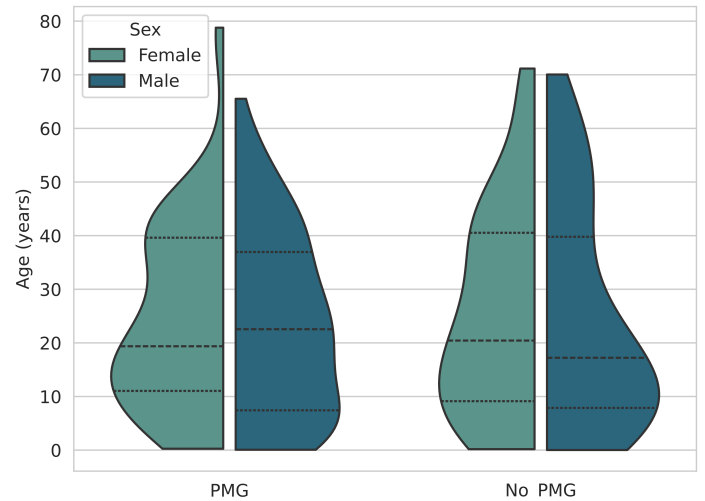


Figure 4: Demographic distribution of sex and age at acquisition, grouped by PMG and No-PMG labels for the cases included in the PMG-BrainDrugs dataset. The samples are considered per acquisition session rather than per unique subject, since age at acquisition varies amongst them.

cludes essential files such as *dataset_description.json*, which describes the dataset, and a *README*, offering additional context. The *participants.tsv* file contains demographic information about the subjects, with *participants.json* defining column descriptions and possible values. Missing information is marked as "n/a", following BIDS conventions.

Patients are identified as subjects with a unique identifier and assigned individual folders, i.e., *sub-EPxxxxxx*. Each subject folder contains session folders numerically indexed and chronologically ordered (e.g., *ses-001*, *ses-002*), where all imaging data is stored. MRI sequences (NIFTI files) are located within the nested *anat* folder, accompanied by a corresponding sidecar JSON file containing metadata. Different acquisitions of the same sequence in the same session are named in accordance with the BIDS format with the respective *run-* number.

To ensure anonymisation, facial features were removed using *pydeface* (Omer Faruk Gulban et al., 2026). After defacing, all images were visually inspected to ensure adequate anonymisation and preservation of brain anatomy. In total, 193 sequences were discarded, including 31 due to failed defacing. Additional exclusions were due to cropped field of view, uncorrectable acquisition artifacts, or scans that were in fact of the spine. The retained dataset may still contain typical acquisition artifacts (e.g. motion-related effects), although the most severe cases were removed during quality control. As the data comprise clinical cases, further limitations include the presence of large ventricles, substantial anatomical abnormalities, or prior resections, which may challenge standard structural MRI preprocessing pipelines.

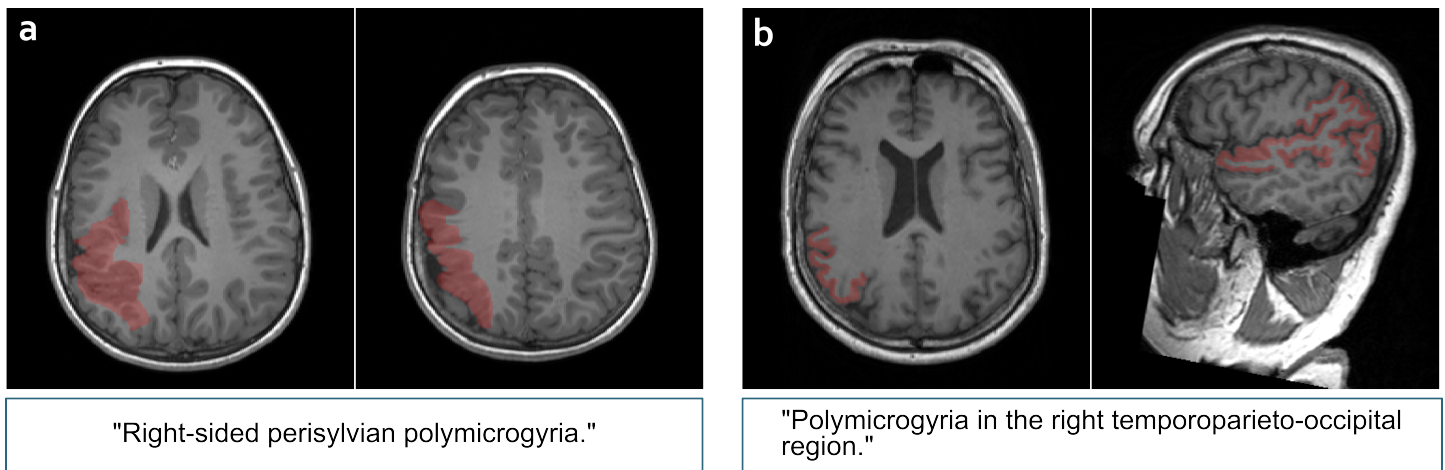


Figure 5: Examples of segmentation masks. (a) Manual segmentation delineated by a neuroradiologist on the axial plane. Two different slices from the same case are shown. (b) Pseudo-mask generated using the cortical parcellation and regions described in the radiology reports. For both cases, the conclusion of the radiology report is also shown, with the location of the abnormality translated from the original Danish report.

The *derivatives* folder contains pseudo regional masks for 31 positive PMG cases, under the "pseudo_masks" subdirectory, and manual PMG masks for the remaining positive cases, in "manual_masks". Manual annotation of PMG abnormalities was performed by a neuroradiologist (AS) with 9 years of experience, 4 of which as a neuroradiology specialist. *3D Slicer* (Fedorov et al., 2012) was used for annotation, using a customised module developed to facilitate the handling of BIDS formatted data (Available at: <https://github.com/suliciac/BIDSIterator>). The 3D annotations were delineated in the axial plane. Additionally, radiology reports were provided to the neuroradiologist responsible for the manual segmentation (AS) to guide the delineation process. Both manual and pseudo-mask filenames include the BIDS *space-* entity, which indicates the scan used as the reference space to delineate on top of. FreeSurfer *recon-all-clinical* (Gopinath et al., 2025) outputs are also provided for the 31 cases where pseudo-masks are provided under the folder "recon-all-clinical".

6. Validation

6.1 General quality of data

Overall, the data quality was rated as acceptable after visual inspection. It was first assured that the field of view comprised the complete brain, that there were no strong motion artifacts, and that there were no obvious field inhomogeneities (see Figure 6 for examples of excluded cases). The quality review was intentionally designed to resemble a visual quality review by a clinical neuroradiologist to ensure realism of the dataset. To create a benchmark dataset that was as close to reality as possible, MRI sequences that contained, e.g., slight motion artifacts were included. It is

noteworthy that the defacing tool used (Omer Faruk Gulban et al., 2026) is a standard algorithm in structural MRI preprocessing. However, it was not designed for clinical data and therefore has failed in some cases, particularly in younger subjects, where intensity distributions differ from those in adult populations, and brain sizes, of course, diverge. Hence, after defacing, all images were again visually inspected to verify that the defacing process was successful. If defacing failed, the scan or even the subject was removed from the dataset to respect the legal requirements for sharing. The dataset conforms to BIDS standards, and its validity was checked using the BIDS Validator v1.15 (Appelhoff et al., 2026).



Figure 6: Examples of cases excluded during visual quality control, such as cropped field of view and acquisition artifacts.

6.2 Manual masks

Manual masks delineated by a neuroradiologist, although increasingly difficult to obtain nowadays, remain the gold standard in most neuroimaging cohorts. In this study, no formal written annotation protocol was used. Instead, annotations followed standard neuroradiological practice, with the neuroradiologist having access to all available MRI se-

quences to delineate the full extent of the PMG lesion. Although multiple sequences were available, the segmentation mask was delineated exclusively on a single reference scan. Radiology reports were also available as a clinical reference to corroborate or refine imaging findings but did not replace expert visual assessment.

Segmentation of PMG is particularly challenging due to variability in cortical morphology, overlapping pathologies (including pachygyria and gray matter heterotopia), and partial-volume effects on 2D T2-w and FLAIR sequences (whenever T1-w was not available for annotating), which at times obscured clear anatomical boundaries. In cases of co-existing malformations, the neuroradiologist relied on expert visual assessment to approximate lesion extent, acknowledging inherent ambiguity in defining transitions between abnormal tissue types. This labour-intensive process reflects the complexity and real-world heterogeneity of the dataset, underscoring its clinical realism. An example of how these annotations look like is shown in Figure 5(a).

6.3 Pseudo-regional masks

For 31 randomly selected PMG cases, pseudo-regional masks were derived from radiology report annotations. Cortical regions associated with PMG were independently identified by a medical student (HK) and a physicist (MG) with 15 years of experience in brain MRI based on the anatomical descriptions provided in the radiology reports. The annotators (HK, MG) were instructed to identify the cortical regions corresponding to the reported PMG location. Their annotations were compared and reconciled through discussion until an agreement was reached. Three additional scans, each already containing a first session, were intended for pseudo-annotation. However, as they are post-surgical scans, their masks were excluded. For each PMG-labeled patient, the first available T1-w sequence was used to perform surface reconstruction using FreeSurfer (Gopinath et al., 2025). From the resulting Desikan-Killiany-Tourville (DKT) atlas parcellation, the cortical regions identified in the radiology reports were extracted to generate pseudo-regional masks, representing the anatomical regions associated with each PMG case.

These pseudo-masks, generated from anatomical information derived from radiology reports, constitute an independent representation of lesion localisation at the cortical-region level. An example of how these pseudo-annotations look like is represented in Figure 5(b).

7. Acknowledgements

This work was supported by the Lundbeck Foundation BrainDrugs grant (R279–2018–1145).

8. Conflicts of Interest

The authors of this paper declare no competing interests.

References

- Stefan Appelhoff, Suyash Bhogawar, Ross Blair, Christian Horea, Joke Durnez, Franklin Feingold, Rémi Gau, Krzysztof J. Gorgolewski, Rohan Goyal, Brian Grass, Charan Sai Guttula, Yaroslav O. Halchenko, Michael Hanke, Nell Hardcastle, Soichi Hayashi, Teal Hobson-Lowther, Chris Holdgraf, Mainak Jas, Alexander Jones, Jakub Kaczmarzyk, Duncan Macleod, Christopher J. Markiewicz, Zack Michael, Mikaël Naveau, David Nishikawa, Gregory Noack, Robert Oostenveld, Eduard Ort, Dimitri Papadopoulos, Dianne Patterson, Russell Poldrack, Travis Riddle, Taylor Salo, Parul Sethi, Adam Thomas, Nicolas Traut, William Triplett, David E. Warren, Joseph Wexler, Matthew Zito, and Wazeer Zulfikar. bids-validator, May 2026. URL <https://zenodo.org/doi/10.5281/zenodo.17902244>.
- Lucia Coll, Abdullah Shafique, Helene Kaas, Martin Prener, Lars Hageman Pinborg, Gitte Moos Knudsen, and Melanie Ganz. A clinical brain MRI dataset of polymicrogyria in patients suspected of epilepsy Data, June 2026. URL <https://datacatalog.publicneuro.eu/dataset/PN000024%20A%20clinical%20brain%20MRI%20dataset%20of%20polymicrogyria%20in%20patients%20suspected%20of%20epilepsy/V1>.
- European Commission. Data Protection explained, a. URL https://commission.europa.eu/law/law-topic/data-protection/data-protection-explained_en.
- European Commission. Standard Contractual Clauses, b. URL https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/standard-contractual-clauses-scc_en.
- Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, John Buatti, Stephen Aylward, James V. Miller, Steve Pieper, and Ron Kikinis. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging*, 30 (9):1323–1341, November 2012. ISSN 1873-5894. .
- Karthik Gopinath, Douglas N. Greve, Colin Magdamo, Steve Arnold, Sudeshna Das, Oula Puonti, and Juan Eugenio Iglesias. “Recon-all-clinical”: Cortical surface reconstruction and analysis of heterogeneous clinical brain MRI.

- Medical Image Analysis*, 103:103608, July 2025. ISSN 13618415. . URL <https://linkinghub.elsevier.com/retrieve/pii/S1361841525001550>.
- Krzysztof J. Gorgolewski, Tibor Auer, Vince D. Calhoun, R. Cameron Craddock, Samir Das, Eugene P. Duff, Guillaume Flandin, Satrajit S. Ghosh, Tristan Glatard, Yaroslav O. Halchenko, Daniel A. Handwerker, Michael Hanke, David Keator, Xiangrui Li, Zachary Michael, Camille Maumet, B. Nolan Nichols, Thomas E. Nichols, John Pellman, Jean-Baptiste Poline, Ariel Rokem, Gunnar Schaefer, Vanessa Sochat, William Triplett, Jessica A. Turner, Gaël Varoquaux, and Russell A. Poldrack. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3(1):160044, June 2016. ISSN 2052-4463. . URL <https://www.nature.com/articles/sdata201644>.
- Sintia Kolbjør, Daniel A. Martín Muñoz, Anne K. Örtqvist, Maria Pettersson, Anna Hammarsjö, Britt-Marie Anderlid, and Maria Dahlin. Polymicrogyria: epidemiology, imaging, and clinical aspects in a population-based cohort. *Brain Communications*, 5(4):fcad213, July 2023. ISSN 2632-1297. . URL <https://academic.oup.com/braincomms/article/doi/10.1093/braincomms/fcad213/7241064>.
- Richard J. Leventer, Anna Jansen, Daniela T. Pilz, Neil Stoodley, Carla Marini, Francois Dubeau, Jodie Malone, L. Anne Mitchell, Simone Mandelstam, Ingrid E. Scheffer, Samuel F. Berkovic, Frederick Andermann, Eva Andermann, Renzo Guerrini, and William B. Dobyns. Clinical and imaging heterogeneity of polymicrogyria: a study of 328 patients. *Brain*, 133(5):1415–1427, May 2010. ISSN 1460-2156, 0006-8950. . URL <https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/awq078>.
- Xiangrui Li, Paul S. Morgan, John Ashburner, Jolinda Smith, and Christopher Rorden. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*, 264:47–56, May 2016. ISSN 01650270. . URL <https://linkinghub.elsevier.com/retrieve/pii/S0165027016300073>.
- Daniel S. Marcus, Anthony F. Fotenos, John G. Csernansky, John C. Morris, and Randy L. Buckner. Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults. *Journal of Cognitive Neuroscience*, 22(12):2677–2684, December 2010. ISSN 0898-929X, 1530-8898. . URL <https://direct.mit.edu/jocn/article/22/12/2677/4983/Open-Access-Series-of-Imaging-Studies-Longitudinal>
- Karla L. Miller, Fidel Alfaro-Almagro, Neal K. Bangerter, David L. Thomas, Essa Yacoub, Junqian Xu, Andreas J. Bartsch, Saad Jbabdi, Stamatios N. Sotiropoulos, Jesper L. R. Andersson, Ludovica Griffanti, Gwenaëlle Douaud, Thomas W. Okell, Peter Weale, Iulius Dragonu, Steve Garratt, Sarah Hudson, Rory Collins, Mark Jenkinson, Paul M. Matthews, and Stephen M. Smith. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, 19(11):1523–1536, November 2016. ISSN 1097-6256, 1546-1726. . URL <https://www.nature.com/articles/nn.4393>.
- Omer Faruk Gulban, Dylan Nielson, Chris Markiewicz, Dimitri Papadopoulos Orfanos, John Lee, Russ Poldrack, Chris Gorgolewski, Vanessa Sochat, Satrajit Ghosh, and Yann Leprince. poldracklab/pydeface: v2.1.0, May 2026. URL <https://zenodo.org/doi/10.5281/zenodo.20090246>.
- R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner. Alzheimer’s Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology*, 74(3):201–209, January 2010. ISSN 0028-3878, 1526-632X. . URL <https://www.neurology.org/doi/10.1212/WNL.0b013e3181cb3e25>.
- Maximilian Rauch, Karsten Lachner, Lea Frickel, Monika Lauer, Simon Jonas Adenauer, Elisabeth Neuhaus, Elke Hattingen, and Luciana Porto. Focally Enlarged Perivascular Spaces in Pediatric and Adolescent Patients with Polymicrogyria—an MRI Study. *Clinical Neuroradiology*, 35(1):87–93, March 2025. ISSN 1869-1439, 1869-1447. . URL <https://link.springer.com/10.1007/s00062-024-01457-5>.
- Mariasavina Severino, Ana Filipa Geraldo, Norbert Utz, Domenico Tortora, Ivana Pogledic, Włodzimierz Klonowski, Fabio Triulzi, Filippo Arrigoni, Kshitij Mankad, Richard J. Leventer, Grazia M. S. Mancini, James A. Barkovich, Maarten H. Lequin, and Andrea Rossi. Definitions and classification of malformations of cortical development: practical guidelines. *Brain*, 143(10):2874–2894, October 2020. ISSN 0006-8950, 1460-2156. . URL <https://academic.oup.com/brain/article/143/10/2874/5890851>.
- Lingfeng Zhang, Nishard Abdeen, and Jochen Lang. pediatric polymicrogyria MRI dataset, 2022. URL <https://www.kaggle.com/datasets/lingfengzhang/pediatric-polymicrogyria-mri-dataset>.
- Lingfeng Zhang, Nishard Abdeen, and Jochen Lang. A novel center-based deep contrastive metric learning

method for the detection of polymicrogyria in pediatric brain MRI. *Computerized Medical Imaging and Graphics*, 114:102373, June 2024. ISSN 0895-6111. . URL <https://www.sciencedirect.com/science/article/pii/S0895611124000508>.