POLITECNICO DI TORINO
Repository ISTITUZIONALE

Deep Learning for MRI Segmentation and Molecular Subtyping in Glioblastoma: Critical Aspects from an Emerging Field

Original
Deep Learning for MRI Segmentation and Molecular Subtyping in Glioblastoma: Critical Aspects from an Emerging Field / Bonada, M., Rossi, L.F., Carone, G., Panico, F., Cofano, F., Fiaschi, P., Garbossa, D., Dimeco, F., Bianconi, A.. - In: BIOMEDICINES. - ISSN 2227-9059. - 12:8(2024). [10.3390/biomedicines12081878]

Availability:
This version is available at: 11583/2996619 since: 2025-01-15T14:09:07Z

Publisher:
Multidisciplinary Digital Publishing Institute (MDPI)

Published
DOI:10.3390/biomedicines12081878

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

04 September 2026

biomedicines

MDPI

Review

Deep Learning for MRI Segmentation and Molecular Subtyping in Glioblastoma: Critical Aspects from an Emerging Field

Marta Bonada [1,2], Luca Francesco Rossi [3], Giovanni Carone [2], Flavio Panico [1,*], Fabio Cofano [1], Pietro Fiaschi [4,5], Diego Garbossa [1], Francesco Di Meco [2] and Andrea Bianconi [1,4,*]

1 Neurosurgery Unit, Department of Neuroscience, University of Turin, Via Cherasco 15, 10126 Turin, Italy; marta.bonada@unimi.it (M.B.); fabio.cofano@unito.it (F.C.); diego.garbossa@unito.it (D.G.)
2 Department of Neurosurgery, Fondazione IRCCS Istituto Neurologico Carlo Besta, Via Celoria 11, 20133 Milan, Italy; giovanni.carone@unimi.it (G.C.)
3 Department of Informatics, Polytechnic University of Turin, Corso Castelfidardo 39, 10129 Turin, Italy; lucafrancesco.rossi@polito.it
4 Division of Neurosurgery, Ospedale Policlinico San Martino, IRCCS for Oncology and Neurosciences, Largo Rosanna Benzi 10, 16132 Genoa, Italy; pietro.fiaschi@unige.it
5 Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics and Maternal and Child Health, University of Genoa, Largo Rosanna Benzi 10, 16132 Genoa, Italy
* Correspondence: flavio.panico@unito.it (F.P.); andrea.bianconi@edu.unito.it (A.B.)

Abstract: Deep learning (DL) has been applied to glioblastoma (GBM) magnetic resonance imaging (MRI) assessment for tumor segmentation and inference of molecular, diagnostic, and prognostic information. We comprehensively overviewed the currently available DL applications, critically examining the limitations that hinder their broader adoption in clinical practice and molecular research. Technical limitations to the routine application of DL include the qualitative heterogeneity of MRI, related to different machinery and protocols, and the absence of informative sequences, possibly compensated by artificial image synthesis. Moreover, taking advantage from the available benchmarks of MRI, algorithms should be trained on large amounts of data. Additionally, the segmentation of postoperative imaging should be further addressed to limit the inaccuracies previously observed for this task. Indeed, molecular information has been promisingly integrated in the most recent DL tools, providing useful prognostic and therapeutic information. Finally, ethical concerns should be carefully addressed and standardized to allow for data protection. DL has provided reliable results for GBM assessment concerning MRI analysis and segmentation, but the routine clinical application is still limited. The current limitations could be prospectively addressed, giving particular attention to data collection, introducing new technical advancements, and carefully regulating ethical issues.

Keywords: artificial intelligence; deep learning; magnetic resonance imaging; glioblastoma; segmentation; molecular data; clinical applicability

check for updates

Citation: Bonada, M.; Rossi, L.F.; Carone, G.; Panico, F.; Cofano, F.; Fiaschi, P.; Garbossa, D.; Di Meco, F.; Bianconi, A. Deep Learning for MRI Segmentation and Molecular Subtyping in Glioblastoma: Critical Aspects from an Emerging Field. *Biomedicines* **2024**, *12*, 1878. https://doi.org/10.3390/biomedicines12081878

Academic Editor: Hung-Pei Tsai

Received: 8 July 2024
Revised: 29 July 2024
Accepted: 31 July 2024
Published: 16 August 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Gliomas, and among them, glioblastoma multiforme (GBM), stands as the most common and deadliest form of primary brain tumor in adults, characterized by its aggressive growth and poor prognosis. According to the Central Brain Tumor Registry of the United States, GBM accounts for 14.6% of all primary brain tumors and an alarming 48.6% of primary malignant brain tumors, with a median survival rate of approximately 15 months post-diagnosis [1]. The complexity of GBM's treatment lies in its highly heterogeneous nature, both genetically and in its response to treatment, necessitating personalized therapeutic approaches [2–4]. In this context, magnetic resonance imaging (MRI) emerges as a cornerstone in the diagnosis, treatment planning, and follow-up of GBM, providing detailed insights into the tumor's location, size, and interaction with surrounding brain structures. The segmentation of GBM from MRI scans, a critical step in delineating the

tumor boundaries, is traditionally performed manually by expert radiologists. This process, however, is not only time-intensive but also prone to variability, underlining the urgent need for more efficient and standardized approaches [5].

Recent advances in artificial intelligence (AI), particularly deep learning, have shown promising potential to address these challenges [6]. Deep learning models, trained on vast datasets of annotated MRI scans, can automate the segmentation process, offering not just speed and efficiency but also the promise of reducing human error [7]. This technological leap could revolutionize the clinical management of GBM, from enhancing diagnostic accuracy to informing the surgical strategy and assessing treatment response. Yet, the path to integrating deep learning into clinical practice is fraught with obstacles. These range from technical challenges, such as the need for large, diverse training datasets and the management of imaging variability, to broader concerns around algorithm validation, integration into clinical workflows, and ethical considerations.

Taking these considerations as a premise, in this article, we intend to provide a comprehensive overview of the current state of deep learning applications in MRI segmentation and molecular subtyping for glioblastoma, critically examining the limitations that hinder their broader adoption in clinical practice and molecular research. Considering the breadth of topics covered, we introduce Table 1 to summarize the various points and facilitate readability.

Table 1. Summary of the presented concerns regarding AI application for glioblastoma MRI segmentation. Each limitation is accompanied by the domain of pertinence, the definition of the problem, and the proposed solution(s).

Section	Limitation	Domain	Definition	Possible Solution(s)
2.1	imaging heterogeneity	technical	scanner-dependent variation in image signal intensity	intensity standardization
2.2	missing MRI sequences	technical	unavaiable modality/ies (T1, T2, FLAIR, T1CE)	rescanning data inter-modality translation knowledge distillation
2.3	deployment issues	technical	limited computational resources and memory constraints	tiling quantization
2.4	performance evaluation	technical	subjective reference standards	cross-validation unsupervised training
3.1	limited number of patients	application	low number of data publicly avaiable	transfer learning
3.2	data quality	application	suboptimal quality of data (non-volumetric scans)	pre-processing inclusion of complex scenarios
3.3	data selection	application	selection bias and reduced applicability	inclusive database
3.4	focus on preoperative scenario	application	logistical and technical issues for postop. MRIs	multi-modality and multi-institutional data
4	exclusion of molecular data	molecular	limited consideration of IDH—1p/19q—MGMT	new coder architecture large-scale data-sharing
5.1	lack of standard guidelines	ethical	scientific integrity not definable	checklist
5.2	lack of transparency	ethical	limited understanding of the results	interpretability methods interdisciplinary collaboration
5.3	privacy and data protection	ethical	difficulty to obtain complete anonimization	skull-stripping

2. Technical Challenges

2.1. MRI Imaging Heterogeneity

Unlike the CT scanner, MRI is characterized by scanner-dependent variation in image signal intensity related to variability in time points, vendors, magnetic field strengths, and acquisition settings [8–10]. Radiomic features are highly sensitive to the level of the signal intensity in the image and, thus, non-biological alterations should be removed [11]. Therefore, to obtain reproducible results in radiomic analysis, MRI signal intensity must be standardized. The standardization should obtain an adequate range and distribution of voxel intensity in the different MRIs [12]. Nowadays, there is no general consensus concerning the most reliable standardization approach to adopt.

AI in the field of medical imaging faces unique challenges due to the innate complexity and diversity of medical data [11]. AI models are intrinsically designed to discern and learn from patterns within their training data. However, these patterns may not purely reflect the biological and pathological information of interest, but also the methodological biases present in the data acquisition process.

When training datasets are predominantly composed of MRI scans from a single medical center, the AI model may inadvertently prioritize the specific characteristics of that center's imaging protocol over the more crucial pathological features of the tumor itself. This situation leads to an overfitting, where the model performs exceptionally well on the training data due to its familiarity with the protocol-specific nuances but fails to generalize this performance to new, unseen data from other centers with different imaging protocols. This issue is exacerbated in the context of gliomas, a highly heterogeneous group of brain tumors, both biologically and morphologically. The heterogeneity is an essential aspect of the disease that AI models must capture to generalize effectively across different patient populations. Therefore, to develop AI models that are robust and generalizable, it is essential to train them on diverse, multi-center datasets that encompass the broad spectrum of imaging techniques and the varied appearances of glioblastoma [13].

For such a reason, for MRI radiomics analysis, a key challenge is to ensure repeatability and reproducibility of the results in the removal of scanner-dependent signal intensity changes [12]. In fact, intensity standardization helps in making the evaluations agnostic to acquisition specifications and allows us to create more reliable models. An interesting tool for MRI harmonization is ComBat, a statistical normalization method for batch-effect correction in genomics that shows promising results in removing scanner-dependent information from extracted features when applied to radiomics [9,10]. Marzi et al. [14] further extended this idea by proposing a harmonizer transformer, an implementation of ComBat allowing its encapsulation as a preprocessing step of a machine learning pipeline, sensitively reducing site effects.

Yet, implementing harmonization in an operative procedure based on the information provided has been shown to have a serious effect on the level of repeatability and redundancy of features [15]. For such a reason, several studies suggest that rescanning data provides the opportunity to assess radiomic feature reproducibility on images from the same patient acquired within a short time delay, despite the minimal modification that a tumor can present within several days [10,15].

2.2. Missing MRI Sequences

In clinical practice, obtaining multiple sequences is time-consuming and expensive [16]. Moreover, MRI examinations may vary among institutions because of different acquisition protocols and/or different hardware with unequal resolution capabilities [17].

Nevertheless, for brain tumor segmentation, multi-contrast MRI modalities such as T1, T2, Fluid-Attenuated Inversion Recovery (FLAIR), and T1 Contrast-Enhanced (T1CE) play an essential role in collecting the informative features [18,19]. Utilizing multimodal data through concatenating multiple MR images as inputs for any machine learning method has demonstrated a proficiency in enhancing the semantic segmentation performances of brain tumors [20,21]. More specifically, each imaging modality enables the deep learning convolu-

tional networks to extract and learn reciprocal knowledge to segment different subregions of glioblastoma. For example, peritumoral edema appears as a hyperintensity area in T2 and FLAIR images, while enhancing tumor is highlighted with hyperintensity in T1CE image [22]. Unfortunately, it is usually required to have the complete four image modalities as inputs (i.e., T1, T2, FLAIR, and T1CE). Thus, the problem of missing modalities from MRI examinations leads to challenges for the segmentation task [23,24]. The complexity increases even more with postoperative segmentation, where the standard sequences listed above should be supplemented with diffusion/ADC maps to better characterize residual areas, which can often be confused with normal postoperative barrier damage, edema changes, or possible vascular alteration.

To tackle this problem, the artificial synthesis of missing target modalities from one or more available modalities has recently attracted increasing attention [25]. However, it remains challenging to achieve accurate segmentation results from the synthesized images [16]. In fact, there is a gap between the synthesized and real target modalities, thus causing the segmentation based on synthesized images to perform worse [26]. Moreover, the model becomes more complex and much deeper than the original independent model and it has a higher risk of overfitting [16]. This issue requires to be addressed with more effective regularization methods to maintain the performance during testing. However, the regularization of these models has rarely been explored in existing works.

Eijgelaar et al. [24] introduced a training method that uses sparsity to enhance model outcomes when working with partial clinical datasets. However, even with these modifications, the highest performance levels were attainable only with the full complement of sequences. In fact, each diagnostic modality, whether it is a variant of MRI, CT, or another type of scan, inherently provides unique information. Consequently, any method that tries to compensate for missing information is essentially attempting to mitigate the impact on the algorithm's effectiveness [26].

Due to the rise in deep learning, image synthesis within the same modality (such as intramodalities: e.g., from MRI to CT) and across modalities (intermodality: e.g., from FLAIR MRI to T1ce MRI), which entails artificially reconstructing missing sequences from available ones, has attracted considerable attention. This area is emerging as a vibrant and promising research domain. In Figure 1, the IMT model proposed by Osman et al. [27], which was able to generate an accurate synthesis result by generating the missing modalities, is shown.

Various network architectures have been proposed for such tasks in medical imaging within the last few years, but three main backbone models achieved the best results: autoencoder, U-Net, and GAN, with the first starting to lose pace compared to the others [26]. Some studies specifically tackled the problem of brain MRI intramodality synthesis. Yang et al. [28] proposed a method to perform image modality translation (IMT) by leveraging conditional generative adversarial networks (cGANs), whose generator follows the U-Net shape by adding skip connections between mirrored layers in the encoder–decoder network and whose discriminator is derived from a PatchGAN classifier. Osman and Tamam [27] instead implemented a U-Net model aimed at learning the non-linear mapping between a source image contrast to a target image contrast.

Generative adversarial networks (GANs) [29,30] are a relatively new type of DL model that have received much attention because of their ability to generate synthetic images. GANs are trained using two neural networks—a generator and a discriminator. The generator learns to create data that resemble examples contained within the training dataset, and the discriminator learns to distinguish real examples from the ones created by the generator [17]. The two networks are trained together until the generated examples are indistinguishable from the real examples. For such reason, from their conception, GANs have found many applications in medical imaging [31,32].

Another interesting example comes from the Tumor Image Synthesis and Segmentation Network (TISS-Net), a dual-task architecture for end-to-end training and inference, where the synthesis and segmentation models are learned synergistically with several novel high-level regularization strategies [16]. TISS-Net leverages not only a GAN-like architecture

comprising a dual-task generator and a dual-task segmentor, but exploits specific domain knowledge while structuring the learning phase, leading to what is known as segmentation-aware target modality image synthesis, where a coarse segmentation is used as an auxiliary task to regularize the synthesis task, and a tumor-aware synthesis loss with perceptibility regularization is introduced to generate segmentation-friendly images in the missing modality. This allows for the improvement and further refinement of the image quality around the tumor region and to reduce the high-level domain gap between synthesized and real target modality images [16].

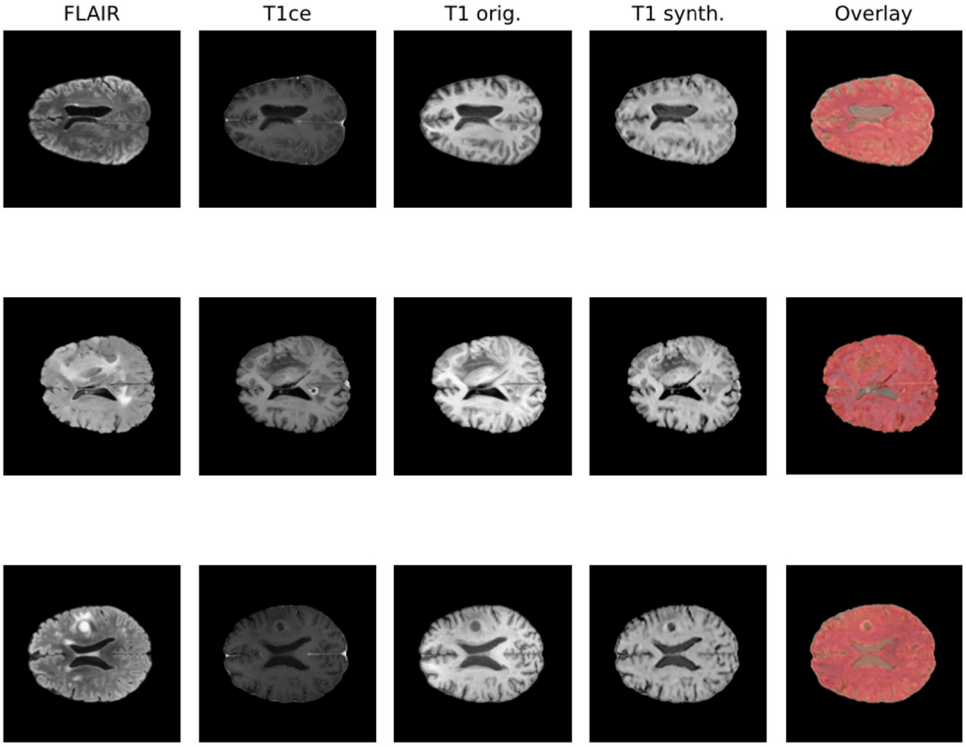

Figure 1. Qualitative comparison of synthesized T1 scans from BraTS 2021 dataset using the IMT technique.

Another promising approach to overcome the limitation of potential missing sequences is knowledge distillation (KD), which utilizes a teacher–student model to compress model architecture from a cumbersome network to a compact one [18]. In medical image analysis domain, especially in brain tumor segmentation, the KD is used to transfer complete multimodal information from the teacher network to a unimodal student network [33]. However, this is a two-stage approach which requires a training phase for the teacher network with full image modalities. Afterwards, the information is transferred to the student network that utilizes limited modalities. This results in additional training costs and extra time to generate the pre-trained model. Moreover, the teacher might need to be updated or fine-tuned during the training of the student [34]. Choi et al. [18] generated a single-stage-learning knowledge distillation algorithm for brain tumor segmentation. In this case, both models are trained simultaneously using a single-stage knowledge distillation algorithm.

2.3. Deployment Issues

Deep learning (DL) recently provided promising results in medical imaging segmentation [35,36]. Nevertheless, the deployment of the available models poses a substantial challenge, mainly related to their computational footprint. In fact, most DL-enabled studies are highly demanding in terms of both energetical and computational resources and

such complexities make them very difficult to deploy, especially in tightly controlled clinical scenarios [37].

Moreover, memory constraints in deep learning accelerator cards have often limited training on large 2D and 3D images due to the size of the activation maps held for the backward pass during gradient descent [38]. Two methods are commonly used to manage these memory limitations: the downsampling of images to a lower resolution and/or the breaking of images into smaller tiles [39,40]. Tiling is often applied in the case of large images to compensate the memory limitations of the hardware [41]. In particular, fully convolutional networks can be taught to be translation-invariant and are a natural fit for tiling methods as they can be trained on images of one size and perform inference on images of a larger size by splitting them into smaller sections and thus performing on the smaller tiles [42]. Nevertheless, tiling methods are used primarily to limit the impact of insufficient memory but they usually do not improve the predictive power of the system [38].

The process of quantization is essential to reduce the memory burden during the time of inference [37]. After the process of quantization, a high precision model is reduced to a lower-bit-resolution model (low-precision floating-point or integer quantization are common choices), thus reducing the size of the model and exploiting SIMD or MIMD computations to make inference faster [43].

In the literature, some models of quantization have been created to reduce computational requirements while keeping the segmentation performance stable, like the one by Thakur et al. [37]. In this light, both quantization-aware training and post-training optimizations present themselves as promising approaches for enabling the execution of advanced DL systems on plain commercial-grade GPUs. Therefore, they contribute to the possible spread of DL-based segmenting applications in clinical environments despite the frequent lack of advanced technological support.

Certainly, an increased access to memory remains essential to make a further step in this direction [43]. This should include both improvements in hardware and computing techniques, such as model parallelism [44] and data parallelism [45].

2.4. Performance Evaluation

AI segmentation performance is gauged against a reference known as the ground-truth, which, in clinical contexts, is often established via manual segmentation by one or more professional radiologists. Various studies have highlighted the subjective nature of reference standards based on radiologists' assessments, noting that model performance can fluctuate when trained on different ground-truths [46].

Due to the intrinsic nature of the medical segmentation task, it is hard to consider different ground-truth strategies other than the manually performed labeling. For such reason, common measuring approaches leverage human error subjectivity mitigation trough the averaging of multiple manual segmentations. Annotation averaging is usually performed through tools such as STAPLE, which is based on expectation–maximization and probabilistically corrects noise elements such as outliers or false positives. Usually, this approach is similarly performed for the AI model, exploiting what is known as cross-validation.

Leveraging multiple trained models to aggregate inferences through STAPLE enhances the robustness of AI predictions. Figure 2 shows that STAPLE was able to correct a misclassification case of the resection cavity. Metrics like the Dice Score and Hausdorff distance percentiles are then used to assess the AI model's true capabilities comprehensively.

Nonetheless, while the significance of human annotations has been widely recognized, the actual process of annotation has received less scrutiny. For instance, Zając et al. [47] highlighted the inherent limitations present in the creation of human-labeled annotations for medical datasets. Sylolypavan et al. [48] also discussed how inherent biases, judgments, and errors from experts could influence AI-driven decision-making in clinical settings.

With the goal of trying to overcome these limitations, some proposals to shift from supervised segmentation tasks to unsupervised ones in the medical environment have been presented. Aganj et al. [49] suggested a whole segmentation approach based on the

local center of mass by grouping pixels iteratively in 3D MRIs, which achieved interesting results. Instead, Kiyasseh et al. [50] tackled the problem from an even wider perspective, proposing a novel framework for evaluating clinical AI systems in the absence of ground-truth annotations theoretically capable of identifying unreliable predictions and of assessing algorithmic biases. A recent study from Yale University introduced a whole framework for unsupervised segmentation [51], which exploits image-specific embedding maps and hierarchical dynamic partitioning at different levels of granularity. These demonstrate an improvement ranging from 10% to 200% on Dice coefficient and Hausdorff distance with respect to previous unsupervised proposals.

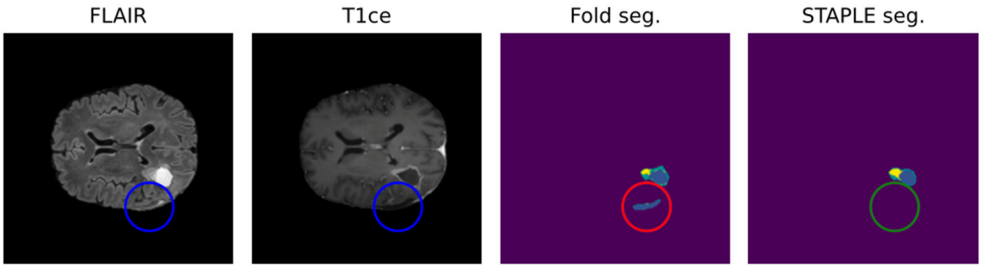

Figure 2. Positive effect of STAPLE fusion for resection cavity segmentation [47]. Results obtained from the fivefold cross-validation process (fold seg.) are merged by the STAPLE algorithm to obtain a final result (STAPLE seg.). The figure shows, as an example, how the STAPLE convergence is able to recognize oversegmentation of a hypointense region misclassified as resection cavity (blue: cavity; yellow: enhancing; green: whole).

Even though the performances achieved do not reach the results from supervised training, there is an undeniable surge in proposals targeting the principal challenges of data labeling in recent years. This trend is paving the way for novel approaches in AI-based model design, specifically tailored for clinical application.

3. Application to a Real-Word Scenario

3.1. Limited Number of Patients

One of the major limitations in the successful translation of AI algorithms into common practice is the limited number of patients included in each study, with a mean of 148.6 patients for each study and a median of 60.5 [52]. Moreover, clinical translation has been significantly hampered due to limited available annotated datasets and decreased performance of algorithms on geographically distinct validation datasets [53,54]. This issue remains in an essential step to obtain an effective training of the algorithm and to lower the risk of overfitting [52].

Paradoxically, the application of complex and sophisticated DL algorithms for segmentation underperforms older ML methods when small datasets are used ($n \leq 15$ patients) [55].

As a first step, the creation of single-center datasets is more suitable to the clinical imaging protocols of the hospital and the patient cohort on site [52]. Nevertheless, the most appealing perspectives are data sharing agreements, the development of image databank consortiums (MIDRC, TCIA, BraTS), and federated learning [56,57].

Since 2012, the Brain Tumor Segmentation (BraTS) challenge has implemented and increased the role of ML in glioma MRI evaluation. The focus has been centered on the evaluation of state-of-the art methods for tumor segmentation, the classification of the lesion and, more recently, the prediction of prognosis [58,59].

Additionally, some informatic tools can be applied to limit the impact of data scarcity on the algorithm performance. Among these tools, the application of transfer learning (TL) has been previously explored and applied [46]. TL derives from the cognitive conception that humans can solve similar tasks by exploiting previously learned knowledge, with such knowledge being therefore transferred across similar tasks to improve performances on a new one. Current transfer learning techniques in medical imaging implement knowledge

transfer from natural imaging. Overall, two main paths to apply transfer learning have historically been delineated: feature extractor and fine-tuning. The main difference lies in the fact that the first freezes the convolutional layers whereas the latter updates parameters during model fitting. Nonetheless, even if some progress is achieved, the knowledge transferred between the two areas can either be insufficient for achieving promising results in the medical task or make the transfer process quite unpredictable [60]. A recent study by Bianconi et al. [46] applied transfer learning from the preoperative brain tumor segmentation task to postoperative segmentation by fine-tuning the model, showing interesting results and promising multi-site generalization leveraging their structural closeness in the knowledge domain.

3.2. Data Quality

A significant obstacle to the efficient growth of AI application in clinical practice is data quality. AI models require large amounts of high-quality images to obtain reliable results, whereas medical data frequently present suboptimal qualities for this task [61]. In fact, suboptimal quality of data is very common in clinical practice, including non-volumetric scans, missing sequences, and artifacts [62]. Low quality may negatively impact the performance because real inter-image variability results are hardly distinguishable from artifacts. In Figure 3, the non-volumetric image negatively affected the result of the segmentation from a DL-based algorithm [47].

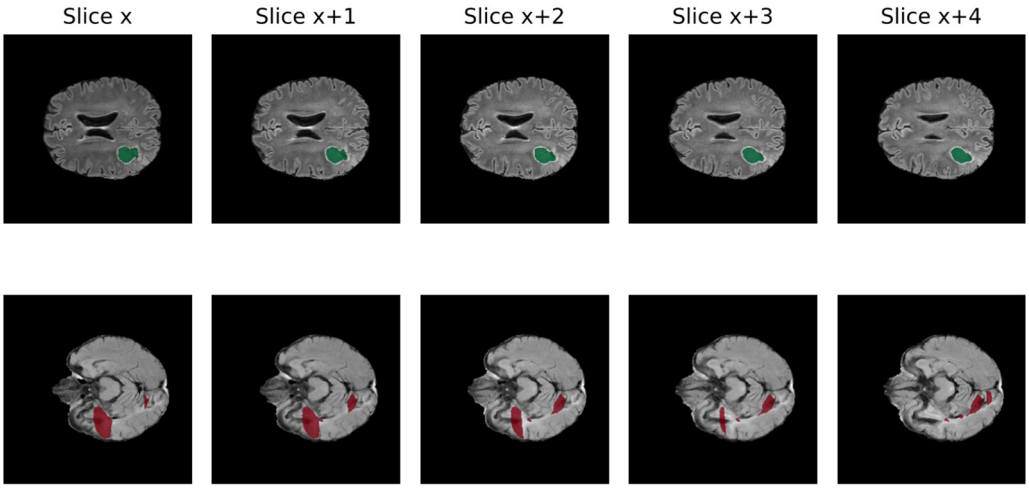

Figure 3. Qualitative comparison of processed FLAIR from volumetric raw input (above) and non-volumetric one in 5 different slices of the same MRI (Green: cavity, Red: FLAIR hyperintensity) [46].

Thus, problems such as bias, improper curation, and low reliability could be introduced [63]. Bias might occur when the AI models are trained on data that are not representative of the target patient population. In particular, DL algorithms are often trained on cured and standardized datasets that do not represent clinical data heterogeneity and quality. Although this selection bias makes the training process easier, it makes the results not as easily transferable to real-world clinical practice [61].

Idealistically, the elimination of non-volumetric scans and low-quality imaging from clinical practice would have the greatest impact for the future clinical application of AI technologies. Nevertheless, at present, it is essential to train the models to perform adequately despite the heterogeneity and the complexity of the cases. Recent efforts from BraTS are aimed at including imaging acquired with lower technologies, such as MRIs from Sub-Saharan Africa [64]. This choice pushes the efforts towards advanced image preprocessing to enhance the resolution and other tools able to support the most accurate segmentation, even in complex scenarios [65].

3.3. Data Selection

The results obtained from most of the algorithms are not easily reproducible in the real world context since they are frequently trained on curated and standardized datasets that do not include suboptimal-quality images. Although this selection bias makes the training process easier, it is not as easily transferable to real-world clinical practice. In fact, suboptimal quality of data is very common in clinical practice, including non-volumetric scans, missing sequences, and artifacts [66]. Moreover, the current literature often presents strong criteria for the inclusion/exclusion of MRI scans in the final datasets. For example, regarding postoperative brain tumor segmentation, criteria generally concern the exclusive inclusion of newly diagnosed GBM, the availability of all imaging modalities, or the defined presence of the resection cavity on visual inspection. An additional selection bias for studies concerning DL methods for the segmentation of glioma MRIs concerns the diffused use of publicly available benchmarks, such as BraTS and TCIA databases for the training phase [52]. On the one hand, their use has a positive role in the development of DL systems, but there is a substantial risk of overfitting. This could explain the high accuracy and reproducibility obtained by different reported algorithms [52].

In a recent study, a DL algorithm was trained on an MRI database that is representative of the real-world scenario, thus including heterogeneous and incomplete data [46]. In this study, inclusion criteria were not restrictive concerning the quality of the available data to avoid selection bias. So, low-quality images (e.g., non-volumetric imaging) and incomplete cases (with missing sequences) were also included. Notably, there is not a benefit in performance from incorporating non-volumetric imaging since its inclusion in the dataset creates a difficult scenario for the algorithm to be correctly classified. Indeed, the benefits resulting from the incorporation of these data are related to clinical applicability of the algorithm. Moreover, an increased heterogeneity of data would derive from a multi-institutional cooperation to create a real clinical database. In fact, having images acquired with different protocols, resolution, and contrast (1.5 T or 3 T) would make the learning process more complex but would probably result in increased adaptability of the algorithm to different clinical scenarios. It is necessary to point out that the positive role of heterogeneity of data for the training phase depend on the final purpose of the algorithm. In case the application concerns a specific scenario with defined features required for the dataset, heterogeneity may not have a positive impact.

Concerning machine learning algorithms which can identify gliomas in datasets containing non-glioma images, some studies have recently been performed in this regard but the algorithms should be further developed to allow for integration into clinical workflow [67].

3.4. Focus on Preoperative Scenario

Notably, most algorithms in the current literature are based on preoperative tumor imaging, whereas most clinical imaging techniques for brain tumors are used after treatment to assess response or to monitor progression [52]. Limitations in postoperative MRI evaluation are partly due to artifacts, caused by blood and air in the resection cavity (RC), and logistical issues in collecting data regularly from the same patient during follow-up [68,69]. Particularly, the RC is frequently a source of artifacts in the MRI because of blood residuals and air bubbles [58,70]. In addition to this, brain anatomy may be partly altered because of the surgical act, the post-surgical edema, and the tumor itself [70]. These problems lower the accuracy of available algorithms in obtaining a postoperative evaluation of MRI. Figure 4 shows an example of the misclassification of postoperative MRI obtained by a reliable DL tool for MRI pre- and postoperative segmentation [46].

Moreover, postoperative images have different acquisition times given the time-course of the disease and the treatment schedule. This means that the postoperative MRI database contains images from different points in time: immediate postoperative, before and after adjuvant treatment, and regular follow-up. Recently, some studies reported good accuracy in postoperative segmentation of MRI, though it is still far from the level of accuracy achieved in preoperative evaluation [66,71]. Moreover, the absence of a wide dataset

like BraTS requires researchers to deploy their model on small private collections, hence reducing comparability and generalizability.

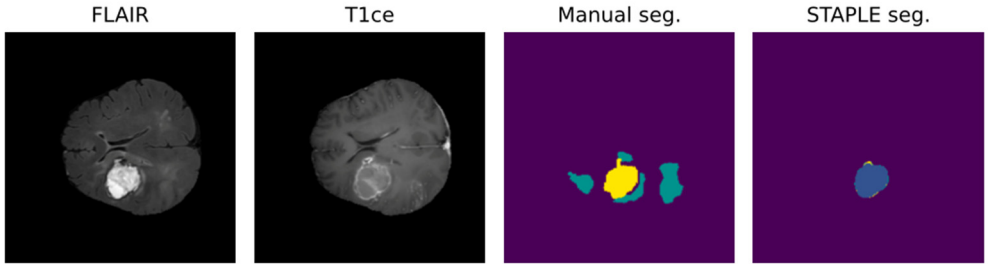

Figure 4. DL-based postoperative segmentation (STAPLE seg.) still obtains non-accurate results in some cases, as the one presented above [46]. In this example, the DL result completely diverged from the manual segmentation.

The future perspective includes the creation of multi-institutional databases including postoperative MRIs. Still, due to the complex nature of such a task, some studies have tried to mimic postoperative MRIs by exploiting generative adversarial network (GAN) capabilities in synthesizing fake-yet-plausible MRIs starting from real scans. A recent study from the State Key Laboratory of Oncology in South China [72] proposed CoCosNet, a neural network showing interesting results in artificially synthesizing postoperative weighted-T1 MRIs from the corresponding preoperative ones and postoperative CTs.

Considering the data-intensive nature of deep DL solutions, the predominant recommendation continues to be the collection of multimodality and multi-institutional MRI data. This strategy aims to normalize and synchronize the postoperative scenario in a manner analogous to how the BraTS challenge has standardized the preoperative phase.

4. Molecular Subtyping

Many recent research articles have reported remarkable success in the use of artificial intelligence to predict the status of 1p19q codeletion, IDH1 mutation, and MGMT promoters. These molecular features acquired an increasing interest as they are related to the prognosis and to the identification of the best treatment options for each singular patient [73]. Also, the BraTS challenge recently gave more relevance to these features. In fact, the second task of BraTS 2021 consisted of the evaluation of methods to predict the MGMT promoter methylation status.

4.1. IDH Mutation

Having IDH1 or IDH2 mutations is associated with improved survival [74,75] as these gliomas respond better to temozolomide therapy [76]. IDH-mutant gliomas demonstrate lower regional cerebral blood volume and flow on MR perfusion, higher apparent diffusion coefficients on diffusion MR imaging, and improved survival [77,78]. In a study by Beiko et al. [79], the resection of non-enhancing gliomas correlated with improvements in PFS in HGG with mutated IDH gliomas as opposed to IDH wild-type tumors. Thus, the knowledge of IDH mutation status before surgical resection may be important. The features that mattered most to predict IDH mutation status include absent or minimal enhancement, central areas with low T1 and FLAIR signal and well-defined tumor margins according to Liang et al. [80]. Their study was performed using the publicly available BraTS 2017 database and finally achieved 84.6% accuracy [77]. More recently, the studies from Chang et al. [81] obtained accuracy levels higher than 90% for the prediction.

4.2. P/19q Codeletion

Indeed, there are a paucity of manuscripts using CNNs to predict 1p19q codeletion. In one of the most successful studies, Chang et al. [81] succeeded in predicting 1p19q codeletion status with an accuracy of 92%. They employed the component analysis to

define the features that were mostly related to the codeletion. According to their analysis, the 1p19q codeletion status is related to frontal lobe location, ill-defined tumor borders, and larger amounts of contrast.

4.3. MGMT Methylation

The hypermethylation of the MGMT promoter is strongly associated with better response to temozolomide chemotherapy and improved prognosis [82,83]. Nevertheless, the prediction of MGMT mutation status from preoperative MRI using AI has achieved modest results by now [84]. More recent studies from Korfiatis et al. [85] and Chang et al. [86] obtained accuracies higher than 80% for the prediction of MGMT status. Again, in this paper, the use of principal component analysis for dimensionality reduction determined that the most important imaging features for the prediction of MGMT status included heterogeneous and nodular enhancement, the presence of eccentric cysts, more mass-like T2/FLAIR signal with cortical involvement, and frontal/temporal lobe locations [86]. These findings confirm the results from prior MR genomics studies [85,87].

In summary, molecular features such as IDH mutation, 1p19q codeletion, and MGMT promoter status are successfully predicted by AI applied to MRI. Moreover, algorithms obtaining accuracies of prediction exceeding 80% to 90% may probably already be superior to human-level performance. This expanding field could have a further impulse and improvement in performance with new coder architecture, large-scale data sharing, and integration of clinical data.

5. Ethical Concerns

5.1. Lack of Standard Guidelines for Clinical Studies

The advent of deep neural networks has engendered many applications in medical imaging [88]. Currently, the field of radiomics lacks standardized evaluation concerning both scientific integrity and clinical relevance of the published radiomics investigations [89]. Rigorous evaluation criteria and reporting guidelines need to soon be established and rigorously respected to obtain clinical applicability [90].

To guarantee a high standard of research and to obtain reproducible results, radiomic investigations should respect the well-defined criteria of reliability and reproducibility concerning both the presented results and the applied methods. Numerous checklists have recently been proposed and came into widespread use [52]. Those checklists include the Standards for Reporting of Diagnostic Accuracy Studies (STARD) [91,92], Strengthening the Reporting of Observational studies in Epidemiology (STROBE) [93], and Consolidated Standards of Reporting Trials (CONSORT) [94,95]. The most recent STARD list was released in 2015 and included 30 items identified by an international group of methodologists, researchers, and editors. Those items were identified so that, when they reported, readers can judge the potential bias in the study, appraise the applicability of the study findings and the validity of the conclusions. Indeed, the Checklist for AI in Medical Imaging (CLAIM) is modeled after the STARD guideline, but it has been extended to address applications in classification, image reconstruction, text analysis, and workflow optimization. These elements are considered as "best practice" elements that should guide authors in presenting their research [96].

5.2. Lack of Transparency

Interpretability of an AI program is defined as the human ability of understanding the link between the initial features extracted by the program and the final prediction obtained. As DL structures are typically complex and composed of numerous hidden layers, this link is difficult to find. This concept is commonly referred to as the "black-box problem" [97]. Nevertheless, the lack of transparency regarding AI techniques is a significant concern. Any medical care system needs to be understandable and explicable for physicians, administrators, and patients. It should ideally be able to fully explain the reasoning behind a decision to all parties concerned [61].

Interpretability methods are approaches designed to explicitly enhance the interpretability of a machine learning algorithm, despite its complexity [97].

Interpretability techniques like guided backpropagation, gradient-weighted class activation mapping (Grad-CAM) [98], and regression concept vectors [99] are being applied to medical images, illustrating the burgeoning interest in Explainable AI (XAI) and Neurosymbolic AI (NeSymAI). These fields are gaining traction as they provide crucial insights into the reasoning behind AI models, particularly in the medical realm, making the interpretation of predictions vital [100,101]. The repertoire of interpretability methods is expanding to keep pace with the complexity of radiology practices, which increasingly integrate different types of patient data, such as imaging, molecular pathways, and clinical scores Hence, interpretability approaches capable of processing this diverse information are seen as highly promising [102].

Also, interdisciplinary collaboration may limit the lack of transparency and, thus, support the acceptance of AI in clinical practice [61]. Surgeons, radiologists, data scientists, AI experts, and ethicists must collaborate to create robust guidelines and standards for the ethical deployment of AI tools in patient care. Moreover, the relationship with the patient plays an essential role in this process of acceptance [60]. In fact, the patients deserve to understand how AI influences their medical care and outcomes. For this reason, a careful and complete informed consent should not be underestimated for the perspective introduction of AI into clinical practice.

5.3. Privacy and Data Protection

Until the past decade, medical imaging included 2D images. However, advances in imaging methods have made high-resolution 3D imaging a reality through smoothing, interpolation, and super-resolution methods, enabling accurate volume rendering [103]. With advances in facial recognition, it is not difficult to match images generated from CT or MRI scans to photographs of an individual. For this reason, in medical imaging research, it is standard practice to modify images using defacing or skull-stripping algorithms to remove facial features [61]. Additionally, this process reduces inter-patient physiological variability, and the pathological aspects can emerge more relevantly. Unfortunately, such modifications can negatively affect the generalizability of machine learning models developed using such data [103]. Anyway, patient privacy is a major health system concern requiring multiple legal quandaries to be addressed prior to uploading and diffusing data [37].

Moreover, when considering advanced healthcare imaging, it is difficult to obtain a complete anonymization, despite the efforts in data protection. The basic task seems straightforward: selectively remove or codify identifiers in the metadata header content of images. Although nearly all radiologic data use a universal format, DICOM, there are a growing number of exceptions, making it more difficult to standardize processes.

A balanced solution likely involves making information about AI systems and data collection understandable for patients, creating relationships of trust between institutions and their patients. At the same time, the aim should be to obtain more effective deidentification models that reduce identifiability as complete anonymization does not seem possibly obtainable in the near future. Current best practices for deidentification in radiology include avoiding the placement of identifiable data in proprietary DICOM fields, optimizing protocols of data management, using validated and tested protocols for deidentification, and investigating safer means of data sharing, such as containerization and blockchain [61]. There are calls for the formation of advisory committees to periodically review the protocols concerning privacy issues and identifiability in imaging [103]. As an example, skull-stripping—also known as brain extraction consisting of the act of removing non-brain signal from MRI data—is usually performed not only to remove redundant information, but especially to avoid facial reconstruction and identification.

6. Conclusions

Deep learning has provided reliable results for GBM assessment concerning MRI analysis and segmentation, including molecular, prognostic, and diagnostic information. We extensively reviewed the issues currently limiting its routine clinical application. Nevertheless, these limitations could be prospectively addressed in the near future. The most important matters concern the introduction and development of new technical advancements, an increased attention to data collection, and the careful addressing of healthcare ethical issues.

Author Contributions: Conceptualization, M.B. and A.B.; writing—original draft preparation, M.B. and L.F.R.; writing—review and editing, F.P., G.C. and A.B.; project administration, F.D.M., P.F. and D.G.; supervision, A.B. and F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ostrom, Q.T.; Gittleman, H.; Liao, P.; Vecchione-Koval, T.; Wolinsky, Y.; Kruchko, C.; Barnholtz-Sloan, J.S. CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010–2014. *Neuro-Oncology* **2017**, *19*, v1–v88. [CrossRef]
2. Martucci, M.; Russo, R.; Giordano, C.; Schiarelli, C.; D'apolito, G.; Tuzza, L.; Lisi, F.; Ferrara, G.; Schimperna, F.; Vassalli, S.; et al. Advanced Magnetic Resonance Imaging in the Evaluation of Treated Glioblastoma: A Pictorial Essay. *Cancers* **2023**, *15*, 3790. [CrossRef]
3. Bianconi, A.; Palmieri, G.; Aruta, G.; Monticelli, M.; Zeppa, P.; Tartara, F.; Melcarne, A.; Garbossa, D.; Cofano, F. Updates in Glioblastoma Immunotherapy: An Overview of the Current Clinical and Translational Scenario. *Biomedicines* **2023**, *11*, 1520. [CrossRef]
4. Morello, A.; Bianconi, A.; Rizzo, F.; Bellomo, J.; Meyer, A.C.; Garbossa, D.; Regli, L.; Cofano, F. Laser Interstitial Thermotherapy (LITT) in Recurrent Glioblastoma: What Window of Opportunity for This Treatment? *Technol. Cancer Res. Treat.* **2024**, *23*. [CrossRef] [PubMed]
5. Li, R.; Ye, J.; Huang, Y.; Jin, W.; Xu, P.; Guo, L. A continuous learning approach to brain tumor segmentation: Integrating multi-scale spatial distillation and pseudo-labeling strategies. *Front. Oncol.* **2024**, *13*, 1247603. [CrossRef] [PubMed]
6. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2016**, *278*, 563–577. [CrossRef]
7. Nakamori, S.; Bui, A.H.; Jang, J.; El-Rewaidy, H.A.; Kato, S.; Ngo, L.H.; Josephson, M.E.; Manning, W.J.; Nezafat, R. Increased myocardial native T_1 relaxation time in patients with nonischemic dilated cardiomyopathy with complex ventricular arrhythmia. *J. Magn. Reson. Imaging* **2017**, *47*, 779–786. [CrossRef]
8. Pandey, U.; Saini, J.; Kumar, M.; Gupta, R.; Ingalhalikar, M. Normative Baseline for Radiomics in BrainMRI: Evaluating the Robustness, Regional Variations, and Reproducibility onFLAIRImages. *J. Magn. Reson. Imaging* **2021**, *53*, 394–407. [CrossRef]
9. Li, Y.; Ammari, S.; Balleyguier, C.; Lassau, N.; Chouzenoux, E. Impact of Preprocessing and Harmonization Methods on the Removal of Scanner Effects in Brain MRI Radiomic Features. *Cancers* **2021**, *13*, 3000. [CrossRef]
10. Orlhac, F.; Lecler, A.; Savatovski, J.; Goya-Outi, J.; Nioche, C.; Charbonneau, F.; Ayache, N.; Frouin, F.; Duron, L.; Buvat, I. How Can We Combat Multicenter Variability in MR Radiomics? Validation of a Correction Procedure. *Eur. Radiol.* **2021**, *31*, 2272–2280. [CrossRef] [PubMed]
11. Fatania, K.; Mohamud, F.; Clark, A.; Nix, M.; Short, S.C.; O'Connor, J.; Scarsbrook, A.F.; Currie, S. Intensity Standardization of MRI Prior to Radiomic Feature Extraction for Artificial Intelligence Research in Glioma—A Systematic Review. *Eur. Radiol.* **2022**, *32*, 7014–7025. [CrossRef] [PubMed]
12. Da-Ano, R.; Visvikis, D.; Hatt, M. Harmonization strategies for multicenter radiomics investigations. *Phys. Med. Biol.* **2020**, *65*, 24TR02. [CrossRef]
13. Renard, F.; Guedria, S.; De Palma, N.; Vuillerme, N. Variability and reproducibility in deep learning for medical image segmentation. *Sci. Rep.* **2020**, *10*, 13724. [CrossRef] [PubMed]
14. Marzi, C.; Giannelli, M.; Barucci, A.; Tessa, C.; Mascalchi, M.; Diciotti, S. Efficacy of MRI data harmonization in the age of machine learning: A multicenter study across 36 datasets. *Sci. Data* **2024**, *11*, 115. [CrossRef]
15. Carré, A.; Klausner, G.; Edjlali, M.; Lerousseau, M.; Briend-Diop, J.; Sun, R.; Ammari, S.; Reuzé, S.; Alvarez Andres, E.; Estienne, T.; et al. Standardization of Brain MR Images across Machines and Protocols: Bridging the Gap for MRI-Based Radiomics. *Sci. Rep.* **2020**, *10*, 12340. [CrossRef] [PubMed]

16. Wu, J.; Guo, D.; Wang, L.; Yang, S.; Zheng, Y.; Shapey, J.; Vercauteren, T.; Bisdas, S.; Bradford, R.; Saeed, S.; et al. TISS-net: Brain tumor image synthesis and segmentation using cascaded dual-task networks and error-prediction consistency. *Neurocomputing* **2023**, *544*, 126295. [CrossRef]
17. Conte, G.M.; Weston, A.D.; Vogelsang, D.C.; Philbrick, K.A.; Cai, J.C.; Barbera, M.; Sanvito, F.; Lachance, D.H.; Jenkins, R.B.; Tobin, W.O.; et al. Generative Adversarial Networks to Synthesize Missing T1 and FLAIR MRI Sequences for Use in a Multisequence Brain Tumor Segmentation Model. *Radiology* **2021**, *299*, 313–323. [CrossRef] [PubMed]
18. Choi, Y.; Al-Masni, M.A.; Jung, K.-J.; Yoo, R.-E.; Lee, S.-Y.; Kim, D.-H. A single stage knowledge distillation network for brain tumor segmentation on limited MR image modalities. *Comput. Methods Programs Biomed.* **2023**, *240*, 107644. [CrossRef]
19. Cofano, F.; Bianconi, A.; De Marco, R.; Consoli, E.; Zeppa, P.; Bruno, F.; Pellerino, A.; Panico, F.; Salvati, L.F.; Rizzo, F.; et al. The Impact of Lateral Ventricular Opening in the Resection of Newly Diagnosed High-Grade Gliomas: A Single Center Experience. *Cancers* **2024**, *16*, 1574. [CrossRef]
20. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **2017**, *36*, 61–78. [CrossRef]
21. Zhou, C.; Ding, C.; Lu, Z.; Wang, X.; Tao, D. One-Pass Multi-Task Convolutional Neural Networks for Efficient Brain Tumor Segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2018, Granada, Spain, 16–20 September 2018; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Volume 11072, pp. 637–645. [CrossRef]
22. Zhou, T.; Ruan, S.; Canu, S. A review: Deep learning for medical image segmentation using multi-modality fusion. *Array* **2019**, *3–4*, 100004. [CrossRef]
23. Al-Masni, M.A.; Kim, D.-H. CMM-Net: Contextual multi-scale multi-level network for efficient biomedical image segmentation. *Sci. Rep.* **2021**, *11*, 10191. [CrossRef] [PubMed]
24. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. In Proceedings of the 4th International Conference on Learning Representations, ICLR 2016-Conference Track Proceedings, San Juan, Puerto Rico, 2–4 May 2016. [CrossRef]
25. Xu, C.; Xu, L.; Ohorodnyk, P.; Roth, M.; Chen, B.; Li, S. Contrast agent-free synthesis and segmentation of ischemic heart disease images using progressive sequential causal GANs. *Med. Image Anal.* **2020**, *62*, 101668. [CrossRef] [PubMed]
26. Wang, G.; Song, T.; Dong, Q.; Cui, M.; Huang, N.; Zhang, S. Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks. *Med. Image Anal.* **2020**, *65*, 101787. [CrossRef] [PubMed]
27. Osman, A.F.I.; Tamam, N.M. Deep learning-based convolutional neural network for intramodality brain MRI synthesis. *J. Appl. Clin. Med. Phys.* **2022**, *23*, e13530. [CrossRef]
28. Yang, F.; Dogan, N.; Stoyanova, R.; Ford, J.C. Evaluation of radiomic texture feature error due to MRI acquisition and reconstruction: A simulation study utilizing ground truth. *Phys. Medica* **2018**, *50*, 26–36. [CrossRef]
29. Erickson, B.J.; Cai, J. Magician's Corner: 5. Generative Adversarial Networks. *Radiol. Artif. Intell.* **2020**, *2*, e190215. [CrossRef] [PubMed]
30. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
31. Marcadent, S.; Hofmeister, J.; Preti, M.G.; Martin, S.P.; Van De Ville, D.; Montet, X. Generative Adversarial Networks Improve the Reproducibility and Discriminative Power of Radiomic Features. *Radiol. Artif. Intell.* **2020**, *2*, e190035. [CrossRef]
32. Sharma, A.; Hamarneh, G. Missing MRI Pulse Sequence Synthesis Using Multi-Modal Generative Adversarial Network. *IEEE Trans. Med. Imaging* **2019**, *39*, 1170–1183. [CrossRef]
33. Chen, C.; Dou, Q.; Jin, Y.; Liu, Q.; Heng, P.A. Learning With Privileged Multimodal Knowledge for Unimodal Segmentation. *IEEE Trans. Med. Imaging* **2021**, *41*, 621–632. [CrossRef] [PubMed]
34. Wang, L.; Yoon, K.-J. Knowledge Distillation and Student-Teacher Learning for Visual Intelligence: A Review and New Outlooks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3048–3068. [CrossRef]
35. Bhalerao, M.; Thakur, S. Brain Tumor Segmentation Based on 3D Residual U-Net. In Proceedings of the 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, 17 October 2019; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Volume 11993, pp. 218–225. [CrossRef]
36. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. NnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nat. Methods* **2021**, *18*, 203–211. [CrossRef]
37. Thakur, S.P.; Pati, S.; Panchumarthy, R.; Karkada, D.; Wu, J.; Kurtaev, D.; Sako, C.; Shah, P.; Bakas, S. Optimization of Deep Learning Based Brain Extraction in MRI for Low Resource Environments. In Proceedings of the 7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021, Virtual Event, 27 September 2021; pp. 151–167. [CrossRef]
38. Ito, Y.; Imai, H.; Le Duc, T.; Negishi, Y.; Kawachiya, K.; Matsumiya, R.; Endo, T. Profiling based out-of-core Hybrid method for large neural networks. In Proceedings of the PPoPP '19: 24th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, Washington, DC, USA, 16–20 February 2019; pp. 399–400. [CrossRef]
39. Huang, B.; Reichman, D.; Collins, L.M.; Bradbury, K.; Malof, J.M. Tiling and Stitching Segmentation Output for Remote Sensing: Basic Challenges and Recommendations. *arXiv* **2018**, arXiv:1805.12219. [CrossRef]

40. Pinckaers, H.; Litjens, G. Training Convolutional Neural Networks with Megapixel Images. *arXiv* **2018**, arXiv:1804.05712. [CrossRef]
41. Roth, H.R.; Oda, H.; Zhou, X.; Shimizu, N.; Yang, Y.; Hayashi, Y.; Oda, M.; Fujiwara, M.; Misawa, K.; Mori, K. An Application of Cascaded 3D Fully Convolutional Networks for Medical Image Segmentation. *Comput. Med. Imaging Graph.* **2018**, *66*, 90–99. [CrossRef] [PubMed]
42. Reina, G.A.; Panchumarthy, R.; Thakur, S.P.; Bastidas, A.; Bakas, S. Systematic Evaluation of Image Tiling Adverse Effects on Deep Learning Semantic Segmentation. *Front. Neurosci.* **2020**, *14*, 65. [CrossRef]
43. Shazeer, N.; Cheng, Y.; Parmar, N.; Tran, D.; Vaswani, A.; Koanantakool, P.; Hawkins, P.; Lee, H.J.; Hong, M.; Young, C.; et al. Mesh-TensorFlow: Deep Learning for Supercomputers. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 10414–10423. [CrossRef]
44. Sergeev, A.; Del Balso, M. Horovod: Fast and Easy Distributed Deep Learning in TensorFlow. *arXiv* **2018**, arXiv:1802.05799. [CrossRef]
45. Revesz, G.; Kundel, H.L.; Bonitatibus, M. The Effect of Verification on the Assessment of Imaging Techniques. *Investig. Radiol.* **1983**, *18*, 194–198. [CrossRef]
46. Bianconi, A.; Rossi, L.F.; Bonada, M.; Zeppa, P.; Nico, E.; De Marco, R.; Lacroce, P.; Cofano, F.; Bruno, F.; Morana, G.; et al. Deep learning-based algorithm for postoperative glioblastoma MRI segmentation: A promising new tool for tumor burden assessment. *Brain Inform.* **2023**, *10*, 26. [CrossRef] [PubMed]
47. Zając, H.D.; Avlona, N.R.; Kensing, F.; Andersen, T.O.; Shklovski, I. Ground Truth or Dare: Factors Affecting the Creation of Medical Datasets for Training AI. In Proceedings of the AIES '23: AAAI/ACM Conference on AI, Ethics, and Society, Montreal, QC, Canada, 8–10 August 2023; pp. 351–362. [CrossRef]
48. Sylolypavan, A.; Sleeman, D.; Wu, H.; Sim, M. The impact of inconsistent human annotations on AI driven clinical decision making. *npj Digit. Med.* **2023**, *6*, 26. [CrossRef]
49. Aganj, I.; Harisinghani, M.G.; Weissleder, R.; Fischl, B. Unsupervised Medical Image Segmentation Based on the Local Center of Mass. *Sci. Rep.* **2018**, *8*, 13012. [CrossRef]
50. Kiyasseh, D.; Cohen, A.; Jiang, C.; Altieri, N. A framework for evaluating clinical artificial intelligence systems without ground-truth annotations. *Nat. Commun.* **2024**, *15*, 1808. [CrossRef] [PubMed]
51. Liu, C.; Amodio, M.; Shen, L.L.; Gao, F.; Avesta, A.; Aneja, S.; Wang, J.C.; Del Priore, L.V.; Krishnaswamy, S. CUTS: A Deep Learning and Topological Framework for Multigranular Unsupervised Medical Image Segmentation. *arXiv* **2022**, arXiv:2209.11359. [CrossRef]
52. Tillmanns, N.; Lum, A.E.; Cassinelli, G.; Merkaj, S.; Verma, T.; Zeevi, T.; Staib, L.; Subramanian, H.; Bahar, R.C.; Brim, W.; et al. Identifying clinically applicable machine learning algorithms for glioma segmentation: Recent advances and discoveries. *Neuro-Oncol. Adv.* **2022**, *4*, vdac093. [CrossRef] [PubMed]
53. Jekel, L.; Brim, W.R.; von Reppert, M.; Staib, L.; Petersen, G.C.; Merkaj, S.; Subramanian, H.; Zeevi, T.; Payabvash, S.; Bousabarah, K.; et al. Machine Learning Applications for Differentiation of Glioma from Brain Metastasis—A Systematic Review. *Cancers* **2022**, *14*, 1369. [CrossRef]
54. Aboian, M.; Bousabarah, K.; Kazarian, E.; Zeevi, T.; Holler, W.; Merkaj, S.; Cassinelli Petersen, G.; Bahar, R.; Subramanian, H.; Sunku, P.; et al. Clinical implementation of artificial intelligence in neuroradiology with development of a novel workflow-efficient picture archiving and communication system-based automated brain tumor segmentation and radiomic feature extraction. *Front. Neurosci.* **2022**, *16*, 860208. [CrossRef]
55. Dieckhaus, H.B.; Meijboom, R.; Okar, S.; Wu, T.; Parvathaneni, P.; Mina, Y.; Chandran, S.; Waldman, A.D.; Reich, D.S.; Nair, G. Logistic Regression–Based Model Is More Efficient Than U-Net Model for Reliable Whole Brain Magnetic Resonance Imaging Segmentation. *Top. Magn. Reson. Imaging* **2022**, *31*, 31–39. [CrossRef]
56. Sheller, M.J.; Edwards, B.; Reina, G.A.; Martin, J.; Pati, S.; Kotrotsou, A.; Milchenko, M.; Xu, W.; Marcus, D.; Colen, R.R.; et al. Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. *Sci. Rep.* **2020**, *10*, 12598. [CrossRef]
57. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.S.; Freymann, J.B.; Farahani, K.; Davatzikos, C. Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **2017**, *4*, 170117. [CrossRef] [PubMed]
58. Anazodo, U.C.; Ng, J.J.; Ehiogu, B.; Obungoloch, J.; Fatade, A.; Mutsaerts, H.J.M.M.; Secca, M.F.; Diop, M.; Opadele, A.; Alexander, D.C.; et al. A framework for advancing sustainable magnetic resonance imaging access in Africa. *NMR Biomed.* **2022**, *36*, e4846. [CrossRef]
59. Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Shinohara, R.T.; Berger, C.; Ha, S.M.; Rozycki, M.; et al. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. *arXiv* **2018**, arXiv:1811.02629. [CrossRef]
60. Raghu; Sriraam, N.; Temel, Y.; Rao, S.V.; Kubben, P.L. A Convolutional Neural Network Based Framework for Classification of Seizure Types. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 2547–2550. [CrossRef]
61. Ibrahim, M.; Muhammad, Q.; Zamarud, A.; Eiman, H.; Fazal, F. Navigating Glioblastoma Diagnosis and Care: Transformative Pathway of Artificial Intelligence in Integrative Oncology. *Cureus* **2023**, *15*, e44214. [CrossRef] [PubMed]

62. Ermiş, E.; Jungo, A.; Poel, R.; Blatti-Moreno, M.; Meier, R.; Knecht, U.; Aebersold, D.M.; Fix, M.K.; Manser, P.; Reyes, M.; et al. Fully automated brain resection cavity delineation for radiation target volume definition in glioblastoma patients using deep learning. *Radiat. Oncol.* **2020**, *15*, 100. [CrossRef]
63. Lin, B.; Tan, Z.; Mo, Y.; Yang, X.; Liu, Y.; Xu, B.; Lin, B.; Tan, Z.; Mo, Y.; Yang, X.; et al. Intelligent oncology: The convergence of artificial intelligence and oncology. *J. Natl. Cancer Cent.* **2023**, *3*, 83–91. [CrossRef] [PubMed]
64. Adewole, M.; Rudie, J.D.; Gbadamosi, A.; Toyobo, O.; Raymond, C.; Zhang, D.; Omidiji, O.; Akinola, R.; Suwaid, M.A.; Emegoakor, A.; et al. The Brain Tumor Segmentation (BraTS) Challenge 2023: Glioma Segmentation in Sub-Saharan Africa Patient Population (BraTS-Africa). *arXiv* **2023**, arXiv:2305.19369v1. Available online: https://pubmed-ncbi-nlm-nih-gov.bibliopass.unito.it/37396608/ (accessed on 16 February 2024).
65. Lin, H.; Figini, M.; Tanno, R.; Blumberg, S.B.; Kaden, E.; Ogbole, G.; Brown, B.J.; D'Arco, F.; Carmichael, D.W.; Lagunju, I.; et al. Deep Learning for Low-Field to High-Field MR: Image Quality Transfer with Probabilistic Decimation Simulator. In Proceedings of the Machine Learning for Medical Image Reconstruction, Second International Workshop, MLMIR 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, 17 October 2019; pp. 58–70. [CrossRef]
66. Chang, K.; Beers, A.L.; Bai, H.X.; Brown, J.M.; Ina Ly, K.; Li, X.; Senders, J.T.; Kavouridis, V.K.; Boaro, A.; Su, C.; et al. Automatic Assessment of Glioma Burden: A Deep Learning Algorithm for Fully Automated Volumetric and Bidimensional Measurement. *Neuro-Oncology* **2019**, *21*, 1412–1422. [CrossRef] [PubMed]
67. Subramanian, H.; Dey, R.; Brim, W.R.; Tillmanns, N.; Cassinelli Petersen, G.; Brackett, A.; Mahajan, A.; Johnson, M.; Malhotra, A.; Aboian, M. Trends in Development of Novel Machine Learning Methods for the Identification of Gliomas in Datasets That Include Non-Glioma Images: A Systematic Review. *Front. Oncol.* **2021**, *11*, 788819. [CrossRef]
68. Zeppa, P.; Neitzert, L.; Mammi, M.; Monticelli, M.; Altieri, R.; Castaldo, M.; Cofano, F.; Borrè, A.; Zenga, F.; Melcarne, A.; et al. How Reliable Are Volumetric Techniques for High-Grade Gliomas? A Comparison Study of Different Available Tools. *Neurosurgery* **2020**, *87*, E672–E679. [CrossRef]
69. Kommers, I.; Bouget, D.; Pedersen, A.; Eijgelaar, R.S.; Ardon, H.; Barkhof, F.; Bello, L.; Berger, M.S.; Conti Nibali, M.; Furtner, J.; et al. Glioblastoma Surgery Imaging-Reporting and Data System: Standardized reporting of tumor volume, location, and resectability based on automated segmentations. *Cancers* **2021**, *13*, 2854. [CrossRef]
70. Visser, M.; Müller, D.; van Duijn, R.; Smits, M.; Verburg, N.; Hendriks, E.; Nabuurs, R.; Bot, J.; Eijgelaar, R.; Witte, M.; et al. Inter-rater agreement in glioma segmentations on longitudinal MRI. *NeuroImage Clin.* **2019**, *22*, 101727. [CrossRef]
71. Cordova, J.S.; Schreibmann, E.; Hadjipanayis, C.G.; Guo, Y.; Shu, H.-K.G.; Shim, H.; Holder, C.A. Quantitative Tumor Segmentation for Evaluation of Extent of Glioblastoma Resection to Facilitate Multisite Clinical Trials. *Transl. Oncol.* **2014**, *7*, 40–47, W1–W5. [CrossRef] [PubMed]
72. Miao, X.; Chen, H.; Tang, M.; Chen, Y. P13.01.A Post-Operative MRI Synthesis from Pre-Operative MRI and Post-Operative CT Using Conditional Gan for the Assessment of Degree of Resection. *Neuro Oncol.* **2023**, *25* (Suppl. 2), ii100. [CrossRef]
73. Bianconi, A.; Bonada, M.; Zeppa, P.; Colonna, S.; Tartara, F.; Melcarne, A.; Garbossa, D.; Cofano, F. How Reliable Is Fluorescence-Guided Surgery in Low-Grade Gliomas? A Systematic Review Concerning Different Fluorophores. *Cancers* **2023**, *15*, 4130. [CrossRef]
74. Ducray, F.; Idbaih, A.; Wang, X.-W.; Cheneau, C.; Labussiere, M.; Sanson, M. Predictive and prognostic factors for gliomas. *Expert Rev. Anticancer. Ther.* **2011**, *11*, 781–789. [CrossRef]
75. Saaid, A.; Monticelli, M.; Ricci, A.A.; Orlando, G.; Botta, C.; Zeppa, P.; Bianconi, A.; Osella-Abate, S.; Bruno, F.; Pellerino, A.; et al. Prognostic Analysis of the IDH1 G105G (rs11554137) SNP in IDH-Wildtype Glioblastoma. *Genes* **2022**, *13*, 1439. [CrossRef] [PubMed]
76. Bianconi, A.; Koumantakis, E.; Gatto, A.; Zeppa, P.; Saaid, A.; Nico, E.; Bruno, F.; Pellerino, A.; Rizzo, F.; Junemann, C.V.; et al. Effects of Levetiracetam and Lacosamide on survival and seizure control in IDH-wild type glioblastoma during temozolomide plus radiation adjuvant therapy. *Brain Spine* **2024**, *4*, 102732. [CrossRef]
77. Kickingereder, P.; Sahm, F.; Radbruch, A.; Wick, W.; Heiland, S.; Von Deimling, A.; Bendszus, M.; Wiestler, B. IDH Mutation Status Is Associated with a Distinct Hypoxia/Angiogenesis Transcriptome Signature Which Is Non-Invasively Predictable with RCBV Imaging in Human Glioma. *Sci. Rep.* **2015**, *5*, 16238. [CrossRef]
78. Law, M.; Young, R.J.; Babb, J.S.; Peccerelli, N.; Chheang, S.; Gruber, M.L.; Miller, D.C.; Golfinos, J.G.; Zagzag, D.; Johnson, G. Gliomas: Predicting Time to Progression or Survival with Cerebral Blood Volume Measurements at Dynamic Susceptibility-weighted Contrast-enhanced Perfusion MR Imaging. *Radiology* **2008**, *247*, 490–498. [CrossRef]
79. Beiko, J.; Suki, D.; Hess, K.R.; Fox, B.D.; Cheung, V.; Cabral, M.; Shonka, N.; Gilbert, M.R.; Sawaya, R.; Prabhu, S.S.; et al. IDH1 Mutant Malignant Astrocytomas Are More Amenable to Surgical Resection and Have a Survival Benefit Associated with Maximal Surgical Resection. *Neuro-Oncology* **2013**, *16*, 81–91. [CrossRef] [PubMed]
80. Liang, S.; Zhang, R.; Liang, D.; Song, T.; Ai, T.; Xia, C.; Xia, L.; Wang, Y. Multimodal 3D DenseNet for *IDH* Genotype Prediction in Gliomas. *Genes* **2018**, *9*, 382. [CrossRef] [PubMed]
81. Zlochower, A.; Chow, D.S.; Chang, P.; Khatri, D.; Boockvar, J.A.; Filippi, C.G. Deep Learning AI Applications in the Imaging of Glioma. *Top. Magn. Reson. Imaging* **2020**, *29*, 115–121. [CrossRef] [PubMed]
82. Hegi, M.E.; Diserens, A.-C.; Gorlia, T.; Hamou, M.-F.; de Tribolet, N.; Weller, M.; Kros, J.M.; Hainfellner, J.A.; Mason, W.; Mariani, L.; et al. MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. *N. Engl. J. Med.* **2005**, *352*, 997–1003. [CrossRef] [PubMed]

83. Bianconi, A.; Prior, A.; Zona, G.; Fiaschi, P. Anticoagulant therapy in high grade gliomas: A systematic review on state of the art and future perspectives. *J. Neurosurg. Sci.* **2023**, *67*, 236–240. [CrossRef] [PubMed]
84. Han, L.; Kamdar, M.R. MRI to MGMT: Predicting Methylation Status in Glioblastoma Patients Using Convolutional Recurrent Neural Networks. *Pac. Symp. Biocomput.* **2018**, *23*, 331–342. [CrossRef] [PubMed]
85. Korfiatis, P.; Kline, T.L.; Lachance, D.H.; Parney, I.F.; Buckner, J.C.; Erickson, B.J. Residual Deep Convolutional Neural Network Predicts MGMT Methylation Status. *J. Digit. Imaging* **2017**, *30*, 622–628. [CrossRef] [PubMed]
86. Chang, P.; Grinband, J.; Weinberg, B.D.; Bardis, M.; Khy, M.; Cadena, G.; Su, M.-Y.; Cha, S.; Filippi, C.G.; Bota, D.; et al. Deep-Learning Convolutional Neural Networks Accurately Classify Genetic Mutations in Gliomas. *Am. J. Neuroradiol.* **2018**, *39*, 1201–1207. [CrossRef]
87. Li, Z.; Wang, Y.; Yu, J.; Guo, Y.; Cao, W. Deep Learning Based Radiomics (DLR) and Its Usage in Noninvasive IDH1 Prediction for Low Grade Glioma. *Sci. Rep.* **2017**, *7*, 5467. [CrossRef]
88. Kahn, C.E. Artificial Intelligence, Real Radiology. *Radiol. Artif. Intell.* **2019**, *1*, e184001. [CrossRef]
89. Bossuyt, P.M.; Reitsma, J.B.; Bruns, D.E.; Gatsonis, C.A.; Glasziou, P.P.; Irwig, L.; Lijmer, J.G.; Moher, D.; Rennie, D.; de Vet, H.C.; et al. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Radiology* **2015**, *277*, 826–832. [CrossRef]
90. Lambin, P.; Leijenaar, R.T.H.; Deist, T.M.; Peerlings, J.; de Jong, E.E.C.; van Timmeren, J.; Sanduleanu, S.; Larue, R.T.H.M.; Even, A.J.G.; Jochems, A.; et al. Radiomics: The Bridge between Medical Imaging and Personalized Medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749–762. [CrossRef] [PubMed]
91. Cohen, J.F.; Korevaar, D.A.; Altman, D.G.; Bruns, D.E.; Gatsonis, C.A.; Hooft, L.; Irwig, L.; Levine, D.; Reitsma, J.B.; de Vet, H.C.W.; et al. STARD 2015 Guidelines for Reporting Diagnostic Accuracy Studies: Explanation and Elaboration. *BMJ Open* **2016**, *6*, e012799. [CrossRef] [PubMed]
92. Bossuyt, P.M.; Reitsma, J.B. The STARD Initiative. *Lancet* **2003**, *361*, 71. [CrossRef] [PubMed]
93. Vandenbroucke, J.P.; von Elm, E.; Altman, D.G.; Gøtzsche, P.C.; Mulrow, C.D.; Pocock, S.J.; Poole, C.; Schlesselman, J.J.; Egger, M. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLoS Med.* **2007**, *4*, e297. [CrossRef] [PubMed]
94. Schulz, K.F.; Altman, U.G.; Moher, D.; CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ* **2010**, *340*, c332. [CrossRef] [PubMed]
95. Begg, C. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* **1996**, *276*, 637–639. [CrossRef] [PubMed]
96. Mongan, J.; Moy, L.; Kahn, C.E. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiol. Artif. Intell.* **2020**, *2*, e200029. [CrossRef]
97. Reyes, M.; Meier, R.; Pereira, S.; Silva, C.A.; Dahlweid, F.M.; von Tengg-Kobligk, H.; Summers, R.M.; Wiest, R. On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. *Radiol. Artif. Intell.* **2020**, *2*, e190043. [CrossRef]
98. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359. [CrossRef]
99. Kim, B.; Wattenberg, M.; Gilmer, J.; Cai, C.; Wexler, J.; Viegas, F.; Sayres, R. Interpretability beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV). In Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholm, Sweden, 10–15 July 2018; Volume 80, pp. 4186–4195.
100. Chaddad, A.; Peng, J.; Xu, J.; Bouridane, A. Survey of Explainable AI Techniques in Healthcare. *Sensors* **2023**, *23*, 634. [CrossRef] [PubMed]
101. Band, S.S.; Yarahmadi, A.; Hsu, C.-C.; Biyari, M.; Sookhak, M.; Ameri, R.; Dehzangi, I.; Chronopoulos, A.T.; Liang, H.-W. Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods. *Inform. Med. Unlocked* **2023**, *40*, 101286. [CrossRef]
102. Geis, J.R.; Brady, A.P.; Wu, C.C.; Spencer, J.; Ranschaert, E.; Jaremko, J.L.; Langer, S.G.; Kitts, A.B.; Birch, J.; Shields, W.F.; et al. Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement. *J. Am. Coll. Radiol.* **2019**, *16*, 1516–1521. [CrossRef] [PubMed]
103. Lotan, E.; Tschider, C.; Sodickson, D.K.; Caplan, A.L.; Bruno, M.; Zhang, B.; Lui, Y.W. Medical Imaging and Privacy in the Era of Artificial Intelligence: Myth, Fallacy, and the Future. *J. Am. Coll. Radiol.* **2020**, *17*, 1159–1162. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.