Journal of *Personalized Medicine*

MDPI

Communication

Explainable Artificial Intelligence for Human-Machine Interaction in Brain Tumor Localization

Morteza Esmaeili [1,2,*], Riyas Vettukattil [3,4], Hasan Banitalebi [1,3], Nina R. Krogh [1] and Jonn Terje Geitung [1,3]

[1] Department of Diagnostic Imaging, Akershus University Hospital, 1478 Lørenskog, Norway; hasan.banitalebi@ahus.no (H.B.); nina.rolland.krogh@ahus.no (N.R.K.); j.t.geitung@medisin.uio.no (J.T.G.)
[2] Department of Electrical Engineering and Computer Science, Faculty of Science and Technology, University of Stavanger, 4021 Stavanger, Norway
[3] Faculty of Medicine, Institute of Clinical Medicine, University of Oslo, 0315 Oslo, Norway; dr.riyas@gmail.com
[4] Division of Paediatric and Adolescent Medicine, Oslo University Hospital, 0372 Oslo, Norway
* Correspondence: mor.esmaeili@gmail.com

Abstract: Primary malignancies in adult brains are globally fatal. Computer vision, especially recent developments in artificial intelligence (AI), have created opportunities to automatically characterize and diagnose tumor lesions in the brain. AI approaches have provided scores of unprecedented accuracy in different image analysis tasks, including differentiating tumor-containing brains from healthy brains. AI models, however, perform as a black box, concealing the rational interpretations that are an essential step towards translating AI imaging tools into clinical routine. An explainable AI approach aims to visualize the high-level features of trained models or integrate into the training process. This study aims to evaluate the performance of selected deep-learning algorithms on localizing tumor lesions and distinguishing the lesion from healthy regions in magnetic resonance imaging contrasts. Despite a significant correlation between classification and lesion localization accuracy ($R = 0.46$, $p = 0.005$), the known AI algorithms, examined in this study, classify some tumor brains based on other non-relevant features. The results suggest that explainable AI approaches can develop an intuition for model interpretability and may play an important role in the performance evaluation of deep learning models. Developing explainable AI approaches will be an essential tool to improve human–machine interactions and assist in the selection of optimal training methods.

Keywords: tumor localization; black box CNN; explainable AI; gliomas; machine learning

check for updates

Citation: Esmaeili, M.; Vettukattil, R.; Banitalebi, H.; Krogh, N.R.; Geitung, J.T. Explainable Artificial Intelligence for Human-Machine Interaction in Brain Tumor Localization. *J. Pers. Med.* **2021**, *11*, 1213. https://doi.org/10.3390/jpm11111213

Academic Editors: Yoshihiro Noda and Niels Bergsland

Received: 27 September 2021
Accepted: 12 November 2021
Published: 16 November 2021

1. Introduction

Artificial intelligence (AI) developments have created opportunities for human life in a wide range of industries, business, education, and healthcare [1,2]. As a part of AI, deep-learning-derived approaches provide convenient autonomous image classification in the medical domain [1]. Traditional modeling techniques such as linear regression and decision trees provide an understandable relationship between input data and the decisions in the model outputs [2]. These models are often called white-box models but are usually not as performant as black-box models such as convolutional neural networks (CNN), complicated ensembles, and other deep learning models. The latest models provide excellent accuracy at the expense of model explainability. Explainable models estimate the importance of each feature on the model predictions, providing interpretable tools for understanding deep learning outcomes [3]. Explainable AI is essential in tackling biases in AI-based decisions. Bias in models can originate long before the training and testing stages [3]. The data used for model training can be entangled in their own set of biases. Therefore, identifying and handling potential biases in datasets is a crucial component of any responsible AI strategy. AI-derived training should aim to build trustworthy, transparent, and bias-free models.

Magnetic resonance imaging (MRI) serves as a gold standard and is the method of choice for the in vivo investigation of most brain diseases, such as different types of brain tumors. In brain tumor examinations, several radiological characteristics, such as morphology, tumor volume and location, composition, and enhancement, can be derived to narrow the differential diagnosis and guide patient management. The most common MRI modalities in brain malignancy examinations include T1-weighted (T1w), T2-weighted (T2w), and Fluid Attenuation Inversion Recovery (FLAIR). These MR modalities provide distinctive imaging contrast to improve lesion identification and pathological delineations.

Deep learning methods have demonstrated unprecedented sensitivity and accuracy in tumor segmentation and malignancy classifications [4]. Several studies have explored explainable AI approaches to visualize the learning evolutions at each neuron layer [5,6]. The methods provide saliency maps to identify the contribution of pixels/features in the learning process. Explainable AI may improve the training performance of deep learning models by intermediate step monitoring at the inter-neuron layers [5,6]. Post hoc explainable evaluations of AI outcomes may profoundly contribute to the understanding of deep learning predictions and eventually enhance their transparency, reliability, and interpretability, and human–machine interactions [7,8]. Furthermore, post hoc methods may contribute to clinical neurophysiology and neuroimaging data analysis, such as the localization of brain sub-regions [9–11].

This study evaluates the high-level features of deep convolutional neural networks, predicting tumor lesion locations in the brain. We used explainable saliency maps to investigate the performance of three established densely deep learning models in accurately identifying tumor lesions.

2. Materials and Methods

The experiments include the TCGA dataset [12,13] retrieved from The Cancer Imaging Archive repositories [13–16]. The dataset consists of lower-grade gliomas and the most aggressive malignancy, glioblastoma (WHO grade IV). With the inclusion criteria of the availability of T2-weighted and FLAIR magnetic resonance (MR) images, we gathered MR data from 354 subjects. We analyzed the axial slices of the T2w images following preprocessing steps: (i) N4BiasCorrection to remove RF inhomogeneity, (ii) intensity normalization, (iii) affine registration to MNI space, and (iv) resizing all images to the resolution of $256 \times 256 \times 128$ using ANTs scripts [16]. A total number of 19,200 and 14,800 slices of brain images with and without tumor lesions, respectively, were prepared for training.

We trained different AI networks with three-fold cross-validation by randomly shuffling imaging data for training, validation, and testing patches. Three identical groups with dataset distributions of 55%, 15%, and 30% for training, validation, and testing, respectively, were generated. Each fold of cross-validation and testing was a new training phase based on an identical combination of the three groups. We held out the testing fraction of the dataset from the in-training step. Thus, the accuracy of the models was calculated using the mean value of the network performance on only the testing dataset.

Selecting our choices from the most deployed AI models in the imaging domain, we included the AI networks DenseNet-121 [17], GoogLeNet [18] (also known as Inception V1, available in Github-GoogLeNet in Keras), MobileNet [19] (retrieved directly from Keras-https://keras.io/api/applications/mobilenet/ (accessed on 11 June 2020)). We incorporated the data from T2w and FLAIR sequences as separate channel inputs to each model, similar to the process of handling RGB inputs. Thus, each input slice contained combined MR images of T2w and FLAIR. Each model had a 2-class output consisting of healthy and tumor. All models employed Adam optimizer with a learning rate of $5e$–4, with a decay rate of 0.9. The batch size and number of epochs were 25 and 100, respectively. All experiments were computed on a computer with two Intel Corei7 CPU, Nvidia GeForce Quadro RTX GPU, and 32 GB of memory. We implemented the Grad-CAM algorithm to visualize each model's performance on tumor lesion localization [6]. The Grad-CAM generates visual explanations of post-processed image space gradients using saliency heatmaps. The normalized heatmap (SanityMap ϵ [0 1]) scales the contribution of each pixel in the learning process, ranging from none ("0") to the most significant ("1") pixels in the 2-dimensional image. This study determined the most important pixels with a cuff-off value greater than "0.5" in the SanityMap. The overlap of the generated heatmaps ([SanityMap] $\geq$ 0.5) and the tumor lesions in one testing group was used to estimate the localization performance as Equation (1):

$$Localization_{hit}(\%) = \frac{hits}{total\ testing\ images} \tag{1}$$

where *hits* was determined as a successful overlap of >50% of pixels of the tumor mask with the Grad-CAM-derived heatmap ($SanityMap_{0.5} \cap Tumor_{mask} > 50\%$). We also calculated the Intersection over Union (IoU) metric as Equation (2) (the Jaccard index) for localization accuracy. IoU is a popular metric for segmentation and object detection [20].

$$IoU(\%) = \frac{SanityMap_{0.5} \cap \ Tumor_{mask}}{SanityMap_{0.5} \ \cup \ Tumor_{mask}} \tag{2}$$

Statistics: The mean difference between the computed localization accuracy of each model was compared using the Mann–Whitney test. Spearman correlation analysis was calculated to investigate the relationships between the model's prediction accuracy and tumor localization. The threshold for statistical significance was defined as $p < 0.05$.

3. Results

The Grad-CAM method provided representative heatmaps, which can be used to quantitively and qualitatively investigate the performance of convolutional neural networks. Examples of heatmaps from four subjects are shown in Figure 1. The DenseNet-121, GoogLeNet, MobileNet achieved a mean cross-validated prediction accuracy of 92.1, 87.3, and 88.9, respectively, on the testing dataset (Table 1). DenseNet-121 provided a significantly higher mean localization accuracy of *hit* = 81.1% and IoU = 79.1% than GoogLeNet (p = 0.01) and MobileNet (p = 0.02) (Table 1). The correlation analysis showed a significant agreement between the models' prediction accuracy and specificity in localizing tumor lesions (R = 0.46, p = 0.005).

Table 1. Mean classification and localization accuracy on the testing database for DenseNet, GoogLeNet, and MobileNet.

Model	Classification Accuracy (%)	Localization	
		Hits (%)	IoU (%)
DenseNet-121	92.1	81.1	79.1
GoogLeNet	87.3	73.7	73.8
MobileNet	88.9	77.8	76.7

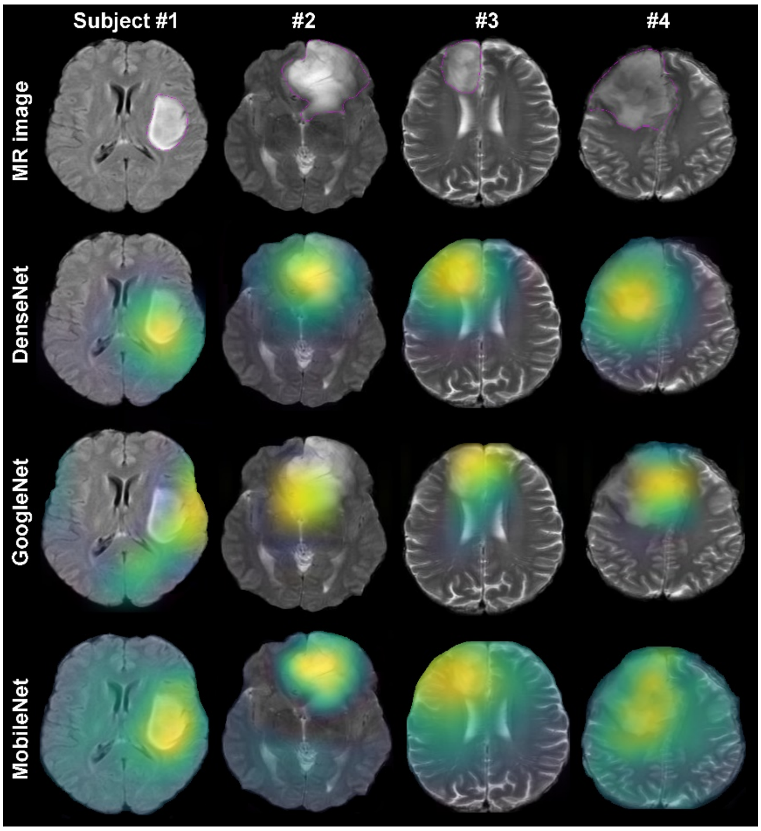

Figure 1. Heatmap visualization overlaid on tumor lesions for different training networks. The top row depicts the original MR image examples from four subjects. The magenta counters indicate the tumor lesion boundaries. The bottom rows show the Grad-CAM visualizations for three different training algorithms on the selected axial slices.

4. Discussion

Our study demonstrated that the explainable Grad-CAM method could visualize the network's performance, distinguishing the images with and without tumor based on the lesion's localization rather than other features in the brain. However, the results showed that a considerable number of tumor brains were classified by false-positive features, identifying other important components in the brain than the tumor lesion. The results indicated that, among the implemented three models, the DenseNet-121 performed better than the other two models, providing more accurate tumor localization (with ~80% hits and IoU). Similarly, previous studies in brain tumor segmentation and classification showed the state-of-the-art performance of the DenseNet [17,21–23], with an accuracy greater than 90%.

Grad-CAM provides post hoc feature attribution heatmaps, aiming to explain the relationship between the models' predictions and in terms of its features. Operationally, the method employs gradients as a measure of importance in the feature space and may be used to explore significant features in the training phase, identifying data skew, or debugging model performance. Several explanatory CAM approaches have recently been proposed in the medical domain [24]. For example, Vinogradova and colleagues [25], produced visual explanations for semantic segmentation by extending Grad-CAM. The authors examined their method on the performance of a U-Net model in feature extraction of Cityscapes dataset and showed that the initial convolutional layers exhibit low-level edge-like feature extractions. Grad-CAM based approaches were also used to visualize 2-dimensional [6,26] and 3-dimensional brain tumor segmentation [27].

Already available in the Keras method libraries, Grad-CAM methods have demonstrated great capabilities in image region discrimination in various clinical and computer vision studies [25–28]. Despite the utility of Grad-CAM, there are some limitations associated with gradient-based explanation and Grad-CAM estimations, especially when targeting multiple objects in an image. Several studies have addressed the gradient-based visualization issues and suggested an extended version of CAM approaches, such as pyramid gradient-based class activation map [29] and high-resolution CAM, to target multiple organs in the computed tomography (CT) scan database [30]. However, Grad-CAM provided acceptable accuracy for our study because we aimed to target only one feature, i.e., tumor lesion, on 2-dimensional MR images.

T2-weighted and FLAIR sequences are MRI acquisition protocols that are recommended by neuro-radiologists for use in brain tumor clinical investigations [31]. The combination of T2w and FLAIR contrast effectively contributed to the performance of deep learning models, as cerebrospinal fluid can be distinguished from edema in these images [32]. The consensus recommendations for standardized brain tumor studies [31] for clinical trials have included functional imaging data acquisitions such as diffusion-weighted imaging. Diffusion imaging provides an apparent diffusion coefficient (ADC) metric, which can be included as an extra layer for deep learning analysis. Since ADC is a sensitive diagnostic metric providing cellular density maps [31], it would be interesting to investigate its contribution to future deep learning training protocols for classification or segmentation tasks.

Despite the close accuracy scores (greater than 90% accuracy) on the ImageNet database achieved by GoogLeNet, MobileNet, DenseNet, and several other complex models, explainable algorithms have shown faulty reasoning and feature extractions [5,6]. In other words, the machine interprets the image features differently, beyond human interpretation and recognition. In cancer clinics, the interpretation of radiological images plays a crucial role in clinical decisions. Thus, explainable AI methods can increase the translational potential of the developed complex models into clinics. The visual explanation of explainable AI provides deep learning algorithms' reasoning and learning process in initiative ways that reflect the model's final output. The visual interpretation will eventually help the investigators and clinicians to examine the true-positive and false-positive outputs achieved by CNN and debug the training pipeline. A platform with an interaction between the end-user and the CNN-based black-box can relate the trained model prototyping to what humans describe as logical components in classifying images. Such a platform generates a prototypical part network to interpret trained models in classification tasks [5].

Visual explanation techniques may potentially contribute to debugging model performance, selecting training strategy or architecture, and training workflow [3]. Wang and colleagues [33] incorporated explainable methods during training. The authors assess the utility of implementing features to increase explainability early in the development process. Chen and colleagues [5] showed that integrating interpretable methods can increase up to 3.5% in the accuracy score of trained models. Zhang and colleagues [34] integrated a Grad-CAM derived heatmap into their deep learning models to develop an effective strategy for identifying the brain areas underlying disease development in multiple sclerosis. A future work could be the inclusion of explainable AI approaches in the training phase to enhance both model classification and localization accuracy.

We performed this study on limited patient data, and thus trained the CNN models on slice-level. In future studies, with access to more patient data, we aim to examine our approach at the patient-level and perform a 3D analysis of tumor localization. Additionally, a prospective investigation may include clinical covariants, such as tumor volume, position, grade, and age, in the deep learning analysis. Indeed, a large range of patient data from multiple centers integrated with novel transfer learning approaches will be essential for such investigation.

5. Conclusions

In summary, we used a well-known explainable AI algorithm to evaluate the performance of three deep learning methods in localizing tumor tissues in the brain. Our results show that the incorporation of explainable AI in deep learning workflows plays an essential role in human–machine communication and may assist in the selection of an optimal training scheme for clinical questions and the AI learning progress.

Author Contributions: Conceptualization, M.E. and J.T.G.; methodology, M.E. and R.V.; validation, M.E. and R.V.; formal analysis, M.E.; writing—original draft preparation, M.E.; writing—review and editing, M.E., R.V., H.B., N.R.K. and J.T.G.; visualization, M.E.; supervision, M.E.; project administration, M.E.; funding acquisition, M.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by HELSE SØR-ØST RHF–Southern Eastern Norway Regional Health Authority, grant number 201804.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: We used publicly available, open-source data, collected and published in accordance with all applicable rules [refs. [13–16]].

Data Availability Statement: The TCGA data are available for research purposes (https://www.cancer.gov/tcga (accessed on 6 February 2020)). The deep learning codes are available on the TensorFlow database (accessed on 11 June 2020).

Acknowledgments: The results shown here are in whole or part based upon data generated by TCGA Research Network: http://cancergenome.nih.gov/ (accessed on 6 February 2020). Specifically, data from TCIA Collections **TCGA-LGG** and **TCGA-GBM**, as they appear within https://www.cancerimagingarchive.net (accessed on 6 February 2020). Authors recognize the contribution of all individuals and research groups involved in TCAI and TCGA studies.

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

References

1. Battineni, G.; Sagaro, G.G.; Chinatalapudi, N.; Amenta, F. Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis. *J. Pers. Med.* **2020**, *10*, 21. [CrossRef]
2. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [CrossRef]
3. Antoniadi, A.; Du, Y.; Guendouz, Y.; Wei, L.; Mazo, C.; Becker, B.; Mooney, C. Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. *Appl. Sci.* **2021**, *11*, 5088. [CrossRef]
4. Muhammad, K.; Khan, S.; Del Ser, J.; de Albuquerque, V.H.C. Deep Learning for Multigrade Brain Tumor Classification in Smart Healthcare Systems: A Prospective Survey. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 507–522. [CrossRef] [PubMed]
5. Chen, C.; Li, O.; Tao, C.; Barnett, A.J.; Su, J.; Rudin, C. This Looks Like That: Deep Learning for Interpretable Image Recognition. *arXiv* **2018**, arXiv:1806.10574.
6. 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.* **2019**, *128*, 336–359. [CrossRef]
7. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* **2020**, *23*, 18. [CrossRef]
8. Knapič, S.; Malhi, A.; Saluja, R.; Främling, K. Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 740–770. [CrossRef]
9. Fellous, J.-M.; Sapiro, G.; Rossi, A.; Mayberg, H.; Ferrante, M. Explainable Artificial Intelligence for Neuroscience: Behavioral Neurostimulation. *Front. Neurosci.* **2019**, *13*, 1346. [CrossRef] [PubMed]
10. Jo, T.; Nho, K.; Saykin, A.J. Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Front. Aging Neurosci.* **2019**, *11*, 220. [CrossRef]
11. Yang, C.; Rangarajan, A.; Ranka, S. Visual Explanations From Deep 3D Convolutional Neural Networks for Alzheimer's Disease Classification. *AMIA Annu. Symp. Proc.* **2018**, *2018*, 1571–1580.
12. Cancer Genome Atlas Research Network. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* **2015**, *372*, 2481–2498. [CrossRef] [PubMed]
13. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [CrossRef] [PubMed]
14. Pedano, N.; Flanders, A.E.; Scarpace, L.; Mikkelsen, T.; Eschbacher, J.M.; Hermes, B.; Sisneros, V.; Barnholtz-Sloan, J.; Ostrom, Q. *Radiology Data from the Cancer Genome Atlas Low Grade Glioma [TCGA-LGG] Collection*; The Cancer Imaging Archive, 2016.

Available online: https://wiki.cancerimagingarchive.net/display/Public/TCGA-LGG#530918864c2b0756f974ab5b574ca3888851202 (accessed on 6 February 2020). [CrossRef]

15. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.; Freymann, J.; Farahani, K.; Davatzikos, C. *Segmentation Labels for the Pre-Operative Scans of the TCGA-GBM Collection [Data Set]*; The Cancer Imaging Archive, 2017; Available online: https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=24282666#242826662c5ce8901dc84f4393fdccced7375a3c (accessed on 6 February 2020). [CrossRef]
16. Scarpace, L.; Mikkelsen, T.; Cha, S.; Rao, S.; Tekchandani, S.; Gutman, D.; Saltz, J.H.; Erickson, B.J.; Pedano, N.; Flanders, A.; et al. *Radiology Data from The Cancer Genome Atlas Glioblastoma Multiforme [TCGA-GBM] Collection [Data Set]*; Archive, T.C.I., Ed.; The Cancer Imaging Archive, 2016. Available online: https://wiki.cancerimagingarchive.net/display/Public/TCGA-GBM#1966258715bed1a14224923b50f1f2e7dae54a1 (accessed on 11 June 2020).
17. Esmaeili, M.; Stensjøen, A.L.; Berntsen, E.M.; Solheim, O.; Reinertsen, I. The Direction of Tumour Growth in Glioblastoma Patients. *Sci. Rep.* **2018**, *8*, 1199. [CrossRef] [PubMed]
18. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* **2016**, arXiv:1608.06993.
19. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. *arXiv* **2014**, arXiv:1409.4842.
20. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
21. Ayachi, R.; Said, Y.; Atri, M. A Convolutional Neural Network to Perform Object Detection and Identification in Visual Large-Scale Data. *Big Data* **2021**, *9*, 41–52. [CrossRef]
22. Fu, X.; Chen, C.; Li, D. Survival prediction of patients suffering from glioblastoma based on two-branch DenseNet using multi-channel features. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 207–217. [CrossRef]
23. Aldoj, N.; Biavati, F.; Michallek, F.; Stober, S.; Dewey, M. Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-net. *Sci. Rep.* **2020**, *10*, 14315. [CrossRef] [PubMed]
24. Tao, Z.; Bingqiang, H.; Huiling, L.; Zaoli, Y.; Hongbin, S. NSCR-Based DenseNet for Lung Tumor Recognition Using Chest CT Image. *BioMed Res. Int.* **2020**, *2020*, 1–9. [CrossRef] [PubMed]
25. Narayanan, M.; Chen, E.; He, J.; Kim, B.; Gershman, S.; Doshi-Velez, F. How do Humans Understand Explanations from Machine Learning Systems? An Evaluation of the Human-Interpretability of Explanation. *arXiv* **2018**, arXiv:1802.00682.
26. Vinogradova, K.; Dibrov, A.; Myers, G. Towards Interpretable Semantic Segmentation via Gradient-Weighted Class Activation Mapping (Student Abstract). *Proc. Conf. AAAI Artif. Intell.* **2020**, *34*, 13943–13944. [CrossRef]
27. Saleem, H.; Shahid, A.R.; Raza, B. Visual interpretability in 3D brain tumor segmentation network. *Comput. Biol. Med.* **2021**, *133*, 104410. [CrossRef]
28. Fernández, I.S.; Yang, E.; Calvachi, P.; Amengual-Gual, M.; Wu, J.Y.; Krueger, D.; Northrup, H.; Bebin, M.E.; Sahin, M.; Yu, K.-H.; et al. Deep learning in rare disease. Detection of tubers in tuberous sclerosis complex. *PLoS ONE* **2020**, *15*, e0232376. [CrossRef] [PubMed]
29. Lee, S.; Lee, J.; Lee, J.; Park, C.K.; Yoon, S. Robust Tumor Localization with Pyramid Grad-CAM. *arXiv* **2018**, arXiv:1805.11393.
30. Draelos, R.L.; Carin, L. HiResCAM: Faithful Location Representation in Visual Attention for Explainable 3D Medical Image Classification. *arXiv* **2021**, arXiv:2011.08891.
31. Ellingson, B.M.; Bendszus, M.; Boxerman, J.; Barboriak, D.; Erickson, B.J.; Smits, M.; Nelson, S.J.; Gerstner, E.; Alexander, B.; Goldmacher, G.; et al. Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. *Neuro-oncology* **2015**, *17*, 1188–1198.
32. Pereira, S.; Meier, R.; McKinley, R.; Wiest, R.; Alves, V.; Silva, C.A.; Reyes, M. Enhancing interpretability of automatically extracted machine learning features: Application to a RBM-Random Forest system on brain lesion segmentation. *Med. Image Anal.* **2018**, *44*, 228–244. [CrossRef]
33. Wang, C.J.; Hamm, C.A.; Savic, L.J.; Ferrante, M.; Schobert, I.; Schlachter, T.; Lin, M.; Weinreb, J.C.; Duncan, J.S.; Chapiro, J.; et al. Deep learning for liver tumor diagnosis part II: Convolutional neural network interpretation using radiologic imaging features. *Eur. Radiol.* **2019**, *29*, 3348–3357. [CrossRef]
34. Zhang, Y.; Hong, D.; McClement, D.; Oladosu, O.; Pridham, G.; Slaney, G. Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging. *J. Neurosci. Methods* **2021**, *353*, 109098. [CrossRef] [PubMed]