2022 2nd International Conference on Advanced Research in Computing (ICARC) | 978-1-6654-0741-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICARC54489.2022.9754131

Interpretable Machine Learning for Brain Tumor Analysis Using MRI

Sasmitha Dasanayaka
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
sasmithanilupul.17@cse.mrt.ac.lk

Sanju Silva
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
sjanupa.17@cse.mrt.ac.lk

Vimuth Shantha
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
vimuth.17@cse.mrt.ac.lk

Dulani Meedeniya
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
dulanim@cse.mrt.ac.lk

Thanuja Ambegoda
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
thanuja@cse.mrt.ac.lk

***Abstract*—A brain tumor is a potentially fatal growth of cells in the central nervous system that can be categorized as benign or malignant. Advancements in deep learning in the recent past and the availability of high computational power have been influencing the automation of diagnosing brain tumors. DenseNet and U-Net are considered state of the art deep learning models for classification and segmentation of MRIs respectively. Despite the progress of deep learning in diagnosing using medical images, generic convolutional neural networks are still not fully adopted in clinical settings as they lack robustness and reliability. Moreover, such black-box models don't offer a human interpretable justification as to why certain classification decisions are made, which makes them less preferable for medical diagnostics. Brain tumor segmentation and classification using deep learning techniques has been a popular research area in the last few decades but still, there are only a few models that are interpretable. In this paper, we have proposed an interpretable deep learning model which is more human understandable than existing black-box models, designed based on U-Net and DenseNet to segment and classify brain tumors using MRI. In our proposed model, we generate a heat map highlighting the contribution of each region of the input to the classification output and have validated the system using the MICCAI 2020 Brain Tumor dataset.**

***Index Terms*—Interpretable machine learning, Brain tumor, Segmentation, Classification, MRI**

I. Introduction

A brain tumor is a collection of abnormal cells growing inside the brain. A Brain tumor is mainly classified as Benign or Malignant. Benign brain tumors are less harmful because they do not invade or spread to other nearby tissues. While on the contrary malignant brain tumors are more invasive and spread to other areas of the brain and spine as well. Brain tumors are the most common central nervous system (CNS) related cancer which account for 85%-90% of all the CNS tumors [1]. Still, medical experts have not been able to discover exact causes for brain tumors or unique symptoms related to brain tumors. The survival rate of patients with a malignant form of brain tumors like Glioblastoma is as low as 6%-9% [2]. Inability to diagnose and identify their types at early stages further reduces the survival rate of brain tumor patients considerably. Currently, there are several brain tumor diagnosis techniques like EEG, CT scan, PET scan. Among them, Magnetic Resonance Imaging (MRI) is the most common diagnostic process followed by medical officers to identify brain tumors. However, the MRI technique generates hundreds of images making the diagnostic process exhaustive which hinders early diagnosis and treatments for tumors [3]. Under existing procedures, medical practitioners have to go through highly time-consuming routine tasks dealing with a large number of images to identify and classify a brain tumor. Making this process more efficient can increase the survivability of patients. The real-world medical diagnostic procedure is to first confirm the existence of a tumor and identify it using MRIs. If the existence of a tumor is confirmed, further analysis is carried out on MRIs and also a biopsy if required, to classify the tumor.

With the advancement of AI and Deep learning in the field of medical imaging, image processing techniques such as image vision, classification, segmentation help radiologists to diagnose diseases as early as possible compared to manual inspections [4]. Since technology is intervening in the medical field, the technological solutions must be reliable and efficient as it keeps human life on the line which mistakes can bring major consequences. Currently, there are many systems based on black-box deep neural networks that aim to improve the accuracy of their predictions [5], [6]. When it comes to high stake fields like medical diagnosis it is not all about accuracy. The system has to make sense at least for domain experts.

978-1-6654-0741-0/22/$31.00 ©2022 IEEE

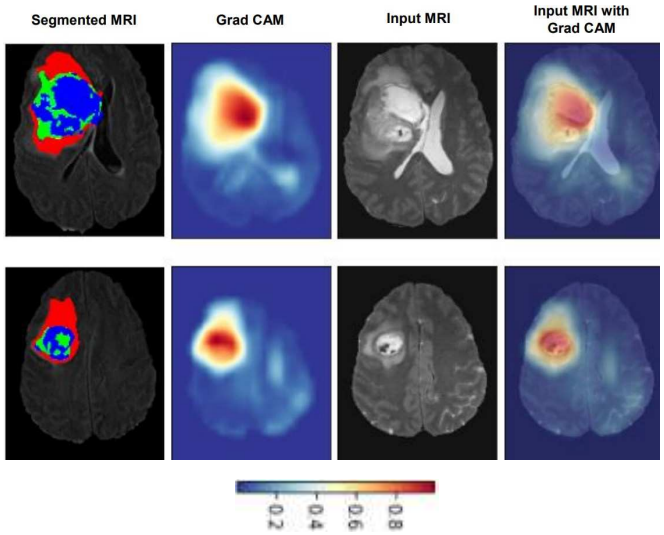

Fig. 1. Heat maps generated at two different depths for test case-I where each column shows segmented results, resulted heat map, original input image, and heat maps are superimposed on top of the original image respectively. Color range at the bottom convince the contributing level of each region where dark blue for less contributing regions while dark red for highly contributing regions for classification.

Hence, in this paper, we have proposed an interpretable model to ensure the final system is more reliable compared to existing models and can be applied in the real world. What we mean by interpretability is the model must be understandable to humans unlike generic black-box models and the decisions made inside the neural network must be transparent. Furthermore, using a segmentation module we make the classification process understandable even for a non-expert in the medical domain. As in Fig. 1 if we output the contribution level of each region of the input image towards the final predicted output of a model, then it is a way of interpreting the decision making process of the model.

Our contribution resides on achieving the decision making process of deep learning models more transparent such that the models become more reliable. To enable this transparency, heat maps are generated automatically specifying the contribution level of each region of the inputs towards the decision making process of the model. Thus, we have provided an interpretable solution to classify brain tumors into 3 malignant tumor classes Glioblastoma, Oligodendroglioma, Astrocytoma, and segment brain tumors using MRI.

II. Related Work

Encoder with decoder CNN architectures such as U-Net [7], 3D U-Net [8] are producing state-of-art results for medical image segmentation as they overcome the gradient vanishing problem through skip connections. The final output of these types of encoder-decoder architectures is a semantic segmented image [9]. Studies have addressed brain tumour identification in different viewpoints such as using radiogenomic [10] and mRNA expressions to predict the survival and risk estimation [11]. Since our model segments out the brain tumor from MRIs, we have utilized VAE 3D U-Net architecture [12] which is an encoder-decoder architecture with a variational auto-encoder branch, and also this architecture won the BraTS 2018 challenge. Variational auto-encoder branch reconstructs the input image itself to regularize the shared decoder and impose additional constraints on its layers.

Classifying brain tumors using only MRI is a challenging task. Generally, both MRI and pathology images are used for better results as in [13], [14]. Since we are focusing on providing an interpretable solution in this paper, we have used only MRI for classification. However, the proposed method can be transferred to interpret classification using pathology images also. Pei *et al.* [15] have suggested a deep convolutional neural network to classify a segmented tumor to achieve a balanced accuracy of 0.749 and F1-score of 0.829. Despite having a segmentation module for completeness of the diagnosis system we have not used segmentation results for classification because the complete brain image is more meaningful when interpreting the classification predictions. In [14], the authors have achieved a kappa coefficient of 0.748 and F1-score of 0.829 for MRI classification using a 3D DenseNet121 model. Both [14], [15] are trained on the CPM-RadPath-2019 dataset which is similar to the dataset [16] used in this paper in size, dimensions, class labels, and class distribution. Most of the existing literature [5], [6], [14], [15] are solely intended to enhance the performance of predictions. Performance focused model development is encouraged by popular competitions irrespective of the ability to apply in the real world. As state-of-art models are reaching peak performance for classification, further research must be carried out to improve applicability of the models.

Neural networks are inherently black boxes in nature. Thus, most of the existing models based on deep neural networks [13], [14], [15] are also less human-understandable. So, in our model, we have overcome this weakness of deep neural networks by introducing interpretable machine learning to the classification module. Gradient-weighted Class Activation Mapping (Grad-CAM) is an interpretable technique proposed in [17] for making CNN-based models more explainable and transparent. The technique is based on gradients of the most dominant logit of the result related to feature maps of the final convolutional layer which eventually produces a localization map highlighting the regions of the input image concerning the importance level of the region for predicted output. Grad CAM has addressed the drawbacks of the most commonly used interpretable technique, Class Activation Maps [18] of having a specific architecture and the need of re-training the model. The generated Grad CAM can be placed as an overlay on top of the input image (like a heat map).

III. Dataset

We have used two separate publicly available datasets for each task segmentation and classification.

A. Dataset for Segmentation

The dataset [19], is annotated in medical image datasets for the development and evaluation of segmentation models. The dataset includes data related to different body organs out of which we have selected the brain tumor dataset. This dataset includes MRI scans stored in the NifTI-1 format, related to patients diagnosed with either glioblastoma or lower-grade glioma. The MRI sequence includes, Fluid Attenuated Inversion Recovery (FLAIR), T1-weighted (T1w), T1-weighted with gadolinium contrast enhancement (T1gd) and T2-weighted (T2w).

The dataset includes 484 images which we split into 400 for training and 84 for validation. The labeled segmented images have four sub-regions of the tumor with integer values 0, 1, 2, and 3 representing background, edema, non-enhancing tumor and enhancing tumor respectively.

B. Dataset for Classification

We have used the dataset published for the MICCAI 2020 Combined Radiology and Pathology Classification Challenge [20], available at [16] for classification. This dataset contains images of 221 cases (70% - training dataset, 20%- validation dataset, 10% - test dataset) with 3D radiology images in NIfTI(.nii) format. The dataset consists of pre-processed images of 3 classes of brain tumors;

1) Lower-grade astrocytoma (A)
2) Oligodendroglioma (O)
3) Glioblastoma and Diffuse astrocytic glioma (G)

MRIs were multi-parametric, which means they have multiple modalities. In this dataset, there were 4 image modalities for each T1-weighted MRI (T1), T1-weighted MRI with contrast enhancement (T1ce), T2-weighted MRI (T2) and T2-weighted MRI with fluid-attenuated inversion recovery (T2-Flair).

IV. Methodology

In this section, we have focused on our approach to design a complete model to replicate the medical procedure to diagnose a brain tumor and introduce interpretability to the model. The design of the whole system will be discussed in a modular approach to get a better insight.

A. Preprocessing

Preprocessing of data can be explained under the following criteria.

1) Segmentation: The MRI dataset for segmentation was already skull-scripted. The original dimensions H×WxDxC of input images are changed into CxH×WxD and we carried out standard normalization depth-wise to reduce noise. We needed to feed a single sequence of MRI (FLAIR, T1w, T1gd, T2w modalities acquired for a patient) in an iteration to U-Net such that the model can extract features from all MRI volumes in the sequence. Since the size of a volume is 240x240x155, we had to crop the volume such that the size is reduced to 160x192x128 to match memory limits. Then this preprocessed dataset is used for segmentation.

2) Classification: The selected MRI dataset for this module was also skull-scripted and denoised. We carried out standard normalization to reduce intensity in-homogeneity. The background was cropped with an intensity level tolerance threshold of 10^{-8}, then MRI volumes were downsampled to 120x148x120. The training set was extended using the following augmentation techniques; random affine, random horizontal flip, random vertical flip, and random elastic deformation.

B. Proposed Solution

The proposed solution in this paper has two main modules for semantic segmentation and classification of brain tumors. The high-level architecture diagram of the complete system is shown in Fig. 2. Grad CAMs are used in the classification module to make the predictions more interpretable.

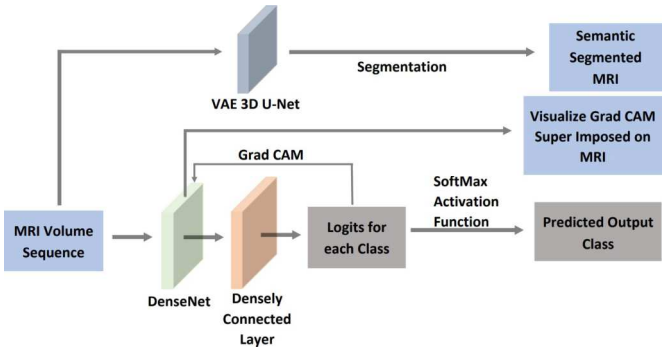

Fig. 2. High-level architecture diagram of the complete solution - The model takes MRI volumes as input and generates three main outputs; segmentation predictions, classification predictions, and Grad CAMs to visualize classification decision making process.

1) Segmentation Module: Our segmentation approach is based on [12] which is an encoder-decoder architecture with three modules - encoder, decoder and variational auto-encoder (VAE) branch. We input a sequence of cropped 3D MRI volumes (4x160x192x128) to the encoder module and finally the model produces three channels (3x160x192x128) of semantic segmented 3D MRI volume with the same spatial size as the input volume for each brain tumor sub-region (here we have considered edema, non-enhancing tumor and enhancing tumor sub-regions without considering the tumor background as a sub-region since we can get the tumor background through the combination of edema, non-enhancing tumor and enhancing tumor sub-regions). We have used a batch size of 1 as specified in [12].

The encoder module is comprised of multiple ResNet blocks [21] where each block consists of two 3D convolutions with a kernel size of 3 followed by a group normalization layer and ReLU layer. At each level of the encoder module, the image dimensions are down-sized by 2 using stridden convolutions instead of using traditional pooling and the feature size is increased by 2.

The decoder module is also structured similar to the encoder module with the same number of spatial levels where each

level includes only one 3x3x3 convolution. Each decoder block begins with reducing the feature size by 2 (using 3D convolution with kernel size 3) and upsizing the spatial dimensions after concatenating the output from the respective equivalent encoder level. At the end of the decoder module, we get an output volume with the same volumetric dimensions as the input to the encoder module. This output is sent via 1x1x1 convolution and a sigmoid function to get 3 channels (because the tumor has 3 sub-regions).

The VAE branch is used only in the training phase to regularize the encoder. As proposed in [12], we were able to achieve better training accuracy by implementing a VAE branch other than without it.

2) MRI Classification Module: Input MRI volumes have dimensions of 120x148x120 after downsampling and each of the four modalities was considered as separate channels. DenseNet [22] architecture was used to implement the classifier because of its ability to perform well with a lower number of parameters compared to other architectures in classification tasks. We have used a DenseNet-BC model with 169 layers and a growth rate of 32. The original DenseNet-BC model was modified to accommodate 3D convolution, 3D pooling, and 3D batch normalization to accord with 4D input volumes. Learning rate and weight decay were set to 10^{-4} and 10^{-5} respectively. The batch size was set to 4 due to limited GPU memory. A dropout layer was added to every transition block and dense block with a drop rate of 0.2.

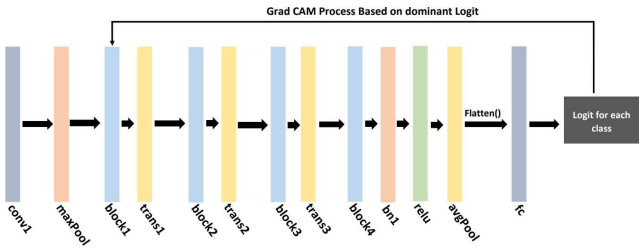

Fig. 3. Classification module has conv (convolution), maxpool (max pooling), block (dense block), trans (transition layer), bn (batch normalization), avgpool (average pooling), fc (fully connected) layers connected sequentially.

Interpretable Approach: We have utilized the Grad CAM technique which finds the gradient of the most dominant logit concerning the feature maps of the final convolution layer in the trained model as the interpretable approach of the proposed solution. So, we have used M3D-CAM library [23] for generating Grad CAMs for classified MRI volumes. The generated Grad CAMS are resized to match with the input volume dimensions (120x148x120). Then Grad CAM slices (like a heat map) of size 148x120 are superimposed on corresponding input slices of the same size as in Fig. 4 which results in 120 superimposed 2D images at different depths. The output heat map depicts which regions of the input MRI slice were informative to predict the corresponding dominant logit by the trained model.

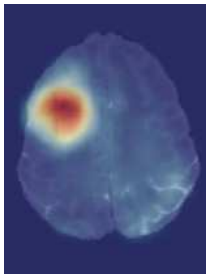

Fig. 4. Grad CAM is superimposed on an MRI slice like a heat map. This interprets that the highlighted areas are highly contributed to give the final output of the model.

V. Results & Discussion

A. Segmentation

After training for 300 epochs as stated in paper [12], the segmented volumes we got as the output were 160x192x128 for each tumor sub-region which is a cropped part of the original MRI volume (240x240x155). Then taking exact locations of voxels in output concerning input volume, we have created predicted labels overlaid over FLAIR sequence on axial, sagittal, and coronal slices. Fig. 5 illustrates a predicted labeled brain tumor from our model with respect to its true labels. The tumor is labeled as edema (red), non-enhancing tumor (green) and enhancing tumor (blue).

The combination of all sub-regions (union of 3 colors) is the whole tumor (WT), combinations of enhancing and non-enhancing regions (union of blue and green) is the tumor core (TC). Here we can see the predicted output is well matching with ground truth.

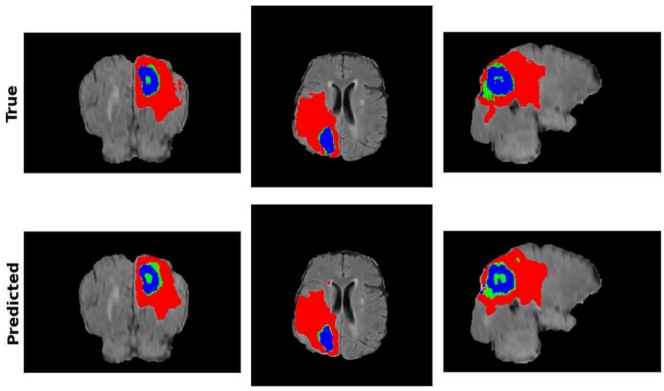

Fig. 5. Segmented MRI for a test image compared with its ground truth label, each image is a cross-section generated at the middle of each axis of MRI volume. Tumors are labeled as edema - red, non-enhancing tumor - green, enhancing tumor - blue.

Dice coefficients of WT, TC, and enhancing tumor (ET) which we used as an evaluation metric of our model are as follows compared with the respected results in the original VAE U-Net in Table I. We were able to achieve interesting results like the original model but using a different dataset could be the reason for a slight difference in values.

TABLE I
SEGMENTATION RESULTS (DICE COEFFICIENTS) ON TEST SET OF OUR MODEL COMPARED WITH ORIGINAL VAE U-NET

	ET	WT	TC
Proposed model	0.7787	0.8849	0.8041
Original VAE U-Net	0.8145	0.9042	0.8596

B. Classification

After completing 20 epochs we were able to achieve the best results for the test dataset. The confusion matrix Fig. 6 and values for performance metrics are shown in Table II. We have used balanced accuracy, f1-score, and Cohen's kappa coefficient to estimate the performance of our classification model to cater for the imbalanced nature of the dataset used which consisted of 133 cases of glioblastoma (G), 54 cases of astrocytoma (A), and 34 cases of oligodendroglioma (O). One major objective of the classification model is to train the hidden layers to an acceptable level that we can accommodate class activation maps to interpret classification predictions.

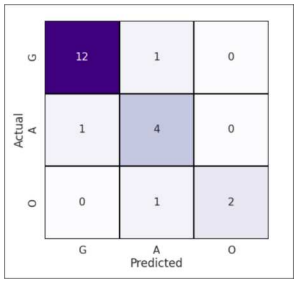

Fig. 6. Confusion matrix for MRI classification for test set consist of 21 cases: G (glioblastoma), A (astrocytoma), O (oligodendroglioma).

TABLE II
CLASSIFICATION RESULTS ON TEST SET

	Class G	Class A	Class O	Micro average
Balanced accuracy	0.899	0.838	0.833	0.893
F1-score	0.923	0.727	0.8	0.857
Kappa coefficient	-	-	-	0.733

C. Interpretable ML for Classification

We used the same test set used in Section V-B and generated heat maps for each test case, from our trained DenseNet model. To confirm the position of brain tumors and check how precise the Grad CAMs are, we have segmented the same images using our trained U-Net model then compared segmented output with heat maps generated from Grad CAM. Although Grad CAM generally considers gradients of the most dominant logit with respect to the feature maps of the final convolution layer, we had to use other layers. Because the Grad CAMs generated using the final convolution layer were too small in dimensions (3x4x3). Hence when we resized the maps back to original input volume dimensions (120x148x120) did not generate acceptable results. Therefore, we tried on different layers and found block1 layer (Fig. 3) as the best layer to generate Grad CAMs.

Fig. 1 and Fig. 7 show segmented results, Grad CAM results, an original input image, Grad CAM superimposed on top of the original image for 2 test cases. Each of these images shows a comparison of the 4 above-mentioned views at two different depths (in each row) of the input volume. The color range of the Grad CAM is from dark blue to dark red as in the bottom of Fig. 1 where dark blue for fewer contributing regions while dark red for highly contributing regions for classification.

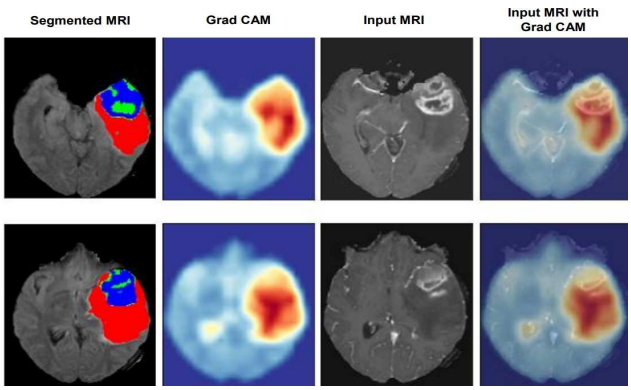

Fig. 7. Grad CAMs were generated at two different depths for test case II where each column shows segmented results, Grad CAM results, an original input image, Grad CAMs superimposed on top of the original image respectively.

By looking at Fig. 1 and Fig. 7, we can see which regions of the MRI were highly contributed and slightly contributed to give the predicted classification output. So, the model is more human-understandable such that anyone can understand which regions of the MRI were used by the model for classification. With the corresponding segmented MRI for each case, we can be confident about Grad CAM heat maps. The generated heat maps can also be used as an alternative for segmentation whenever pixel-wise correctness is not required. Labelling MRI datasets and training models for brain tumor segmentation are highly computationally expensive tasks. Therefore, using heat maps as an alternative is a trade-off between accuracy and resource requirement.

Another interesting observation we came across in our study is, for certain depths (topmost and bottom-most) of MRI volume where tumor does not exist, Grad CAM indicates the whole brain is equally important for predictions as shown in the first and third rows of Fig. 8.

VI. CONCLUSION

In this paper, we have proposed an interpretable machine learning pipeline to diagnose brain tumor using MRI inspired by the real-world diagnosis procedures. Interpretability is critical for deep learning solutions applied in high-stake domains. Despite the large number of papers published in the domain of brain tumor classification and segmentation using ML, we

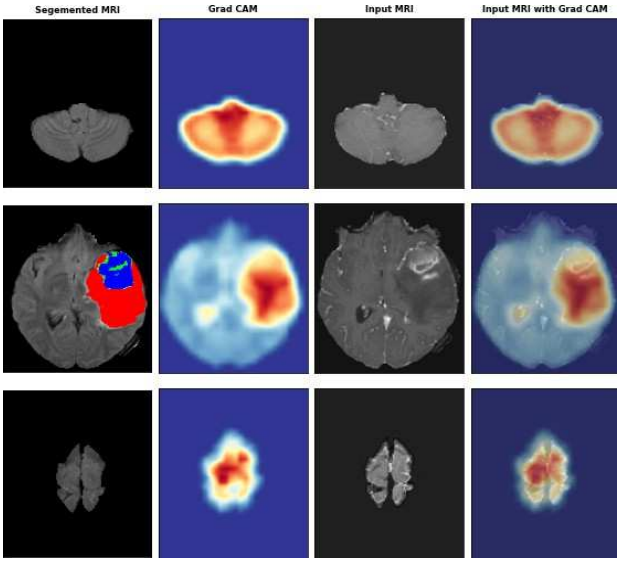

Fig. 8. Grad CAM of regions generated at different depths of same MRI volume where tumor exists (row 2) and does not exist from bottom view (row 1) and from top view (row 3) of brain.

did not find much research work on applying interpretable ML. The proposed solution combines a U-Net model for segmentation and a DenseNet model for classification. The main focus of this study is to apply interpretable machine learning to understand how state-of-art models produce results and ensure that the produced results are based on important features that define the target. The ability to explain decision-making mechanisms can enhance the potential to apply existing systems in real-world scenarios. We have applied Grad CAM at different levels of DenseNet and different depths of MRI volumes. During this study, we observed that deeper layers of DenseNet cannot be used to generate Grad CAMs because generated Grad CAMs at those layers are too small to map back to the input size. We also observed how the classification model reacts for parts of the input image where the tumor does not exist. In conclusion, we were able to interpret the contribution of each region of the input towards the predicted output of DenseNet which is a state-of-art classification model.

REFERENCES

[1] "Brain tumor - statistics," 2021. [Online]. Available: https://www.cancer.net/cancer-types/brain-tumor/statistics

[2] "Brain tumor survival rate," Moffitt Cancer Center, 2021. [Online]. Available: https://moffitt.org/cancers/brain-cancer/survival-rate/

[3] K. Nael, E. Gibson, C. Yang, P. Ceccaldi, Y. Yoo, J. Das, A. Doshi, B. Georgescu, N. Janardhanan, B. Odry *et al.*, "Automated Detection of Critical Findings in Multi-Parametric Brain MRI using A System of 3D Neural Networks," *Scientific reports*, vol. 11, no. 1, pp. 1–10, 2021, doi: 10.1038/s41598-021-86022-7.

[4] N. Wijethilake, D. Meedeniya, C. Chitraranjan, I. Perera, M. Islam, and H. Ren, "Glioma Survival Analysis Empowered With Data Engineering—A Survey," *IEEE Access*, vol. 9, pp. 43 168–43 191, 2021, doi: 10.1109/ACCESS.2021.3065965.

[5] J. Seetha and S. S. Raja, "Brain Tumor Classification using Convolutional Neural Networks," *Biomedical & Pharmacology Journal*, vol. 11, no. 3, p. 1457, 2018, doi: 10.13005/bpj/1511.

[6] N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, "Brain Tumor Classification using Convolutional Neural Network," in *World Congress on Medical Physics and Biomedical Engineering*. Springer, 2019, pp. 183–189, doi: 10.1007/978-981-10-9035-6_33.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[8] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2016, pp. 424–432, doi: 10.1007/978-3-319-46723-8_49.

[9] I. Rubasinghe and D. Meedeniya, "Ultrasound Nerve Segmentation Using Deep Probabilistic Programming," *Journal of ICT Research and Applications*, vol. 13, pp. 241–256, 2019, doi: 10.5614/itbj.ict.res.appl.2019.13.3.5.

[10] N. Wijethilake, M. Islam, D. Meedeniya, C. Chitraranjan, I. Perera, and H. Ren, "Radiogenomics of glioblastoma: Identification of radiomics associated with molecular subtypes," in *2nd MICCAI Workshop on Radiomics and Radiogenomics in Neuro-oncology Using AI (MLCN), RNO-AI* , USA, 2020, pp. 229–239.

[11] N. Wijethilake, D. Meedeniya, C. Chitraranjan, and I. Perera, "Survival Prediction and Risk Estimation of Glioma Patients Using mRNA Expressions," in *Proc. of the IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, Cincinnati, USA, 2020, pp. 35–42, doi: 10.1109/BIBE50027.2020.00014.

[12] A. Myronenko, "3D MRI Brain Tumor Segmentation using Autoencoder Regularization," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer, 2019, pp. 311–320, doi: 10.1007/978-3-030-11726-9_28.

[13] M. Lerousseau, E. Deutsh, and N. Paragios, "Multimodal Brain Tumor Classification," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Rimi and S. Bakas, Eds. Cham: Springer International Publishing, 2021, pp. 475–486, doi: 10.1007/978-3-030-72087-2_42.

[14] X. Ma and F. Jia, "Brain Tumor Classification with Multimodal MR and Pathology Images," in *International MICCAI Brainlesion Workshop*. Springer, 2019, pp. 343–352, doi: 10.1007/978-3-030-46643-5_34.

[15] L. Pei, L. Vidyaratne, W.-W. Hsu, M. M. Rahman, and K. M. Iftekharuddin, "Brain Tumor Classification Using 3D Convolutional Neural Network," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi and S. Bakas, Eds. Cham: Springer International Publishing, 2020, pp. 335–342, doi: 10.1007/978-3-030-46643-5_33.

[16] "Miccai 2020 cpm-radpath," 2020. [Online]. Available: http://miccai2020-data.eastus.cloudapp.azure.com/

[17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.

[18] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2921–2929, doi: 10.1109/CVPR.2016.319.

[19] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze *et al.*, "A Large Annotated Medical Image Dataset for The Development and Evaluation of Segmentation Algorithms," *arXiv preprint arXiv:1902.09063*, 2019.

[20] "Codalab-competition," 2020. [Online]. Available: https://miccai.westus2.cloudapp.azure.com/competitions/1

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[23] K. Gotkowski, C. Gonzalez, A. Bucher, and A. Mukhopadhyay, "M3d-CAM," in *Bildverarbeitung für die Medizin 2021*, C. Palm, T. M. Deserno, H. Handels, A. Maier, K. Maier-Hein, and T. Tolxdorff, Eds. Wiesbaden: Springer Fachmedien Wiesbaden, 2021, pp. 217–222, doi: 10.1007/978-3-658-33198-6_52.