



Original software publication

# Interpretable machine learning for brain tumour analysis using MRI and whole slide images

Sasmitha Dasanayaka, Vimuth Shantha, Sanju Silva, Dulani Meedeniya \*, Thanuja Ambegoda

Department of Computer Science and Engineering, University of Moratuwa, Moratuwa, Sri Lanka



## ARTICLE INFO

### Keywords:

Deep learning  
Interpretability  
Brain tumour  
Classification  
Grad-CAM  
Multimodal

## ABSTRACT

Tumour-Analyser is a web application that classifies a brain tumour into three classes, namely, lower-grade astrocytoma (A), oligodendroglioma (O), glioblastoma & diffuse astrocytic glioma (G). We use a magnetic resonance imaging (MRI) sequence and a whole slide imaging (WSI) that are classified using DenseNet and ResNet, respectively. The tool interprets the decision-making process of each classification model. Tumour-Analyser provides a viable solution to the less human understandability of existing models due to the inherent black-box nature of deep learning models and less transparency, by applying interpretability.

## Code metadata

Current code version	v1.0
Permanent link to code/repository used for this code version	<a href="https://github.com/SoftwareImpacts/SIMPAC-2022-81">https://github.com/SoftwareImpacts/SIMPAC-2022-81</a>
Permanent link to reproducible capsule	
Legal code license	GNU General Public License (GPL)
Code versioning system used	Git
Software code languages, tools and services used	python, PyTorch, Flask, Google Colab, Google Cloud Platform
Compilation requirements, operating environments and dependencies	Not available, because all the requirements are fulfilled in the docker image.
If available, link to developer documentation/manual	<a href="https://drive.google.com/file/d/1qNFXyBUizSNvGLwvlsZhwuYP0AIO8igY/view?usp=sharing">https://drive.google.com/file/d/1qNFXyBUizSNvGLwvlsZhwuYP0AIO8igY/view?usp=sharing</a>
Support email for questions	<a href="mailto:teamcuresaders.17@gmail.com">teamcuresaders.17@gmail.com</a>

## 1. Introduction

Brain tumour is an abnormal mass of cells spreading within the brain. With the assistance of current biomedical solutions, medical experts are using brain tumours diagnosis strategies such as magnetic resonance imaging (MRI), electroencephalogram (EEG), positron emission tomography (PET) scan and computerized tomography (CT) scan. However, these strategies produce many images obstructing and delaying the analysis and diagnostic process of brain tumours. This directly affects the survival rate of brain tumour patients since it hinders the diagnosis process. The typical real-world practice of the diagnostic mechanism of brain tumours is to use MRI and histopathology imaging modalities. First, the tumour is diagnosed using the MRI, and then it is confirmed and classified using the histopathology images. This traditional real-world practice is a highly time-consuming routine.

With the immense development of artificial intelligence and deep learning within the discipline of the medical domain, image processing techniques assist medical experts in diagnosing diseases as early as viable as compared to human-guided inspections [1]. These solutions automatically detect unusual symptoms of different diseases and support the manual process to increase effectiveness and efficiency. Since technology is stepping into the medical field, technological solutions must be reliable and efficient as they are tightly coupled with human life, which cannot be paid off for any mistake. Thus, there is a need for an automated system to support the medical image diagnosis process.

Although there are many existing solutions based on deep learning techniques, the majority of them are black boxes. Thus, they are less understandable to humans. In addition, most of the existing systems focus on improving the prediction accuracy [2–5], irrespective of human understandability. In high-stake fields such as the medical

\* Corresponding author.

E-mail addresses: [sasmithanilupul.17@cse.mrt.ac.lk](mailto:sasmithanilupul.17@cse.mrt.ac.lk) (S. Dasanayaka), [vimuth.17@cse.mrt.ac.lk](mailto:vimuth.17@cse.mrt.ac.lk) (V. Shantha), [sjanupa.17@cse.mrt.ac.lk](mailto:sjanupa.17@cse.mrt.ac.lk) (S. Silva), [dulanim@cse.mrt.ac.lk](mailto:dulanim@cse.mrt.ac.lk) (D. Meedeniya), [thanuja@cse.mrt.ac.lk](mailto:thanuja@cse.mrt.ac.lk) (T. Ambegoda).

<https://doi.org/10.1016/j.simpa.2022.100340>

Received 13 June 2022; Accepted 23 June 2022

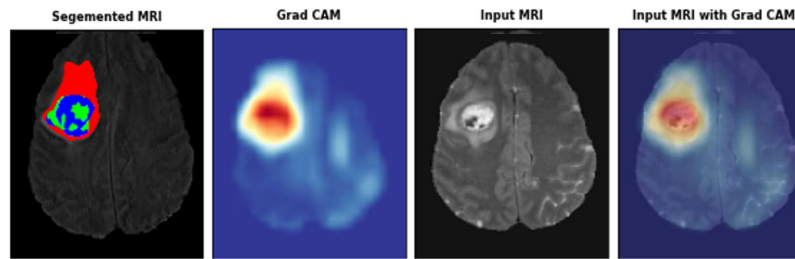


Fig. 1. Heatmap generated at a particular depth of an MRI.

domain, other than the accuracy, the model should be interpretable for the domain experts. This study presents the Tumour-Analyser web application using interpretable deep learning [6], to address the issues in model understandability.

## 2. System model

The Tumour-Analyser tool interprets the decision-making process of the considered deep learning models. Our model uses magnetic resonance imaging (MRI) and whole slide image (WSI) modalities to classify brain tumours. In this application, we replicate the real-world practice of the diagnostic mechanism of brain tumours using bi-modal. A user can input an MRI sequence to the application and get the classified class of the tumour with a heatmap, that indicates the contributing level of each of the regions of the input MRI towards the decision-making process of the MRI classification model and the segmented tumour image. Fig. 1 illustrates the output for a single input MRI with segmented tumour, GradCAM heatmap, input MRI and heatmap overlaid on the input image. Users can use the segmented tumour image to check whether the highly contributing regions of the classification model are residing inside the tumour region and self-evaluate the classification model. Thus, the internal process of the model becomes transparent and human-understandable.

Moreover, the users can input the WSI images to confirm the correctness of the classification of MRIs. At the same time, The Tumour-Analyser tool produces a heatmap denoting the contributing level of each input WSI region for the prediction. This approach interprets the WSI classification module. Accordingly, we achieved the interpretability in the presented solution. Fig. 2 shows the high-level design of the web application.

The tumour analyzer comprises 3 main modules that operate independently. They are the MRI segmentation module, MRI classification module, and WSI classification module.

### 2.1. MRI segmentation module

We used the segmentation module as an approximate evaluation approach of the interpretable results of the MRI classification model. The U-Net based encoder and decoder architectures produce better results for image segmentation in the medical domain [7]. Since it avoids gradient vanishing via skip connections, we used Variational AutoEncoder (VAE) 3D U-Net [8] for the MRI segmentation functionality. VAE 3D U-Net has an auxiliary variational auto-encoder (VAE) branch in addition to the encoder and decoder modules. This model takes a series of 3D MRI volumes as input and produces semantic segmented 3D MRI volumes. This model is trained to segment edema, enhancing and non-enhancing tumour sub-regions of a brain tumour. These sub-regions are used to generate the tumour background (combined expanse of edema, enhancing and non-enhancing tumour regions). By comparing this brain tumour segmented MRI and MRI heatmaps, we can approximately evaluate the heatmaps.

### 2.2. MRI classification module

Input for the MRI classification module is a sequence of 3D volumes, where each channel corresponds to a different MRI modality. This sub-module is an extension of our previous work [6]. The MRI classifier was implemented based on DenseNet, because of its capacity to perform effectively in classification problems with a smaller number of parameters than alternative designs. We employed a DenseNet-BC model with 4 dense blocks and a total of 169 hidden layers, the growth rate was set to 32. The original DenseNet design was modified with 3D convolution to comply with 4D input volumes. The main purpose of the MRI classifier is to predict the most probable class for each input MRI sequence. This model performs with an accuracy of 86% during classification, while interpreting its internal decision-making process with a human-understandable visualization.

### 2.3. WSI classification module

Initially, preprocessing techniques are applied to calculate a low dimensional feature depiction for patched WSIs as proposed in [9], before the classification of WSI. After the preprocessing, the feature extraction phase is carried out for patches of size  $256 \times 256$  pixels in the identified region of interest of input WSI. Feature extraction is performed using a pre-trained ResNet-50 model with ImageNet [10] dataset. The ResNet model outputs a feature vector of dimensions  $1024 \times 1$  for each patch, which is later directed to the classification phase carried out using a model composed of a chain of densely connected layers. The complete pipeline of WSI preprocessing, feature extraction, and classification is referred to as the WSI module in the web application. The purpose of the WSI classification module is to classify WSI images into one of the aforementioned classes. Similar to the MRI classification module, interpretability is introduced to the WSI classifier also. These two classifiers function independently, thus, a user with only either MRI or WSI images can still utilize the web application.

## 3. Evaluation

### 3.1. Interpretable ML for MRI classification module

We used Gradient-weighted Class Activation Mapping (Grad-CAM) [11], to interpret the MRI classification module. Grad-CAM finds the most dominant logit concerning the feature maps of the final convolution layer in the trained model. The eventual output is a heatmap as in Fig. 3, highlighting the contribution level of individual pixels of the input MRI towards the predicted label. The dark blue and dark red regions indicate the contribution level of each region to the classification results in less and high strength, respectively. Thus, the output is more human-understandable and avoids the black-box nature of deep learning models.

The Grad-CAM technique is originally used for two-dimensional images. However, MRIs are three-dimensional volumes. These volumes are a stack of two-dimensional images. Therefore, we utilized the theory and mathematical expressions of the Grad-CAM technique to make MRI volumes interpretable. Also, the Grad-CAM uses feature

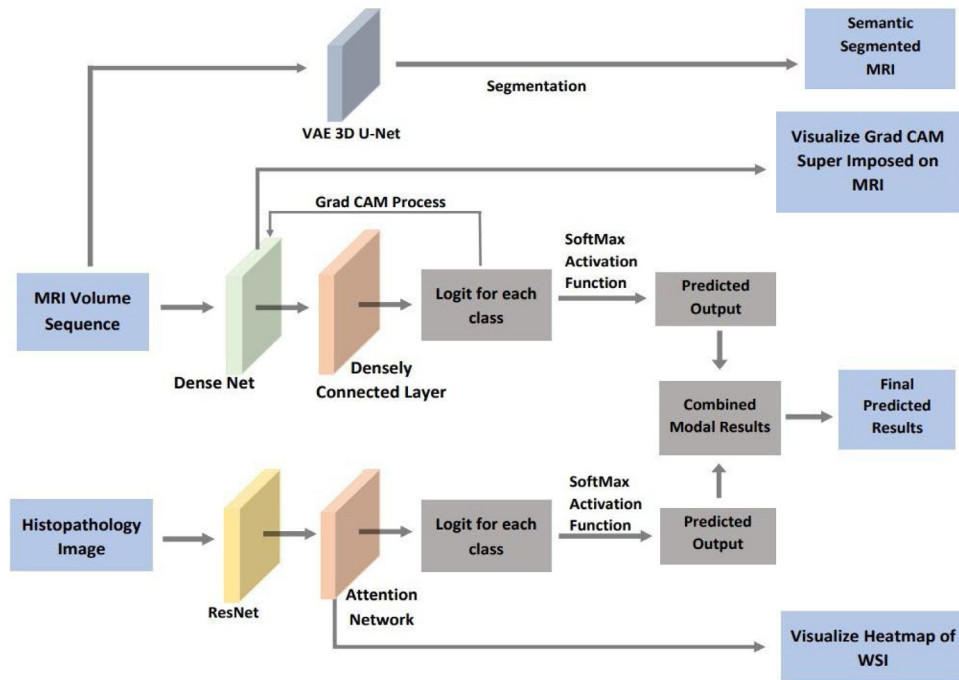


Fig. 2. High-level Design of the web application.

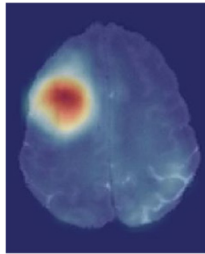


Fig. 3. Heatmap resulted in Grad-CAM process superimposed on input MRI.

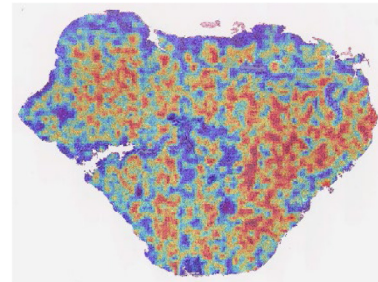


Fig. 4. A heatmap generated for a single WSI using attention scores.

maps of the final convolution layer to create heat maps. But when it was applied to MRI volumes, we found that the heat maps were very small in dimension, as a result, we lost much information from the heatmaps. Thus, we tried generating heatmaps for all the layers in the convolution neural network and found the best matching layer is not the final convolution layer feature maps to generate heatmaps.

### 3.2. Interpretable ML for WSI classification module

GradCAM approach cannot be applied to the WSI classification module because unlike the MRI classifier, the input images for the WSI classification module are asymmetric across channels. Therefore, to introduce interpretability to the WSI classification module, we implemented a sigmoid gated attention network [9,12]. The attention network outputs an attention score for each patch of the input image. This attention score is proportional to the amount of contribution of the corresponding patch to the predicted class. Hence, it generates heatmaps similar to Fig. 4, using normalized attention scores for every input WSI. This can be used to visualize the prediction of the classification model. Regions with higher attention scores are highly contributed to the prediction, while regions with lower attention scores have less contribution to the prediction. Accordingly, we have addressed the complex decision-making process inside the WSI classification model, by making it transparent and human-understandable.

## 4. Impact

Currently, there are many studies designed to classify brain tumours using deep learning models. However, those models are occasionally applied in the real world. As deep learning techniques are advanced enough to generate results with high accuracy, applicability is not hindered by performance. It is a matter of transparency or human understandability. The proposed Tumour-Analyser tool provides the same functionality with the same performance but presents human-understandable (interpretable) models. Therefore, the domain experts can use heatmap visualizations to decide whether a model operates as intended or is generating results just by random chance. This transparent nature builds confidence in the system among domain experts shaping the system with more reliability and applicability. Subsequently, the Tumour-Analyser tool can be used to replace traditional, tedious brain tumour diagnosis procedures.

Interpretability is not the only factor that makes our system more applicable in the real world. MRI segmentation module, MRI classification module, and WSI classification modules are arranged in such a way that the Tumour-Analyser tool replicated real-world diagnosis procedures. The current version of the Tumour-Analyser tool only supports a limited number of imaging modalities and MRI sequences; however, it can be integrated with other imaging modalities easily. The

concept of making black-box models transparent using interpretable machine learning has a much wider scope not only in other medical applications but also in many other high-stake domains. Furthermore, heat maps created are plausible substitutes for segmented tumours whenever pixel-wise precision is not essential, because label annotation and training segmentation models demand more resources. Understanding the decision-making process inside classification models helps to identify training models with irrelevant features which extricate developers from unwanted resource usage.

## 5. Further enhancements

The scope of this study focused on MRI and WSI modalities, as they are the most popular imaging techniques used to diagnose brain tumours. The presented approach can be extended to other imaging modalities such as PET, CT, and EEG. We created a functional web application that combined all three modules in the first phase considering the unavailability of necessary hardware resources. In the next phase, we expect to deliver a desktop application as it is more realistic for the context of medical practitioners having access to high-end GPUs on their premises. The current application relies on fluid-attenuated inversion recovery (FLAIR), which is an MRI technique that shows areas of tissues, T1w, T1gd, and T2w modalities as input MRIs. However, users may not have access to all of these modes in practice. Consequently, we plan to enhance the system to allow users to be more flexible when providing input for the MRI classifier.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] N. Wijethilake, D. Meedeniya, C. Chitraranjan, I. Perera, M. Islam, H. Ren, Glioma survival analysis empowered with data engineering—A survey, *IEEE Access* 9 (2021) 43168–43191, <http://dx.doi.org/10.1109/ACCESS.2021.3065965>.
- [2] J. Amin, M. Sharif, N. Gul, M. Yasmin, S.A. Shad, Brain tumour classification based on DWT fusion of MRI sequences using convolutional neural network, *Pattern Recognit. Lett.* 129 (2020) 115–122, <http://dx.doi.org/10.1016/J.PATREC.2019.11.016>.
- [3] A. Rehman, M.A. Khan, T. Saba, Z. Mehmood, U. Tariq, N. Ayesha, Microscopic brain tumour detection and classification using 3D CNN and feature selection architecture, *Microsc. Res. Tech.* 84 (1) (2021) 133–149, <http://dx.doi.org/10.1002/JEMT.23597>.
- [4] N. Wijethilake, M. Islam, D. Meedeniya, C. Chitraranjan, I. Perera, H. Ren, Radiogenomics of glioblastoma: Identification of radiomics associated with molecular subtypes, in: *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-Oncology*, Springer, Cham, 2020, pp. 229–239, [http://dx.doi.org/10.1007/978-3-030-66843-3\\_22](http://dx.doi.org/10.1007/978-3-030-66843-3_22).
- [5] N. Wijethilake, D. Meedeniya, C. Chitraranjan, I. Perera, Survival prediction and risk estimation of glioma patients using mRNA expressions, in: *Proc. of the 20th IEEE Conference on Bioinformatics and Bioengineering, BIBE*, 2020, pp. 35–42, <http://dx.doi.org/10.1109/BIBE50027.2020.00014>.
- [6] S. Dasanayaka, S. Silva, V. Shantha, D. Meedeniya, T. Ambegoda, Interpretable machine learning for brain tumour analysis using MRI, in: *2nd International Conference on Advanced Research in Computing, ICARC*, 2022, pp. 212–217, <http://dx.doi.org/10.1109/ICARC54489.2022.9754131>.
- [7] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015, <http://dx.doi.org/10.48550/arXiv.1505.04597>, Arxiv.org.
- [8] A. Myronenko, 3D MRI Brain Tumour Segmentation using Autoencoder Regularization, 2018, <http://dx.doi.org/10.48550/arXiv.1810.11654>, Arxiv.org.
- [9] M.Y. Lu, D.F.K. Williamson, T.Y. Chen, R.J. Chen, M. Barbieri, En F. Mahmood, Data-efficient and weakly supervised computational pathology on whole-slide images, *Nat. Biomed. Eng.* 5 (6) (2021) 555–570, bll.
- [10] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai. Li, Li. Fei-Fei, ImageNet: A large-scale hierarchical image database, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255, <http://dx.doi.org/10.1109/CVPR.2009.5206848>.
- [11] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, 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, <http://dx.doi.org/10.1109/ICCV.2017.74>.
- [12] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual attention network for image classification, in: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3156–3164.

**Sasmitha Dasanayaka** is a B.Sc. Eng. (Hons) graduate from the Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka. He has research experience in bio-health informatics and deep learning.  
Email: [sasmithanilupul.17@cse.mrt.ac.lk](mailto:sasmithanilupul.17@cse.mrt.ac.lk)

**Vimuth Shantha** is a B.Sc. Eng. (Hons) graduate from the Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka. He has research experience in bio-health informatics and deep learning.  
Email: [vimuth.17@cse.mrt.ac.lk](mailto:vimuth.17@cse.mrt.ac.lk)

**Sanju Silva** is a B.Sc. Eng. (Hons) graduate from the Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka. He has research experience in bio-health informatics and deep learning.  
Email: [sjanupa.17@cse.mrt.ac.lk](mailto:sjanupa.17@cse.mrt.ac.lk)

**Prof. Dulani Meedeniya** is a Professor in Computer Science and Engineering at the University of Moratuwa, Sri Lanka. She holds a Ph.D. in Computer Science from the University of St Andrews, United Kingdom. She is the director of the Bio-Health Informatics group at her department and engages in many collaborative research. She is a co-author of 100+ publications in indexed journals, peer-reviewed conferences and book chapters. Prof. Dulani has received several awards and grants for her contribution in research. She serves as a reviewer, programme committee and editorial team member in many international conferences and journals. Her main research interests are Software modelling and design, Bio-Health Informatics, Deep Learning and Technology enhanced learning. She is a Fellow of HEA (UK), MIET, Senior Member of IEEE, Member of ACM and a Chartered Engineer registered at EC (UK).  
Email: [dulanim@cse.mrt.ac.lk](mailto:dulanim@cse.mrt.ac.lk)

**Dr. Thanuja Ambegoda** is a senior lecturer in the Department of Computer Science and Engineering at the University of Moratuwa, Sri Lanka. He holds a Ph.D. from ETH Zürich. His research interests are in the field of Computer Vision, Robotics, and Bio-Health Informatics.  
Email: [thanuja@cse.mrt.ac.lk](mailto:thanuja@cse.mrt.ac.lk)