Brain Tumor Prediction Model's Workflow through Layer-wise Relevance Propagation (LRP)

Understanding the Prediction Model:

Data Preparation

Dataset (Brain Tumor Classification Dataset)

- It is Magnetic Resonance Imaging (MRI) data.
- It has been divided into Testing and Training Folder.
- There are a total of four classifications for the diagnosis results in this dataset, including 'glicoma tumor', 'meningioma tumor', 'pituitary tumor' and lastly 'no tumor'.

Data Pre-Processing

- Tasks including resizing images, converting them to tensors, and normalizing pixel values.

Purpose of this Model

- It is used to make predictions on new MRI images and LRP visualizations can be made for these predictions to help people understand why each classification is made.
- The goal is to make correct predictions while also showing how the model makes decisions, especially when it comes to medical imaging data.

Prediction Model

Model Selection (VGG16 Model)

- It is a convolutional neural network (CNN) architecture used for image classification tasks by learning hierarchical features from images.
- It is pretrained on a large dataset, then adapted for the brain tumor classification task.
- The output layer is modified to have four classes corresponding to different tumor types.

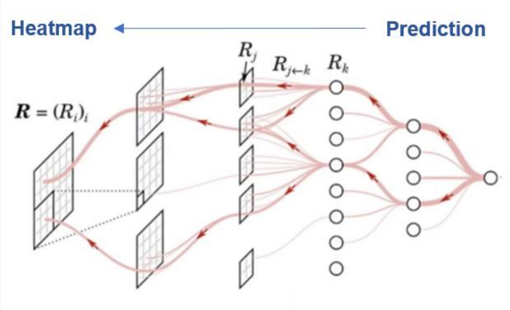
Model Training

- It is trained using the training dataset with labeled images.
- The loss function used is Cross-Entropy Loss and Adam optimizer is utilized for updating model weights during training.
- The training loop goes through epochs to keep adjusting the model parameters to minimize the loss.

Model Evaluation

- It is evaluated on a separate testing dataset.
- Accuracy is the metric to measure how well the model works. The batch accuracy is measured and shown during each training round. The batch accuracy is done by comparing the predicted labels with the ground truth labels for the current batch. For the given codes, we know that the printed batch accuracy is from first batch of the test dataset.

Interpretation Using XAI Method



Source: http://www.heatmapping.org/

- LRP is applied as an XAI method to 'explain' the model's predictions by enhancing the interpretability in the outcomes of the predictions.
- LRP gives relevance scores to different parts of the raw MRI images. This helps us figure out which parts of the images help the model make decisions.
- The outcome is the LRP visualizations (a plot showing the heatmap of relevance scores alongside the original image). They provide insights into the features the model considers important for tumor classification. From the given codes, the LRP visualization is applied to the specific image from the test batch with index of image ID equals to 31 (last image in that batch).

- The data scientist wrote program scripts that describe a series of steps that include preparing the data, building the model, training it, and interpretation (slide 1). This brain tumor prediction model moves on to an important stage: fine-tuning and continuous improvement.

- In this stage, two other stakeholders are involved. The model may undergo further fine-tuning depending on comments from medical professionals (doctors) and additional data (from new patients) before getting deployed. The data scientist should regularly monitor and analyze the model to ensure its long-term efficacy and correctness. In addition, regular examinations of the faithfulness, stability and fairness of the model using XAI technique is also crucial.

- In this ppt, we will focus on discussing how to explain the given Brain Tumor Prediction Model using LRP methods, which have already been coded, from the perspectives of important stakeholders: data scientist and doctor.

- For the given task, there are two scenarios have to be discussed. The first scenario involves the data scientist explaining the model to the doctor, while the second one is as a doctor explains the diagnosis to the patient. Before delving into these two scenarios, let's talk about the kind of output we can get after running the given codes.

Three Stakeholders


Data Scientist


Doctor


Patient

Output of the Code:

i. Batch Accuracy Value

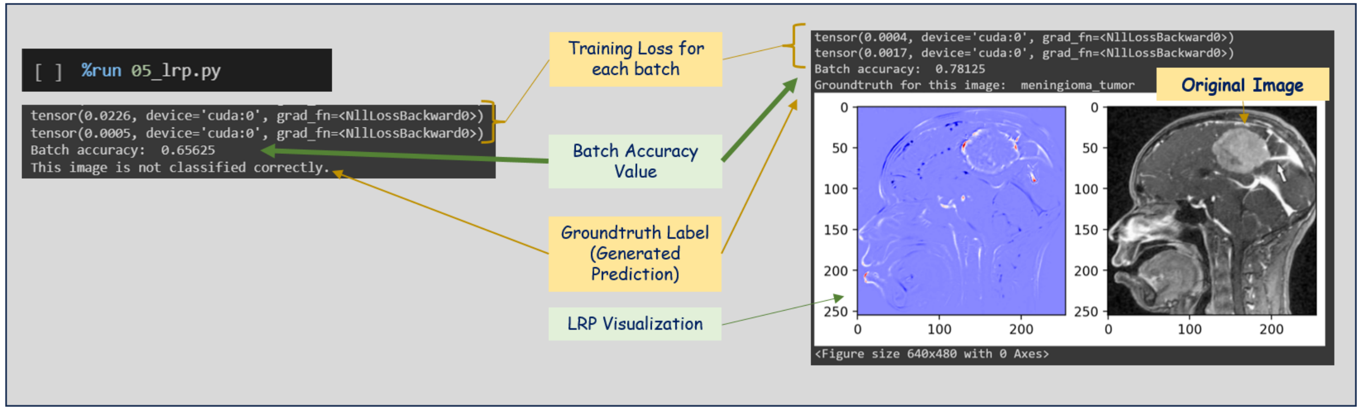
- For each epoch, the training loss for each batch will be printed. The loss is a measure of how well the model is performing on the training data. After that, the batch accuracy value that calculated by comparing the predicted labels to the groundtruth labels for the entire batch will be shown.

ii. Groundtruth Label for the Specific Image

- The groundtruth label for the specific image in the test batch will be shown. It is noted that the statement "Groundtruth for this image:" and the LRP visualization will only be printed if the generated prediction matches the ground truth value for that specific image. Otherwise, the result "This image is not classified correctly" will be printed. This implies that an incorrect prediction for that particular image (misclassification) will not getting any LRP visualization in the output.

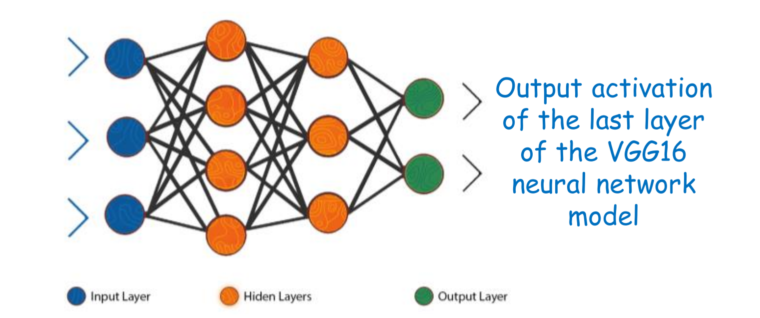
iii. LRP Visualization and Image Comparison

- The LRP visualization generates relevance maps, which emphasize parts of the input image that are important to the model's decision. These maps can be visualized alongside the original image and the ground truth label. A colormap ("seismic") is used to emphasize positive and negative contributions on the relevance map.
- The regions in red color indicates the positive contributions, which means that areas positively influenced the model prediction, while the regions in blue color indicates the negative regions, indicating that areas have negative impact on the model prediction.
- The comparison is facilitated through the side-by-side display of the LRP visualization and the original input image from the test batch. It is to help understand which features the model found important for the prediction.



Small Discussion on the Relationships:

Brain Tumor Prediction Model



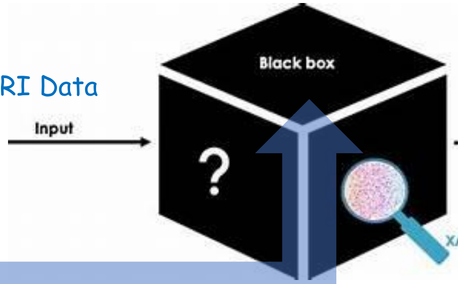
Source: https://www.analyticsvidhya.com

Output activation of the last layer of the VGG16 neural network model

MRI Data

CNN Model (VGG16)

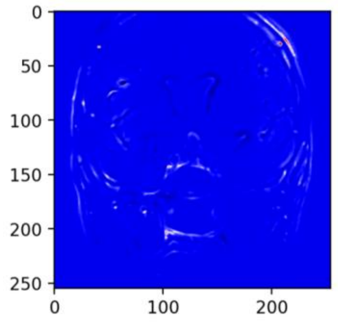
LRP operates backward through the layers of the network, assigning relevance scores to each neuron or feature map based on their contribution to the final prediction



Layer-wise Relevance Propagation (LRP) Algorithm

A set of relevance scores assigned to different regions or features of the input image

Heatmap



Develop & Train & Fine-Tune

Validate & Verify

Provide Transparency

Data Scientist:

- Develop and training model.
- Implement interpretability techniques namely LRP to make prediction more understandable to medical professionals.
- Evaluate the model performance to ensure model meets the necessary standards for medical applications.

Collaboration

Diagnose

Prediction Model:

- Serves as a decision support tool for doctors by providing prediction & LRP interpretations based on patterns learned from a large dataset of medical images.
- Enhances the transparency of the model, help building trust among doctors and patients.

Doctor:

- Validate & verify model predictions with their medical knowledge and examine false positives or false negatives.
- Review & validate the LRP visualizations, whether the output is make sense from the view of medicine.

Doctor:

- Use this model (including prediction & LRP interpretation) to aid in their decision-making while diagnosis.
- Bring domain expertise to the brain tumor diagnosis.
- Benefit from LRP visualizations to understand which regions of the brain images are considered most relevant to the diagnosis result.

Patient:

- Understand the predictions and interpretations provided by the brain tumor prediction model, with the assistance of the doctor.

Scenario 1: Data Scientist & Doctor

- Upon effective development and validation, the brain tumor prediction model has the potential to enhance diagnostic capabilities by offering medical professionals with further insights and information that facilitate well-informed decision-making.

- Collaboration between data scientists and doctors is critical in this regard, as the ultimate goal is to enhance patient outcomes. Early and accurate diagnosis of brain tumors can result in earlier treatments and improved treatment planning. The effectiveness and precision of brain tumor diagnosis can be increased by collaboration.

- Regular meetings between the doctor and data scientist provide a continuous feedback loop that guarantees the model's improvement based on medical insights and maintains its relevance to clinical demands. To bridge the gap between the technical parts of machine learning and the clinical requirements of medical professionals, effective communication is required.

- Doctors are particularly concerned with the prediction model's validity and dependability. As a result, they have to know how the model generates its predictions. They are also worried about the model's outputs' practical usefulness in real-world medical circumstances, as the model's predictions should be clinically relevant.

Meeting

Hello there! Today, I'd like to provide insights into our brain tumor prediction model as well as how LRP advances our knowledge of its decision-making.

First and foremost, consider the model architecture. Our model is based on the VGG16 architecture, a convolutional neural network that is well-known for its performance in picture classification tasks. We fine-tuned the model for brain tumor prediction by changing the final classification layer to fit the specific characteristics of different tumor types.

Furthermore, for data preparation and training, we employed a dataset of brain MRI images classified into four classes: glicoma tumor, meningioma tumor, pituitary tumor, and no tumor. The model was rigorously trained, with the weights optimized using the Adam optimizer and the cross-entropy loss function minimized. To improve the model's ability to generalize, data augmentation approaches were used.

But here's where it gets interesting. We employed Layer-wise Relevance Propagation (LRP) to crank up the interpretability of our predictions. This technique assigns relevance scores to each pixel in an input image, indicating their impact on the final prediction. Positive relevance highlights areas supporting the prediction, while negative relevance signals regions that may have influenced the model negatively. To make it clear, we've generated visualizations that overlay the LRP relevance scores onto the original brain MRI images. These visuals let you peek into the brain regions that our model finds most relevant for its predictions.

As you review these scans, you'll find that spots that light red with positive relevance correspond to features that match the model's forecast. It assisted the model in identifying a specific tumor. On the other hand, blue zones with negative relevance indicate locations where the model discovered traits that were discordant with the expected class or condition. It does not necessary imply that something is wrong, but rather that the model saw some features as less relevant to the given diagnosis.

However, our model is just a decision support tool, but the final diagnosis is in your hands. Your expertise in the medical domain is gold to us. Your feedback on the model's performance and the insights from LRP interpretations are crucial for the ongoing refinement and improvement of our system! We're in this together.

Scenario 2: Doctor & Patient

- The model is intended to help doctors make decisions by providing additional insights on brain tumor detection. It aids doctors by providing predictions and LRP interpretations based on patterns learned in a huge sample of medical images. LRP improves model transparency by displaying visualizations that illustrate the model's decision-making process. This transparency is critical for establishing confidence between doctors and patients.

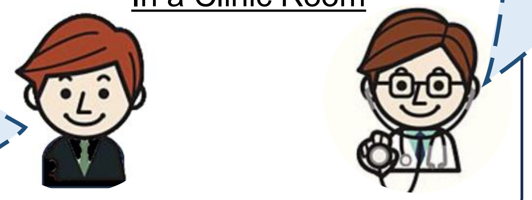
- Doctors are crucial in communicating the model's output to patients. Discussing the diagnosis result of the model, the significance of LRP visualizations, and how these insights contribute to the entire diagnosis and treatment plan are all part of this. Patients take an important part in their own healthcare. Understanding the predictions and interpretations provided by the brain tumor prediction model, with the help of a doctor, enables patients to make informed treatment decisions.

- It is critical that patients and doctors communicate well. Patients should feel comfortable speaking with their healthcare providers about the model's projections, LRP interpretations, and any concerns they may have.

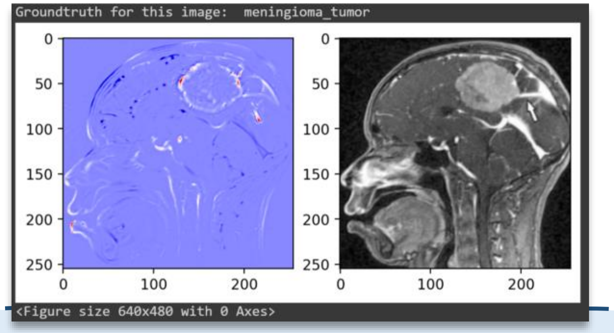
- Patients are more concerned about the accuracy as well as reliability of the diagnosis. They want to know that the model's predictions are accurate and correspond to their actual medical state. Besides, they want to understand the diagnosis and its implications for their health without being overwhelmed by technical jargon, as a straightforward and obvious explanation is what they need.

In a Clinic Room

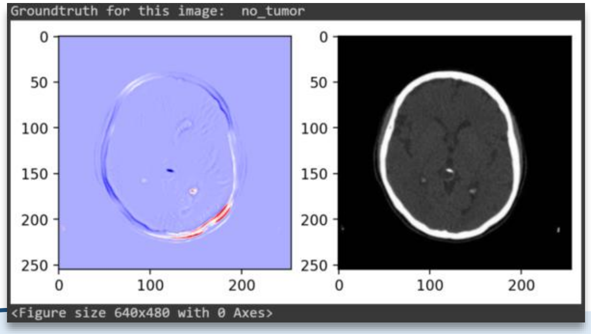
Morning, Doc! I have questions. Why the brain tumor diagnosis? Is it correct? Is the system performing well in terms of accuracy?



Diagnosis of Patient 1:



Diagnosis of Patient 2:



Hello and good morning. I have noticed your inquiries, therefore today I want to take a moment to discuss the recent brain tumor diagnosis we received based on medical evaluation, including advanced imaging techniques and the assistance of an artificial intelligence (AI) model.

If you still remember, about two weeks ago, we conducted the brain magnetic resonance imaging (MRI) scan, to examine your overall brain health. The findings highlighted certain traits that warranted further investigation. In addition to traditional diagnostic methods, we used an advanced AI model designed to assist in brain tumor detection. This AI model has been trained on a large dataset of medical images and has shown promising results in aiding clinicians with assessments. To enhance transparency and provide insights into the AI model's decision-making process, developer utilized a technique called Layer-wise Relevance Propagation (LRP). LRP helps us understand which specific regions of the brain images were considered most relevant by the AI model when making its prediction. I also use this model to help me in diagnosising your medical state.

Alright, now I want to share with you a visualization generated by LRP. You can see two pictures. The right one is your brain MRI and the left graph is the LRP visualization. The graph overlays the LRP relevance scores onto the original brain MRI images. The red spots show where the model focused and found features suggesting a tumor. It helps us understand why the AI flagged it.

[Patient 1: As you can see, the visual here, these red areas, they're like the brain's way of pointing out something. The model has detected that something abnormal in the frontal lobe of your brain. In your case, it's indicating the presence of a meningioma tumor.**]**

[Patient 2: Now, for this result with no tumor, the red spots shows where the AI sees patterns that match the absence of a tumor. Meanwhile, the blue regions are where the model thinks these features aren't strong indicators of a tumor. The overall pattern, with more red and less intense blue, supports the model's high confidence in the "no tumor" diagnosis. These visualizations help us understand why the model made that call and assure us it's looking at the right features.**]**

Remember, these visuals are just serve as a guide, showing us where the model is focusing. It's not the final say, your medical history and our experience are combined to make the diagnosis. The AI's just helping us better understand stuff. Don’t worry, I will thoroughly examine the MRI image to verify the model diagnosis and I’ll get back to you soon by next week with the final result. Today is just a preliminary diagnosis. See you later.

Additional Discussion

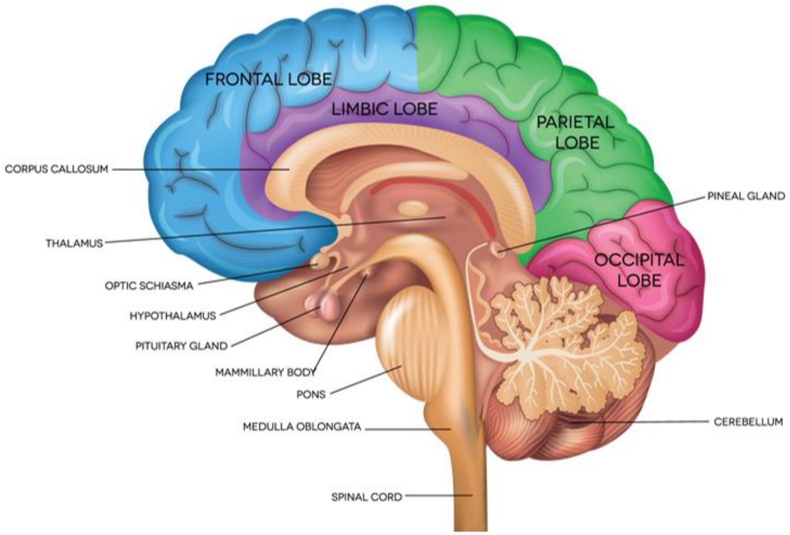
- The use of a pre-trained VGG16 model provides a robust foundation. The accuracy of the model is used to assess its performance, which is a popular indicator for classification jobs. However, considering the medical nature of the application, additional metrics such as precision, recall, and F1 score could be included for a more comprehensive review.

- The usage of LRP improves the model's overall interpretability. This is especially important in healthcare applications where trust and comprehension of the model's judgments are critical. LRP contributes to greater transparency in decision-making.

- The brain tumor prediction model is not intended to replace clinical expertise, but rather to act as a decision support tool. LRP enables doctors to incorporate AI findings into their clinical judgments.

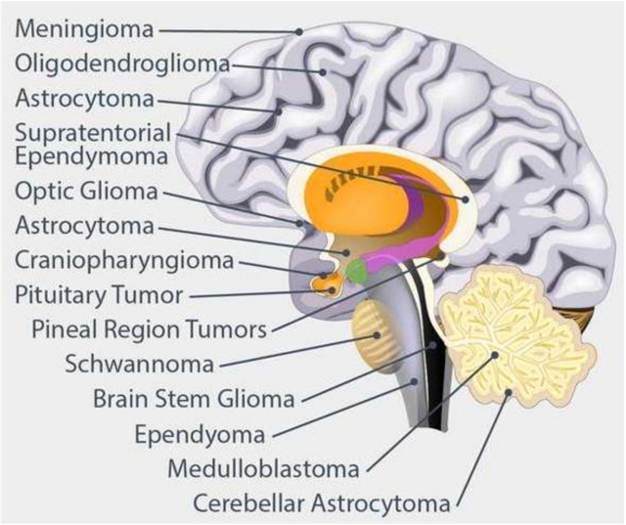
- Exploration of alternative XAI methodologies to compare and supplement the insights supplied by LRP may be examined for areas of improvement.

Anatomy of the Brain



Source: https://solvingkidscancer.org

Brain Tumor Types



Source: https://miamineurosciencecenter.com