Multimodal Convolutional Neural Network Models Allow for the Accurate Classification and Grading of Preoperative Meningioma Brain Tumors

Mihir Rane

March 15, 2023

Abstract

Magnetic resonance imaging (MRI) scanning and Computed Tomography (CT) proves to be a reliable form of imaging for modern medical use, providing clear images for physician and radiologist diagnosis. MRI and CT scans are especially important for neuroimaging of tumors for neuro-oncology after a patient lists symptoms indicating brain cancer. Although imaging does produce a lucid depiction of possible cancerous growth in the brain, inspection by a physician could be challenging due to subtleties in the image or human error. A diagnosis could also never be exact, as a biopsy is the only diagnostic test that can ascertain meningioma growth. A physician could confuse a noncancerous cyst located near the meninges of the brain for a meningioma tumor. Furthermore, World Health Organization (WHO) grading of each tumor could be complicated to differentiate. One possible solution to the human handicap is a Convolutional Neural Network (CNN), a commonly used machine learning method for image extrapolation and classification. For the purposes of this primary research, a multimodal CNN was given testing and training data of different types of brain cancers to test if it could properly classify different forms of CT and MRI scans of meningioma compared to glioma, pituitary, and scans with no tumor. The no tumor dataset included noncancerous cysts, as mentioned before, that could be confused with meningioma. Furthermore, a separate CNN was given different testing and training data on meningioma tumors with WHO grades one to three. The CNNs were run on a private GPU environment on Visual Studio Jupyter Notebook and were given input data in the form of standardized JPEG image files from research institutes around the world. The patient data came from various ages, different nationalities, and both genders. The concept of transfer learning was used to train the model, where the solution to one problem is used to solve another problem. The results of the models show high accuracies above 98% with an upward trend through the twelve epochs ran, indicating stability. The recall and precision scores were also high, indicating quality. Finally, the AUC scores were all above .99, describing the CNN's capability to include threshold-invariance and scale-invariance. Finally, an attention study demonstrated the CNN's tendency to apply most attention to the tumor mass itself rather than extraneous variables.

1 Introduction and Related Work

Brain tumor image classification and accurate diagnosis has been a perennial problem for physicians, stemming from the current limited knowledge on the brain [4]. This is especially true for meningioma, which is tough to detect in early stages of growth. The Convolutional Neural Network, a type of machine learning method, could be a beneficial tool to aid in the accurate grading and diagnosis of meningioma [4, 8]. Complications with present diagnostic testing of meningioma will be discussed along with the proposed CNN to diagnose meningioma and with its grade according to WHO.

1.1 Deep Learning

Deep learning is a set of techniques and algorithms that excel in discovering complicated patterns in large data sets, now surpassing human performance in certain tasks [41]. One example is the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), in which humans were outperformed by a neural network in multiple image processing tasks [52]. Deep learning also assists in language processing, speech recognition, and analyzing unstructured untabulated data [41]. Healthcare could reap benefits from deep learning due to the large amounts of data produced, with models making predictions based on patient history and large datasets. Machine learning can help analyze and make predictions based on data such as imaging or medical records. Such predictions and analyses include diagnosis, prognosis, health, and treatment [9, 59]. Deep Learning uses feature learning, in which algorithms automatically learn features or patterns of data. Using this technique, discovering features and performing a task are merged into one problem, and therefore both improved during the same training process [7, 41].

1.2 Feed forward Neural Networks

Feed forward Neural Networks include an input that is fed through multiple non-linear functions for transformations. The output for one network layer becomes the input for the next. The model is given data and makes predictions at the output layer matching the known labels. The function is typically a sigmoid or ReLU, multiplied by a weight of matrices of numbers [64]. The weight is a parameter within a neural network that transforms input data within the network's hidden layers. Weights determine the amount of impact an input can have on a specific output, and can be changed to optimize the output [5]. Optimizing the outputs is accomplished through a gradient descent optimization algorithm, evaluating how correct an output is using cost and loss functions. Loss is a value that shows the difference between the predicted and actual value. The higher the loss, the more off the prediction [58, 64, 72]. A loss function is for a single training input. In contrast, a cost function is the average loss over the entire training data set [34, 72]. During back propagation, the weights are adjusted to reduce loss. The bias value then shifts the functions used to adjust the output, like an intercept in a linear equation. The bias value is typically set to one [33, 42].

1.3 Convolutional Neural Networks

One node sharing connections to all nodes of the succeeding layer in the neural network would be inefficient. Convolutional Neural Networks (CNNs), types of feed forward neural networks, aim to have the least amount of connections possible while still conserving the spatial relationships in the network [51]. Below is a list of layers included in a CNN:

- Convolutional Layer: The convolution layer is the main part of the CNN and does most of the computational work [51, 69]. This layer performs a dot product between a set of a kernel (the learning parameter) and a predetermined fraction of the receptive field. The kernel is smaller than the original image in height and width, but reaches all three (RGB) channels [11, 40, 69]. During a step in the convolutional layer called the forward pass, the kernel slides across the height and width of the receptive region, resulting in a two-dimensional representation of the image, known as an activation map. The activation map is highly specific and includes every spatial position of the original image. These outputs then pass on to the pooling layer [11, 40, 51].

- Pooling Layer: The pooling layer statistically summarizes nearby outputs wherever it is placed in the CNN. Because of the smaller spatial size, this summary eases the computational and weight workload [51]. Different pooling functions include the average of the rectangular neighborhood, the L2 norm of the rectangular neighborhood, and the weighted average distance from the central pixel. Finally, there is max pooling, which is the most commonly used and uses the neighborhood's maximum output [49, 70].

- Fully Connected (FC) Layer: All the neurons in the fully connected layer are connected with all the neurons in the preceding and succeeding layers. FC layers are computed with normal matrix multiplication and bias. All neurons are connected in order to illustrate the full relationship between the input and output [44, 51].

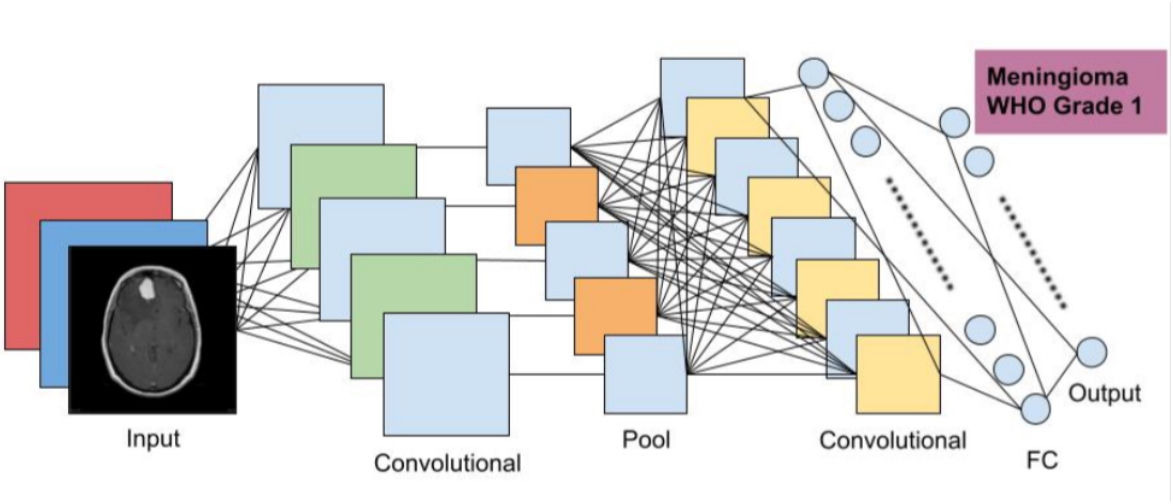

Figure 1: Model showing basic CNN architecture with an example input and output based on this work.

- Non-Linearity Layers: The images used in image segmentation and classification CNNs are not linear, contradicting with the linear operations used in convolution. Non-linearity layers are layers often used directly after the convolutional layer so that the CNN can include non-linearity in is activation map and non-linear functions to the CNN [19, 51, 71].

1.4 Convolutional Neural Networks in Medicine

Artificial intelligence and machine learning models were first applied to the medical field in 1972 at Stanford University. The model MYCIN was given patient and laboratory data on blood tests and was tasked to diagnose the patient and then provide the best course of treatment [60, 66]. Although still primitive and not preferred over doctors due to its low accuracy of sixty-five percent, MYCIN was the start of machine learning's introduction to the medical field [60].

1.5 Meningioma

A meningioma is a primary central nervous system (CNS) tumor, beginning in the brain or spinal cord. Meningiomas are the second most common type of primary brain tumor, accounting for around thirty percent of all CNS tumors. Meningioma is typically noncancerous in early stages and occurs in the meninges, which are the membranes surrounding the brain and spinal cord [10, 55, 68]. For the purposes of this work, data contains meningioma tumors only are present in the brain.

The layers of the meninges are split into three different layers: the tough dura matter (upper layer), the coushin-like arachnoid, and the delicate pia matter (inside layer). The space between the arachnoid and pia matter is the largest, and is filled with cerebrospinal fluid (CSF) [10, 12]. This area is called the subarachnoid space. In the case of meningioma, the arachnoid cells are the cells that develop the tumor, explaining why meningioma is typically on the periphery of the brain or spinal cord if not metastasized [12].

Monosomy 22 and inactivating mutations of NF2 have been connected to meningioma, although the pathophysiology and etiology are not well-known [68]. Meningioma can cause seizures or loss of sensory function, depending on the localization of the tumor. Because meningioma is present throughout the brain and sometimes in the spinal cord, it can also cause memory loss, motor loss, and other neurodegenerative issues [55]. For this reason, although benign meningiomas are typically non-cancerous and slow growing, physicians recommend immediate surgery. Early resection and treatment of meningioma is vital before malignancy and metastasis [20, 48].

Meningioma has three different grades based on Word Health Organization (WHO) standardization. This is the basis for diagnosis and grading of meningioma across the world [23, 36, 47]. The grading of cancer differs from the staging of cancer. Staging explains the spread of cancer, while grading depicts the appearance and abnormalities of cancer cells [39]. Although grading is typically done through histopathological means under a microscope [36, 39], complications with biopsy will be discussed later.

- Grade 1: This type of Meningioma grows slowly. It can be detected through solid borders surrounding the tumor. Grade 1 Meningioma makes up the majority of meningioma cases (around 78 to 81 percent of patients). These tumors are harmless and are noncancerous [23, 25]. The different forms include but are not limited to Meningiothelial, Fibrous, Transitional, Psammomatous, Angiomatous, Microcytic, Secretory, Lymphoplastmacyte-rich, or Metaplastic meningioma [25].

- Grade 2 (Chordoid, Clear-Cell, or Atypical Meningioma): This type of meningioma is detected by its namesake, in which the tumors look atypical compared to normal meningioma tumors. These types of tumors grow quickly and can become malignant quickly, although they are not cancerous. Grade 2 meningiomas account for 15 to 20 percent of patients [1, 23, 47]. Grade 2 meningioma is diagnosed with higher mitoses rates (4-19 per 10 HPF (high power field) along with high brain invasion. This grade of meningioma is officially diagnosed as per the WHO with three or more of these five features: increased cellularity, small cells with a high nuclear-cytoplasmic ratio, prominent nucleoli, sheeting, and foci of spontaneous necrosis [1, 15, 30].

- Grade 3 (Anaplastic (malignant), Rhaboid, or Papillary Meningioma): This is the quickest growing tumor and spreads across the brain. Grade 3 meningiomas are cancerous and account for the smallest number of cases (one to four percent of patients) [23, 47, 63]. Meningioma graded this way is diagnosed based on how far it has spread throughout the brain [36, 63]. According to the WHO, Grade 3 meningioma is officially diagnosed through cytology indicating overt malignancy, which can resemble melanoma, carcinoma, and higher grades of sarcoma. Grade 3 meningiomas also have mitotic rates higher than 20 mitoses per 10 HPF [15, 30, 63].

- Neurofibromatosis type 2 (NF2): While not a grade, NF2 is a syndrome including more than one meningioma tumor growth in the brain [22, 43].

1.6 Diagnostic Testing for Meningioma

In order to test for meningioma growth and grading in the brain, physicians use various diagnostic testing techniques including a Magnetic Resonance Imaging (MRI) scan, a Computed Tomography (CT) scan, a Cerebral Angiogram, or a biopsy [28, 57]. Observation could also serve as a diagnostic method, but is highly qualitative [28]. Diagnostic tests are essential for an accurate diagnosis and grading/staging of meningioma growth in the CNS [68].

During a Magnetic Resonance Imaging (MRI) scan, the patient is surrounded by a strong magnetic field, compelling oppositely charged electrons in the targeted area to align with the transmitted field. The second step of an MRI includes a radio frequency wave that is sent through the patient's body. During this period, the electrons are stimulated and no longer stay in equilibrium due to the two straining forces. When the frequency is turned off, the electrons move back to equilibrium. The energy released during this step changes based on the chemical nature of the molecules surrounding. The image picks up this energy release and accurately illustrates an image based on this data [32]. Visualization is possible through the different chemical properties in the body's molecules, e.g., tumor cells have more metabolic activity; therefore, electrons passing through them release more energy. This becomes discernible on the scan by specialized doctors or radiologists. Radiologists and physicians look for foreign growth in the brain or any neurodegenerative signs [56]. MRI imaging have variables called T1 and T2 times, demonstrating different tissue relaxation times. T1 (longitudinal relaxation time) is the measure of the amount of time for spinning protons to realign with the magnetic field. T2 (transverse relaxation time) is the amount of time for spinning protons to reach equilibrium and stop having phase coherence [27, 62]. With the function of MRI discussed, there are three types of MRI imaging with varying applications:

- **Structural MRI:** This type of MRI is a visual image of the target organ or tissue that is used for examination of anatomy and physiology [67].
- **Functional MRI:** This type of MRI detects subtle changes of blood flow in the cerebral cortex. The patient is made to do simple tasks and a visual representation [54].
- **Perfusion MRI:** This form of magnetic resonance imaging identifies the portion of the brain that receives less blood flow. This could be an indicator of a tumorous growth blocking the blood vessels [21].

There are many types of MRI imaging sequences, most commonly being T1-weighted, T2-weighted, and FLAIR scans. During T1-weighted image scans, the T1 aspects of the tissue are exposed using shorter TE and TR times. On the other hand, T2-weighted images expose the T2 properties of the scanned tissue by having longer TE and TR limes [13]. Fluid Attenuated Inversion Recovery (Flair) use prolonged TE and TR times, eclipsing the timings of even T2 scans. This method makes it so the cerebrospinal fluid (CSF) and other extraneous variables are dark while the neoplasms in the image are bright [3, 26].

Along with these various types of scans, various additional substances like contrast agents can be used to better visualize abnormalities. One such contrast agent used with T1-weighted imaging is Gadolinium (Gad). Paramagnetic Gad is non-toxic and shortens T1 times. Gad shows up brightly. Most practitioners use Gad to look for abnormalities and neoplasms in vascular structures as well as the blood brain barrier. This makes Gad useful for inflammation, tumors, and abscesses [61].

Diffusion weighted imaging (DWI) for MRI is primarily used to detect any forms of stroke or ischemia. Normally, water diffuses freely intracellularly and extracellularly through the plasma membrane of cells and through aquaporins. Diffusion of water attempts to maintain an isotonic concentration of solutes inside and outside the cell. The free diffusion in ischemic brain tissue becomes restricted, however. During ischemia, the Na/K pump shuts down and sodium starts to build up intracellularly. Due to the intracellular hypertonic nature, water diffuses into the cell. Because the intracellular space is more confined than the extracellular, the movement of water molecules inside the cell is more restricted [38, 35]. DWI picks up on this restriction, and becomes a bright signal on the MRI. For this reason, DWI can accurately diagnose strokes and ischemia that potentially could be from a tumor mass [38].

During a Computed Tomography Imaging (CT) scan, the patient is put in a tubular machine, much like the MRI discussed earlier. In the case of CT scanning, however, uses X-ray waves as its modality for imaging [17]. X-ray waves are cylindrically sent all around the patient in order to generate a cross-section view on the transverse axis of the patient. The x-rays are a type of electromagnetic radio wave that passes through most of the body's tissues but gets absorbed by bones and most foreign objects, e.g., tumorous or cystic growth. The radiation that passes through the targeted area gets reabsorbed by special photosensors. In CT, the photosensors are placed circularly around the transverse axis. All data is then amalgamated into one cross-section image of the targeted area. On CT images, the slices are at an angle. Thus, structures near the top of the slice are more rostral than structures near the bottom of the slice. This is different from MRI, which has true axial slices [29].

Different contrast agents, imaging methods, or sections (Sagittal, Coronal, or Axial) can be used to better visualize CT or MRI scans. Sagittal sections take images from the sagittal plane, coronal sections take images from the frontal plane, and axial sections take images from the transverse plane, the most popular physician choice. One example of a contrast fluid used in MRI and CT images is C+ contrast, which is given through a venous catheter. The fluid consists of iodine, which becomes visible on images and helps with image clarity [16, 31, 46].

Cerebral angiograms use X-rays just like CT scans, but target for the blood vessels and blood flow in the cerebrum. Normally, X-ray radio waves are not absorbed by blood in the cerebrum. In the case of cerebral angiograms, a catheter is typically placed into the arm or leg (femoral artery is considered standard practice). An iodine dye in injected into the blood to reach the cerebrum. Because X-ray radio waves cannot pass through iodine, an image of blood flow displays on the scan [50]. Doctors can

use this data in order to diagnose possible blockages in blood flow (tumors), blood clots, or aneurysms. Cerebral angiograms are a commonly used diagnostic technique for diagnosing neoplasms, but are often alone inefficient for a definitive diagnosis of cancer. [18, 50].

A needle biopsy is typically performed for brain samples. During this procedure, a hole is drilled into the skull and a sample of abnormal growth is taken from the patient. This sample is then tested and seen under a microscope by pathology, with a respective diagnosis [6, 53]. Needle biopsies are typically guided by ultrasound or other fluorescence substances [6, 14].

1.7 Meningioma Complications

Although the biopsy does serve as the only definitive form of diagnosis for meningioma growth, this form of diagnostic testing does introduce some complications and risks. Possible risks of a biopsy include subdural hematomas, infections, brain swelling, strokes, seizures, and blood clots [24]. For this reason, brain biopsies are typically avoided and replaced with imaging diagnostic tools like the CT or MRI [51, 65]. Although practitioners are the best at differentiating abnormal growth to normal brain matter in CNS imaging, they often need biopsies in order to gain a definitive diagnosis. Apart from biopsy, the patient is subject to a multitude of other diagnostic tests that have their own risks. These include MRI, CT, and angiograms that expose patients to harmful radiation [2, 45]. For this reason, machine learning and CNNs could replace biopsies as a definitive diagnostic test for meningioma by reaching high accuracies just using CT and MRI. This paper hopes to illustrate how a multimodal CNN can diagnose and grade meningioma brain tumors with a single MRI or CT scan, limiting radiation and risks with other diagnostic procedures.

2 Methods

2.1 Dataset

The dataset used in this primary research was an amalgamation of previously documented datasets as well as data from a variety of documented cases from both inpatients and outpatients in public hospitals and private practitioners. The MRI images were taken from both 1.5t and 3.0t scanners and the CT images were taken by special X-ray scanners at 12 different hospitals and institutions in the United States, Europe, and India. All data from CT and MRI are grayscale and with only one channel. Patient age ranges from 8 years to 82 years, gender includes 54.8% female and 45.2% male, and racial demographics are unknown.

Tumor Classification Model:

	Meningioma	Glioma	Pituitary	No Tumor
Training Data	822	826	827	395
Testing Data	115	100	74	105

Table 1: A table illustrating the number of sample images used for each type of tumor in the tumor classification model before data augmentation. Images are from multiple modalities. Training and testing data amounts are reported.

Meningioma WHO Grading Model:

	Grade 1	Grade 2	Grade 3
Training Data	1286	1278	1282
Testing Data	201	202	197

Table 2: A table illustrating the number of sample images used for each WHO grade of meningioma in the sub-grading model before data augmentation. Images are from multiple modalities. Training and testing data amounts are reported.

All data collected are preoperative, before treatment such as surgery, chemotherapy, radiotherapy, or other therapeutic measures. Overall volume of tumor resection was greater than a volume of 1ml

Figure 2: Examples of brain tumors from the raw MRI volumes collected in this study. All images are representations of meningioma tumor growth. The far left is an axial cross-section, the middle is a sagittal section, and the far right is a coronal section of meningioma. Images were not altered. For the purposes of this figure, the tumor masses were manually annotated in red. Annotated images were not used in the training steps of this model.

for those surgical patients. MRI and CT volume dimensions cover [192; 576] × [240; 640] × [16; 400] voxels, and the voxel size range [0.34; 1.17] × [0.34; 1.17] × [0.50; 8.0] mm^3. These values include all tumors in the classification model's datasets. In the meningioma grading dataset, MRI and CT volume dimensions cover [192; 512] × [224; 512] × [11; 290] voxels, and the voxel size range [0.41; 1.05] × [0.41; 1.05] × [0.60; 7.00] mm^3. For reference, an average MRI or CT volume is [349 × 363 × 85] pixels with a spacing of [0.72 × 0.72 × 4.21] mm^3. The matrix sizes for all images include (128x, 256x, 320x, 512x).

This work's CNN models employ multimodal CT and MRI data. As mentioned previously, there are many forms of traditional MRI and CT scans that can form from different dyes, methods of taking the scan, or plane of section. Images were taken from axial, coronal, and sagittal sections. In addition, 20% of samples use C+ contrast fluid.

2.2 Patch Sampling with Data Augmentation

As small datasets are prone to model overfitting and low accuracies, data augmentation was used in order to generate more data with the same images. The techniques used to create a variety of images from one sample include mirroring, random scaling, random rotations [-20,20] degrees, elastic deformations, resampling, random contrast, random brightness, and gamma correction. All data augmentation was accomplished alongside a python package from Medical Image Computing at the German Cancer Research Center (DKFZ) and the Applied Computer Vision Lab of the Helmholtz Imaging Platform [37].

All patches resulting from data processing are generated randomly during the training step of the model. Each batch has a foreground and a background class for easier image classification that are input to the training model. The foreground class allows for proper data augmentation through all transformations.

2.3 Image Preprocessing

The image preprocessing is independent of the CNN, and various steps were taken in order to optimize the input and resulting output for the training model. All preprocessing steps were done through ANTsX/ANTs, WCHN/CTseg, FSL FLIRT, and FSL BET. Here is a list of preprocessing methods to reduce non-uniformity:

- Image Registration: CT and MRI images are of different modalities and each have sub-weights. Image registration is vital in order to transform these different data sets into one dataset on one coordinate system.

- Bias field Correction: MRI images of all weights contain variance in the low frequency intensity in both the bias and gain fields. During this prepossessing step, bias was corrected in MRI images.
- Skull Stripping: Normal MRI and CT images contain not only brain matter, but the surrounding bone and tissue. Because the skull and other surrounding tissues introduce unfavorable bias and variables, skull stripping removes the skull from the end image while attempting to not distort brain matter.
- Image Normalization: Image intensity is standardized to a range of [0,1] and image size is normalized to (224, 224, 3).

2.4 Architecture Design and Attention Mechanisms

In this work, EfficientNetB0 was used as the backbone for the CNN. This was used as the backbone for this work because of its employment of compound scaling instead of normal random scaling. Instead of the traditional method of balancing the scale in one dimension, compound scaling targets three different dimensions: width, depth, and image resolution. Under sufficient standardized testing by the authors of this CNN, it surpassed other methods without compound scaling in accuracy and efficiency. EfficientNetB0 was trained by over one million images and pre-trained weights from the ImageNet database, with 237 layers of varying filer sizes and over 11 million trainable parameters. EfficientNetB0 was used over its newer counterpart B7 because this light work CNN can still accomplish high accuracies.

This model's attention mechanisms are deployed directly from the keras implementation of EfficientNetB0, and this work will therefore not go into depth into encoder and decoder pathways nor linear transformations. EfficientNet uses composite coefficients to uniformly scale all dimensions of depth, width, and resolution. A simple Self Attention network structure is trained in advance to process data images. This will enable the pre-training network to focus as much as possible on the data and image features with a large amount of information and a large gap before entering EfficientNet training. After that, the pre-training network will train the model.

The GlobalAveragePooling2D layer was included in this work. This layer acts similar to the Max Pooling layer in this CNN, the only difference is that it uses the Average values instead of the Max value while pooling. This aids in decreasing the computational load on the machine while training. The dropout layer omits some neurons at each step from the layer, making the neurons more independent of the neighboring neurons. It helps in avoiding overfitting. Neurons to be omitted are selected at random. The rate parameter is the likelihood of a neuron activation being set to 0, thus dropping out the neuron. The included dense layer is the output layer, which classifies the image into 1 of the 4 possible classes. This work uses the softmax function, which is a generalization of the sigmoid function. Finally, callbacks can help fix bugs more quickly, and can help build better models. Callbacks help visualize how your model's training is going, and can even help prevent overfitting by implementing early stopping or customizing the learning rate on each iteration. This work will be using TensorBoard, ModelCheckpoint and ReduceLROnPlateau callback functions.

2.5 Training Strategies

The CNN was run in 12 consecutive epochs. There were two loss fucntions used in these CNNs: class-average loss and Focal Tversky Los (FTL). Class-average loss was chosen in order to exlude the background and improve overall CNN computation. The FTL was employed due to its ability (Tversky index) and flexibility (Focal parameter). FTL is capable of efficiently balancing false positive and false negative predictions. FTL's flexibility in focal parameter also allows it to leverage original data image volumes to account for loss calculation. The typical values for FTL were used: $= 0.7$ and $= 0.3$ for the Tversky index and $= 2.0$ as the focal parameter. These values attempt to limit the number of predicted false negatives by the model. During the training step of the model, in order to determine the model with the highest quality, the one with the least overall validation loss was chosen.

All models were trained using two samples in a batch due to the large memory footprint. The models

in this work use mini-batch sizes of 30 elements, which was accomplished through accumulated gradients. After various tests, batch sizes of up to 30 have proven to improve generalization and model quality. During accumulated gradients, each batch of two was run in an order of 15 steps with the same weights for each step. The gradients sum up and accomplish the same goal as a larger batch size. This ensures that the weights are updated at each step, instead of relying on one batch size. This also accomplished the total batch size of 30 elements.

3 Results

3.1 Implementation Details

Results were obtained using an HP desktop: Intel®Core™i9-9900K CPU @ 3.60 GHz, 16.0 GB of RAM @2666 MHz, 512 SSD, 2 TB HDD, NVIDIA GeForce RTX 2080 Ti GPU, and a regular hard-drive. Implementation was done in Python using TensorFlow v1.13.1, and PyTorch lightning v0.7.3 with PyTorch back-end v1.3 on JupyterNotebook via Visual Studio Code.

3.2 Overall Performance Study

Tumor Type	F1	Recall	Precision	Accuracy
No Tumor	97.40 ± 1.02	96.89 ± 1.51	98.21 ± 1.23	99.21 ± 0.54
Meningioma Tumor	98.70 ± 1.01	95.59 ± 1.12	98.78 ± 1.48	99.79 ± 0.23
Pituitary Tumor	96.23 ± 1.14	95.54 ± 1.07	98.01 ± 1.33	98.82 ± 0.66
Glioma Tumor	98.16 ± 1.76	97.23 ± 1.24	98.25 ± 1.55	99.31 ± 0.44

Table 3: Performance Summary for Each Tumor Type Averaged Across Five Folds

WHO Meningioma Grade	F1	Recall	Precision	Accuracy
Grade I	97.14 ± 1.13	96.83 ± 1.03	98.71 ± 1.82	99.19 ± 0.57
Grade II	98.59 ± 1.11	97.71 ± 1.87	98.64 ± 1.43	99.64 ± 0.34
Grade III	98.09 ± 1.93	97.24 ± 1.00	98.33 ± 1.71	99.06 ± 0.43

Table 4: Performance Summary for Each Meningioma Grade Averaged Across Five Folds.

Regarding the pituitary tumor subtype in the classification task, the diffuse nature of the tumors and the less defined gradients are probable explanations for the lower classification performance. For the meningioma category, the reason for the lower recall values can be attributed to the larger number of small tumors (<3ml) compared to other subtypes. In addition, outliers have been identified in this dataset where a small amount of the tumors were enhanced due to calcification.

While tumor volumes and outlier MR and CT scans are reasons for the discrepancy in recall values across the board, precision is rather unaffected and more stable. The nature of the convolutional neural network architecture and training strategy used can explain those results. Given GPU memory limitation, the preprocessed MR and CT scans have undergone a significant down sampling, and as such small tumors are reduced to very few voxels, impacting mainly recall performance.

The high precision and recall values also indicate high quality of the model. Average precision scores range from 98.01% to 98.78% in the tumor classification task and 98.33% to 98.71% in the sub-grading task. The results indicate very few false positive values and the return of more relevant results than negative. The model is stricter in classifying a sample as positive. Average recall scores range between 95.54% to 97.23% in the classification model and 96.83% to 97.71% in the sub-grading model. High recall scores indicate the return of most of the relevant results and true positives. The overall study performance indicates high quality and performance of both classification tasks in this work. The accuracies also reach high scores, being above 98% for all, indicating reliability of the CNNs.

3.3 Accuracy and Loss Analysis

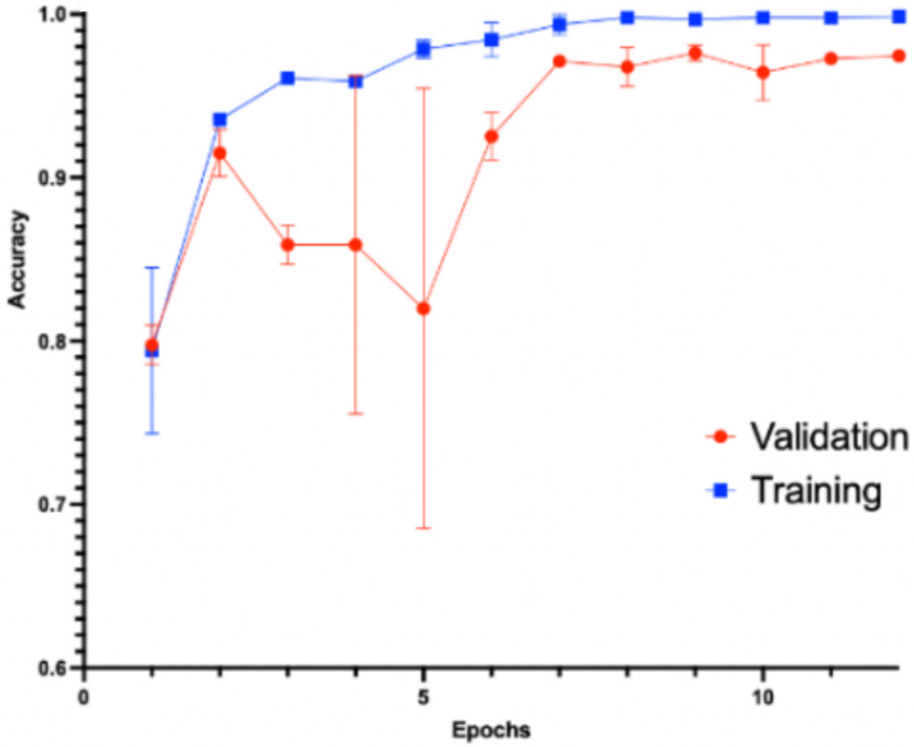

Figure 3: Graph showing the validation and training accuracy averaged over five splits during 12 epochs for the classification model. Error bars represent a 95% confidence interval, while the solid line represents the average.

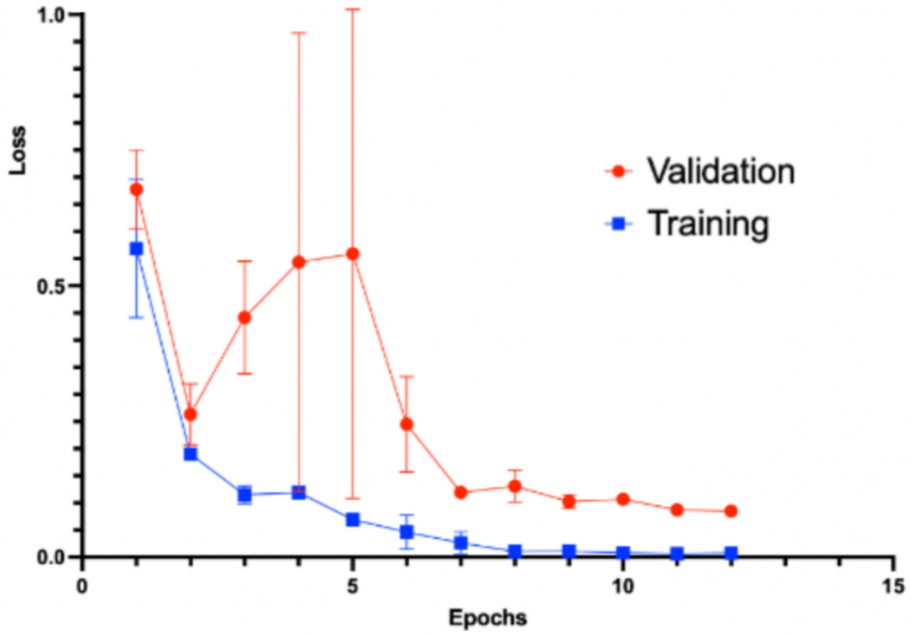

Figure 4: Graph showing the validation and training loss averaged over five splits during 12 epochs for the classification model. Error bars represent a 95% confidence interval, while the solid line represents the average.

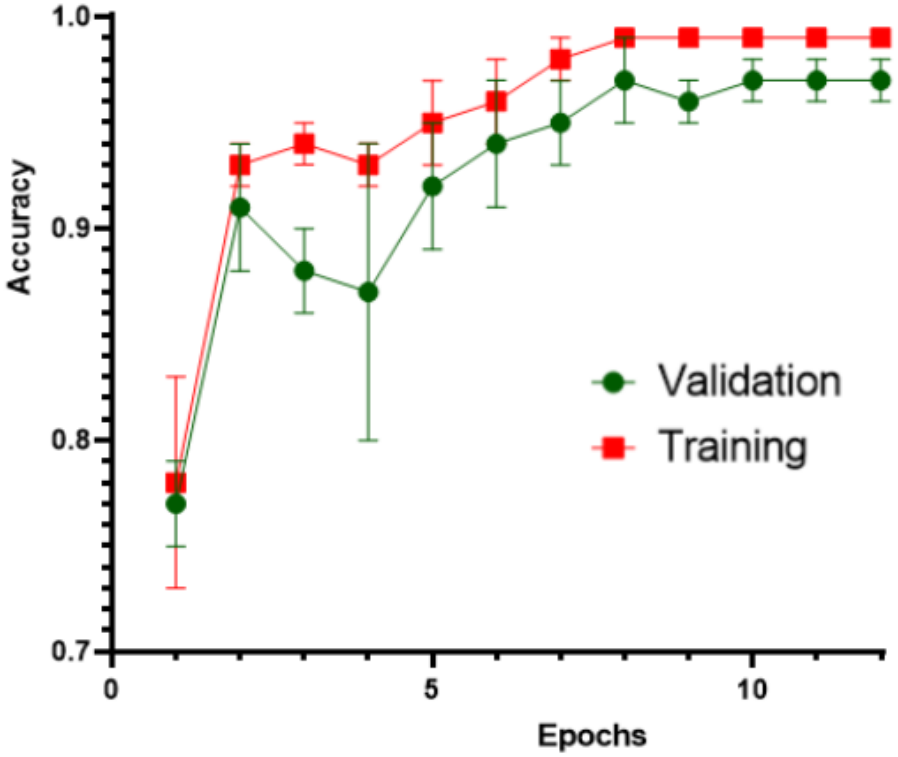

Figure 5: Graph showing the validation and training accuracy averaged over five splits during 12 epochs for the grading model. Error bars represent a 95% confidence interval, while the solid line represents the average.

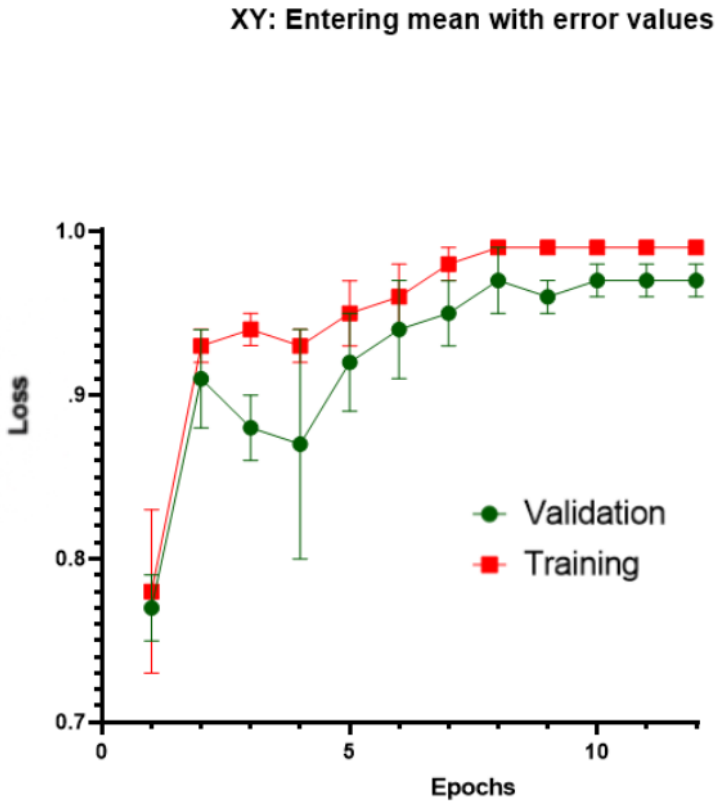

Figure 6: Graph showing the validation and training accuracy averaged over five splits during 12

epochs for the grading model. Error bars represent a 95% confidence interval, while the solid line represents the average.

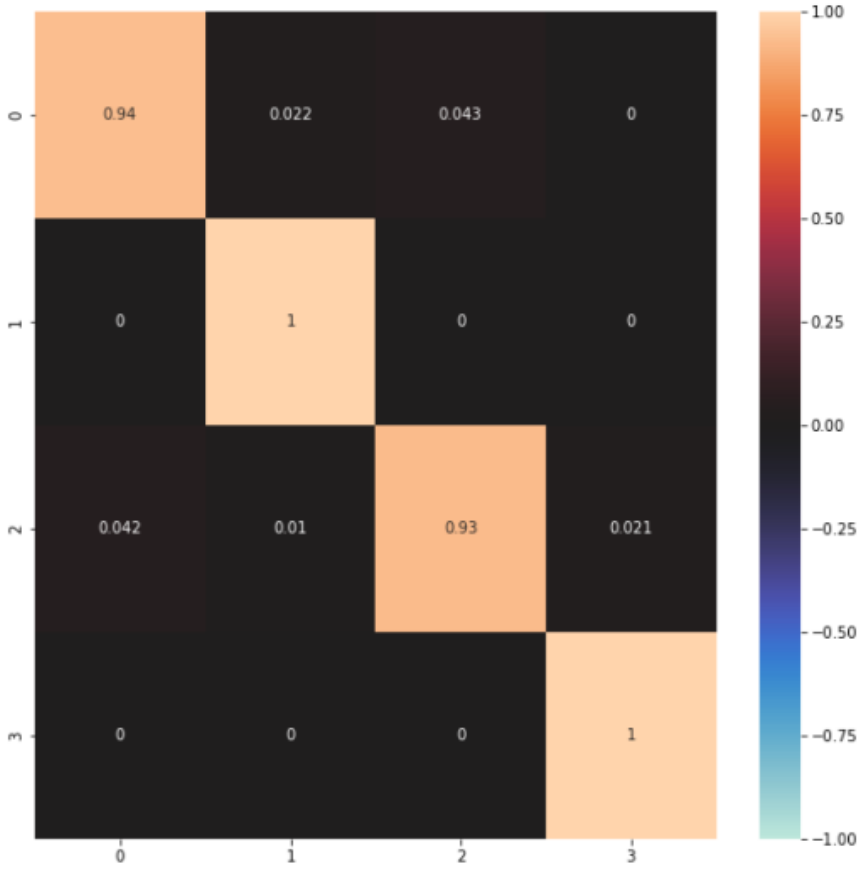

Figure 7: Heat Map of the confusion matrix regarding accuracies of tumor classification. The color scale is shown on the right, with 1.00 representing the highest accuracy.

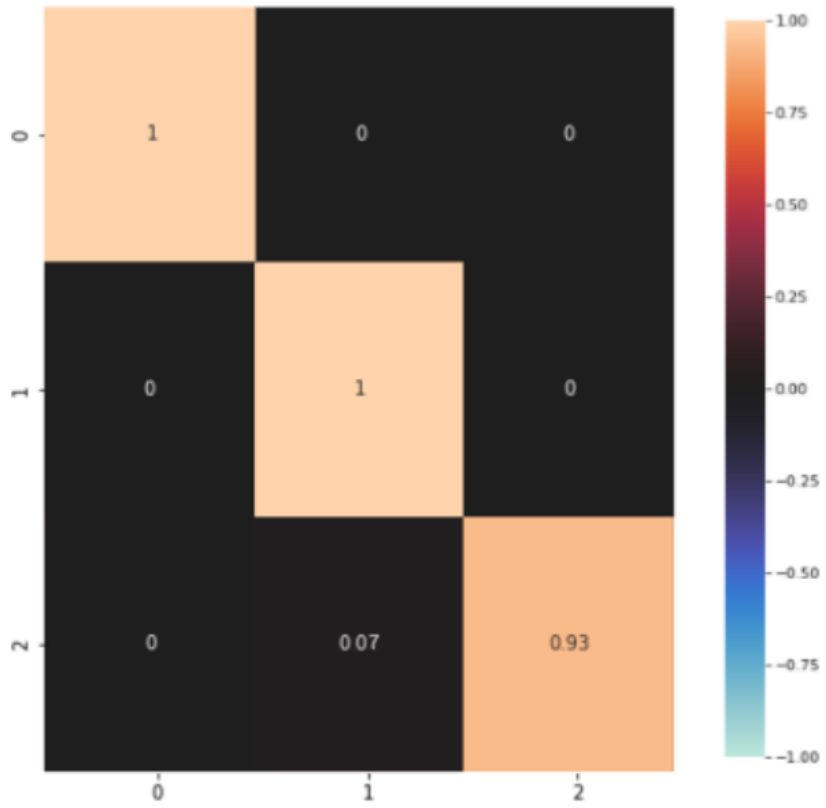

Figure 8: Heat Map of the confusion matrix regarding accuracies of meningioma tumor grading. The color scale is shown on the right, with 1.00 representing the highest accuracy.

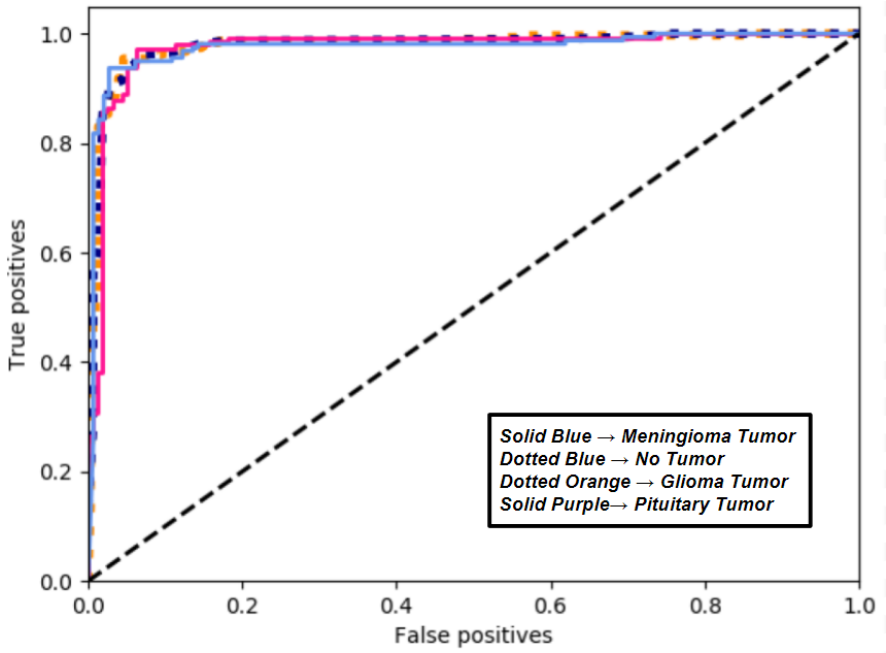

Figure 9: ROC Curve depicting False Positive versus True Positive Rates given threshold values between 0 and 1 with intervals of 0.1 for the classification task. The graph was generated through the Sklearn utility on python. A baseline of .5 AUC is demonstrated by the black line.

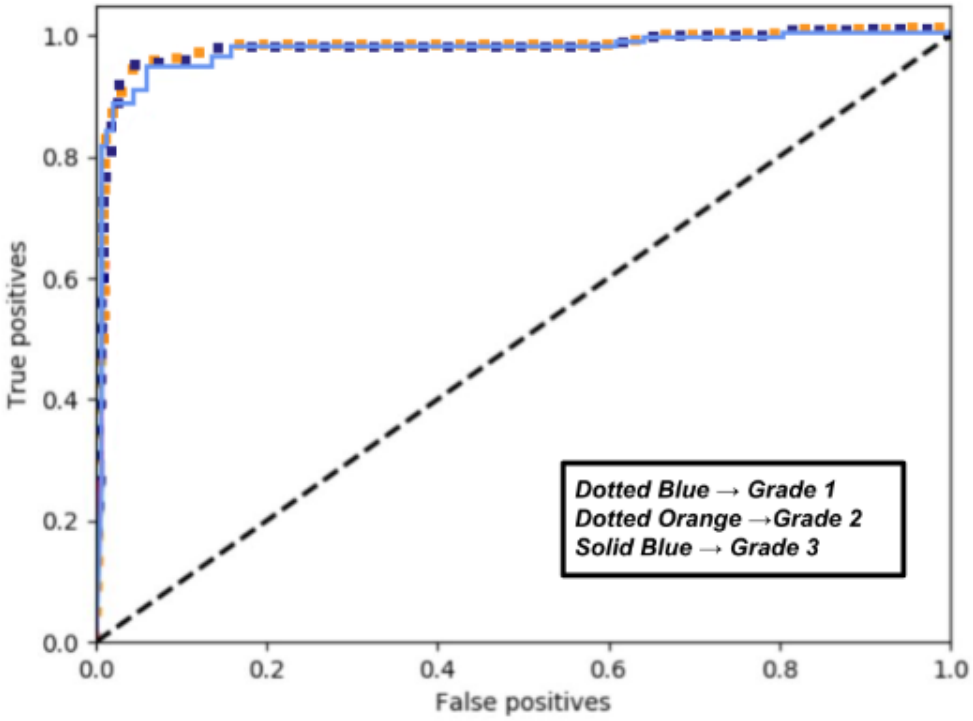

Figure 10: ROC Curve depicting False Positive versus True Positive Rates given threshold values between 0 and 1 with intervals of 0.1 for the grading task. The graph was generated through the Sklearn utility on python. A baseline of .5 AUC is demonstrated by the black line.

As indicated through figures 3 through 10, high accuracy and AUC scores were achieved during an upwards trend through 12 epochs during the validation and training steps of the model. Below are tables listing the AUC scores of the classification and grading tasks demonstrated in Figures 9 and 10.

Classification Task:

	Meningioma Tumor	Pituitary Tumor	Glioma Tumor	No Tumor
AUC Score	.987	.981	.982	.985

Grading Task:

	Grade 1	Grade 2	Grade 3
AUC Score	.984	.989	.978

In concordance to the high accuracies revealed in previous tests, both models demonstrate exemplary AUC scores. During classification, the models were able to accurately distinguish positive and negative classes. Furthermore, AUC was calculated using a Receiver Operative Characteristic (ROC) in order to include variables such as threshold-invariance and scale-invariance, two variables absent if accuracy is used alone. The positive instance is ranked higher than the negative instance on both CNNs more than 98% of the time, indicating high quality of the model, the quality of the loss functions, and its reliability to a practitioner.

The confusion matrices corroborate the results shown in the ROC curve. True-positive classes reach high accuracies, ranging from 93% up to 100% in the classification task and grading model. In both models, the extremes are 0, furthermore indicating high quality and stability. The Confusion Matrix displays the tendency to rank positive instances higher than negative instances almost all the time. Figures 7 and 8 provide a useful visual model for the accuracies seen in Tables 3-4 and Figures 3-6.

Some discrepancies and false-positives seen in both models can be attributed to tumor mass less than

3 ml, also affecting recall and precision scores described earlier. The low percentage of true-positives could also be due to neoplastic growth included in the dataset that are not tumors, like cysts. These extreme cases, however, do not seem to drastically affect scores across the board. In figures 3-6, the validation accuracies range greatly and are slightly lower than the training accuracies for each epoch. There are two plausible explanations for this discrepancy: slight overfitting of the model or data splits drastically affecting validation performance. The latter instance proves more likely due to the standard deviation (SD) of the validation step ranging greatly. Because splits were altered in order to get values, the variance in results can be attributed to splits unfavorable for validation accuracy. Overfitting is less likely due to the high quality and stability of the training step.

3.4 Attention Study

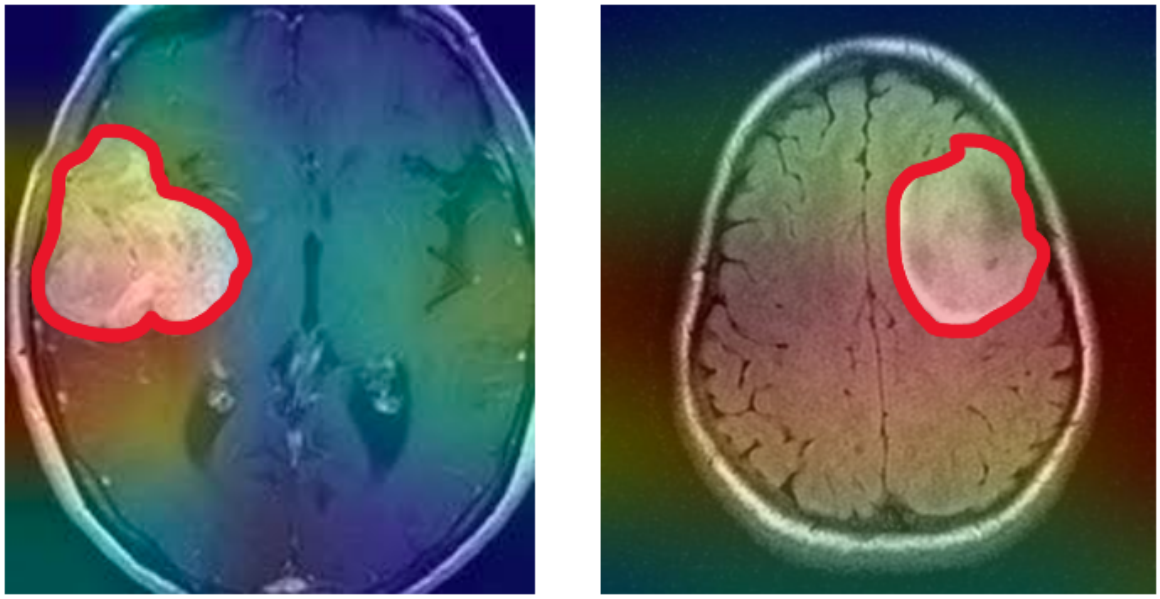

Figure 4: Images taken with GradCam TensorFlow extension, showing a heat map of the areas with the highest attention in two T1-weighted MRI images. Highest attention approaches red from blue up the RGB scale to red. Manual delineation is shown in solid red. These samples are training data in the WHO grading of meningioma task.

Attention, allocated importance during classification, in the images seem to surround the manually marked tumor mass, with the model giving most importance to these contrasting sections. This study rules out any underlying possibility of psuedo-accuracy generated from falsified attention to extraneous variables, such as barcodes or markings outside the brain mass itself. Attention patterns in this study corroborate the high accuracies noted previously.

Although attention is fairly concentrated in the image on the left, attention seems to spread across the middle of the MRI image on the right. This discrepancy can be attributed to displacement of neighboring brain mass in response to the growth of a foreign mass. Another explanation for superfluous attention would be the inclusion of outlier non-cancerous cystial images to the training and testing datasets. Cystial growth has been attributed to excess brain deformation when compared to tumor growth. Small amounts of attention could have formed due to patterns in tumor localization as well.

4 Discussion

This study investigated convolutional neural networks' application to various CT and MR modalities. The dataset included in this work includes images from various scanners from 12 different hospitals around the world. This variance in data source collection eliminates possible bias or overfitting of overtly similar data. Inclusion of a large dataset, smaller brain samples (<3ml), and extraneous samples of cystial growth in the brain makes the model more robust. Overall, the multimodal approach to

the models in this work did not cause lower scores across the board. Dice Scores, recall, precision, and accuracy all indicate high quality and reliance of both models. Data augmentation proves useful for increasing total image count and more contrast variables for classification tasks. Image preprocessing also likely contributed to the high precision and recall scores due to the lower variance in image samples.

Convolutional Neural Networks (CNNs) can be used to assist neurosurgical and radiology practitioners in accurately diagnosing and grading meningioma tumor growth in the brain. This multimodal CNN can look at CT and MRI images of different weights, contrasts, and positioning for diagnosis. During this study, tumor volumes of greater than 1 ml were used for accurate classification of tumor and sub-grading. This was accomplished through an EfficientNetB0 backbone along with a multitude of image augmentation and preprocessing methods for the data. In further studies, false-positive, true-negative, and more complex samples of meningiomas should be used in more practical classification for healthcare. Nevertheless, proper training strategies and tested attention mechanisms allowed for exemplary diagnosis accuracies and scores. The CNN was capable of segmenting valid tumor mass and did not involve other extraneous variables or biases.

Overall, EfficientNetB0 architecture provided exemplary performance with no adjustments. Training this model with native MR and CT resolutions proved as a challenge, however. Although training was done via a high-end GPU environment with the Nvidia 2080ti, high resolutions lead to a sizable memory footprint. Large datasets likely exacerbated this issue. Although down sampled MR and CT samples were used in this work, both models maintained high precision, accuracy, and recall scores. The figures displayed an increasing trend in accuracy during the validation and training steps. Furthermore, the high AUC scores all above .99 display the high quality and stability of the model, rarely predicting extreme values.

The current performances based on the scores mentioned and accuracy prove as beneficial for surgical use. Applications include but are not limited to diagnosis, surgical planning, and clinical use. Increasingly, after thorough clinical trial, machine learning models and CNNs such as those displayed in this work can be implemented into the Medical field. Although not highly complex, both classification and grading can assist improve accuracy for medical use or can also be a precursor for more specified and deeper models in the future. Improvement on this specific CNN could be to test exactly which specific images predictions were erroneous. This study could be important in both seeing which specific images the CNNs could be most useful in and making future improvements on the model. Improvements to the dataset could incorporate various different types o neoplasms that make the model more robust. Although other modalities do have their respective downsides, modalities like histology and angiography could increase accuracy for more demanding models.

Overall, this work displays how CNNs can be useful in healthcare. Although advancements are being made to ease surgery and diagnosis, they are not yet fully implemented into everyday clinics. Light-work CNNs that can be run on a standard GPU environment, like the model in this work, demonstrate how easily machine learning can be integrated into the medical field, even in countries with less access to medical funding. Similar to MYCIN, projects like this one serve as useful precursors for future machine learning projects that can have a more expansive impact on global life. As of now, this work's models display how it is possible to accurately diagnose and classify different WHO grades of meningioma with a light-work CNN.

5 Disclosures

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Rights and consent were obtained for all imaging scans of patients and external programs.
No funding was required for this research.

6 Acknowledgements

This work was written and edited alongside Harvard University PhD Doctor of Philosophy candidate Sharifa Sahai.

References

[1] Rabih Aboukais, Fahed Zairi, Jean-Paul Lejeune, Emile Le Rhun, Maximilien Vermandel, Serge Blond, Patrick Devos, and Nicolas Reyns. Grade 2 meningioma and radiosurgery. *Journal of Neurosurgery*, 122(5):1157–1162, 2015.

[2] Duncan Anderson and Mourad Khalil. Meningioma and the ophthalmologist: a review of 80 cases. *Ophthalmology*, 88(10):1004–1009, 1981.

[3] R Ashikaga, Y Araki, and O Ishida. Mri of head injury using flair. *Neuroradiology*, 39(4):239–242, 1997.

[4] Karla Batista-García-Ramó and Caridad Ivette Fernández-Verdecia. What we know about the brain structure–function relationship. *Behavioral Sciences*, 8(4):39, 2018.

[5] George Bebis and Michael Georgiopoulos. Feed-forward neural networks. *IEEE Potentials*, 13(4):27–31, 1994.

[6] H Benediktsson, T Andersson, U Sjölander, M Hartman, and PG Lindgren. Ultrasound guided needle biopsy of brain tumors using an automatic sampling instrument. *Acta Radiologica*, 33(6):512–517, 1992.

[7] Yoshua Bengio, Aaron C Courville, and Pascal Vincent. Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR, abs/1206.5538*, 1(2665):2012, 2012.

[8] Abhishta Bhandari, Jarrad Koppen, and Marc Agzarian. Convolutional neural networks for brain tumour segmentation. *Insights into Imaging*, 11(1):1–9, 2020.

[9] Rohan Bhardwaj, Ankita R Nambiar, and Debojyoti Dutta. A study of machine learning in healthcare. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, volume 2, pages 236–241. IEEE, 2017.

[10] Peter McL Black. Meningiomas. *Neurosurgery*, 32(4):643–657, 1993.

[11] Jake Bouvrie. Notes on convolutional neural networks. 2006.

[12] Robin A Buerki, Craig M Horbinski, Timothy Kruser, Peleg M Horowitz, Charles David James, and Rimas V Lukas. An overview of meningiomas. *Future Oncology*, 14(21):2161–2177, 2018.

[13] Richard B Buxton, Robert R Edelman, Bruce R Rosen, Gary L Wismer, and Thomas J Brady. Contrast in rapid mr imaging: T1- and t2-weighted imaging. *Journal of computer assisted tomography*, 11(1):7–16, 1987.

[14] Giuseppe Catapano, Francesco Giovanni Sgulo, Vincenzo Seneca, Giuseppina Iorio, Matteo de Notaris, and Giuseppe di Nuzzo. Fluorescein-assisted stereotactic needle biopsy of brain tumors: a single-center experience and systematic review. *Neurosurgical review*, 42(2):309–318, 2019.

[15] Marc C Chamberlain. Hydroxyurea for recurrent surgery and radiation refractory high-grade meningioma. *Journal of neuro-oncology*, 107(2):315–321, 2012.

[16] Sang-Hee Choi, Sooho Bae, Suk Kyeong Ji, and Moon Jong Chang. The mri findings of meniscal root tear of the medial meniscus: emphasis on coronal, sagittal and axial images. *Knee Surgery, Sports Traumatology, Arthroscopy*, 20(10):2098–2103, 2012.

[17] Antonio F Corno and Gigi P Festa. Introduction to ct scan and mri. *Congenital Heart Defects: Decision Making for Cardiac Surgery Volume 3 CT-Scan and MRI*, pages 1–17, 2009.

[18] Christopher F Dowd. Cerebral angiography: techniques and practice. *Handbook of Clinical Neurology*, 176:107–119, 2021.

[19] Shiv Ram Dubey, Satish Kumar Singh, and Bidyut Baran Chaudhuri. Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, 2022.

[20] Anthony L D'Ambrosio and Jeffrey N Bruce. Treatment of meningioma: an update. *Current Neurology and Neuroscience Reports*, 3(3):206–214, 2003.

[21] Marco Essig, Mark S Shiroishi, Thanh Binh Nguyen, Marc Saake, James M Provenzale, David Enterline, Nicoletta Anzalone, Arnd Dörfler, Àlex Rovira, Max Wintermark, et al. Perfusion mri: the five most frequently asked technical questions. *AJR. American journal of roentgenology*, 200(1):24, 2013.

[22] DGareth R Evans. Neurofibromatosis type 2 (nf2): a clinical and molecular review. *Orphanet journal of rare diseases*, 4(1):1–11, 2009.

[23] Ali-Reza Fathi and Ulrich Roelcke. Meningioma. *Current neurology and neuroscience reports*, 13(4):1–8, 2013.

[24] Melvin Field, Timothy F Witham, John C Flickinger, Douglas Kondziolka, and L Dade Lunsford. Comprehensive assessment of hemorrhage risks and outcomes after stereotactic brain biopsy. *Journal of neurosurgery*, 94(4):545–551, 2001.

[25] Mathew J Gallagher, Michael D Jenkinson, Andrew R Brodbelt, Samantha J Mills, and Emmanuel Chavredakis. Who grade 1 meningioma recurrence: are location and simpson grade still relevant? *Clinical neurology and neurosurgery*, 141:117–121, 2016.

[26] Erin Gibson, Fuqiang Gao, Sandra E Black, and Nancy J Lobaugh. Automatic segmentation of white matter hyperintensities in the elderly using flair images at 3t. *Journal of magnetic resonance imaging*, 31(6):1311–1322, 2010.

[27] Mark A Goldberg, Peter F Hahn, SANJAY Saini, MS Cohen, P Reimer, TJ Brady, and PR Mueller. Value of t1 and t2 relaxation times from echoplanar mr imaging in the characterization of focal hepatic lesions. *AJR. American journal of roentgenology*, 160(5):1011–1017, 1993.

[28] Roland Goldbrunner, Giuseppe Minniti, Matthias Preusser, Michael D Jenkinson, Kita Sallabanda, Emmanuel Houdart, Andreas von Deimling, Pantelis Stavrinou, Florence Lefranc, Morten Lund-Johansen, et al. Eano guidelines for the diagnosis and treatment of meningiomas. *The Lancet Oncology*, 17(9):e383–e391, 2016.

[29] Keith J Goulden. An introduction to the ct scan. *CT Scan: New Frontiers*, page 3.

[30] Andrew T Hale, Li Wang, Megan K Strother, and Lola B Chambless. Differentiating meningioma grade by imaging features on magnetic resonance imaging. *Journal of Clinical Neuroscience*, 48:71–75, 2018.

[31] Kendra M Hasebroock and Natalie J Serkova. Toxicity of mri and ct contrast agents. *Expert opinion on drug metabolism & toxicology*, 5(4):403–416, 2009.

[32] Ray Hashman Hashemi, William G Bradley, and Christopher J Lisanti. *MRI: the basics: The Basics*. Lippincott Williams & Wilkins, 2012.

[33] Thomas Hellström, Virginia Dignum, and Suna Bensch. Bias in machine learning–what is it good for? *arXiv preprint arXiv:2004.00686*, 2020.

[34] Irmela Herzog. Theory and practice of cost functions. In *Fusion of cultures. Proceedings of the 38th annual conference on computer applications and quantitative methods in archaeology, Granada, Spain*, pages 375–382, 2010.

[35] Samantha J Holdsworth and Roland Bammer. Magnetic resonance imaging techniques: fmri, dwi, and pwi. In *Seminars in neurology*, volume 28, pages 395–406. © Thieme Medical Publishers, 2008.

[36] Tibor Hortobágyi, János Bencze, Gréta Varkoly, Mahan C Kouhsari, and Álmos Klekner. Meningioma recurrence. *Open Medicine*, 11(1):168–173, 2016.

[37] Wasserthal Jakob Zimmerer David Petersen Jens Kohl Simon Schock Justus Klein Andre Roß Tobias Wirkert Sebastian Neher Peter Dinkelacker Stefan Köhler Gregor Maier-Hein Klaus Isensee Fabian, Jäger Paul. Mic-dkfz. https://github.com/MIC-DKFZ/batchgenerators, 2020.

[38] Michael MY Khoo, Philippa A Tyler, Asif Saifuddin, and Anwar R Padhani. Diffusion-weighted imaging (dwi) in musculoskeletal mri: a critical review. *Skeletal radiology*, 40(6):665–681, 2011.

[39] Ziya Kirkali, Theresa Chan, Murugesan Manoharan, Ferran Algaba, Christer Busch, Liang Cheng, Lambertus Kiemeney, Martin Kriegmair, R Montironi, William M Murphy, et al. Bladder cancer: epidemiology, staging and grading, and diagnosis. *Urology*, 66(6):4–34, 2005.

[40] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[41] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[42] Timothy P Lillicrap, Adam Santoro, Luke Marris, Colin J Akerman, and Geoffrey Hinton. Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6):335–346, 2020.

[43] Simon KW Lloyd and D Gareth R Evans. Neurofibromatosis type 2 (nf2): diagnosis and management. *Handbook of clinical neurology*, 115:957–967, 2013.

[44] Wei Ma and Jun Lu. An equivalence of fully connected layer and convolutional layer. *arXiv preprint arXiv:1712.01252*, 2017.

[45] Nicolas Macagno, Dominique Figarella-Branger, Karima Mokthari, Philippe Metellus, Anne Jouvet, Alexandre Vasiljevic, Anderson Loundou, and Corinne Bouvier. Differential diagnosis of meningeal sft-hpc and meningioma. *The American journal of surgical pathology*, 40(2):270–278, 2016.

[46] Hidenobu Maegawa, Kazuo Sano, Yoshimasa Kitagawa, Toshiyuki Ogasawara, Kazuki Miyauchi, Joji Sekine, and Tsugio Inokuchi. Preoperative assessment of the relationship between the mandibular third molar and the mandibular canal by axial computed tomography with coronal and sagittal reconstruction. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, 96(5):639–646, 2003.

[47] Stephen T Magill, Jacob S Young, Ricky Chae, Manish K Aghi, Philip V Theodosopoulos, and Michael W McDermott. Relationship between tumor location, size, and who grade in meningioma. *Neurosurgical focus*, 44(4):E4, 2018.

[48] Rickey Miller Jr, Michele L DeCandio, Yaenette Dixon-Mah, Pierre Giglio, W Alex Vandergrift III, Naren L Banik, Sunil J Patel, Abhay K Varma, and Arabinda Das. Molecular targets and treatment of meningioma. *Journal of neurology and neurosurgery*, 1(1), 2014.

[49] Naila Murray and Florent Perronnin. Generalized max pooling. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2473–2480, 2014.

[50] Anne G Osborn. *Diagnostic cerebral angiography*. Lippincott Williams & Wilkins, 1999.

[51] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[52] Myeongsuk Pak and Sanghoon Kim. A review of deep learning in image recognition. In *2017 4th international conference on computer applications and information processing technology (CAIPT)*, pages 1–3. IEEE, 2017.

[53] Steve H Parker, A Thomas Stavros, and Mark A Dennis. Needle biopsy techniques. *Radiologic Clinics of North America*, 33(6):1171–1186, 1995.

[54] James J Pekar. A brief introduction to functional mri. *IEEE Engineering in Medicine and Biology Magazine*, 25(2):24–26, 2006.

[55] Arie Perry. Meningiomas. In *Practical surgical neuropathology: a diagnostic approach*, pages 259–298. Elsevier, 2018.

[56] Sunder S Rajan. Mri: a conceptual overview. 1997.

[57] Jason Rockhill, Maciej Mrugala, and Marc C Chamberlain. Intracranial meningiomas: an overview of diagnosis and treatment. *Neurosurgical focus*, 23(4):E1, 2007.

[58] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[59] K Shailaja, B Seetharamulu, and MA Jabbar. Machine learning in healthcare: A review. In *2018 Second international conference on electronics, communication and aerospace technology (ICECA)*, pages 910–914. IEEE, 2018.

[60] Edward H Shortliffe. Mycin: A knowledge-based computer program applied to infectious diseases. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 66. American Medical Informatics Association, 1977.

[61] Henning Steen and Vedat Schwenger. Good mri images: to gad or not to gad? *Pediatric Nephrology*, 22(9):1239–1242, 2007.

[62] VL Stevenson, GJM Parker, GJ Barker, K Birnie, PS Tofts, DH Miller, and AJ Thompson. Variations in t1 and t2 relaxation times of normal appearing white matter and lesions in multiple sclerosis. *Journal of the neurological sciences*, 178(2):81–87, 2000.

[63] Alexey Surov, Sebastian Gottschling, Julia Bolz, Malte Kornhuber, Alex Alfieri, Hans-Jürgen Holzhausen, Jasmin Abbas, and Sabrina Kösling. Distant metastases in meningioma: an underestimated problem. *Journal of neuro-oncology*, 112(3):323–327, 2013.

[64] Daniel Svozil, Vladimir Kvasnicka, and Jiri Pospichal. Introduction to multi-layer feed-forward neural networks. *Chemometrics and intelligent laboratory systems*, 39(1):43–62, 1997.

[65] AF Thornton Jr, HM Sandler, RK Ten Haken, DL McShan, BA Fraass, ML LaVigne, and BR Yanks. The clinical utility of magnetic resonance imaging in 3-dimensional treatment planning of brain neoplasms. *International Journal of Radiation Oncology* Biology* Physics*, 24(4):767–775, 1992.

[66] William Van Melle. Mycin: a knowledge-based consultation program for infectious disease diagnosis. *International journal of man-machine studies*, 10(3):313–322, 1978.

[67] Mike P Wattjes. Structural mri. *International psychogeriatrics*, 23(S2):S13–S24, 2011.

[68] Ian R Whittle, Colin Smith, Parthiban Navoo, and Donald Collie. Meningiomas. *The Lancet*, 363(9420):1535–1543, 2004.

[69] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4):611–629, 2018.

[70] Dingjun Yu, Hanli Wang, Peiqiu Chen, and Zhihua Wei. Mixed pooling for convolutional neural networks. In *International conference on rough sets and knowledge technology*, pages 364–375. Springer, 2014.

[71] Chen-Lin Zhang and Jianxin Wu. Improving cnn linear layers with power mean non-linearity. *Pattern Recognition*, 89:12–21, 2019.

[72] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for neural networks for image processing. *arXiv preprint arXiv:1511.08861*, 2015.