Student: Stanislav Hakkarainen
LT2326 Machine Learning for Statistical Data Linguistics
Professor: Asad Basheer Sayeed

Project Final Report
"Brain Tumor Detection Model"

ABSTRACT

Ever since the creation of the first camera, one was allured to taking images for various reasons: to keep them as souvenirs, to hold on to, or to depict an entity and freeze it in a frame to study later (Rose, 2000). Nowadays, there is a great number of images and even a greater number of ways of taking a picture. One of them is an MRI machine (Dreizen, 2004, p.78). The discovery of the MRI, which stands for Magnetic Resonance Imaging, took place in two stages. Firstly, it was proposed to benefit from using external nuclear magnetic resonance (NMR) to scan human tissue hoping to find the difference between normal and cancer tissue. Afterward, Lauterbur, Mansfield, and Damadian contributed to the development of a better way of depicting the insides of a human by using magnetic field gradients. Consequentially, the first MR image of a human saw the light of day in 1972 (Dreizen, 2004, p.78).

Since the first successful MRI image, the desire to get the quality of the images produced by the system only arose. In the era of technological advances, the field of MR image-taking has developed and prospered to such heights that it is almost impossible to have a hospital without an MRI machine (Plewes & Kucharczyk, 2012, p.1039). Without any doubt, a magnetic resonance image facilitated the development of modern medicine by allowing professionals to quickly and effortlessly detect a fast amount of physical or physiologic phenomena residing in one's body. One of the benefits of getting an MRI is the non-invasive approach to tissue scanning. Moreover, an MRI could be accompanied by such modifications as contrast reflecting proton density, fields of motion, tissue perfusion, and many more (Plewes & Kucharczyk, 2012, p.1039).

Even though MR imaging could be used on the entire patient's body, one part of it that gets more attention is using MRI for detecting brain tumors (Sudputa & Bandyopadhaya, 2012, p. 477). Having a brain tumor refers to the notion of having a collection of neoplasms. It is a body of tissue that is known for growing rapidly and without any control (DeAngelis, 2001; Sudputa & Bandyopadhaya, 2012).

In this report, a topic regarding brain tumor detection and its classification is challenged. The report raises such notions as what a brain tumor is, what its characteristics could be, and how it could put one's life at risk. The report also highlights the development of a CNN-powered machine learning model that is trained on a vast amount of magnetic resonance images to detect and classify a patient's brain MRI in order to detect whether a certain MR image contains a brain tumor or not. Additionally, if a selected image does in fact contain a brain tumor, then a question of successful type classification arises. The model ought to be able to classify the type of tumor amongst the three given options: Glioma, Meningioma, and Pituitary tumor. The Brain Tumor Detection Model produces several metrics that help detect its accuracy over the course of model training and validation. The outcomes of the model training are highlighted and discussed in the report as well. Additionally, space to discuss possible setbacks of the model, as well as its future modifications that could improve and/or broaden up the range of the model's use has also been allocated. Moreover, the report puts forth the

idea of using machine learning image classification models in the field of medicine to gain faster results, prognosis, and course of treatment for the patient.

INTRODUCTION

In the era of technological advances, the field of medicine is one of the most fast-paced fields of development. Every field of it strives for new achievements and tools to help professionals to diagnose a patient faster (Fanu, 1999). Consequently, the field of neurology is not an exception. Neurology is a field of medicine that mainly targets a patient's brain. One of the main distinctions of the following field is the emphasis on localization and phenomenology. In order to put forward an effective course of treatment, one ought to undergo a series of precise medical examinations (Daroff & Bradley, 2015). One of the main cases that the aforementioned field has to deal with is tumors, brain tumors in particular.

As it has been stated above, a tumor is a collection of neoplasms in various parts of the human body. A brain tumor can be further classified based on the following criteria: origin, enlargement, and malignancy. Mostly, brain tumors commence brain cells or the tissue around a patient's brain. In some severe cases, a brain tumor can arise from cancer cells in other parts of one's body (Sudputa & Bandyopadhaya, 2012). However, as it has been previously believed that tumors consist entirely of cancer cells, it has been proven that not only "corrupted" cells can form a tumor, but also "normal" neighboring cells can be drawn into it as well (Ariztia et al., 2006).

According to the American Cancer Society, approximately 16,800 cases of brain tumors were detected in 1999. Additionally, it has been stated that around 13,100 people suffered from primary cancer of the CNS (Central Nervous System), which resulted in premature deaths (DeAngelis, 2001, p. 114). There are several factors for the successful brain tumor detection at play. One of the most crucial factors is the location and the rate of growth of the tissue. Additionally, specialists must be on the lookout for the displacement of neural structures and the presence of additional pathologies, such as edema, hemorrhage, vascular compromise, and cerebrospinal fluid obstruction (Alentorn et al., 2016).

In order for a professional to successfully detect a tumor, one has to undergo various examinations, and one of them may be getting an MRI. An MRI enables one to get soft tissue contrast of high spatial resolution in a non-invasive way (Plewes & and Kucharczyk, 2012, p. 1039). It is crucial to underline the significance of MRI's approach, as brain tumors can appear in any part of a patient's brain, and it is arduous to predict its size without slicing one's brain.

A vast majority of professionals would agree with the statement that it is pivotal for a patient's well-being and course of treatment to detect an abnormality as soon as possible before the patient crosses the line of being terminal (Ariztia et al., 2006). Alas, the field of medicine in various countries faces shortages of professionals, which leads to the possibility of not having enough workforce to attend to all the patients. According to the data presented by the World Health Organization (WHO) in 2014, 57 countries are considered to be in a health crisis (Hoyler et al.,2013). Therefore, a question of how to accelerate the ability to give a faster prognosis arises.

One of the solutions to this dilemma might be the use of machine learning techniques to classify and analyze MRI of patients with a view to assigning a course of treatment.

The objective of this paper is to dwell on the use of a machine learning model to predict the

possibility of a brain tumor on a patient's MRI and classify it according to three available options: Glioma Tumor, Meningioma Tumor, and Pituitary Tumor. The model ought to perform two classifications: binary and multiclass. The first one is supposed to detect whether an MRI contains signs of a tumor or not. Consequentially, if the image in fact contains a brain tumor, then the model is supposed to classify the type of tumor. The main idea behind the implementation of image classification ML models in the field of medicine is to advance the possibility of early anomaly detection and to reduce the cost of a patient's medical treatment.

DATA RESOURCES

For the model to be successfully trained, a question of getting relevant data has been raised. It has been decided to search for MR images on Kaggle.com. In particular, a dataset named "Brain Tumor Classification (MRI)" by a user "sartaj" has been chosen as a potential dataset for the model's training. The dataset contains 3264 magnetic resonance images that present three main types of brain tumors: Glioma Tumor, Meningioma Tumor, and Pituitary Tumor. Additionally, the dataset contains images of brain scans that do not show qualities of containing any kind of tumor.

After the acquisition of the dataset, the next step is to prepare the data for the model to be trained and validated on. The dataset has been restructured in the following way: all the images have been split into an 80/20 ratio and compartmentalized into training and testing folders. Both folders contain two more folders that are used separately for binary and multiclass classifications. Each of the aforementioned folders contains MRIs that are separated in a specific way. The binary classification folder only contains MR images of brains with no tumor and with tumor, but they are not separated into what kind of tumor is presented in the image, whereas the multiclass classification folder contains only images of brains with tumors separated based on the factor what of kind of tumor is presented in the MRI. Having completed all the steps, the dataset is ready to be used on the model.

METHODS

For the sake of the experiment and potential learning, it has been decided to create a new machine-learning model instead of using the ones that have already been created and fine-tuned. The following part of the paper elucidates every aspect of the created model. As it has been mentioned before, the model supports two stages of classification: Binary Classification (BC) (Tumor vs. No Tumor) and Multiclass Classification (MC) (Glioma Tumor, Meningioma Tumor, and Pituitary Tumor). Once the dataset is ready to be used on the model, it is loaded and preprocessed. The "ImageGenerator" presented by Keras is used for such a task. Additionally, the images in the dataset are augmented, which means that such techniques as rotation, width and height shifting, shearing, zooming, and horizontal flipping are applied to create variations of the used data. The decision behind using image augmentation on the dataset lies within the inability to acquire a bigger dataset for the training and validation loops, and to prevent the model from overfitting (Shorten & Khoshgoftaar, 2019). Additionally, the model is accompanied by the class weights calculation. The calculation is used both in BC and MC. Class weights are primarily used to address the issue of class-imbalanced data by giving more importance to the underrepresented classes during the training loop (Cardie & Nowe, 1997).

When it comes to the model's architecture, it has been decided to use a pre-trained VGG16 as the base model in order to minimize the amount of time and GPU resources to train the model.

VGG16 is a convolutional neural network model that has been deemed popular for its high performance in the field of image classification. The model's architecture contains the following characteristics: input layer, convolutional layers, max-pooling layers, fully connected layers, and output layers (Simonyan & Zusserman, 2014). Additionally, supplemental dense layers on top of the VGG16 base model were added for the sake of better learning of specific features. The model also uses a learning rate scheduler, which specifies how the learning rate should be changed over the course of training. Dropout layers are also added to the model's architecture out of regard for overfitting. Moreover, the L2 regularization, also known as weight decay, is applied to the model's dense layers. The following regularization is also responsible for the control of overfitting by adding a penalty to the loss function based on the magnitude of the weights (Van Laarhoven, 2017).

For binary classification, the model is executed with the binary cross-entropy loss, whilst for multiclass classification, it has been decided to implement the categorical cross-entropy loss. Furthermore, the Stochastic Gradient Descent (SGD) optimizer has also been added to the training loop. Model training is performed separately for BC and MC and the data is presented to the training loops in batches using the data generators.

As soon as the training loops are finished, the model's performance is evaluated on the test data using various metrics, such as accuracy, ROC AUC, Matthews Correlation Coefficient (MCC), Cohen's Kappa, and classification reports. Speaking particularly about classification reports, such information as precision, recall, F1-score, and support are presented there.

RESULTS

By taking a look at the results of the model, it is possible to determine how well it performs with the given dataset. Talking about the results of the binary classification first (see Appendix 1, Table 1,2), the overall accuracy score is 58.88%, which is a fairly good result for the model. The results indicates that there is an almost 60% chance for the model to determine whether a patient has any sort of brain tumor or not. The precision of the BC is 0.45 for "No Tumor" and 0.57 for "Tumor" respectively. Recall measures the ability of the model to correctly identify instances of a specific class given, and in the model's case, the results are 0.69 for "No Tumor" and 0.64 for "Tumor". The final metric in the classification report is the F1-score, which is the mean of precision and recall. In terms of the Binary Classification model, the results are 0.54 for "No Tumor" and 0.56 for "Tumor", which shows that the balance between precision and recall is better in the "Tumor" class. Additionally, it is worth looking at secondary metrics that have been calculated for BC. The Area Under the Receiver Operating Characteristic Curve is a measure taken to distinguish between the positive and negative classes. The AUC-ROC for BC is 0.56, which indicates that the ability of the model to differentiate between classes is slightly higher than a random chance. The next metric is the Matthews Correlation Coefficient, which is a measure of the quality of the BC model, taking into account true and false positives and negatives. The results of the metrics are 0.51, which indicates that the model is slightly better than a random classifier. The last produced metric for the Binary Classification Model is Cohen's Kappa, which is the agreement between the model's predictions and the actual data. The results of the mentioned metric are 0.48, which indicates a slightly low but still present agreement.

Lastly, once the results of the Multiclass Classification are produced (see Appendix 1, Table 3.4), it is possible to determine how well the model detects various kinds of brain tumors. The first metric is accuracy. For each type of tumor, the accuracy results are 56.33%, 47.64%, and 45,61%, and they indicate that the model deals the best with the Glioma tumor, however, the results would still not be considered to be perfect. Speaking about precision, the results are the following, 0.45 for the Glioma tumor, 0.48 for the Meningioma tumor, and 0.49 for the Pituitary tumor. The results indicate that the model performs the best with the Pituitary tumor. The next metric is recall: 0.89, 0.8, and 0.75. The last metric in the MC report is the F1-score, which is 0.51 for the Glioma tumor, 0.49 for the Meningioma tumor, and 0.5 for the Pituitary tumor. The F1 scores show that the classes are relatively close, which indicates a balanced performance in terms of precision and recall.

Taking additional metrics into account, the MCC score is 0.56, which is slightly above a random choice and Kohen's Kappa is 0.51 respectively.

ANALYSIS

By interpreting the results of the model, it is possible to say that there is a 60% likelihood that the model is able to predict whether a person has a tumor or not and, potentially, it is able to predict which kind of tumor, although the precision of the latter prediction falls notably below that of general binary classification.

By summarizing the general performance of the model, one could conclude that whilst the model still produces a reasonably competent result, it has not reached the level of fruition suitable for serving as a definitive diagnostic tool for medical purposes.

Additionally, it is worth highlighting the challenges that could be the reason for the model's poor performance. One of the reasons is the model's poor fine-tuning. Picking parameters for image classification models is a delicate job and it takes an undeniably big amount of time to find the right parameters and settings so that the model achieves higher results. Another setback that was faced during the creation of the model was the scarcity of the data to train and validate the model on. Alas, it is rather onerous to find MR images without tickling ethical concerns, thus it has been decided to proceed with the number of images that was already gathered.

CONCLUSION

Over the course of developing the model, it became possible to detect which techniques work for the brain tumor classification task better and which should be left behind. It also became possible to practice various approaches to data augmentation and collection in order to gain the best results possible.

Notwithstanding the modest performance exhibited by the model, it nonetheless is a contribution to the field of medicine, as by developing more advanced image classification machine learning models and obtaining more recent data for its training and validation, it is possible to boost the field of neurology by locating tumors faster and starting the patient's treatment sooner in order to amplify their faster recovery.

Additionally, it is pertinent to underscore that it is worth trying to improve the model by using other approaches. One of the approaches that should be tried is using the ViT model from HuggingFace. Ever since the model was released, it has shown impeccable results in the field of binary and multiclass classifications, which may lead one to wonder, whether it could also be used to improve

the results of not only brain tumor classification but any other type of tumor in one's body.

References:

1. Alentorn, A., Hoang-Xuan, K., & Mikkelsen, T. (2016). Presenting signs and symptoms in brain tumors. In *Handbook of Clinical Neurology*(pp. 19–26). https://doi.org/10.1016/b978-0-12-802997-8.00002-5
2. Ariztia, E. V., Lee, C. J., Gogoi, R., & Fishman, D. A. (2006). The tumor microenvironment: key to early detection. *Critical Reviews in Clinical Laboratory Sciences*, *43*(5–6), 393–425. https://doi.org/10.1080/10408360600778836
3. Cardie, C., & Nowe, N. (1997). Improving minority class prediction using Case-Specific feature weights. *International Conference on Machine Learning*, 57–65. https://dblp.uni-trier.de/db/conf/icml/icml1997.html#CardieN97
4. Daroff, R. B., & Bradley, W. G. (2015). *Bradley's Neurology in clinical practice*. http://ci.nii.ac.jp/ncid/BB10756208
5. DeAngelis, L. M. (2001). Brain tumors. *The New England Journal of Medicine*, *344*(2), 114–123. https://doi.org/10.1056/nejm200101113440207
6. Dreizen, P. (2004). The Nobel prize for MRI: a wonderful discovery and a sad controversy. *The Lancet*, *363*(9402), 78. https://doi.org/10.1016/s0140-6736(03)15182-3
7. Fanu, J. L. (1999). Rise and fall of modern medicine. *The Lancet*, *354*(9177), 518. https://doi.org/10.1016/s0140-6736(05)75559-8
8. Hoyler, M., Finlayson, S. R., McClain, C. D., Meara, J. G., & Hagander, L. (2013). Shortage of Doctors, Shortage of Data: A review of the Global Surgery, Obstetrics, and Anesthesia Workforce literature. *World Journal of Surgery*, *38*(2), 269–280. https://doi.org/10.1007/s00268-013-2324-y
9. Plewes, D. B., & Kucharczyk, W. (2012). Physics of MRI: A primer. *Journal of Magnetic Resonance Imaging*, *35*(5), 1038–1054. https://doi.org/10.1002/jmri.23642
10. Rose, G. (2000). Practising photography: an archive, a study, some photographs and a researcher. *Journal of Historical Geography*, *26*(4), 555–571. https://doi.org/10.1006/jhge.2000.0247
11. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0197-0
12. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for Large-Scale image recognition. *Computer Vision and Pattern Recognition*. http://export.arxiv.org/pdf/1409.1556
13. Sudipta, R. S., & Bandyopadhyay, S. B. (2013). Detection and Quantification of Brain Tumor from MRI of Brain and it's Symmetric Analysis. *International Journal of Information and Communication Technology Research*, *2*(6), 477–483.
14. Van Laarhoven, T. (2017). L2 Regularization versus Batch and Weight Normalization. *arXiv (Cornell University)*. https://arxiv.org/pdf/1706.05350.pdf

Appendix 1:

Label/Metric	Precision	Recall	F1-score	Support
No Tumor	0.45	0.69	0.54	105
Tumor	0.57	0.64	0.49	289
Accuracy	-	-	0.56	394

Metric	Score
AUC-ROC	0.5685
MCC	0.51
Cohen's Kappa	0.4892

Table 1,2: Results of the Binary Classification

Label/Metric	Precision	Recall	F1-socre	Support	Accuracy
Glioma Tumor	0.453265	0.89	0.5159	100	56.33%
Meningioma Tumor	0.48762	0.8	0.4913	115	47.64%
Pituitary Tumor	0.492795	0.75	0.5015	74	45,91%

Metric	Score
MCC	0.56
Kohen's Kappa	0.5143

Table 3,4: Results of the Multiclass Classification