Developing a Convolutional Neural Network Based Tool to Aid in the Diagnosis of Brain Tumors in a Rural Context

Maxwell Klaiman1,2

1. Northwestern University School of Professional Studies

Master’s of Science in Data Science Program

1. Church St, Evanston, IL 60201

2. Allina Health System

2925 Chicago Avenue, Minneapolis, MN 55407

Address to which correspondence should be addressed:

[MaxwellKlaiman2023@u.northwestern.edu](mailto:MaxwellKlaiman2023@u.northwestern.edu)

**Abstract**

Diagnosis of brain tumors – the most aggressive and lethal form of adult cancer – presents a unique challenge to physicians and machine learning algorithms alike. The presence of over 150 types of brain cancer, each with its own treatment plan and genetic profile, necessitates rapid and accurate diagnoses. In a rural setting, which lacks consistent access to specialized medical practitioners, this diagnostic process is often delayed, resulting in a corresponding loss of overall survival in brain tumor patients from rural zip codes. Convolutional Neural Networks (CNNs) have sought to leverage existing medical imaging and histopathological slides to aid physicians in making brain tumor diagnoses with great success, but the utility of these algorithms is currently suboptimal, where diagnoses cannot be made beyond high- versus low-grade tumors.

Although this information is useful, a more specific tumor grade will have higher implications for initial treatment, prognosis, and follow-up treatment. Seeking to improve imaging analysis tools, multiple deep CNN-based algorithms were developed and analyzed to determine the capacity of these classifiers to grade brain tumors. These models were trained and tested on an amalgamation of brain tumor Magnetic Resonance Imaging (MRI) datasets, and compared to both state-of-the-art neural network models and other machine learning techniques, showing up to 95% accuracy in diagnosing tumors from MRIs. Successful development of a model which can diagnose tumors may have a crucial impact on improving rural brain tumor diagnosis, and ultimately patient survival.

**Keywords:** Convolutional Neural Network, Deep Learning, Magnetic Resonance Imaging, Artificial Intelligence, Brain Tumor, Glioblastoma, Rural Medicine

1. **Introduction**

Brain tumors are among the most aggressive and lethal cancers, affecting nearly 25,000 people each year in the United States (cancer.net). Most common amongst those diagnosed tumors is glioblastoma (GBM), a terminal cancer with a 5-year survival rate of 6.9% (braintumor.org). However, other tumor types, such as a low-grade pilocytic astrocytoma, have 10-year survival rates over 90% (Park, et al). With such a variable difference in prognosis, specific diagnoses of brain tumor type, including histopathological grading, molecular profiling, and imaging analysis are all performed in search of a precise diagnosis. Yet due to the labor of conducting these analyses, a patient at a state-of-the-art medical center may have to wait 6-8 weeks to have a confirmed diagnosis – adding to patient agony as well as needlessly delaying life-extending treatment (Redlich Dunsmar and Brody 1948).

This months-long delay can be partially addressed by the implementation of AI-derived tools to aid hospitals in making diagnoses from available information. All neurological workups involve obtaining magnetic resonance imaging (MRI), which can help physicians understand the type of health problem a patient may be experiencing. Convolutional Neural Networks (CNNs) can be developed to analyze possible tumor magnetic resonance imaging (MRI) conducted in the triaging of possible brain tumors. These CNNs may be able to match or surpass the diagnostic accuracy of a general practitioner, and can aid a physician in the diagnosis of a brain tumor from an MRI without consulting a specialist.

This issue is only furthered in a non-urban setting, where access to healthcare facilities, let alone specialized neuro-oncologists or surgeons, may be limited. Delevar, et al. (2019), investigated the difference between outcomes by demographics. The results were stark: in the case of glioblastoma, median overall GBM survival was 14 months in an urban setting, compared to 11 months in a rural location. This is exacerbated further in the most-rural and most-urban zip codes, where overall survival was 9 months and 15 months, respectively. It is no stretch to posit that a major contributor to that difference is a delay in diagnosis, with a corresponding delay in receiving treatment. Even if rural patients are airlifted to neurosurgeons as soon as a diagnosis is confirmed, by that point, months of life may be lost. To provide rural hospitals and practitioners with a validated, CNN-based, diagnostic tool may have a crucial impact in connecting rural patients to care faster – possibly extending overall survival significantly.

1. **Literature Review**

Existing algorithms are able to accomplish impressive diagnostic and predictive tasks in glioblastoma specifically. Chatterjee et al. (2022) conducted an impressive CNN-driven analysis of 3-D, volumetric, magnetic resonance (MR) images to identify high- and low-grade gliomas with astounding results – achieving nearly 97% accuracy in determining whether a tumor was of a high grade, such as a GBM, or low-grade. The authors also investigated the impact of pre-training a CNN on non-pathological MR images, which further augmented the success of their algorithms. Additional tools, especially those developed by the University of Pennsylvania, allow for prediction of key outcomes, such as recurrence location, overall survival, and even histological makeup of affected cells (i.e. active tumor vs. necrosis) (Menze et al. 2015).

These existing algorithms each have a shared pitfall: a lack of utility in a non-complex medical environment, such as a community hospital in a rural area. In the case of Chatterjee, a specialized 3D volumetric MRI protocol is required, yet many underfunded rural hospitals lack the capacity to conduct this state-of-the-art imaging, which only became prevalent in 2018 (Shakoor et al. 2018). Almost all rural centers, however, have the capacity to conduct standard MR imaging, which contains several imaging modalities.

Further consideration must be given to the subtype of MRI being performed. In tumor diagnosis, 5 main imaging subtypes are conducted: T1, T1-contrast enhanced (T1ce), T2, FLAIR, and Perfusion. Chatterjee utilized only T1ce images, but perhaps other imaging modalities offer equivalent or improved performance, and a mixed approach (i.e. utilizing T1ce and FLAIR images) would yield better results. Rural hospitals may lack the capacity to conduct the newly-developed Perfusion technique, which would preclude the use of that imaging subtype in any CNN-based analytic tool for implementation in a rural hospital.

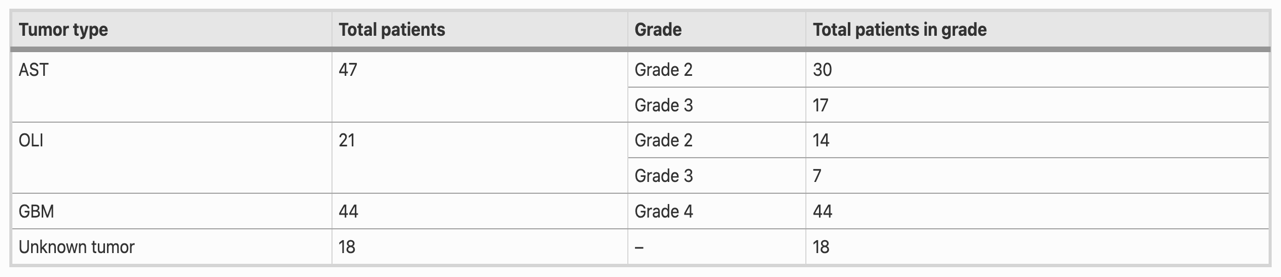
In the multitude of attempts to classify various aspects of brain tumors, a wide variety of methods have been employed (Irmak 2021). Undoubtedly, the majority of methods rely on convolutional neural network (CNN) derived techniques. To begin, popular state-of-the-art CNN models, including AlexNet, Inception v3, VGG-16, ResNet-50 and GoogleNet have been deployed to attempt tumor classification with varied results. Due to the specialized nature of this task, it is perhaps apparent that a specialized CNN could be more successful. Badza and Barjaktarovic (2020) describe a 22-layer deep convolutional neural network that was able to achieve nearly 97% success in a basic brain tumor classification task (determining tumor type). Other specialized CNN-based approaches have had similar success, such as Ayadi et al. (2021) who was able to outperform Badza and Barjaktarovic by reducing kernel size from 9x9 to 3x3. This is a key finding – where a higher-powered magnification is able to produce better accuracy results. The drawback of this is higher computational power, which may require resources that are not present in rural hospital systems, so investigation is still needed into the computation cost to accuracy benefit ratio of using a smaller kernel size. Finally, in S. Kevin Zhou’s book *Deep Learning for Medical Image Analysis*, the unbounded rectifier linear unit (ReLU) is recommended as the optimal activation function to leverage, which is confirmed by the use of ReLU exclusively in the top-performing algorithms.

1. **Data**

There does not exist a singular brain tumor image dataset that lends itself nicely to MRI analysis. As such, an amalgamation of publicly available datasets containing the same imaging modalities will be created for the sake of CNN development. Since Chatterjee et al. showed that pre-training of the neural network on a non-pathological dataset showed improvement in accuracy, the IXI non-tumor MRI dataset will be employed for that purpose.

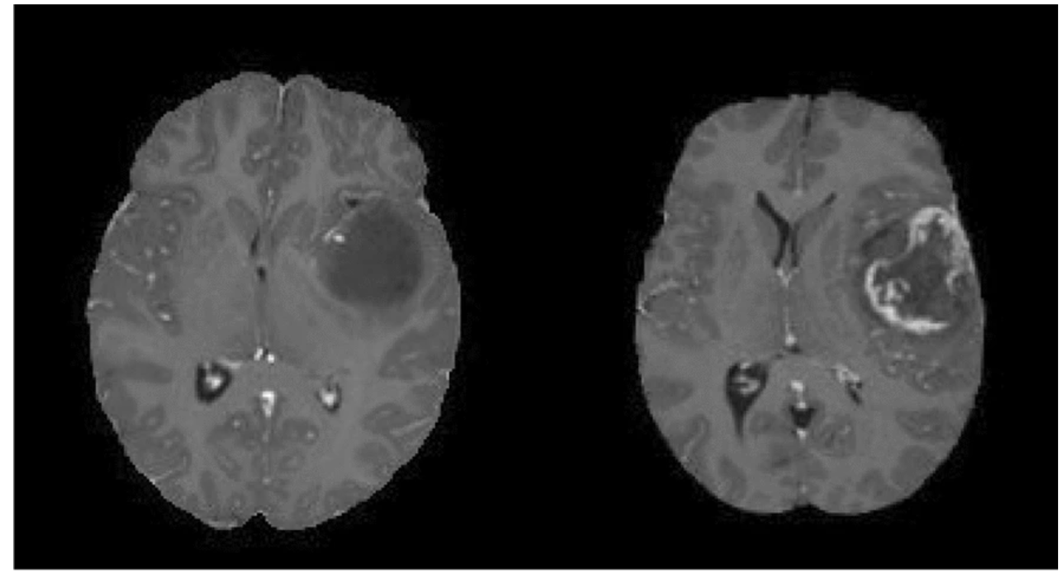
In looking at brain tumor MRI datasets, a couple of selection parameters must be employed to ensure that data is consistent across the board. First, data must contain specific MRI modalities, including T1, T1-contrast enhanced, T2, and FLAIR sequences. Due to many rural hospitals lacking the capacity to perform perfusion imaging, that imaging type will not be included in the dataset. Additionally, the dataset contains information only about tumor type (GBM, Astrocytoma, or Oligodendroglioma) and grade (grades II, III, or IV). From there, additional data modifications (i.e. classifying grades III and IV into “high grade”) are performed ad hoc. Finally, images must be substantial enough that inclusion has an impact on the overall dataset. For this reason, imaging sets must exceed 50 MRIs.

Specifically, the following brain tumor imaging datasets will be used. First, the REMBRANDT Dataset, containing 130 brain tumor images across grades II, III, and IV, will be used in the dataset. A breakdown of the patient demographics is below:



The 18 unknown tumors will not be excluded from any training or testing data, but may be retained for future algorithm validation once those tumoro types become available.

Additionally, the BRaTS dataset, which is managed and curated by the University of Pennsylvania, will be utilized in this database. The BRaTS 2020 dataset contains images of low- and high-grade gliomas, of which an example is included in Image 2 below. The high-grade glioma is on the right panel of this T1-contrast enhanced image.



1. **Methods**

In an attempt to most-accurately classify brain tumor images, a broad approach must be taken to address the options that have the capacity to be most successful. These modeling approaches can be categorized into two areas: (1) proven algorithms and (2) specialized algorithms. In each case, at least one modelling approach will be made. For the first cases, existing ImageNet and AlexNet models, which have been proven to be successful in both brain tumor image classification and a wide variety of other computer vision tasks, will be implemented. These models have already demonstrated accuracy in the high-90s in simpler classification tasks (i.e. in high-grade vs. low-grade glioma), but are generally unproven in grading tumors.

The second class of algorithms that will be investigated are those which are specifically designed to complete this task, and do not fall into a standard model (such as AlexNet or VGG-16). In line with Irmak’s 2021 paper, these algorithms could be capable of achieving the highest success rates. For this analysis, three separate CNNs will be developed. Since it has been shown that a smaller kernel size (i.e. 3x3) outperforms larger alternatives, each model will have a minimum 3x3 filter size. Additional investigation will occur in the presence of fully-connected layers, convolutional layers, and max-pooling layers. The proposed structure of each of the algorithms is below:

1. Kernel size: 5x5. Total Layers: 18. Activation Function: ReLU
2. Kernel size: 3x3. Total Layers: 18. Activation Function: Leaky ReLU
3. Kernel size: 3x3. Total Layers: 22. Activation Function: ReLU

Additionally, based on the results, an optional fourth model may be created with the parameters that result in the highest-performing algorithm. For example, if model 3 outperforms model 1, a Leaky ReLU activation function would be used in the fourth model. While it is not expected that a larger kernel size (model 1) would outperform a higher-powered 3x3 kernel, the computational cost is substantially lower and comparable results could show benefits in future model development. Should the larger kernel size perform comparably, future development should be focused on algorithms with lower computational expense. It would be expected that these specialized models are able to outperform the more general AlexNet and ImageNet models. However, if this is not the case, future investigation should be directed to those models.

1. **Results**

*Data Preparation and Model Development*

Initially, MRIs were obtained, processed, and formatted into 4 CSVs (one for each specific imaging modality). All MRIs have a 240 x 240 x 155 image shape, these CSVs contained many patient’s worth of MRIs, each having 8928000 rows of information. The resulting CSVs were nearly 10 GB in size, or 40 GB altogether. The massive size of these databases proved to be a hurdle, which many personal computers may not be able to overcome. It was not feasible to assess tumor type using more than one imaging modality. As such, each imaging modality was run through the first AlexNet CNN to determine the optimal MRI subtype – through which all future model training occurred.

In all, five models were developed to attempt characterization of brain tumor MRIs. These models were tasked with assessing the (1) grade and (2) tumor type from each MRI. In general, the models can be described as such:

Model 1: AlexNet. 19 total layers with 9,201,674 total parameters.

Model 2: ImageNet. 22 total layers with 134,356,290 total parameters.

Model 3: AlexNet-derived. Same structure, but with a Leaky ReLU Activation Function

Model 4: AlexNet-derived. Same activation function, but with a 5x5 kernel size

Model 5: AlexNet-derived, but with 3 additional layers and 13,270,026 total parameters.

*Imaging Modality Determination*

First, analysis was done to assess the magnetic resonance imaging modality that is most effective for brain tumor diagnoses. T1, T1-contrast enhanced (T1ce), T2, and FLAIR sequences were run through the AlexNet CNN. Table 1, below, shows the overall accuracy scores for each MRI sequence:

|  |  |
| --- | --- |
| MRI Type | Maximum Accuracy |
| **T1** | 83.05% |
| **T1 Contrast Enhanced** | 89.83% |
| **T2** | 79.31% |
| **FLAIR** | 91.53% |

From Table 1, it was determined that FLAIR MRI sequences would be the most useful in predicting brain tumor diagnoses, and all future results indicate accuracy from FLAIR MRIs.

*Model Performance*

Finally, the five CNN models were trained on each of the FLAIR MRI sequences compiled from the dataset. There were 6 different tumor types that could be classified:

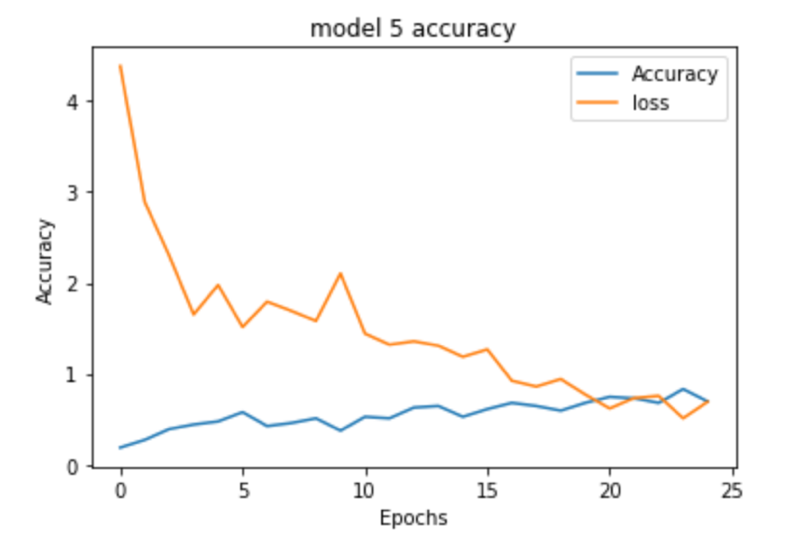
* Grade 2 Astrocytoma
* Grade 3 Astrocytoma
* Grade 4 Astrocytoma
* Grade 2 Oligodendroglioma
* Grade 3 Oligodendroglioma
* Grade 4 Glioblastoma

Table 2, below, summarizes the results of all tumor classification efforts:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model Name | Kernel Size | Activation Function | Number of Layers | Number of Epochs | Maximum Accuracy | Val Accuracy | Approx. Runtime |
| AlexNet | 3x3 | ReLU | 19 | 25 | 91.53% | 60% | 550 s |
| ImageNet | 3x3 | ReLU | 22 | 25 | 94.92% | 60% | 1625 s |
| Test 1 – Leaky Alex | 3x3 | Leaky ReLU | 19 | 25 | 89.83% | 60% | 475 s |
| Test 2 – Simplified ImageNet | 5x5 | ReLU | 22 | 25 | 83.05% | 40% | 450 s |
| Test 3 – Expanded Alex | 3x3 | ReLU | 22 | 25 | 83.05% | 40% | 500 s |

Overall model accuracies ranged from 83.05% up to 94.92%, with the ImageNet model having the highest total accuracy and tied for the highest validation accuracy.

Image 1 shows the general accuracy and loss over the 25 training epochs. While this image contains only the information pertinent to Model 5 (Test 3), each of the training models followed a similar pattern, starting with high loss and low accuracy, and rapidly approaching 80% accuracy with loss around 0.75:



1. **Analysis and Interpretation**

In this 5-model design of experiments, several interesting trends become apparent. Perhaps of greatest importance is the capacity of specialized CNNs to interpret brain MRIs and put forth a likely diagnosis. With nearly 95% accuracy, the ImageNet CNN is able to most accurately diagnose a patient’s cancer from a FLAIR MRI sequence, but AlexNet (91%) and one specialized CNN (89%) are able to perform at a similar level.

In looking at the imaging modalities to determine the optimal MRI subtype for analysis, FLAIR does have a non-trivial accuracy benefit over the other three subtypes. As such, it is reasonable to consider the use of FLAIR imaging sequences to conduct future analyses. Interestingly, the utility of the T2 sequence alone was quite low. Radiologically, the T2 sequence plays a key part in guiding diagnoses (along with the other imaging subtypes), but it appears to have diminished capacity as a diagnostic tool on its own.

Perhaps the most useful analysis can be conducted on the three test cases, each which pertains to a hyperparameter modification or minor structural change. The AlexNet and ImageNet CNNs – the two highest performing algorithms – were not easily improved upon by modification of certain parameters. Fundamentally, the parameters that were modified were (1) number of layers, (2) kernel size and (3) activation function. The increase in kernel size saw the expected effect on both model runtime and accuracy. When the kernel is increased from 3x3 to 5x5, the accuracy of the model drops by nearly 12% over the ImageNet model. Logically, this makes some sense, as a larger kernel size is going to have a lower capacity to pick out microscopic features in the image. One upside of the larger kernel size is a diminished runtime of the algorithm, but the accuracy cost of such a model is too great to consider clinical implementation.

Additional feature modification provided some interesting insight into designing optimal model performance. While it may seem apparent, the Test 3 model, which added additional layers to the AlexNet architecture to determine the impact of additional training parameters, did not outperform the simpler AlexNet CNN. As such, it is no stretch to posit that the success of a CNN is not directly correlated to the number of layers present in the model, and in many cases a more complex model will not guarantee more accurate results. Finally, Test 1, which modified the activation function of the CNN, did not have a positive impact on model performance (although the impact was not as consequential as the other changes). Functionally, this makes sense as well, and confirms the well-established AlexNet and ImageNet CNNs cannot be improved in this specialized case by modification of the activation function.

It is important to also consider the areas in which the models encounter difficulties. In general, the greatest challenge posed to each of the five models was distinguishing between grade II and grade III in both astrocytomas and oligodendrogliomas. There were no cases where GBMs or grade IV astrocytomas were misdiagnosed, which is reasonable as those cases have the clearest appearance on MRI (and are the easiest to diagnose). However, within the categories of astrocytoma and oligodendroglioma, there was some model confusion between grades II and III. The difference between grades II and III astrocytoma, for example, may be so miniscule (the level of a single DNA codon) that advanced analysis is needed clinically regardless.

1. **Conclusions**

There are a couple takeaways from this experimental design, both at the general and granular levels. At its crux, this design of experiments proved that convolutional neural networks do have the capacity to assess brain tumor MRIs, which could greatly aid in the care for patients in a rural setting. While these CNN-based algorithms can be somewhat variable in their accuracy, the ImageNet CNN that was developed and trained for this task was able to diagnose one of the 6 most common brain tumor types from FLAIR MRIs with 95% accuracy. This is more than sufficient for proof-of-concept confirmation, and shows that the tried-and-tested CNN frameworks can be implemented in such a specialized case.

On the more model-specific level, a couple of implementation-level questions arise based on the results obtained. Arguably the largest question is at what point the increased runtime of a high-performing CNN would be preferred to a much less computationally expensive (and slightly less accurate) model. In its current state, where the model training is only around 30 minutes, and the additional time carries a 4% accuracy benefit, there is no reason to not seek a clinical deployment of these ImageNet neural networks. However, as models continue to develop, there is the capacity for a computationally simpler model to begin outperforming both trained models and practicing radiologists.

1. **Directions for future work**

There are three directions in which future work can be undertaken: improving on the accuracy of developed models, expanding the diagnostic capacity of CNN MRI analysis into novel areas, and implementing these developed tools in rural hospital and clinic settings. Making more than marginal improvements to the algorithm’s accuracy likely necessitates additional data, either in the form of additional patient MRIs and diagnoses, or in the incorporation of all MRI imaging modalities into one model. In either case, the additional of further data is likely to improve the diagnostic capacity of a CNN, but will almost certainly require specialized resources beyond personal computers. Additionally, future work can be aimed at answering more complex precision medicine questions. While attempts have been made to infer tumor molecular profile from an MRI, there remains much room for growth in that area, and difficult questions (such as an imaging-level differentiation between molecular and histopathological GBMs) can be undertaken by these models.

However, the direction that may prove most impactful for future work is in the clinical implementation of these CNN-derived diagnostic tools. The problem remains that individuals living in rural zip codes with brain cancer are likelier to have shorter overall survival than those from urban locations, and a diagnostic tool with sufficient accuracy such as this may be one step towards bridging the gap in care between the two demographics. Key steps include gathering provider and institutional buy-in for these types of imaging analyses, and while future work may only rely slightly on fine-tuning the algorithms, the efforts that need to be undertaken to implement these models in a rural hospital or clinic setting are those that will provide great benefits to both general practitioners and their patients.

1. **References**

“About Glioblastoma.” n.d. National Brain Tumor Society. <https://braintumor.org/events/glioblastoma-awareness-day/about-glioblastoma/>.

Aldape, Kenneth, Kevin M. Brindle, Louis Chesler, Rajesh Chopra, Amar Gajjar, Mark R. Gilbert, Nicholas Gottardo, et al. 2019. “Challenges to Curing Primary Brain Tumours.” *Nature Reviews Clinical Oncology* 16 (8): 509–20. <https://doi.org/10.1038/s41571-019-0177-5>.

Arash Ghorbannia Delavar, Omar M Al, Kathleen Maguire, Arvin R Wali, and Martin H Pham. 2019. “The Impact of Rural Residence on Adult Brain Cancer Survival in the United States” 144 (3): 535–43. <https://doi.org/10.1007/s11060-019-03254-4>.

ASCO. 2018. “Brain Tumor - Statistics.” Cancer.net. March 20, 2018. <https://www.cancer.net/cancer-types/brain-tumor/statistics>.

Ayadi, Wadhah, Wajdi Elhamzi, Imen Charfi, and Mohamed Atri. 2021. “Deep CNN for Brain Tumor Classification.” *Neural Processing Letters* 53 (1): 671–700. <https://doi.org/10.1007/s11063-020-10398-2>.

Badža, Milica M., and Marko Č. Barjaktarović. 2020. “Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network.” *Applied Sciences* 10 (6): 1999. <https://doi.org/10.3390/app10061999>.

Bakas, Spyridon, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S. Kirby, John B. Freymann, Keyvan Farahani, and Christos Davatzikos. 2017. “Advancing the Cancer Genome Atlas Glioma MRI Collections with Expert Segmentation Labels and Radiomic Features.” *Scientific Data* 4 (1). <https://doi.org/10.1038/sdata.2017.117>.

Chatterjee, Soumick, Faraz Ahmed Nizamani, Andreas Nürnberger, and Oliver Speck. 2022. “Classification of Brain Tumours in MR Images Using Deep Spatiospatial Models.” *Scientific Reports* 12 (1): 1505. <https://doi.org/10.1038/s41598-022-05572-6>.

Grant, Robin, Therese Dowswell, Eve Tomlinson, Paul M Brennan, Fiona M Walter, Yoav Ben-Shlomo, David William Hunt, et al. 2020. “Interventions to Reduce the Time to Diagnosis of Brain Tumours.” *Cochrane Database of Systematic Reviews* 2020 (9). <https://doi.org/10.1002/14651858.cd013564.pub2>.

Hollon, Todd, Cheng Jiang, Asadur Chowdury, Mustafa Nasir-Moin, Akhil Kondepudi, Alexander Aabedi, Arjun Adapa, et al. 2023. “Artificial-Intelligence-Based Molecular Classification of Diffuse Gliomas Using Rapid, Label-Free Optical Imaging.” *Nature Medicine*, March. <https://doi.org/10.1038/s41591-023-02252-4>.

Irmak, Emrah. 2021. “Multi-Classification of Brain Tumor MRI Images Using Deep Convolutional Neural Network with Fully Optimized Framework.” *Iranian Journal of Science and Technology, Transactions of Electrical Engineering* 45 (3): 1015–36. <https://doi.org/10.1007/s40998-021-00426-9>.

Menze, Bjoern H., Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, et al. 2015. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS).” *IEEE Transactions on Medical Imaging* 34 (10): 1993–2024. <https://doi.org/10.1109/tmi.2014.2377694>.

Park, Jae Hui, Nani Jung, Seok Jin Kang, Heung Sik Kim, El Kim, Hee Jung Lee, Hye Ra Jung, Misun Choe, and Ye Jee Shim. 2019. “Survival and Prognosis of Patients with Pilocytic Astrocytoma: A Single-Center Study.” *Brain Tumor Research and Treatment* 7 (2): 92–97. <https://doi.org/10.14791/btrt.2019.7.e36>.

Redlich, Fredrick C, Rembrandt H Dunsmore, and Eugene B Brody. 1948. “Delays and Errors in the Diagnosis of Brain Tumor” 239 (25): 945–50. <https://doi.org/10.1056/nejm194812162392501>.

Shakoor, Delaram, Ali Guermazi, Richard Kijowski, Jan Fritz, Sahar Jalali-Farahani, Bahram Mohajer, John Eng, and Shadpour Demehri. 2018. “Diagnostic Performance of Three-Dimensional MRI for Depicting Cartilage Defects in the Knee: A Meta-Analysis.” *Radiology* 289 (1): 71–82. <https://doi.org/10.1148/radiol.2018180426>.

Yu Wei Heng, Kia Hooi Tan, and Nelson Kok. 2023. “Brain Tumor: A Review of Its Demographic in a Rural Hospital of Sibu in Sarawak, Malaysia,” March. <https://doi.org/10.1055/s-0043-1760855>.

Zhou, S. Kevin, Hayit Greenspan, and Dinggang Shen. 2017. *Deep Learning for Medical Image Analysis*. London, United Kingdom: Academic Press is an imprint of Elsevier.

‌