A Computationally Efficient Convolutional Neural Network for Multiclass Brain Tumour Detection in MRI Scans

Ian Tatlow Whatley
Dept. of Mathematics & Statistics
McGill University
Montreal, Canada
ian.whatley@mail.mcgill.ca

Mohammad Khabiir Khodabaccus
Dept. of Computer Science
McGill University
Montreal, Canada
khabiir.khodabaccus@mail.mcgill.ca

***Abstract*— We propose an efficient CNN-based model for the detection and classification of brain tumours in MRI scans that achieves 93% validation accuracy with better regularization than existing approaches using less than half the parameters. Our model misclassifies tumourous scans as normal in less than 1% of cases in our test data, thereby demonstrating clinical efficacy.**

Keywords— medical imaging, computer vision, convolutional neural networks, brain tumours

I. Introduction

Brain tumours are commonly misdiagnosed due to their wide range of symptoms and their similarities with more common conditions [1]. Of the varieties of brain tumours, three common varieties include glioma, meningioma, and pituitary tumours; of which meningioma is commonly benign. For many brain tumours, diagnostic accuracy remains around 87%, often due to the propensity for tumours to be misdiagnosed as other forms of inflammation, even with the use of Magnetic Resonance Imaging (MRI) scans [2][3].

A particularly consequential boon from MRI scans is the detection of brain tumours *en masse* without any invasive procedures. With further digitisation, MRI scans are easily stored in hard drives and the cloud across hospital databases and potentially sent across the internet. With this digitisation, there is great potential for machine learning and data science methods for the detection of brain tumours in MRI scans.

Premier among such methods would be computer vision: a machine learning technique that utilises Convolutional Neural Networks (CNNs) to scan images and detect patterns, objects, or properties within images [4]. Computer vision applications for the detection of brain tumours is useful considering the rapid advancements of research in machine learning and computer vision, their ease of application, and the vast stores of digitised MRI scans.

II. Literature Review

A large body of research has been concerned with the application of CNNs for brain tumour recognition in MRI scans [5]. One of the earliest high-accuracy results came with Ghosal et al.'s 2019 paper using ResNet-101 [6] which yielded 93.83% accuracy and used a validation set to provide evidence against overfitting in their model. This in-depth paper showed outcomes for multiclass brain tumour classification; results which have rarely been replicated since.

In 2020, Huang et al. introduced CNNBCN for brain tumours, which examines random graphs of CNNs and converting them into an optimally accurate network that can then be selected for further training [7]. They found 95.8% accuracy in their highest case.

Recently, many authors have been set on finding higher accuracy while simplifying the complexity of the model. In 2022, Balaji et al. [8] used EfficientNetB0 to create a model with fewer parameters than Ghosal et al. yet increasing test accuracy to 97.6%, the highest seen since Huang et al. Furthermore, Balaji also provides for a "no tumour" class in the model as well, showing better clinical potential. However, Balaji's learning curve shows signs of overfitting, as the validation accuracy never truly converges with the training curve.

Here, we present only the papers that present learning curves and training-test splits. Unfortunately, few researchers on this topic present such information. In machine learning, this should be of high priority as validation and loss curves that follow a similar trajectory to the training set imply low overfitting which is vital in ensuring that the model can be used in a wide variety of circumstances.

Furthermore, of those that have presented valid research, few of the multiclass models included an option for no tumour being present. In the context of medical diagnosis, this is incredibly important as the model needs to have an option not to include a tumour in a medical context as most MRI scans will find no such tumour. As we discussed, other abnormalities in MRI scans may be misclassified as brain tumours.

III. Methodology

In this section, we describe the methodology for our CNN model from dataset preparation to the neural network's architecture. Our model was created using the Sci-Kit Learn library for the Python 3 language.

A. The Dataset

Our dataset comes from Masoud Nickparvar's Brain Tumour MRI Dataset [9], containing JPEG images of MRI scans consisting of 512x512 pixels per image. The scans include transverse, coronal, and sagittal scans of patients and were

classed into four groups: scans of pituitary tumours, glioma, meningioma, and those without tumours. The data were already split between training and test datasets, with a training dataset containing 5,712 images. The corresponding portion of each group was not evenly distributed, with no tumours consisting of 1,595 images, while glioma, meningioma, and pituitary consisted of 1321, 1339, and 1456 images respectively.

We randomly shuffle and split the data into three sets. The training data set was 70% of our original dataset, being used to train the model for its eventual use. The validation set is then 10% of our original dataset, being used to assess the model's estimators throughout the test through cross-validation. Next, our test set is used come to final conclusions about the model, consisting of the remaining 20%.

B. *Data Preparation*

Keras models generally work better with numeric data. While our images can be represented as a matrix of numeric pixel values, the categorical text labels of "glioma", "meningioma", "no tumour", and "pituitary" need to be encoded. There are multiple ways in which categorical labels can be encoded but we use one hot encoding as it is one of the most widely used encoding techniques for machine learning [10].

We divided the process into two stages. First, we use the SciKit-Learn Labelencoder function to assign an integer value to each label. One may argue that this should be enough for the Keras model however, our categorical labels do not possess any ordinal relationship between them, but label encoding does assign an order. Hence, there is a high probability that the model might misunderstand the relationship between the labels. We thus resort to the final stage where we use the Scikit-Learn toCategorical function to transform the integer labels to a binary vector which consists of the class label (being given the value 1) and the other elements (being 0).

C. *The Model*

Our model consists of a convolutional neural network trained in mini-batches of size 32 as batches greater than 32 tend to perform worse [11]. However, for our model, batches smaller than this also yielded worse results. Since our dataset was evenly distributed by class, we changed the initial class weights in training to account for this, using formula 1 to calculate the class weights (denoted w_i). We find the best overall results in the training, validation, and test phases of our model by using the test set's class weights. Our weights are 1.08, 1.06, 0.87, and 1.02 for glioma, meningioma, no tumour, and pituitary, respectively.

$$\text{for class i: } w_i = \frac{5712}{4(\text{n-samples}_i)} \quad (1)$$

Our model begins with three convolutional layers consisting of 8, 16, and 32 filters respectively. For each convolution, we have a 3x3 kernel to detect the edges of the tumours; meaning that the output contains only relevant information. Furthermore, the 3x3 kernel is a common and efficient size for CNNs [12]. After each convolution, we include a ReLU activation function with batch normalisation to normalise the output from our convolution; preventing overfitting and allowing faster optimisation in fewer epochs.

After the convolutional layers, we flatten the data into a one-dimensional array, activate the dropout layer, and feed the remaining data into the dense layers. The dropout layer further regularises our model. We chose a dropout rate of 60% to ensure a conservative estimate of avoiding overfitting as we found this best regularises our model.

The dense layer includes a batch normalised ReLU activation function consisting of 512 nodes. We then feed the data into the final activation layer consisting of four nodes (one for each class of tumour) activated by a Softmax function to ensure our outputs are probabilities between 0 and 1.

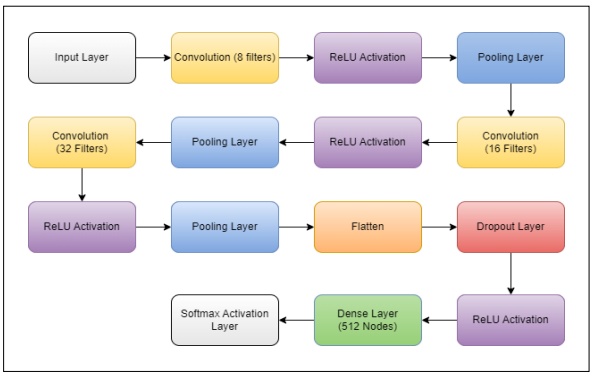


Fig. 1: Model Architecture

Finally, we chose optimise using ADAM with a learning rate of 0.0001. ADAM has routinely proven the best optimiser for the classification of brain tumours in MRI scans [13].

IV. Results

A. *Limitation of the Dataset*

We limit our analysis to only four of the 120 classes of brain tumours [14] due to the availability of public datasets. A larger dataset would enable a wider variety of tumour types to be classified.

TABLE I. ATTEMPTED AND FINAL PARAMETERS

Factor	Value(s)
Convolutional Layers	2, **3,** 4
Dropout Layer	0, **1**
Dense Layer	**1**
No. Filters	**8, 16, 32,** 64, 128
Kernel Sizes	2x2, **3x3.** 5x5
Dropout Rate	0.2, 0.5, **0.6**
Mini Batch Size	**32**
Adam Learning Rate	**0.0001,** 0.005, 0.001, 0.05

Note: Bold indicates final model value.

Secondly, the dataset could have been larger. In general, the more data with which one works, the better. Given the open-source datasets available, we believe that this dataset, having included those with no tumour, is the most practical for actual medical use because it allows for the model to train in avoiding misclassifying objects as tumours.

Likewise, our dataset only includes scans of normal brains in the "no tumor" category. It is common in MRI tumour detection to mistake other abnormalities such as inflammation for tumours. Since we trained our model on only tumorous and normal scans, it will be prone to misclassify abnormalities which appear similar to tumours as such.

B. *Previous Iterations*

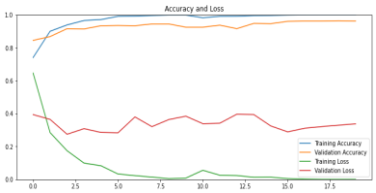


Fig. 2: Original Learning Curves

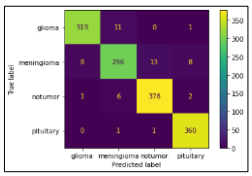


Fig. 3: Original Model Confusion Matrix

Originally, we had a complex model consisting of over 19 million parameters, no batch normalisation, no dropout layer, convolutional filters twice as large, and a learning rate higher than our final model. Our original model used early stopping and the same mini-bath size, resulting in 15 epochs but an expensive computation time of 2,640 seconds.

This model had an accurate confusion matrix as well and an accuracy of 96%. One would assume that this model would be our preferred choice. However, the major flaw in this model is that it was computationally expensive and terribly overfit judging by figure 2. When comparing the validation loss and training loss, we found a median difference between training and validation loss was -0.3104, implying overfitting.

For this reason, we decided to tweak the model to increase computational efficiency while minimising loss in accuracy. In order to increase computational efficiency, we would simply decrease the size of each convolutional filter, whereas in order to prevent overfitting, we included both batch normalisation and dropout layers. These additions decreased computation time by over 50% and reduced the median difference between training and validation loss, but reduced accuracy by 3%. Despite the reduction in accuracy, the model is preferable because of the decreased overfitting seen in the final model.

C. *Assessment of the Final Model*

Using early stopping with a patience of 10 epochs, the final model took 1,108 seconds and 32 epochs to reach optimality. The training was performed on an 11th Generation Intel Core i7-11700 at 2.50 GHz with 16 CPUs and 32,768 MB of RAM.

Regularisation is vital in proving a model's efficacy beyond selective data. Our model shows minimal signs of overfitting. The validation and training accuracy follow nearly identical trajectories (see figure 3) with small dips below at some points. The validation loss never differs far from the training loss, resulting in a median validation loss of only 0.0386 lower than the training loss. In terms of accuracy, our training accuracy reaches 99%. Our validation accuracy reached 93%, slightly lower than ResNet-101. Overall, there was an average difference between training and validation accuracy of 6%.

We achieve an accuracy of 93% in the test set. Though lower than in our original model, we believe that considering

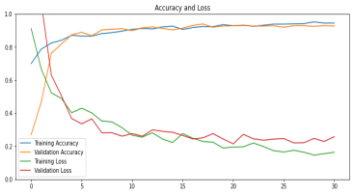


Fig. 4: Model Learning Curves

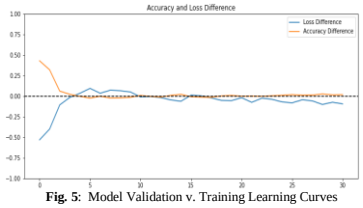


Fig. 5: Model Validation v. Training Learning Curves

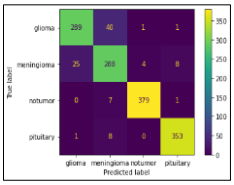


Fig. 7a: Confusion Matrix

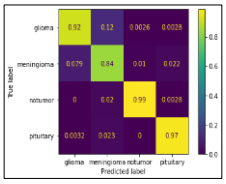
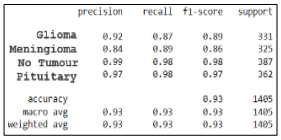


Fig. 7b: Confusion Matrix by Precision



	precision	recall	f1-score	support
Glioma	0.92	0.87	0.89	331
Meningioma	0.84	0.89	0.86	325
No Tumour	0.99	0.98	0.98	387
Pituitary	0.97	0.98	0.97	362
accuracy			0.93	1405
macro avg	0.93	0.93	0.93	1405
weighted avg	0.93	0.93	0.93	1405

Fig. 8: Model Test Summary

the evidence against overfitting, this reduction in accuracy is necessary since the increased regularisation implies our model can be used on scans with disparate appearances.

Finally, our model implies medical efficacy due to the high precision rate for no tumour. In a real medical context, a false diagnosis of the absence of a tumour could put a patient at substantial risk. For this reason, we wanted our model to have a high precision rate for "no tumour". In our model, the precision rate for "no tumour" rises to 99% in the test set and 100% in the training set as seen in figure 7b.

TABLE II. COMPARISON TO PREVIOUS LITERATURE

Paper	No Tumour	Acc.	Test Acc.	Test Rec.	Prec.	Test Prec.
Ghosal et al., 2019	No	100	94.7	93.84		
Huang et al., 2020	No	100	95.49			
Balaji et al., 2022	**Yes**	100	**97.6**	**98.54**		**98.56**
Proposed	**Yes**	**99**	93	98	**99**	97

TABLE III. COMPARISON OF MODEL COMPLEXITY WITH OTHER REGULARISED MODELS

Paper	Number of Layers	Number of Parameters
Ghosal et al., 2019	101	44.65 M
Huang et al., 2020	*Unknown*	*Unknown*
Balaji et al., 2022	237	11 M
Proposed	22	4.7 M

V. DISCUSSION

To optimise our model for clinical efficacy, the model could be improved with a larger dataset complete with blurry, tilted, and altered data to ensure the model can handle a variety of anomalies in MRI scans. Furthermore, the dataset could include inflammation of the brain or other abnormalities that appear similar to tumours to ensure the model can distinguish between these.

As for the model itself, should the precision for no tumour decrease with more numerous and altered data, a threshold could be applied to the no tumour class. This would ensure that the model is less likely to detect no while sacrificing the recall. In a clinical setting, this would be valuable as a doctor would diagnose after a tumour is detected.

Our model could also experiment with greater run time. We use a liberal patience of 10 epochs in early stopping which forces our model to stop after 32 epochs. If we had more computational power and time, we could allow for greater leniency in epochs in the hope this would lead to better convergence in both training and validation estimators.

VI. Conclusion

Ultimately, our model achieves a training set accuracy of 99% and a test set accuracy of 93%. In addition to these high estimators, we avoided overfitting and ended with a validation accuracy of 93% and a validation learning curve that follows a nearly identical trajectory to the training set's learning curve.

We show clinical efficacy thanks to the high precision for detecting no tumour, implying a conservative approach for tumour diagnosis; appreciated in a medical setting. In addition, our model indicates minimal overfitting thanks to the convergence in our validation vs training estimators (see fig. 4 & 5), implying that our model will be effective in a wide variety of brain tumour MRI images.

Overall, the model works on a wide variety of common tumour types and minimises the risk of misclassifying scans with tumours as being absent of tumours. With further research, a medically approved CNN computer vision model can be used to detect tumours more accurately, faster, and in greater numbers than is currently possible via human labour.

VII. References

[1] Walter, F. M., Penfold, C., Joannides, A., Saji, S., Johnson, M., Watts, C., Brodbelt, A., Jenkinson, M. D., Price, S. J., Hamilton, W., & Scott, S. E. Missed opportunities for diagnosing brain tumours in primary care: a qualitative study of patient experiences. *The British journal of general practice : the journal of the Royal College of General Practitioners* 69, 224–235 (2019). https://doi.org/10.3399/bjgp19X701861

[2] Munir, S., Khan, S. A., Hanif, H., & Khan, M. Diagnostic accuracy of magnetic resonance imaging in detection of intra-axial gliomas. *Pakistan journal of medical sciences* 25, pp. 125–130 (2021). https://doi.org/10.12669/pjms.37.1.2489

[3] Dixon, L., Jandu, G., Sidpra, J., & Mankad, K. Diagnostic accuracy of qualitative MRI in 550 paediatric brain tumours: evaluating current practice in the computational era. *Quantitative Imaging In Medicine And Surgery* 12, pp. 131-143 (2021). https://doi.org/10.21037/qims-20-1388

[4] Wiley, Victor & Lucas, Thomas. Computer Vision and Image Processing: A Paper Review. *International Journal of Artificial Intelligence Research* 2, 22 (2018). https://doi.org/10.29099/ijair.v2i1.42.

[5] Arabahmadi, M., Farahbakhsh, R., & Rezazadeh, J. Deep Learning for Smart Healthcare-A Survey on Brain Tumor Detection from Medical Imaging. *Sensors* (Basel, Switzerland) 22, 5 (2022). https://doi.org/10.3390/s22051960

[6] Ghosal, P., Nandanwar, L., Kanchan, S., Bhadra, A., Chakraborty, J., & Nandi, D. Brain tumor classification using ResNet-101 based squeeze and excitation deep neural network. In 2019 *Second International Conference on Advanced Computational and Communication Paradigms* (ICACCP), 1-6 (2019). IEEE.

[7] Z. Huang et al., "Convolutional Neural Network Based on Complex Networks for Brain Tumor Image Classification With a Modified Activation Function," in IEEE Access, vol. 8, pp. 89281-89290, 2020, doi: 10.1109/ACCESS.2020.2993618.

[8] Balaji, Gopinath, Ranit Sen, and Harsh Kirty. "Detection and Classification of Brain tumors Using Deep Convolutional Neural Networks." arXiv preprint arXiv:2208.13264 (2022).

[9] Masoud Nickparvar. Brain Tumor MRI Dataset. https://doi.org/10.34740/KAGGLE/DSV/2645886 (2021).

[10] Dahouda, M. K., & Joe, I. (2021). A Deep-Learned Embedding Technique for Categorical Features Encoding. IEEE Access 9, pp. 114381–114391. Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/access.2021.3104357

[11] Masters, D., & Luschi, C. Revisiting small batch training for deep neural networks. (2018). arXiv preprint arXiv:1804.07612.

[12] Coletti, Mark & Lunga, Dalton & Bassett, Jeffrey & Rose, Amy. Evolving Larger Convolutional Layer Kernel Sizes for a Settlement Detection Deep-Learner on Summit. pp. 36-44 (2019). https://doi.org/10.1109/DLS49591.2019.00010

[13] Yaqub, M., Jinchao, F., Zia, M. S., Arshid, K., Jia, K., Rehman, Z. U., & Mehmood, A. State-of-the-Art CNN Optimizer for Brain Tumor Segmentation in Magnetic Resonance Images. Brain sciences 10, p. 427 (2022). https://doi.org/10.3390/brainsci10070427

[14] Kleihues P, Cavenee WK, eds. World Health Organization classification of tumours: Vol. 1. Pathology and genetics of tumours of the nervous system (Lyon: IARC Press, 2000)