BrainLens! Precision Diagnosis in Brain Tumour Imaging

Ahmad El-Zein
Faculty of Engineering
Western University
London, Ontario
aelzein2@uwo.ca

Amer Al Rawashdeh
Faculty of Engineering
Western University
London, Ontario
aalrawas@uwo.ca

Abdelrahman Alhalbouni
Faculty of Engineering
Western University
London, Ontario
aalhalb3@uwo.ca

Khanh Dang
Faculty of Engineering
Western Engineering
London, Ontario
kdang28@uwo.ca

***Abstract*—In the intersection of artificial intelligence and medicine, BrainLens advances brain tumour diagnostics through MRI scans with a significant accuracy improvement. Leveraging a composite dataset from figshare, SARTAJ, and Br35H, BrainLens utilizes an innovative deep convolutional neural network architecture (CNN) optimized through data augmentation and normalization techniques to achieve 96.64% accuracy, tailored specifically for the complex nature of brain tumour imaging. This system addresses the need for precision in diagnostic imaging, countering the traditional compromise between speed and accuracy. Our results indicate that the effectiveness of the model not only provides a high degree of significant improvements in rapid and accurate tumour diagnosis but also opens avenues for broader applications in medical image analysis, marking a step forward in the integration of AI technologies into clinical settings. Our engineered solution serves as a potential method for revolutionizing patient care through accurate diagnosis and classification.**

I. Introduction

In the field of medical diagnostics, specifically in neuro-oncology, the precise classification of brain tumours via MRI scans remains a critical challenge, directly impacting patient outcomes and treatment strategies. Recognizing the pivotal role of technology in medicine, we strategized and implemented a method to achieve high precision in brain tumour diagnostics. Leveraging a rich dataset from sources such as figshare, SARTAJ, and Br35H, available on Kaggle [1], our project sought to address the limitations of current methodologies and offer a more reliable alternative.

Before implementing our strategy, we confronted a prevalent research gap: existing methods for tumour classification, while advanced, were not fully leveraging AI capabilities, particularly in handling the four key classifications of ‘glioma’, ‘meningioma’, ‘notumor’, and ‘pituitary’ tumours. Notable studies such as those by Ahmad et al. have demonstrated the promise of deep learning for medical imaging, specifically in the automatic diagnosis of cancers [2]. Yet, these studies also underscored the challenge of limited data availability, a barrier our solution aimed to overcome.

In the formative stages of our research, we examined existing and integrated academic insights from recent advances in AI-based brain tumour diagnosis. Works such as those published in *Biomedicines* by Cè et al. and in *Diagnostics* informed our understanding of the need for refined classification systems and fueled our determination to develop a resolution [3] [4].

Our goals with this research are diverse: develop a deep learning model that could accurately process and classify MRI images into the correct tumour categories, contribute to the field of knowledge in AI diagnostics, and provide a tool that could assist medical professionals in making informed decisions on what a specific tumour is classified as by their way of inputting their existing data. Using a deep convolutional neural network, informed by the existing methodologies, we engineered a system capable of determining a tumour by its type with high accuracy.

The model achieved impressive accuracy rates, reflecting the potential of AI to transform medical diagnostics. Our approach's innovation poses in the intricate design of the deep neural network and the strategic use of data preprocessing and augmentation techniques, which ensured the model's effectiveness across various tumour types.

The purpose of this report is to highlight and discuss our proposed advancement in applying artificial intelligence to the domain of medical diagnostics. This report presents the evolution of our research, BrainLens, first by discussing existing work within the field, and challenges in brain tumour classification (Section II). It progresses to describe our research goals and the methodological framework employed to develop our AI model (Section III). The subsequent sections detail the observed and experimental outcomes of our model (Section IV), and the contributions of our research to the domain of AI in medicine, and reflect on the potential future developments inspired by our findings (Section V).

With BrainLens, we aim to enrich the toolkit of diagnostic resources, aiding clinicians in the precise identification of brain tumours. Our decision and motivations align with the broader

movement towards precision medicine, showcasing AI's potential to refine diagnostic accuracy and efficiency.

II. BACKGROUND

A. Previous Works

The integration of deep learning into the field of MRI-based brain tumour classification marks a significant evolution from traditional methodologies. Historically, machine learning techniques required extensive manual intervention for feature extraction, which often failed to capture the complexity inherent in high-dimensional MRI data. This limitation prompted a shift towards convolutional neural networks (CNNs), which are proficient at automatic feature extraction due to their deep, hierarchical structure that mimics the human visual system's processing layers.

An influential piece of work by Hossain et al. illustrates this transition, showcasing a collective approach that leverages the shared strengths of multiple pre-trained models like VGG16, InceptionV3, and Xception. This strategy not only broadened the scope of feature extraction but also significantly improved classification accuracy, achieving rates as high as 96.94%, which is almost identical to the accuracy of our proposed model [5]. Such an approach underscores the potential of CNNs in transcending the limitations of traditional machine learning by utilizing the power of deep features and complex model architectures for more accurate and robust classification outcomes.

Furthering this method, Kang et al. explored the interaction of deep features from an array of pre-trained CNNs combined with advanced machine learning classifiers. This ensemble method was particularly effective with large datasets, where the depth and variety of features extracted by CNNs could be fully leveraged, demonstrating the critical role of deep learning in enhancing the performance of classification systems [6].

B. Research Gaps

Despite these advances, deep learning's reliance on large, annotated datasets presents a significant challenge. The insufficient amount of precise datasets, especially in medical imaging, limits the applicability of deep learning models, as they are prone to overfitting when trained on smaller datasets. This issue of dataset adequacy versus model complexity remains a balance to achieve optimal performance without compromising the model's ability to generalize to new, unseen data.

Tandel et al. highlighted the potential of majority-voting-based methods to mitigate the risks of overfitting, suggesting a direction for future research in model architecture and training methodologies that can maintain high accuracy without extensive datasets. Moreover, while transfer learning has emerged as a promising strategy to utilize pre-trained models for new tasks, the fine-tuning of these models to specific applications such as brain tumour classification from MRI images poses challenges. Achieving an optimal balance between model adaptability and specificity without yielding to overfitting is a complex task, as indicated by Charron et al., who achieved an 81% classification accuracy but acknowledged the need for further improvements in model generalization [7].

These gaps underscore the necessity for innovative approaches that can enhance model performance and generalization, even in the face of limited data availability. Our research aims to contribute to this field by developing a custom deep-learning solution tailored for MRI-based brain tumour classification, drawing from the strengths and addressing the limitations of existing methodologies.

C. Improvements

Our research is situated at this intersection, offering a solution specifically architected for brain tumour MRI classification. By using TensorFlow and Keras libraries, our model employs a deep custom CNN architecture designed to learn directly from the complex patterns within the imaging data. The preprocessing steps implemented in our code, including image resizing, normalization, and augmentation techniques such as random flips, crops, and contrast adjustments, are directly aimed at resolving overfitting issues. This was to sustain the model's generalizability across imaging datasets.

Slight image improvements were carried out by creating a versatile dataset using a Python tool called ‘glob’ to consolidate a large volume of MRI images, followed by label extraction to prepare for a training approach. The augmentation strategies were carefully selected to reflect real-world variations, thereby enhancing the model's ability to generalize from the training data to unseen test cases. This attention to preprocessing and data augmentation aligns with the noted gaps in research within the field, where these steps are critical for the success of deep learning models in medical imaging applications.

By adopting best practices in data handling and modelling as seen in our code, our research advances in addressing the noted gaps. The neural network's structure, featuring multiple convolutional, pooling, and dense layers, followed by batch normalization and dropout for regularization, was precisely designed to optimize feature learning. Our training process, detailed in the subsequent section (Section III), illustrates a keen focus on achieving high classification performance without sacrificing the model's ability to generalize.

III. METHODOLOGY

A. Research Objectives

Within our research, BrainLens, there are two primary research objectives for this study:

I. Implement data preprocessing and augmentation to enhance model generalization, addressing dataset

variability and size constraints on the dataset composed of other sources such as figshare SARTAJ, and Br35H.

II. Develop a deep CNN for accurate classification of MRI images into ‘glioma’, ‘meningioma’, ‘notumor’, and ‘pituitary’ tumours. This will aid in diagnostic precision.

In addressing the prevalent challenge of dataset diversity and limitations in size, particularly when integrating data from sources like figshare, SARTAJ, and Br35H, this objective underscores the critical role of advanced preprocessing and augmentation techniques. By improving our dataset through methods such as image normalization, rotation, flipping, and scaling, we aim to simulate a broader range of real-world variations. This approach is anticipated to significantly bolster the model's ability to generalize, thereby improving its performance across unseen data and mitigating overfitting risks. Enhanced generalization is crucial for developing a model that remains robust and reliable across diverse clinical settings and imaging conditions.

Central to our research is the development of a specialized deep CNN architecture tailored to the intricacies of brain tumour classification from MRI scans. This model is not just a technical aspect but a strategic response to the urgent need for high-precision diagnostic tools in neuro-oncology. By leveraging the depth and complexity of CNNs to distinguish subtle patterns unique to different tumour types, we aim to provide a powerful aid to radiologists and medical professionals, potentially revolutionizing the speed and accuracy of brain tumour diagnostics.

B. Research Methodology

Initiating the development of our solution, the first task was to load the images, comprising of training and testing data sourced from figshare, SATAJ, and Br35H datasets. The entire dataset contained 7023 MRI images and so they were extracted and split into different distributions of training (4569 images), testing (1311 images), and validation (1143 images) datasets. The class distribution was fairly equal in both the training and testing datasets. Fig. 1 shows that 23.1% of the training data represent ‘glioma’ tumours, 27.9% being ‘notumor’, 23.4% being of type ‘meningioma’, and finally 25.5% identified as ‘pituitary’. Fig. 2 represents the distribution of the training data with similar ratios being 22.9% ‘glioma’, 30.9% being ‘notumor’, 23.3% of type ‘meningioma’, and finally, ‘pituitary’ tumours comprising 22.9% of the testing set. By observing this, it is evident that the data across both sets is diverse which ensures that the model we created would be exposed to a wide variety of cases – essential for learning generalized features that are indicative of each tumour type. Also, it allows for a reliable evaluation of the model’s performance across all the different classes, which assures that the accuracy is not skewed by an overrepresentation of any single class.

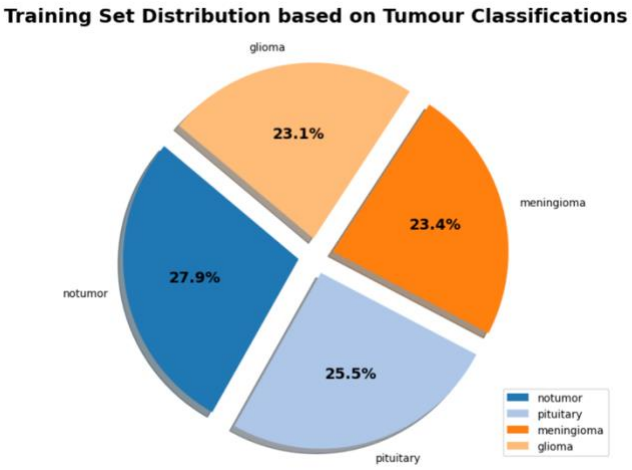

Fig. 1. Training data distribution based on tumour classifications.

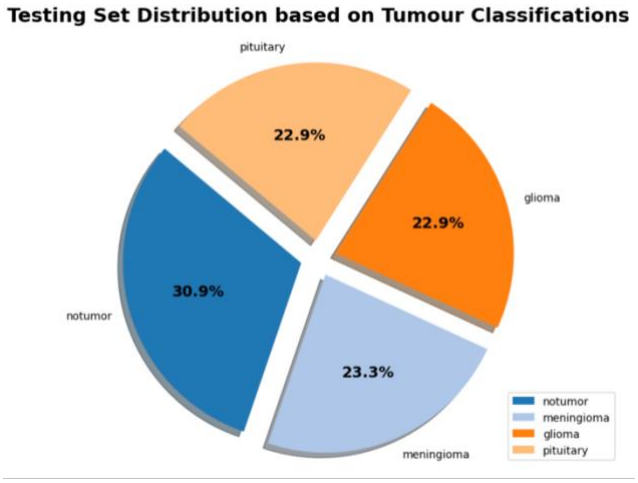

Fig. 2. Testing data distribution based on tumour classifications.

After loading the data, preprocessing was needed. We assigned a consistent image size of 256 x 256 pixels, which was a decision driven by the need to balance the computational efficiency, with factoring in the retention of the critical image details that are found and are essential for tumour classification. Preprocessing also involved normalizing the pixel values of each image to a [0, 1] range. Equation 1 was used to normalize the pixels – crucial in deep learning as it ensures that our model’s training process was stable and converged at a faster rate, and by also keeping the input values small. Furthermore, the decision before preprocessing and applying the augmented and normalization methods were chosen after we examined histograms of sample images. The range and distribution of the pixels in the sample images guided us in choosing the most efficient techniques. It also allowed us to evaluate the contrast of the image, which was essential in allowing us to distinguish between tumour tissues and normal brain tissues. The higher contrast regions in the sample images helped indicate potential areas of interest for tumour detection, ultimately contributing to

its proper classification. Fig. 3 is an example of a sample image alongside its corresponding histogram containing its pixel intensity.

$$\text{Normalized Pixel Value} = \frac{Pixel\ Value}{255} \quad (1)$$

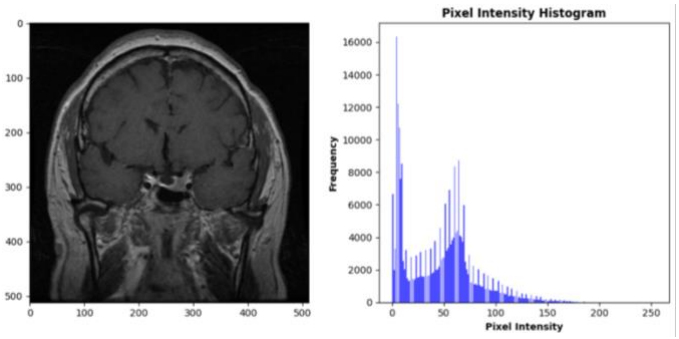

Fig. 3. Pixel intensity histogram of a sample image.

To enhance the robustness of our model and address the potential for overfitting given our dataset's size constraints, we implemented data augmentation techniques. This included random horizontal flips, slight rotations, crops, and zooms to artificially expand our training dataset, simulating a broader range of imaging scenarios that the model might encounter in practice. This approach is grounded in the principle that a more varied training dataset can improve the model's ability to generalize from the training data to unseen data.

The next task was to implement data generators. We created custom data generators to increase efficient memory management and streamline the training process. The generators batched the entire dataset into three different sets, being the training, validation, and testing sets. We decided to use a regular-sized batch (16) as we believed that it was a balance between model performance and the predicted computational load. The batched datasets from the data generators would then be inputted into our model during training and evaluation. Fig. 4 shows a sample image from the training generator set classified as 'meningioma' that has been preprocessed, normalized and augmented. All in all, these components in our approach were carried out to address and satisfy research objective I.

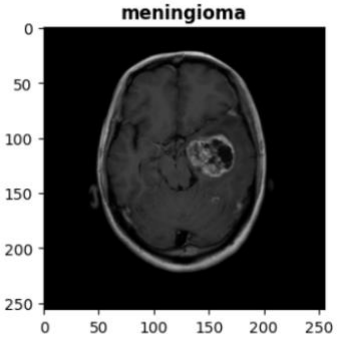

Fig. 4. Preprocessed, normalized, and augmented sample image.

Moving forward, the next task that needed to be completed was the most critical component of our solution, which was the creation of our 'Convolutional Neural Network' (CNN). This component is how the data is trained and evaluated – model creation. Our strategy is primarily centralized around the goal of getting the highest accuracy, and the most significant factor to ensure this was through the addition of more layers to the neural network (although other factors can impact the accuracy and complexity of a model). From this, our focus was to add different components to the model to help yield an increase in robustness and performance for the brain tumour MRI classification. The first convolutional layer of our neural network consisted of 32 filters with a kernel size of 3 x 3. This decision was based on the understanding that brain tumour features are spatially localized, and so the small kernel size would allow the network to detect and extract finer details that are critical for distinguishing between different types of brain tissue and details that may indicate different tumours. The convolutional operation is given in (2), where *F* is the feature map, *I* is the input image, and *K* is the kernel. In addition to the first layer, the 'ReLU' (Rectified Linear Unit) activation function was used to introduce nonlinear properties to the model, allowing the model to learn more complex patterns found within the data. The ReLU function is given in (3).

$$F_{ij} = \sum_{u=0}^{2}\sum_{u=0}^{2} I_{i+u,j+v} \cdot K_{uv} \quad (2)$$

$$f(x) = \max(0, x) \quad (3)$$

MaxPooling2D layers with a pool size of 2 x 2 were added to reduce the spacial dimensions and to compute the most relevant details. It was added to essentially condense the feature map and to minimize computational complexity. The first convolutional layer also consisted of batch normalization and dropout layers where batch normalization was to accelerate the training and stabilize the learning process by normalizing the inputs to the layer. The dropout layers were to improve the regularization and stability of the network by randomly dropping a proportion of the nodes during training to make sure that the network did not depend on any single node. Also, we utilized dropout layers to mitigate the risk of our model from overfitting the data to improve the generalization capabilities of the network. The pooling layer, defined mathematically, is represented in (4), where *L* is the input to the pooling layer, and *M* is the pooled output.

$$M_{ij} = \max(L_{2i:2i+1,2j:2j+1}) \quad (4)$$

Following the first convolutional layer, another layer with a filter size of 64 was added, also being paired with subsequent

MaxPooling, dropout layers, and batch normalization. Moving further into the model, the complexity increased as all subsequent hidden convolutional layers increased from filter sizes of 128 to 512. The hierarchical deepening of the neural network was intentionally constructed, as we understood that it would help equip our model with the ability to recognize more intricate and abstract features and patterns of the brain tumour MRI images as the depth increased. The addition of five convolutional layers also had batch normalization and drop-out layers added which was essential to assist in accurately classifying the MRI scans.

Post-convolution, the data was flattened to transition between the convolutional layers to the fully connected layers. The transitional shift is critical for the model for extracting the learned spatial features from the brain tumour MRI scans in the trained set generator, into a format that is suitable for classification. Our model's first fully connected layer contained 256 neurons, serving as a dense representation of the newly extracted features, which then had a dropout layer to further reduce overfitting, alongside a ReLU activation function and batch normalization.

Finally, the network concluded with dense layers, resulting in an accumulated output layer that was used to classify the MRI images into one of the four classes of tumours. To facilitate a clear and probabilistic classification to classify the images, we used a softmax activation function, defined in (5) on the final dense layer. We selected the 'ADAM' optimizer for its adaptive learning rate capabilities (stochastic gradient descent for adjusting as training progresses), knowing that it was well-suited for datasets with variable image characteristics such as MRIs of brain tumours. In addition to this, it was also used to improve our model's convergence speed and performance. Moreover, categorical-cross entropy, defined in (6) was used for the loss function which is the most efficient option for multi-class classification – direct correspondence to BrainLens. In (6), *M* is the number of classes, *y* is the binary indicator of class *c* in observation *o*, and *p* is the predicted probability of observation *o* being in class *c*.

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \tag{5}$$

$$\text{Categorical Cross Entropy} = -\sum_{c=1}^{M} y_{o,c}\, log(p_{o,c}) \tag{6}$$

Our model underwent training for 20 epochs, with each epoch consisting of 286 steps. This number of steps was calculated based on the number of samples and the batch size, ensuring that the entire dataset in each epoch was seen by the model. A batch size of 16 was chosen as a compromise between training speed and model convergence quality, as previously mentioned.

Every step taken within our methodology was comprised of strategic rationale to optimize our model's performance and applicability to real-world diagnostic challenges. The preprocessing and augmenting techniques used were tailored specifically for the best practice in deep learning for brain tumour classification. The architectural choices for our CNN were focused on the need to extract and process the different complex features that were found in the MRI images, also factoring in the model's interpretability and computational efficiency. Moving forward, the training parameters were chosen specifically to ensure that our proposed model could achieve very high performance with minimal to zero overfitting on the training data. All in all, this was completed to fulfill research objective II.

In the process of developing our research, BrainLens, there were numerous amounts of technological and process challenges that were addressed through the different iterative cycles of building, training, and testing. The most detrimental limitation we faced was with computational resources. As our model consists of many hidden layers, the deep network required significant processing power to train the model, especially with a large dataset used. To mitigate this, we had to constantly refine our model to bypass this limitation while still maintaining the necessary factors that contributed to the complexity of our model. We also increased the batch size from 8 to 16 to reduce memory consumption. Moreover, another complication that we faced was with data preprocessing. The variability in the MRI image quality and format in the dataset posed challenges in standardizing the dataset for our model's training. We developed a preprocessing method that normalized image intensities and resized images (256 x 256) to ensure consistency. However, fine-tuning these preprocessing steps required multiple iterations to find the optimal configuration that preserved essential diagnostic features of the different classes of brain tumours.

IV. Results

Our proposed brain tumour classification model achieved a very strong and impressive accuracy of 96.64% on the test set used. This indicates that the model has exceptional abilities in distinguishing the different complex features of the MRI scans to classify them into their respective type of brain tumour.

As discussed in Section III, the most significant contributing factor to the high accuracy was the model's architecture and methodology for its design. The deep neural network was structured to capture the complex features of the MRI scans from the tumour types of 'pituitary', 'meningioma', 'glioma' and 'notumor'. The initial layers of the CNN, with their smaller convolutional filters, were designed to detect the finer details and edges as the network progressed. The hidden convolutional layers increased with filter sizes to adapt and capture the higher-level features such as the texture and shape of the tumours. With MaxPooling, dense layers, batch normalization, and the utilization of a softmax activation function, this all ensured that the probability and model prediction could be interpreted with confidence. An example of a CNN architecture can be seen in Fig. 5, which is not fully identical to our model, but is similar as

it represents the typical structure and flow of our BrainLens model.

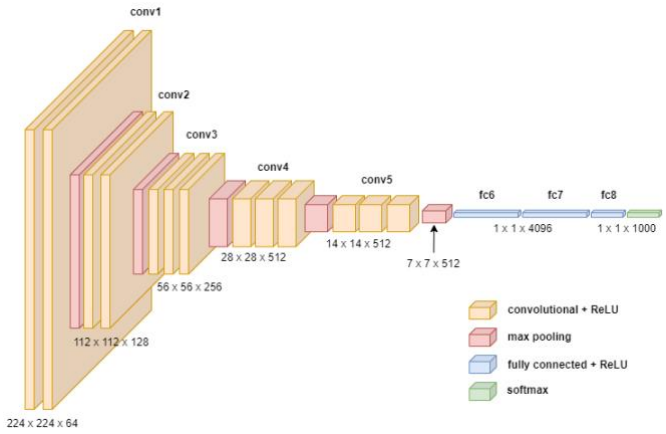

Fig. 5. Convolutional Neural Network (CNN) architecture (adapted from [8]).

The model's performance over training epochs is depicted in two key figures: model accuracy and loss. Starting with the accuracy plot (Fig. 6), we see an initial sharp rise in accuracy within the first few epochs, a clear indication that the model quickly learned to classify the tumour types from the MRI scans. After the initial surge, the accuracy curve begins to level around the 10th epoch, hovering close to the achieved 96.64% accuracy, with the training accuracy slightly above the validation. This trend suggests that the model had effectively captured the underlying patterns in the dataset and converged on an optimal set of weights for classification.

The initial rapid improvement in model accuracy reflects the successful feature extraction capabilities of the initial convolutional layers in the neural network. The immediate identification and learning of critical imaging features like edges, contrasts, and textures are crucial for differentiating tumour types, which speaks to the robustness of the early feature detection strategies integrated into the model's design.

The levelled phase in the accuracy figure indicates a matured learning phase where the model has learned the defining characteristics of the data. This stabilization is likely due to the deeper layers of the CNN, which process more abstract features, reaching an optimal state that correctly identifies MRI patterns associated with each tumour class. The convergence of the training and validation accuracy implies that while the model fits well with the training data, it also maintains a high level of performance on data it has not seen before, showcasing the model's ability to generalize.

Moving forward, in Fig. 7, the loss plot presents a descending trajectory, illustrating that the model steadily reduced the prediction error as it learned. The training loss descends smoothly, reflecting the model's growing confidence in its predictions. Meanwhile, the validation loss shows greater fluctuations, particularly with pronounced peaks at around epochs 3, 6, and 9. These peaks could indicate that the model encountered a few instances where the validation set deviated from the learned patterns, prompting the model to adjust its weights to better generalize to these new data points. In addition to this, the fluctuations in the validation loss challenge the model's current state, triggering necessary adjustments in the model's parameters. Despite these peaks, the general trend is downward, converging closely with the training loss as epochs advance, which aligns with the increasing accuracy and signifies effective learning.

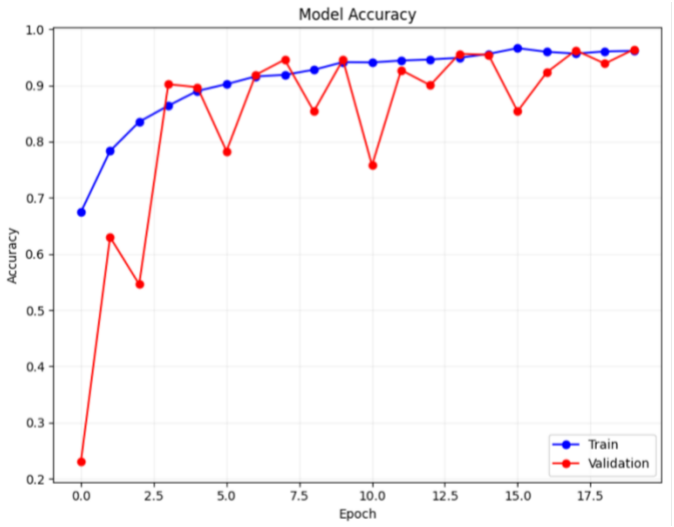

Fig. 6. Model accuracy plot with training and validation sets as a function of training epochs.

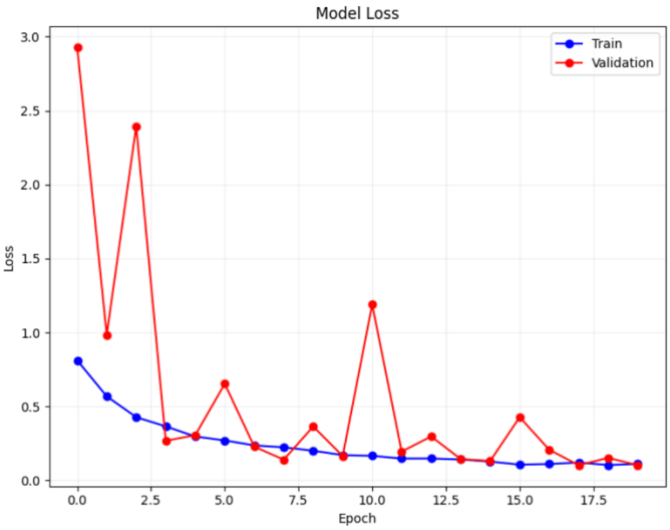

Fig. 7. Model loss plot with training and validation sets as a function of training epochs.

Further examining the results of the classification model, individual sample classifications (Figs. 8 and 9) gave us direct evidence of the model's predictive ability. Correct predictions with high confidence (Conf.) scores, like 'glioma', 'pituitary' and 'notumor' cases with scores of 1.00, 0.99, and 0.98 respectively, underscore the model's precision. These instances, where the model demonstrates near-perfect confidence, highlight the effectiveness of our layered CNN architecture in extracting and interpreting the fine features specific to different

types of brain tumours. The high confidence scores in these correct classifications suggest that the model had not only learned the general patterns associated with each tumour type but was also able to apply this knowledge to accurately classify new, unseen images.

However, out of the sample predictions, there was an incorrect classification where the predicted tumour, with a confidence of 0.84, was ‘meningioma’ when its true class was ‘glioma’. This indicated a case where the model may have struggled due to similarities in tumour presentation or insufficient learning from the training data. These misclassifications are particularly insightful, as they highlight the challenges inherent in medical imaging, where different tumour types can often present with overlapping features. The confidence score of 0.84 in this misclassification case reveals a level of uncertainty in the model's decision, suggesting areas where the training data may have needed to be refined with additional hyperparameters, or further augmentation. This also indicates that the model could have been refined slightly better to distinguish between similar tumour presentations. However, the sample predictions as a whole present a clear representation of the model’s strength to classify correctly.

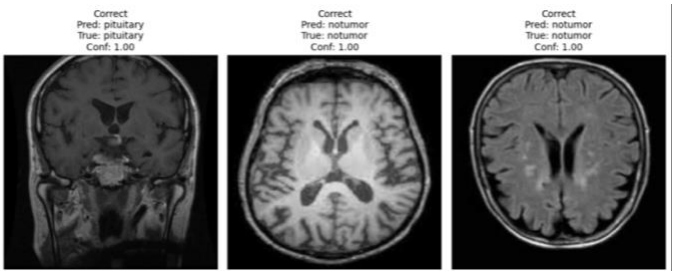

Fig. 8. Sample predictions comparing MRI scans with their predicted and true classes, alongside confidence scores.

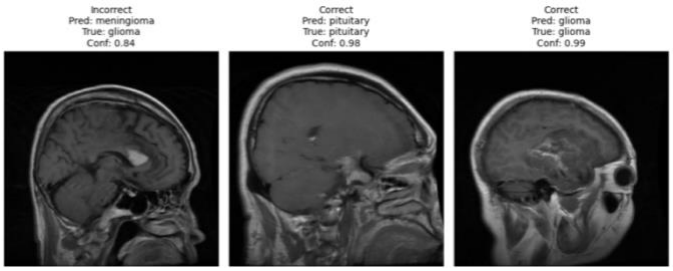

Fig. 9. Sample predictions comparing MRI scans with their predicted and true classes, alongside confidence scores.

Moving forward, another metric we used to analyze the performance of our model was a receiver operating characteristic (ROC) curve, which can be referenced in Fig. 10. We wanted to evaluate the model’s performance across the different classes of brain tumours. After performing the metric, it appeared that the area under the ROC curve (AUC) for each class – ‘glioma’, ‘meningioma’, ‘notumor’, and ‘pituitary’ were exceptional, all nearing 1.00, demonstrating the model's excellent sensitivity and specificity. In essence, for almost every instance, the model correctly discerned the presence or absence of each tumour type without being confused by false positives. The notable AUC values near 1.00 for all classes indicate not just the model's accuracy, but also its reliability across diverse scenarios. This uniform high performance across different tumour types suggests a well-balanced model that does not favour one class over another, a common challenge in medical diagnostics where imbalanced datasets can skew a model's predictions. The high AUC values across the board reflect comprehensive learning from the dataset, where the model has effectively internalized the distinguishing features of each tumour type. The depth and complexity of the CNN’s architecture helped contribute to the results present in the ROC curve.

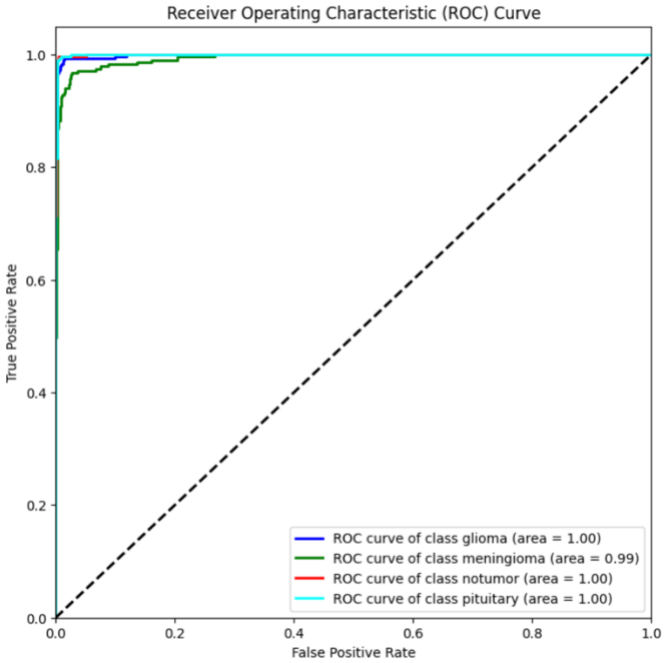

Fig. 10. Receiver Operating Characteristic Curve (ROC) based on the performance of the four brain tumour classes.

Another metric used was a confusion matrix (Fig. 11), to identify the number of accurate, false positive and negative classifications. Fig. 11 provides the results of our model through the different tumour classes. The model performed best with classifying ‘notumor’ and ‘pituitary’ cases at 404 and 299 true positives respectively, which shows the model’s strengths in recognizing the distinct features of these tumour types. The ‘glioma’ class had 295 true positives and the ‘meningioma’ class had 269 true positives. Consequently, the confusion matrix seemed to have revealed a slight challenge in differentiating between ‘glioma’ and ‘meningioma’, with 12 instances of each being misclassified as the other, ‘notumor’ with ‘meningioma’ with 14 misclassified instances, and finally ‘pituitary’ with ‘meningioma’ with 11 misclassifications. These areas, where the accuracy is the lowest point to the complexities in distinguishing these tumour times, potentially being the overlap of radiological features in MRI scans. The misclassification cases between these types suggest a potential need to adopt more thorough feature extraction capabilities in our CNN. The confusion matrix was an essential way for us to understand and address areas of higher and lower accuracy.

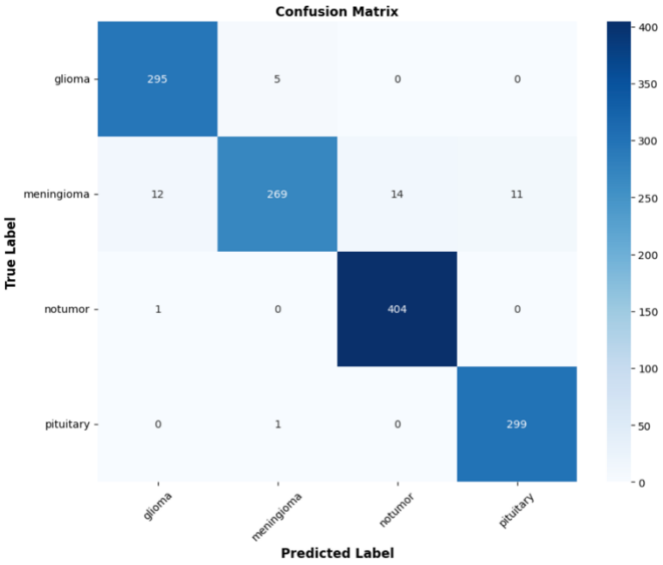

Fig. 11. Confusion matrix for true positive, false positive and negative classifications of tumour types from the test set.

Finally, the last method used to examine the performance of our model was through precision, recall, and F1-score metrics. We wanted to observe how often our neural network performed with true positives, false positives, and negatives. Figs. 12 and 13 represent these performance tests as a plot and classification report respectively. Precision, which measures the number of true positives against the sum of true positives and false positives, is measured at a score of 0.96 for the ‘glioma’ class, indicating that the model predicted this category 96% percent of the time. Noting this high precision is representative of our model having a low rate of either false positives or negatives, which can translate to a low likelihood of misdiagnosing a patient as having a glioma tumour. Moreover, the recall measure was to determine the model’s ability to identify all true positives correctly. With scores of 1.00 for both 'notumor' and 'pituitary,' the model demonstrated perfect recall, meaning it correctly identified all cases within these categories without missing any. This is particularly vital in ensuring that no cases of these tumour types go undetected. Finally, the use of the F1-score was to consider both the false positives and false negatives (balancing the two). Observing the results, all classes had high F1-scores – ‘glioma’ with a score of 0.97, ‘meningioma’ being 0.93, ‘notumor’ with 0.98, and ‘pituitary’ having a score of 0.98. This tells us that the model is not only accurate but is very consistent. It maintains a balance between not overlooking actual cases of tumours (high recall) and not over-diagnosing tumours where they aren't present (high precision).

Our model's uniqueness lies in its specialized CNN architecture, data handling, and augmentation strategies specifically designed for the domain of brain MRI analysis. This custom approach, unlike generalized models, caters to the intricate patterns present in medical imaging data, which is evident in the high accuracy rates and robust performance metrics observed in all results presented.

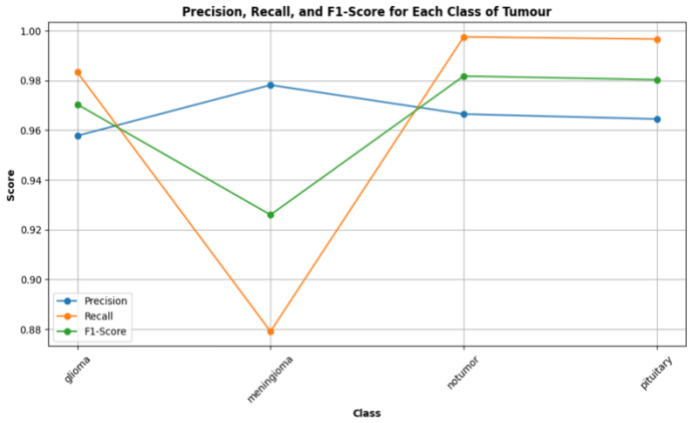

Fig. 12. Precision, Recall, and F1-Score plot based on the test set inputted into the brain tumour classification model.

	precision	recall	f1-score	support
glioma	0.96	0.98	0.97	300
meningioma	0.98	0.88	0.93	306
notumor	0.97	1.00	0.98	405
pituitary	0.96	1.00	0.98	300
accuracy			0.97	1311
macro avg	0.97	0.96	0.96	1311
weighted avg	0.97	0.97	0.97	1311

Fig. 13. Classification report with Precision, Recall, and F1-Scores for each tumour class.

V. Conclusions

In concluding our work on our research, BrainLens, it is confident to say that our objectives have been met with significant success. The project aimed to enhance the precision of brain tumour MRI classification using advanced convolutional neural network (CNN) techniques, and the results have been highly promising. With an achieved accuracy of 96.64%, the effectiveness of the model in classifying tumours into ‘glioma’, ‘meningioma’, ‘notumor’, and ‘pituitary’ classes is supported.

The high precision, recall, and F1-scores achieved across all classes confirm the model’s robustness and its capability to handle diverse datasets, a common challenge in medical imaging. Particularly, the perfect recall scores for 'notumor' and 'pituitary' tumour categories underscore the model's precision in identifying these classes. These results bring us a step closer to integrating artificial intelligence into clinical settings, offering a tool that could potentially streamline diagnostics and treatment planning.

Many valuable lessons have been learned from this project. Firstly, we understood the importance of a diverse and detailed dataset, as well as the effectiveness of data augmentation in enhancing model generalization. We also learned that while our model performs exceptionally well in certain classifications, it can slightly struggle with others, indicating the need for continuous refinement.

Future work based on BrainLens can take several directions. One area is to further improve the model's ability to distinguish between ‘glioma’ and ‘meningioma’ tumours, potentially through the integration of additional imaging modalities or advanced feature extraction techniques. Another path is to expand the dataset to include a broader range of tumour presentations, including rarer subtypes, to improve the model's completeness.

Additionally, while our model is optimized for high accuracy, the computational efficiency and speed also need to be considered, especially for deployment in real-world clinical environments. Therefore, future iterations could focus on model pruning and optimization for faster inference without compromising accuracy.

Lastly, potential studies could explore the model's transferability to other forms of medical imaging data, testing its adaptability and learning capability across different imaging methods.

In essence, BrainLens has set a foundational benchmark in brain tumour classification using deep learning. It represents a significant development in the field of medical artificial intelligence, opening numerous possibilities for future exploration and development in precision medicine.

REFERENCES

[1] “Brain Tumor MRI Dataset,” *www.kaggle.com*. https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

[2] B. Ahmad, J. Sun, Q. You, V. Palade, and Z. Mao, “Brain Tumor Classification Using a Combination of Variational Autoencoders and Generative Adversarial Networks,” *Biomedicines*, vol. 10, no. 2, p. 223, Jan. 2022, doi: https://doi.org/10.3390/biomedicines10020223.

[3] M. Cè *et al.*, “Artificial Intelligence in Brain Tumor Imaging: A Step toward Personalized Medicine,” *Current Oncology*, vol. 30, no. 3, pp. 2673–2701, Mar. 2023, doi: https://doi.org/10.3390/curroncol30030203.

[4] R. Kaifi, “A Review of Recent Advances in Brain Tumor Diagnosis Based on AI-Based Classification,” *Diagnostics*, vol. 13, no. 18, p. 3007, Jan. 2023, doi: https://doi.org/10.3390/diagnostics13183007.

[5] W. Chen, X. Tan, J. Zhang, G. Du, Q. Fu, and H. Jiang, “A robust approach for multi-type classification of brain tumor using deep feature fusion,” *Frontiers in neuroscience*, vol. 18, Feb. 2024, doi: https://doi.org/10.3389/fnins.2024.1288274.

[6] J. Kang, Z. Ullah, and J. Gwak, “MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers,” *Sensors*, vol. 21, no. 6, p. 2222, Mar. 2021, doi: https://doi.org/10.3390/s21062222.

[7] S. Krishnapriya and Y. Karuna, “Pre-trained deep learning models for brain MRI image classification,” *Frontiers in Human Neuroscience*, vol. 17, Apr. 2023, doi: https://doi.org/10.3389/fnhum.2023.1150120.

[8] K. Leung, “How to Easily Draw Neural Network Architecture Diagrams,” *Medium*, Sep. 23, 2021. https://towardsdatascience.com/how-to-easily-draw-neural-network-architecture-diagrams-a6b6138ed875