INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 17 No. 5 (2025) 389-401
https://publisher.uthm.edu.my/ojs/index.php/ijie

A Lightweight CNN Model Using Depthwise Separable Convolutions for Brain Tumour Classification

Aifian Adi Sufian Chan[1,2], Saizalmursidi Md Mustam[2]*, Mohammad Faiz Liew Abdullah[2], Farhana Ahmad Po'ad[2], Ariffuddin Joret[2], Hashim Safdar[3]

[1] *Penang - Robert Bosch Malaysia Sdn. Bhd.,*
Bayan Lepas Free Industrial Zone Phase 1, 11900 Bayan Lepas, Pulau Pinang, MALAYSIA
[2] *Faculty of Electrical and Electronic Engineering,*
Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, MALAYSIA
[3] *Department of Electrical Engineering,*
Federal Urdu University of Arts Science and Technology, Islamabad 44000, PAKISTAN

*Corresponding Author: saizal@uthm.edu.my
DOI: https://doi.org/10.30880/ijie.2025.17.05.030

Article Info

Received: 28 February 2025
Accepted: 21 April 2025
Available online: 30 August 2025

Keywords

Depthwise separable convolution, brain tumour, convolutional neural network, deep learning, image classification

Abstract

Every year, the number of patients with brain cancers (BCs) or brain tumours (BTs) increases. This trend emphasises the necessity of a computerised system for rapid and accurate detection during the diagnosis of BTs. This paper presents a lightweight deep learning (DL) model based on a convolutional neural network (CNN) for a fast and accurate BC detector. The core component of the BC detector is a depthwise separable convolution (DSConv) on top of the 24-layer CNN architectures. The usage of DSConv with Adam's optimiser achieves comparable effectiveness to conventional convolutional layers, although using fewer parameters. Additionally, L2 regularisation, dropout, and data augmentation were implemented to mitigate the issues of overfitting. The proposed model was trained and tested using the publicly available dataset consisting of MRI images collected from 233 patients in Nanfang Hospital and General Hospital, with 3063 images in total. In summary, the DSConv-based CNN model demonstrates an average accuracy of 97.50% and has an average inference time of 2.1 milliseconds per classification. It consistently surpasses 96.50% accuracy in the classification of the three types of BTs. These findings indicate that the model is well-suited for accurate BTs classification, particularly for glioma, meningioma, and pituitary tumours from MRI images.

1. Introduction

Brain cancer (BC), also known as a brain tumour (BT), is one of the deadliest types of tumours that arise from brain cells growing out of control [1-3]. Global cancer data from 2022 revealed approximately 321,476 new cases of BC, accounting for 1.6% of all new cases [4]. Overall, roughly 248,305 deaths were attributable to BC. This figure represents 2.6% of all cancer-related deaths. Additionally, in 2021, BTs accounted for the majority of cancer deaths, particularly among those below the age of 20 [5]. For this reason, earlier detection of cancer is necessary for enhancing survival rates and maximising the effectiveness of treatment [6].

Therefore, the detection of cancer as early as possible is essential for increasing survival rates and maximising effective treatment [6]. When diagnosing BTs, the most common and reliable method is magnetic resonance imaging (MRI). Due to its excellent soft tissue contrast [7], MRI offers precise visualisation of the brain tissue [8],

improving detection. While MRI scans significantly enhance neurologists' diagnostic capabilities, the increasing patient population poses problems for the healthcare system, such as misdiagnosis.

Numerous studies in the literature have recommended the integration of a computer-aided diagnostic (CAD) system to improve the accuracy, efficiency, and reliability of medical diagnosis. The combination of CAD systems with deep learning (DL) techniques, like convolutional neural networks (CNN) [9], can significantly enhance detection and classification capabilities, leading to faster and more accurate diagnoses [3, 7, 8, 10-12]. Some of the existing architectures and their corresponding accuracy include the following: a 23-layer CNN (97.8%) [3]; convolutional dictionary learning with local constraints (96.39%) [7]; a separable CNN (97.52%) [8]; a deep neural network (96.13%) [10]; a three-hybrid approach combining dense speeded-up robust features (DSURF), histogram of orientated gradients (HoG), and support vector machines (SVM) (90.27%) [11]; and CNNs enhanced by genetic algorithms (94.2%) [12].

While the accuracy of these methods is promising, each approach has its limitations. For example, some methods rely heavily on large datasets and incur high computational costs [3, 10]. The iterative optimisation process can be computationally expensive, which may result in overfitting when tested with smaller datasets [7]. Additionally, simpler architectures may have difficulty in capturing complex patterns, especially with more challenging datasets [8]. Although hybrid models are versatile, they can be complicated and often require longer training times [11, 12], which increases the risk of overfitting [11]. Overfitting is a common issue when the model learns not only the patterns in the training data but also the noise present in the data, causing poor performance on the new dataset. Additionally, using a complex model in comparison to the available training data can also contribute to overfitting. To prevent overfitting, several methods can be employed, including regularisation techniques: dropout [1, 3, 12] and batch normalization [13], data augmentation [3, 14, 15] and early stopping [1].

In terms of datasets, all researchers used publicly accessible MRI-scanned images for the model's training and testing. The contrast-enhanced (CE)-MRI Figshare dataset, published in [16], contains 3064 MRI images from 233 patients diagnosed with meningioma, glioma, or pituitary tumours and was referenced in [3, 8, 10-12]. In addition, [3] used a Harvard dataset comprising 152 images: 71 healthy brains and 81 with tumours, such as glioma, metastatic adenocarcinoma, metastatic bronchogenic carcinoma, meningioma, and sarcoma. The repository of molecular brain neoplasia data (REMBRANDT) dataset from the cancer imaging archive (TCIA) contains over 110,000 MRI images from 130 patients with various tumours, including glioblastoma and astrocytoma, and was used to study glioma grades [7, 10, 12]. The IXI dataset provides normal brain images for comparison and was used in [12]. To sum up, each dataset serves different research needs, with CE-MRI being the most commonly referred dataset.

Although all the aforementioned methods are promising, challenges such as computational cost, overfitting, and dataset size remain and still need to be addressed. The review emphasises that the use of CAD systems can significantly enhance the skills of neurologists and speed up the period of diagnosis. This enables the healthcare system to manage the continuously growing influx of patients. Thus, this paper intends to come up with a CAD system utilising DL technology to assist neurologists in identifying and classifying BTs.

Since a lightweight DL framework with high accuracy is vital for practical applications in clinical settings, this paper presents a lightweight CNN model called BC detector using depth-wise separable convolution (DSConv). The proposed model aims to accurately categorise three common types of BTs: glioma, meningioma, and pituitary tumours, using a publicly available dataset referenced in [16], which encompasses a diverse variety of tumours. To combat overfitting, the model uses a combination of several methods, such as L2 regularisation, dropout, and data augmentation. These combination techniques can improve the model's ability to generalise to new datasets. In this paper, the following contributions are highlighted:

- The development of a DSConv-based CNN model with the ability to categorise three types of BTs.
- The model is lightweight and can operate on low-spec hardware.
- It demonstrates performance that is on par with or better than previous methods [3, 7, 8, 10-12].

The rest of this paper is organised as follows: Section 2 outlines the methodology, which includes the block diagram of the proposed models, image acquisition and preprocessing procedures, the architecture of the proposed BC detector, hyperparameter fine-tuning settings, and the performance metrics. Section 3 demonstrates the outcomes and discusses the performance thoroughly. In Section 4, the model's performance is concluded, its limitations are highlighted, and suggestions for future work are offered.

2. Methodology

Fig. 1 depicts the operation of the BC detector using DSConv for classifying three types of BTs. It starts with image acquisition from the selected dataset. Afterwards, the dataset is preprocessed and then split into training and testing sets. Both datasets are essential during the model's training and testing phases. Throughout the training phase, a lightweight BT classification model (BTcm), consisting of a BC detector using DSConv, is developed. Then, hyperparameters are fine-tuned to enhance the model's stability and speed.

Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 389-401

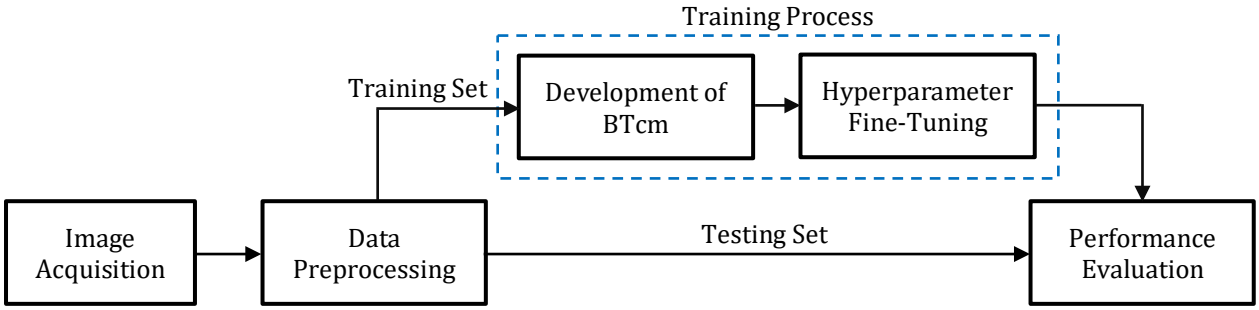


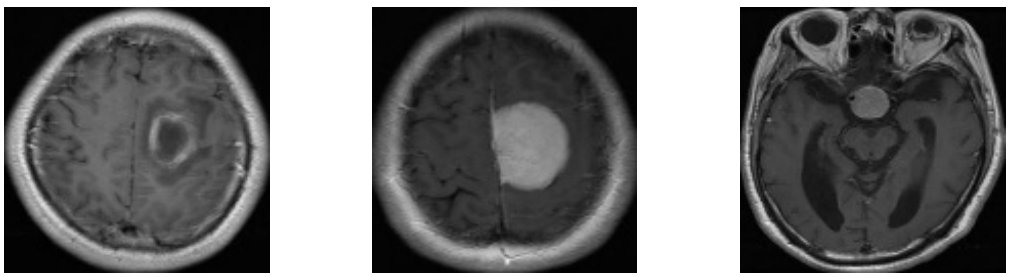
Fig. 1 *A block diagram of the BC detector using DSConv*

2.1 Image Acquisition: Dataset of BT and Preprocessing Procedures

Image acquisition is a crucial step in developing any DL model, as both the qualitative and quantitative aspects of a dataset directly influence the model's accuracy and reliability. In this paper, the BT dataset compiled in [16] is utilised for training and testing the proposed BC detector using DSConv. This dataset has MRI images, which were T1-weighted contrast-enhanced, from 233 patients who were treated at Nanfang Hospital and General Hospital in China from 2005 to 2010. They were diagnosed with three types of BTs: gliomas, meningiomas, and pituitary tumours. The total number of MRI images contained in the dataset is 3,064, which includes 1,426 images of glioma tumours, 708 images of meningioma tumours, and 930 images of pituitary tumours. This dataset is publicly accessible at http://dx.doi.org/10.6084/m9.figshare.1512427; however, specific demographic information such as age, race, region, and other patient records has been anonymised due to medical security and is classified as confidential medical data. Thus, the dataset diversity in this paper focuses solely on the three main primary types of brain tumours, using MRI images collected from two hospitals in China. Table 1 tabulates the distribution of the BT dataset's image. Meanwhile, Fig. 2 demonstrates sample images of the three types of BTs.

Table 1 *Images distribution in the BT dataset according to the type of BTs* [16]

Types of BTs	Number of patients	Total images	Type of view	Number of images
Glioma	91	708	Coronal	268
			Sagittal	231
			Transverse	209
Meningioma	82	930	Coronal	319
			Sagittal	320
			Transverse	291
Pituitary	60	1426	Coronal	437
			Sagittal	495
			Transverse	494
Total	233	3064	-	3064

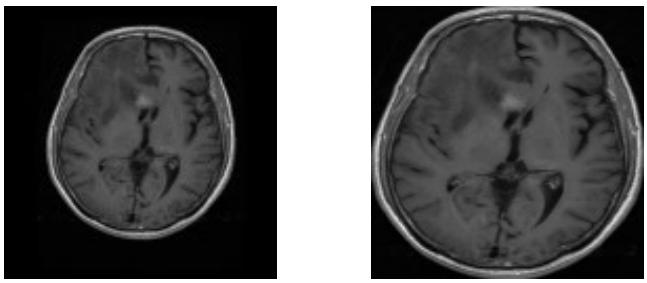


(a) **(b)** **(c)**

Fig. 2 *Three different types of BTs (a) Glioma tumour; (b) Meningioma tumour; (c) Pituitary tumour* [16]

Initially, the images in the BT dataset were provided in MATLAB data format (.mat file), with each file containing an image of 512 × 512 pixels along with its labels. The .mat files dataset was then changed to JPEG to understand the BT dataset better and assist in the training phase. After this conversion, the preprocessing processes were implemented, which included cropping the photos to eliminate any undesired areas and applying Gaussian blur for noise reduction in the MRI images [3]. The reason for selecting Gauss blur is because it strikes the right balance between noise reduction, detail preservation, and computational efficiency. The images were then resized from 512 × 512 pixels to 150 × 150 pixels. The main reason for this resizing is to reduce the computational load, which helps to facilitate faster processing and analysis.

These preprocessing procedures are considered for several reasons. First, it helps minimise or eliminate noise from the image, which can limit the model's training ability and may affect its accuracy. Second, these steps reduce the computational burden by downsizing the image and eliminating unneeded areas to increase computational efficiency. Finally, these steps enhance the model's learning capability as it focuses more on learning essential features from the image, which leads to an increase in the model's accuracy. Fig. 3 shows a sample image before and after the preprocessing steps.



(a) **(b)**

Fig. 3 *A sample image (a) Before preprocessing; (b) After preprocessing*

After the preprocessing steps, the images in the BT dataset were shuffled and divided into two separate sets, namely the training set and the testing set. The training set and testing set images are allocated at an 80% and 20% ratio, respectively. Table 2 provides the detailed distribution for both the training and testing sets.

Table 2 *The number of images for training and testing sets*

Type of BTs	Training set	Testing set	Total
Glioma	1140	286	1426
Meningioma	566	142	708
Pituitary	744	186	930
Total	2450	614	3064

According to Table 2, the BT dataset comprises 2,450 images designated for training the proposed model. Using a smaller dataset might lead to a high risk of overfitting during the training process [14, 15], where the model's generalisation capability is limited to its training data and does not generalise well to a newly discovered dataset. To mitigate overfitting, both the size and diversity of the dataset are increased through data augmentation techniques. This approach aims to improve the generalisation capabilities of the models. Numerous data augmentation techniques are available, including geometrical and colour space transformations [15]. However, only specific data augmentation techniques were deemed suitable for this dataset and were implemented accordingly.

In order to reduce overfitting and increase the dataset for better diversity and variations, several data augmentation techniques were implemented. First, 10° rotation was applied to help the models recognise tumours at slightly different angles. Next, the brightness intensity range was set from 0.85–1.15. This was done to ensure the model could handle the differences in scanner intensity. The third augmentation is applying small width and height shifts (0.002). This will make the model more resilient as it introduces tumours in different positions. Next, 12.5° shear was applied to help the model adapt to subtle distortions in MRI scans. Finally, horizontal flipping was implemented to improve generalisation by training the model to recognise tumours even if they appear mirrored. Table 3 tabulates the data augmentation techniques used and their corresponding settings to generate new variations of images for the training set.

Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 389-401

Table 3 *The parameter settings used for data augmentation techniques*

Data augmentation techniques	Parameter setting
Rotation range	10
Brightness range	0.85-1.15
Width shift range	0.002
Height shift range	0.002
Shear range	12.5
Horizontal flip	True

2.2 The Proposed BC Detector Model Architecture

Fig. 4 illustrates the architecture of the proposed model. This configuration aims to classify three types of tumours: gliomas, meningiomas, and pituitary tumours. The model starts by processing the MRI slice, which serves as input to the input layer. The training process will help the model learn to differentiate each image and categorise it accordingly.

The proposed architecture consists of 24 layers, primarily utilising the DSConv layer. The DSConv layer was introduced by Challet *et al.* [17], but this layer is not fully functional. It was used in various famous models, including Xception [17], MobileNet [18], and GhostNet [19]. The main advantage of using the DSConv layer instead of standard convolution is its computational efficiency and its use of fewer parameters.

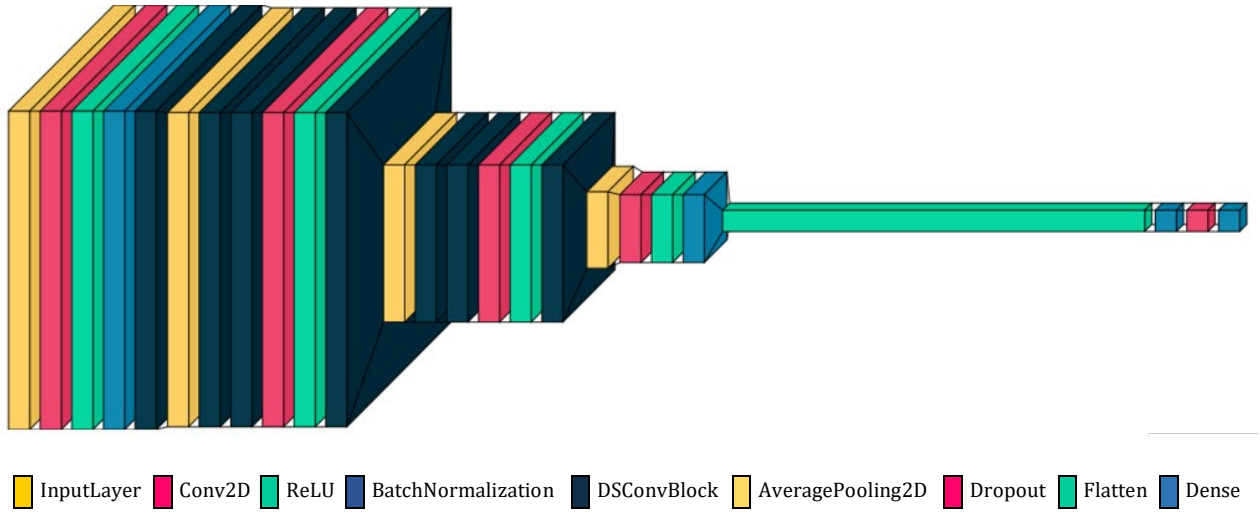


Fig. 4 *Visualisation of the proposed model architecture for the BC detector using DSConv*

The DSConv layer enhances computational efficiency by utilising two types of convolutions: depthwise convolution (DC) and pointwise convolution (PC). The PC combines the output of the DC linearly, while the DC applies a single convolutional filter to each input channel. The proposed architecture is now both lightweight and practical. Because of this, it works effectively on devices with minimal specifications and little processing power. The total parameters involved in the proposed design are less than one million.

Additionally, the activation function used is a rectified linear unit (ReLU). The ReLU introduces non-linear components to solve the gradient vanishing problem during backpropagation. This can be represented mathematically as follows:

$$f(x) = max(0, x) \tag{1}$$

Where $f(x)$ represents the output after the ReLU function is applied, and x is the input to the activation function. If $x > 0$, the output is equal to the input, $f(x) = x$. If $x < 0$, the output is equal to zero, $f(x) = 0$.

Furthermore, a pooling layer was used to down-sample a feature map, which can be calculated as follows [18]:

$$f_{avg}(x) = \frac{1}{N}\sum_{i=1}^{N}|x_i| \tag{2}$$

Where $f_{avg}(x)$ represents the average absolute value of the inputs, N is the total number of inputs, x_1, x_2, …, x_N are the inputs, and $|x_i|$ is the absolute value of the inputs.

Basically, average pooling and max pooling are two commonly used types of pooling layers in CNN. In this paper, average pooling was utilised because it yields superior results. The proposed architecture, illustrated in Fig. 4, incorporates three average pooling layers. Each layer uses a 2 × 2 kernel size with a stride of 2. In this case, the average pooling layer down-samples a feature map by calculating the average values of each specified region. Fig. 5 provides a visual representation of the working process of an average pooling layer operation.

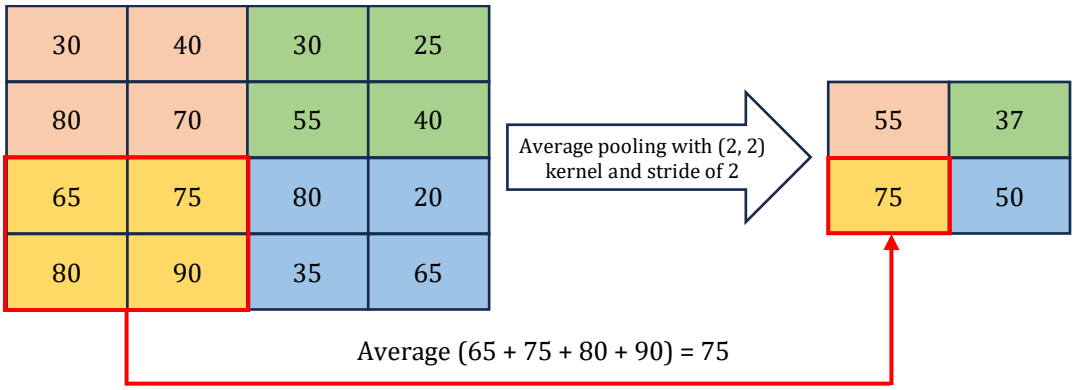


Fig. 5 *Visualization of an average pooling operation*

Moreover, the model employed two batch normalisation (BN) layers for optimal outcomes. These BN layers boost model performance by using a higher learning rate during training [13], which promotes faster convergence and improves the model's generalisation capabilities. Additionally, two techniques were used to reduce the problem of overfitting, particularly in a small dataset. First, a dropout layer was incorporated to randomly deactivate units [20] according to specified probabilities (0.35, 0.35, and 0.45). This strategy prevents the occurrence of co-adaptation and encourages robust feature learning that leads to better generalisation of the model. Second, the BC detector incorporated L2 regularisation, which adds a penalty term to the loss function. This penalty encourages the BC detector to allocate weights across the features uniformly. These techniques mitigate the risk of overfitting, and the BC detector does not excessively rely on any single feature.

2.3 Hyperparameter Fine-Tuning

Hyperparameter fine-tuning is an effective technique used to improve the overall performance of the proposed model [21]. By adjusting various hyperparameters, the training process' stability and speed can be enhanced, leading to improved performance in classifying BTs. Table 4 presents the hyperparameter settings used in the proposed model.

Table 4 *Fine-tuned hyperparameter settings*

Hyperparameter	Setting
Batch size	32
Learning rate	0.0003
Number of epochs	200
Optimiser function	Adam
Loss function	Categorical cross-entropy

In general, there are several optimisers available, but Adam Optimiser, with a batch size of 32, has demonstrated the best performance. The learning rate used is 0.0003. It assists the model in achieving better convergence. Categorical cross-entropy is the preferred loss function for multiclass image classification problems. Additionally, early stopping was implemented by monitoring both validation and training losses. For example, the training phase will end if the model achieves a plateau after 10 epochs in both losses, as it is evident that additional training will not enhance the performance of the model or its generalisation capability. This helps prevent overfitting and saves training time.

Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 389-401

2.4 Performance Evaluation Metrics

In this paper, the BC detector was developed and evaluated using Python, TensorFlow, and Keras. The steps explained in the previous sub-sections were implemented on hardware equipped with an RTX 3060 GPU and 32 GB of RAM. The proposed model's performance was assessed using four metrics: accuracy, precision, recall, and F1-score. However, accuracy has become the primary performance metric under study to determine its fitness in assisting neurologists in identifying three different types of BTs. All these four metrics can be defined mathematically as follows:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{5}$$

$$\text{F1-score} = \frac{2\ (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{6}$$

In the context of classification, TP represents true positive, FP stands for false positive, TN indicates true negative, and FN refers to false negative.

3. Results and Discussion

In this section, the performance evaluation results of the developed BC detector are presented. Generally, the results are organised in the following order: model performance analysis, confusion matrix analysis, precision-recall curve analysis, and a comparative analysis with the previously proposed models. By performing these four perspectives of analysis, the result and discussion will be able to fully demonstrate the proposed model's strengths, limitations, and potential for further refinement, providing a balanced and comprehensive performance evaluation.

In the analysis of model performance, parameters such as accuracy, precision, recall, and F1-score are thoroughly evaluated and discussed. These performance metrics give a basic understanding of the model's robustness and reliability, as well as its practical implications and relevancy in real-world applications. Next, the confusion matrix is used to analyse the classification outcomes of the models. The evaluation of outcomes, such as TP, FP, TN, and FN, is essential. This analysis discovers specific patterns in the model's predictions, which provides valuable insights for improvement, especially in areas such as data representation. The precision-recall (PR) curve analysis evaluates the trade-offs between precision and recall. From this PR curve, the model performance can be assessed as either performing as a no-skill classifier or approaching the perfect classifier. Finally, a comparative study with the existing models will help to evaluate the proposed models' strengths and weaknesses. It will also give a better understanding of its performance compared to other models.

3.1 Performance Analysis

Table 5 and Table 6 tabulate a detailed result of the proposed BC detector (DSConv-based CNN model) performance based on the four previously described metrics when evaluated using the testing set consisting of three types of BTs. In summary, the model successfully classified the BTs with average accuracy, average precision, and recall scores of 97.50%, 96.46%, and 95.71%, respectively. The results indicate its reliable performance in detecting and classifying three distinct tumour categories.

Table 5 *The performance of the proposed in terms of TP, TN, FP, and FN*

Type of BTs	TP	TN	FP	FN
Glioma	278	315	13	8
Meningioma	130	468	4	12
Pituitary	183	422	6	3

Table 6 *The performance of the proposed model in terms of accuracy, precision, recall, and F1-score*

Type of BTs	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
Glioma	96.58	95.53	97.20	0.9636
Meningioma	97.39	97.01	91.55	0.9420
Pituitary	98.53	96.83	98.38	0.9760
Average scores	97.50	96.46	95.71	0.9605

Based on Table 6, the data also provides a clear indication of the model's strengths, weaknesses, and room for improvement. The model's strength can effectively and reliably detect pituitary tumours with minimal errors. Based on the result, the BC detector can perform with the highest accuracy (98.53%), highest precision (96.83%), highest recall (98.38%), and F1-score close to 1 (0.9760). It has the highest accuracy, which means the model can correctly predict pituitary tumours most of the time. Additionally, it also has the highest recall and precision, implying that the model successfully detected 98.53% of actual tumours and correctly predicted almost all (96.83%) of the detected pituitary types of BTs, respectively. The very low false negative (FN = 3), as shown in Table 5, indicates that the model rarely misses in classifying pituitary tumour cases. The F1-score close to 1 (0.9760) indicates that the model is well-balanced between precision and recall.

In general, two important factors contribute to the success of the proposed model. First, there are different features of pituitary tumour images compared to glioma and meningioma tumours. The distinction between these features in images can be observed in Fig. 2(a), Fig. 2(b), and Fig. 2(c). The distinct features help the model to learn better about the pituitary, which can be proved with low FN. The second reason is that it has sufficient training data, which ensures the model learns diverse and representative features of the pituitary glands.

In classifying glioma types of BT, the BC detector performs with an accuracy of 96.58%, a precision of 95.53%, and a recall of 97.20%, showing a balanced result with good recall and relatively low false positive rates. The F1 score of 0.9636 validates this balanced performance by displaying the model's reliability in the categorisation of the glioma type of BTs.

Meanwhile, for the meningioma type of BT classification, the proposed model performs with an accuracy of 97.39% and a precision of 97.01%. Nonetheless, its recall is the lowest compared to the other two types of BTs, which is 91.55%. The lowest recall in Table 6 means the model has more misclassification cases (FN = 12 in meningioma compared to the other two types of BTs); see Table 5.

Nevertheless, its performance is still impressive. The underperformance of this BT type can be attributed to three reasons. The first reason can be attributed to its being the smallest size for both the training set (566) and the testing set (142), as tabulated in Table 2. The small training set might not be sufficient for the model to learn the representative feature of meningioma tumours, making it more challenging to effectively distinguish between meningioma and glioma classes, as there are 10 misclassifications, as can be observed by the confusion matrix in Fig. 6. The second reason is due to the shared similarity between glioma and meningioma tumours highlighted in Fig. 2(a) and Fig. 2(b). The overlap feature, or similarity, makes the model struggle a bit to distinguish between these two classes, where it can be observed from the confusion matrix Fig. 6 that there is a total of 14 misclassifications between them.

Overall, the BC detector demonstrates strong performance on pituitary and glioma tumours while showing room for improvement on meningioma tumours. There are a few steps that can be taken to enhance the performance of the proposed model, which are related to the datasets and the model's architecture. The dataset can be improved via the following:

- Raise the number of dataset images, which helps to provide more representative features, allowing the model to have better generalisation capabilities [22-23].
- Ensures a balanced data distribution among the three classes to prevent any bias toward a specific category.

The model can be improved via the following steps:

- To address misclassifications caused by overlapping features, employing contrastive learning [24] or triplet loss can improve class separability [25].
- Further fine-tuned hyperparameters [26], such as learning rate or batch size, to improve the generalisation capability.

Hence, by incorporating these steps from two different ends of the development, it is possible to improve its performance further.

3.2 Confusion Matrix Analysis

Fig. 6 shows a confusion matrix that summarises the model performance evaluated using the testing set of the BT dataset. The *x*-axis displays the predicted labels, while the *y*-axis shows the ground truth labels. Based on Table 2, there are a total of 614 testing set images: 286 (glioma), 142 (meningioma), and 186 (pituitary). According to Fig.

Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 389-401

6, the proposed model accurately classified 278 out of 286 glioma images, 130 out of 142 meningioma images, and 183 out of 186 pituitary images. Even though the model achieved an overall accuracy of 97.50%, it misclassified a total of 23 images out of 614: eight glioma images as meningioma (4) and pituitary (4), 12 meningioma images as glioma (10) and pituitary (2), and three pituitary images as glioma.

A closer investigation found that misclassification mainly occurs when classifying the glioma and meningioma types of BTs and vice versa. Four images from the glioma type of BT are misclassified as the meningioma, while 10 images from the meningioma type of BT are classified as glioma. These results implied that the model has some difficulty in distinguishing between glioma and meningioma. This is due to the similar image characteristics of the two types of BTs, as shown in Fig. 2(a) and Fig. 2(b). This similarity presents a challenge to the proposed models in classifying these two tumours accurately. To enhance detection, the dataset needs further refinement, incorporating more samples from both classes with distinct features. By implementing techniques such as triplet loss or contrastive learning, misclassification occurrence can be reduced. These two techniques can improve classification performance by learning a feature space where similar tumours are clustered together. In contrast, dissimilar tumours are pushed apart, creating a distinct separation between gliomas and meningiomas tumours.

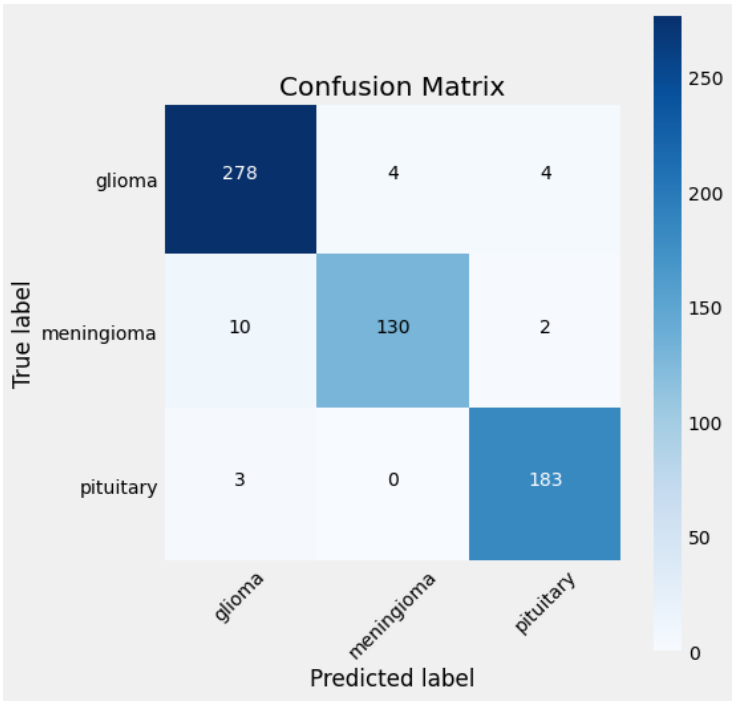


Fig. 6 *Confusion matrix for the proposed BC detector model*

3.3 Precision-Recall Curve

The PR curve shown in Fig. 7 evaluates the BC detector model performance in terms of the trade-offs between precision and recall. The three classes of BTs are sketched by the blue line, green line, and red line to indicate the glioma, meningioma, and pituitary, respectively. This curve provides a great insight into understanding the balance between recall and precision. Ideally, a well-performing or perfect model should have curves that remain close to the top right corner, indicating high precision and recall across different thresholds.

According to the graphs in Fig. 7, it can be observed that the overall model performance is awe-inspiring. This is because all three curves managed to retain high precision for most recall values. However, if the graph is examined, it can be seen that there are differences in performance for the three types of BT classification. The pituitary tumour (red line) presents the best performance, where the precision remains almost perfect across all recall values; this is an excellent indication that the BC detector can effectively classify pituitary tumours with minimal error. However, glioma and meningioma show slight variations in precision, particularly at high recall values. Their curves slightly dip during high recall values, which shows that they are struggling to make good decisions and start making misclassification errors. As mentioned earlier, this can be attributed to the feature similarities between glioma and meningioma, making them harder to distinguish.

It can be concluded that the BC detector performs exceptionally well despite the challenges it encounters. The BC detector can effectively classify pituitary tumours with high confidence, while glioma and meningioma tumours

require further refinement to reduce misclassification at high recall levels. However, by incorporating the step outline in previous sections, the model performance is expected to be further improved.

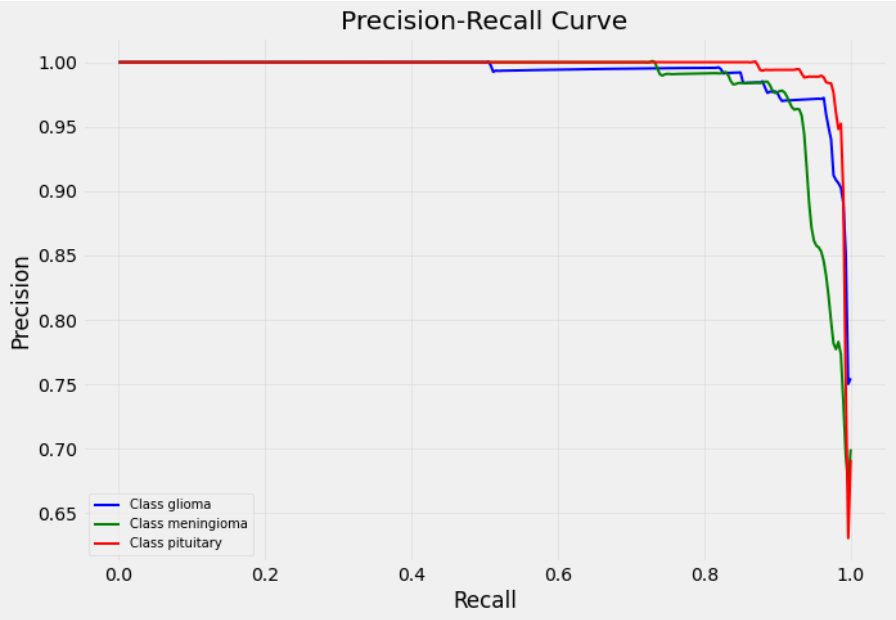


Fig. 7 *Precision-recall curve of the proposed BC detector model for classifying the glioma, meningioma, and pituitary types of BTs*

3.4 Comparative Analysis with the Previous Models

Finally, the proposed BC detector was compared to various previously developed models. Table 7 tabulates a comparison of seven different models, using accuracy as the primary performance metric. The results indicate that the proposed model performs competitively, achieving third place out of the seven models evaluated.

However, it may not be the top-performing model in terms of accuracy; the proposed BC detector significantly improves upon most previous studies, where, in some instances, it outperforms earlier models by as much as 7.23%, whereas the minor improvement recorded is around 1.11%. It falls behind two exception models developed by Isunuri *et al.* [8] and Khan *et al*. [3]. However, these models' performance does not differ too much. In the second place, the model introduced in [8] outperforms the proposed model by a small margin of 0.02%. Meanwhile, the first presented in [3] exceeds the proposed model by just 0.30%. These slight differences in accuracy highlight that the proposed BC detector remains a viable option for this task.

Even more impressive, the developed model managed to achieve a competitive performance and will be one of the lightest models out of these three development models; our model is relatively lighter compared to these two models. For example, the model in [8] has approximately 3.4 million parameters, whereas the proposed BC detector has fewer than one million parameters. This reduced complexity makes the proposed model a better choice for real-world applications where computational resources may be limited while still providing the high accuracy needed to detect and classify the three classes of BTs.

Table 7 *An average accuracy performance in comparison to previous related works*

No	Model or approach	Average accuracy (%)
1	Three-step hybrid approach, DSURF + HoG + SVM [11]	90.27
2	CNN with genetic algorithm [12]	94.20
3	Deep neural network [10]	96.13
4	Convolutional dictionary learning with local constraint [7]	96.39
5	DSConv-based CNN model (the proposed model)	97.50
6	Separable convolution-based neural network [8]	97.52
7	23-layer CNN architecture [3]	97.80

4. Conclusion

In conclusion, a DSConv-based CNN model has been developed to detect and classify three types of BTs: gliomas, meningiomas, and pituitary tumours. The architecture consists of 24 layers and primarily utilises the DSConv layer. To address overfitting due to the small dataset, several techniques were implemented, including dropout layers, L2 regularisation, and data augmentation. To enhance its performance, several hyperparameters were fine-tuned to achieve optimal results. The performance evaluation was conducted using a testing set derived from the BT dataset. The model achieved a high classification accuracy of over 96.50% across three different BT classes. Furthermore, the proposed model can classify the BTs with 97.50% accuracy on average, surpassing most previous works, except for the two proposed models by Isunuri *et al.* and Khan *et al.* However, due to the small margin in accuracy, the proposed model remains a strong competitor for this task. While the proposed model has achieved excellent results, some issues remain. The main problem is the lack of quality brain tumour images to form large datasets. For example, the BT dataset provides 3064 MRI slices of three different types of BTs, but 120 more types of brain tumours can be diagnosed. There are rare BTs, such as atypical teratoid/rhabdoid tumours (AT/RT), which are not represented in the BT dataset, which hinders the model's ability to generalise effectively. Without sufficient data, the model may struggle to distinguish between tumours, where under-representation and overlapping features can increase the risk of misclassification. In addition, the annotation process for brain tumours is often complicated. This is because only qualified and certified medical professionals can accurately diagnose these types of brain tumours. Hence, building a large dataset not only required good-quality image scans but also qualified medical professionals to annotate the images to ensure that they were correctly annotated. In future work, the goal should be to enhance the proposed model's performance by addressing the issues mentioned earlier while also making it lighter. This will enable quicker diagnoses and allow the model to run on various devices, as it will require less computational power.

Acknowledgement

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1 (Vot Q336) and RE-GG (Vot H884) funding.

Conflict of Interest

The authors declare that they have no conflict of interest related to this publication.

Author Contribution

The authors acknowledge their contribution to the paper as follows: ***Conceptualisation, methodology, coding development and simulation, writing-original draft preparation, revision, and editing****: Aifian Adi Sufian Chan, Saizalmursidi Md Mustam;* ***Data curation, validation, and supervision****: Mohammed Faiz Liew Abdullah;* ***Coding verification, validation, writing-reviewing, and editing:*** *Farhana Ahmad Po'ad, Ariffuddin Joret;* ***Revision, writing-reviewing, and editing:*** *Hashim Safdar. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Nazir, M., Shakil, S., & Khurshid, K. (2021). Role of deep learning in brain tumour detection and classification (2015 to 2020): A review. *Computerized Medical Imaging and Graphics, 91,* 1-45. https://doi.org/10.1016/j.compmedimag.2021.101940

[2] Çinar, A., & Yildirim, M. (2020). Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture. *Medical Hypotheses*, *139,* 1-8. https://doi.org/10.1016/j.mehy.2020.109684

[3] Khan, M.S.I., Rahman, A., Debnath, T., Karim, M.R., Nasir, M.K., Band, S.S., Mosavi, A., & Dehzangi, I. (2022). Accurate brain tumor detection using deep convolutional neural network. *Computational and Structural Biotechnology Journal*, *20*, 4733–4745. https://doi.org/10.1016/j.csbj.2022.08.039

[4] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer Journal for Clinicians, 74(3),* 229–263. https://doi.org/10.3322/caac.21834

[5] Siegel, R. L., Giaquinto, A. N., & Jemal, A. (2024). Cancer statistics, 2024. *CA Cancer Journal for Clinicians, 74(1),* 12–49. https://doi.org/10.3322/caac.21820

[6] Abd-Ellah, M. K., Awad, A. I., Khalaf, A. A. M., & Hamed, H. F. A. (2019). A review on brain tumor diagnosis from MRI images: practical implications, key achievements, and lessons learned. *Magnetic Resonance Imaging, 61,* 300–318. https://doi.org/10.1016/j.mri.2019.05.028

[7] Gu, X., Shen, Z., Xue, J., Fan, Y., & Ni, T. (2021). Brain tumor MR image classification using convolutional dictionary learning with local constraint. *Frontiers in Neuroscience, 15,* 1-12. https://doi.org/10.3389/fnins.2021.679847

[8] Isunuri, B. V., & Kakarla, J. (2022). Three-class brain tumor classification from magnetic resonance images using separable convolution based neural network. *Concurrency and Computation: Practice and Experience 34(1),* 1-9. https://doi.org/10.1002/cpe.6541

[9] Bhat, S. S., Ananth, A., & Venugopala P. S., (2023). Design and evolution of deep convolutional neural networks in image classification – a review. *International Journal of Integrated Engineering*, 15(1), 213-225. https://publisher.uthm.edu.my/ojs/index.php/ijie/article/view/9893

[10] Sultan, H. H., Salem, N. M., & Al-Atabany, W. (2019). Multi-classification of brain tumor images using deep neural network. *IEEE Access, 7,* 69215–69225. https://doi.org/10.1109/ACCESS.2019.2919122

[11] Ayadi, W., Charfi, I., Elhamzi, W., & Atri, M. (2020). Brain tumor classification based on hybrid approach. *The Visual Computer, 38,* 107–117. https://doi.org/10.1007/s00371-020-02005-1

[12] Anaraki, A. K., Ayati, M., & Kazemi, F. (2019). Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *Biocybernetics and Biomedical Engineering, 39(1),* 63-74. https://doi.org/10.1016/j.bbe.2018.10.004

[13] Bjorck, B., Gomes, C., Selman, B., & Weinberger, K. Q. (2018). *Understanding Batch Normalization*. 32nd International Conference on Neural Information Processing Systems, Montréal, Canada, 1-12. https://proceedings.neurips.cc/paper_files/paper/2018/file/36072923bfc3cf47745d704feb489480-Paper.pdf

[14] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6(60),* 1–48. https://doi.org/10.1186/s40537-019-0197-0

[15] Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen, F. (2022). Image data augmentation for deep learning: a survey. arXiv. https://doi.org/10.48550/arXiv.2204.08610

[16] Cheng, J., Yang, W., Huang, M., Huang, W., Jiang, J., Zhou, Y., Yang, R., Zhao, J., Feng, Y., Feng, Q., & Chen, W. (2016). Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation. *PLoS One, 11(6)*, 1-15. https://doi.org/10.1371/journal.pone.0157112

[17] Chollet, F. (2017, July 21-26). *Xception: deep learning with depthwise separable convolutions*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 1800–1807. https://doi.org/10.1109/CVPR.2017.195

[18] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020, June 13-19). *GhostNet: more features from cheap operations*. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 1577-1586. https://doi.org/10.1109/CVPR42600.2020.00165

[19] Nirthika, R., Manivannan, S., Ramanan, A., & Wang, R. (2022). Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. *Neural Computing Applications, 34,* 5321–5347. https://doi.org/10.1007/s00521-022-06953-8

[20] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15(1),* 1929–1958. https://dl.acm.org/doi/abs/10.5555/2627435.2670313

[21] Yu, T., & Zhu, H. (2020). Hyper-parameter optimization: a review of algorithms and applications. arXiv. https://doi.org/10.48550/arXiv.2003.05689

[22] Dawson, H. L., Dubrule O., & John, C. M. (2023). Geosciences, impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification. *Computer and Geosciences*, 171, 1-11. https://doi.org/10.1016/j.cageo.2022.105284

[23] Luo, C., Li, X., Wang, L., He, J., Li, D., & Zhou, J. (2018). *How does the data set affect CNN-based image classification performance?.* 5th International Conference on Systems and Informatics (ICSAI), Nanjing, China, 361-366. https://doi.org/10.1109/ICSAI.2018.8599448

[24] Hsu, B. W. -Y., & Tseng, V. S. (2022). Hierarchy-aware contrastive learning with late fusion for skin lesion classification. *Computer Methods and Programs in Biomedicine*, 216, 1-9. https://doi.org/10.1016/j.cmpb.2022.106666

[25] Yun, J., & Lee, J. -S. (2023). Learning from imbalanced data using triplet adversarial samples. *IEEE Access*, 11, 31467-31478. https://doi.org/10.1109/ACCESS.2023.3262604

Int. Journal of Integrated Engineering Vol. 17 No. 5 (2025) p. 389-401

[26] Ismail, A., Ahmad, S. A., Che Soh, A., Hassan, K., & Harith, H. H. (2019). Improving convolutional neural network (CNN) architecture (miniVGGNet) with batch normalisation and learning rate decay factor for image classification. *International Journal of Integrated Engineering*, 11(4), 51-59. https://publisher.uthm.edu.my/ojs/index.php/ijie/article/view/4558