تشخيص سرطان الدماغ في الصور بناءاً على التعلم العميق

م.م هند ابراهيم محمد

قسم الرياضيات، كلية المقداد، جامعة ديالى، ديالى، العراق.

hindim@uodiyala.edu.iq

الملخص

تعتمد الطرق التقليدية لتشخيص سرطان الدماغ على التفسير اليدوي للصور الطبية، والتي لها قيود مثل الذاتية والخطأ البشري. في السنوات الأخيرة، أظهر استخدام التعلم العميق في تشخيص سرطان الدماغ نتائج واعدة، وخاصة في الكشف عن أنواع مختلفة من أورام المخ وتصنيفها. يهدف هذا البحث إلى تطوير طريقة دقيقة وفعالة لتشخيص سرطان الدماغ في الصور الطبية باستخدام ست تقنيات للتعلم العميق. تم تدريب الخوارزميات على مجموعتين كبيرتين من البيانات من موقع Kaggle العالمي. كانت نتائج دقة التشخيص مختلطة. عند اختبارها على مجموعة البيانات الأولى، حققت CNN و99 CNN-GRU٪، بينما حققت94 VGG19 ٪، وحققت CNN-LSTM 92٪، وكانت CNN-SVM مقنعة إلى حد ما بدقة 87٪، وتأخر DNN بشكل منخفض جدا مقارنة ببقية النماذج بدقة 61٪. عند تطبيقها على مجموعة البيانات الثانية، أظهرت النتائج أن CNN حافظت على كفاءتها بدقة 99٪، وفشلت CNN-GRU عند مقارنتها بالنتائج المحققة على مجموعة البيانات الأولى بدقة 60٪. حققت نماذج CNN-LSTM و VGG19و CNN-SVMنتائج بنسبة 97% و96% و95% على التوالي. بالإضافة إلى ذلك، لم تحقق DNN أيضا نتائج جيدة بدقة 62٪.

الكلمات المفتاحية: الذكاء الاصطناعي، سرطان الدماغ، الشبكات العصبية التلافيفية، التعلم العميق، الشبكة العصبية العميقة، الوحدة المتكررة المغلقة، الذاكرة طويلة المدى، التصوير بالرنين المغناطيسي، آلة الدعم المتجهي

Brain Cancer Diagnostics in Images Based on Deep Learning

Hind I. Mohammed
Department of Mathematics, Al- Muqdad, College of Education, University of Diyala
hindim@uodiyala.edu.iq

ABSTRACT

Traditional methods for brain cancer diagnosis rely on manual interpretation of medical images, which have limitations such as subjectivity and human error. In recent years, the use of deep learning (DL) in brain cancer diagnosis has shown promising results, especially in detecting and classifying various types of brain tumors. This research aims to develop an accurate and effective method for brain cancer diagnosis in medical images using six deep learning techniques. The algorithms were trained on two large datasets from the global Kaggle website. The results of the diagnosis accuracy were mixed. When tested on the first dataset, CNN and CNN-GRU achieved 99%, while VGG19 achieved 94%, CNN-LSTM achieved 92%, CNN-SVM was somewhat convincing with 87% accuracy, and DNN lagged very low compared to the rest of the models with 61% accuracy. When applied to the second dataset, the results showed that CNN maintained its efficiency with 99% accuracy, and CNN-GRU failed when compared to the results achieved on the first dataset with 60% accuracy. CNN-LSTM, VGG19, and CNN-SVM models achieved 97%, 96%, and 95% results, respectively. In addition, DNN also did not achieve good results with an accuracy of 62%.

Keywords: Artificial Intelligence (AI), **B**rain cancer**,** Convolutional Neural Networks (CNNs), Deep Learning (DL)**,** Deep Neural Network (DNN), Gated Recurrent Unit (GRU)**,** Long short-term memory (LSTM)**,** Magnetic Resonance Imaging (MRI), Support Vector Machine (SVM)**.**

INTRODUCTION

The human brain is a major organ that regulates complex cognitive and physical processes. It is located within the protective skull and is divided into three main regions: the cerebellum, the brainstem, and the cerebrum [1]. As shown in Figure (1). The cerebellum is the largest region of the brain, and it is positioned below the cerebrum. The brainstem is the main connection between the spinal cord and the cerebellum. The brain is vulnerable to a number of illnesses, including the formation of brain tumors, even though it is very important. These tumors are abnormal lumps or swellings that are formed by genetic abnormalities in the chromosomes of cells, which disturb their normal functioning. Every year on June 8, World Brain Tumor Day is observed to increase awareness of this fatal illness and to provide assistance to those in the community who are impacted by brain tumors. The World Health Organization (WHO) reports that there are more than 120 distinct forms of brain tumors [3], some of which grow slowly and others rapidly [4].

However, manually classifying brain cancers using MRI data is time-consuming. A number of AI techniques have been proposed recently with the goal of automating the difficult work of diagnosing brain tumors in the medical industry. However, the capacity of earlier methods to diagnose brain tumors in a variety of circumstances has been limited since they frequently depended on single datasets [5].

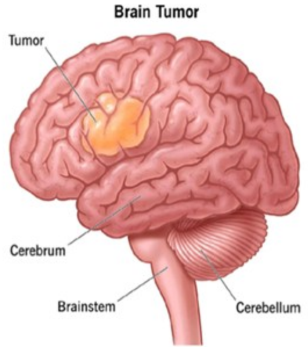


Figure 1: Growth of a tumor in the human brain [2].

DL is a type of AI that use artificial neural networks (ANNs) to mimic the human brain and learn from enormous volumes of data. DL architectures are chosen over typical machine learning methods because they can learn and detect complicated picture information on their own. The constant creation of new models improves feature extraction in DL methodologies used in numerous medical disciplines. These methods are commonly employed in image processing, classification, and segmentation problems. Radiologists, neurologists, and physicians perform numerous steps to correctly diagnose a brain tumor, including a physical examination, a medical history review, the use of contrast agents, and biopsy testing. The objective is to accurately detect aberrant tissue and to pinpoint its precise position, area, and orientation. To develop digital pictures of the brain, physical examination and history analysis come after the imaging scan assessment and interpretation phase. It is highly effective to use magnetic resonance imaging. Early and accurate brain tumor diagnosis is crucial, as is precise tumor area measurement for targeted therapy. Manual brain tumor diagnosis is a complex, time-consuming, and laborious task that can be prone to human error due to factors like fatigue and information overload. Automated methods to diagnose brain tumors have Only regular imaging can identify them, The paper seeks to develop more accurate and reliable methods using deep learning techniques to process medical images and diagnose brain tumors with higher accuracy.

2. LITERATURE REVIEW

In the past ten years, many researchers in the field of AI have conducted scientific and software experiments using machine and DL to predict the risk of brain cancer or its spread, which was developed as a result of recent advances in DL. These systems provide essential assistance to doctors in the initial diagnosis of the disease, starting with [7] , the authors' experiments with 3D stacked GANs for glioma prediction were based on GP-GAN, a new objective function consisting of l1 and Dice losses. In addition, they used segmented feature maps to guide the generator to produce better quality generated images. Regarding the generator, it was built using a 3D U-Net architecture that combines hierarchical features to provide an improved image, in [8] the DL-based diffusion technique used anatomical registration of MRI more accurately than dMRI registration, according to the Human Connectome Project (HCP), which is the basis for introducing segmentation of MRI (DDSeg) for segmentation from high-quality imaging data. A new enhanced objective loss function was applied to train a CNN in order to improve the accuracy of the segmented tissue boundary region. Therefore, in order to improve the resolution, the non-Gaussian diffusion of water molecules was characterized using diffusion kinetic imaging (DKI) properties in conjunction with conventional diffusion tensor imaging parameters.

The authors Brain illness diagnosis using neuroimaging modalities with DL patterns such auto encoders (AEs), pertained CNNs, and GANs was thoroughly analyzed [9]. The suggested approach identifies brain tumors using a CNN model and multiple pre-trained models: efficientnetb0, googlenet, mobilenetv2, nasnetlarge, resnet50, resnet10, shufflenet, vgg16, and vgg19 It then categorized image deep characteristics using an SVM classifier. The classification accuracy of linear, Gaussian, cubic, and quadratic kernel functions is compared. Deep features from the efficient model classified brain cancers precisely. The SVM classifier with Gaussian kernel function had 99.78% classification accuracy [10].

In [11], CNN architecture uses the VGG19 model to detect brain cancers in MRI scans quickly enough to save lives. Classifying glioma images with preprocessed VGG-19 and HE data, this study tests the model's accuracy, precision, recall, and f1-score. The original data had the highest accuracy, precision, recall, and F1-score (97%, 100%, 97%, and 98%). Instead, HE preprocessing data was 92%, 100%, 92%, and 96% accurate. The authors Proposed using an appropriate filter to remove noise to preserve image edges without losing information needed to interpret brain tumor images, augmentation to generate dataset variations for synthetic data for training, and DL to classify brain x-ray images as tumor or non-tumor [12].

After that the authors Suggested to apply SVM algorithm and CNN to brain tumor classification (benign or malignant tumor) by using brain Magnetic Resonance Images. The authors suggested a system based on the novel concept to detect tumor that uses techniques of feature extraction, segmentation algorithm and classification, To recognize patients who have brain tumor, to ascertain the tumor type and tumor sizes. CNN, SVM methods obtain high accuracy, CNN accuracy of training and classification get 96.33% while SVM accuracy of training and classification get 95% Although SVM manages the brain tumor process categorization in MRI scans, as compared to CNN, SVM provides less accuracy and consumes more time for execution[13], The authors proposed a hybrid deep CNN model that might be utilized in the diagnosis and classification of several brain tumor cases based on the datasets of medical imaging , proposed three CNN models for the classification task. It performs detection and classification of MR images into glioma, meningioma, pituitary tumors, normal tissues, and metastatic tumors. The accuracy of the glioma grading is 99.53%, 93.81%, and 98.56%, respectively, and in [14] proposed models were compared to classical ones: Alex Net, DenseNet121, ResNet-101, VGG-19, and Google Net. The CNN model proved to enhance early diagnostic accuracy and reliability for various brain tumors.

In [15] A 10-layer CNN model is presented for segmenting 3D CT images, which can predict the types of brain tumors. This work used some preprocessing methods to improve the quality of CT images on BraTS2020 and BraTS2021 datasets. In comparison, the proposed scheme outperformed other methods of 3D segmentation. The metrics of recall, precision, Hausdorff, and F1-Score reached 89.63% and 88.96%, respectively, in [16] suggested a technique of an effective feature selection using a lightweight deep CNN. the main goal of using this technique to use fewer parameters and get the highest accuracy. The applied method is divided into three sections: preprocessing, feature extraction step, and classification. In the section of preprocessing, data augmentation is used to improve the number of the dataset images and K-fold cross-validation is utilized for the strong classification. In the section of feature extraction, by using modified CNN, active features are extracted. And lastly, SVM is utilized to Noticeably classify the different types of brain tumors from MRI scans. The suggested model CNN-SVM can activitly detects and classifies the different brain tumors types and achieved the best accuracy of 96.7%,

Finally, in [17] proposed SVM methodology classifies the MRI images to detect whether the tumor is malignant or benign. Noise is removed during preprocessing by comparison, the tumors are classified into benign or malignant using the proposed methodology using Butterworth, wavelet, ideal, and median filters with 95% accuracy, 88% precision, and 91% specificity.

In the past decade, significant progress has been made in the field of AI, especially in the field of machine and DL, to predict the risk of brain cancer incidence and spread. These studies use different DL models and algorithms such as generative generative networks, CNNs, and SVMs, along with advanced imaging techniques such as magnetic resonance imaging and computed tomography, to aid in early diagnosis. However, [22] the performance of generative generative networks in generating medical images has been inconsistent, and the complexity of 3D stacked generative generative networks may increase computation time and resource consumption. [24] The weakness lies in the reliance of DKI on specific imaging parameters from its generalization across different imaging systems or conditions. [25] Although high accuracy has been achieved, the reliance on pre-trained models may limit the system's ability to adapt to unseen or novel types of brain tumors without fine-tuning. While [30] preprocessing steps such as image enhancement may degrade image quality in some cases, which may affect the model's performance on certain MRI datasets.

[33] A hybrid model combining CNN and SVM provides good classification performance, but it is more time-consuming to execute, [34] training such deep models requires intensive computational resources, and the complexity of the model may reduce interpretability and increase processing time.

While [35] the complexity of 3D image processing can increase computational requirements, the model may face challenges with less quality-controlled CT datasets, the low number of parameters may not be sufficient to detect more subtle features in certain brain tumors, and the model may not perform well on very large or complex datasets [36], [37]

Strengths: The SVM-based approach has shown promising results (95% accuracy) with different noise removal filters, which enhances the reliability of tumor detection.

From the above, we conclude that CNNs, especially when combined with pre-trained models and SVM classifiers, usually achieve the highest classification accuracy. However, the accuracy of SVMs is slightly lower compared to CNNs, especially on complex datasets.

In terms of image processing, many studies have used advanced preprocessing techniques (e.g., denoising, image upscaling), which have significantly improved model performance. However, these steps can sometimes lead to image distortions or loss of important data features.

As for algorithmic complexity, GANs and deep CNNs provide high accuracy but come with increased computational complexity. Models such as SVMs and lightweight CNNs provide faster execution times but with slightly lower accuracy.

In conclusion, DL models like CNNs (especially VGG-19, ResNet, EfficientNet) paired with SVM classifiers seem to provide the most promising results for brain tumor classification. However, challenges like computational complexity, image artifacts, and the generalizability of models remain issues to address for broader applications in clinical settings.

3. DATASETS

The first dataset consisted of 7022 brain MRI images in 4 categories: glioma, meningioma, no tumor, and pituitary gland. 7000 images were used for our research [18].

The second dataset contains [3264] images, of which only [3000] images were used which are also classified into 4 categories: glioma, meningioma, no tumor and pituitary gland [19].

4. RESEARCH METHODS

Our comparative study contains six methods. These methods description is short presented in this section as in:

A. Convolutional Neural Network (CNN)

CNN is a common type of deep neural networks that basically uses convolution filters to produce feature maps targeting to complete a particular task such as object detection, image recognition, image classification or segmentation or Medical image analysis. CNN use convolutional layers to learn a spatial hierarchy of features from input data automatically and adaptively. With the advancement of graphical user interface (GPUs')[20], The figure(2) shows the general structure of the CNN

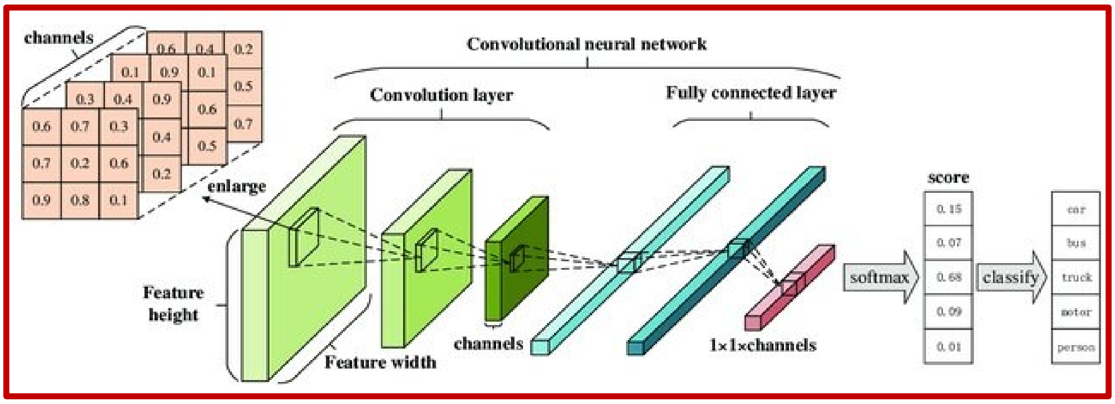


Figure (2): CNN Structure [21].

B. Support vector machines (SVMs)

SVMs are kinds of machine learning and AI technique that is mostly utilized in supervised learning. It has proven distinct success in a broad set of classification and regression tasks. SVMs aim to find the best super level that separates various classes of data points in a space of high-dimensional feature [22].

.

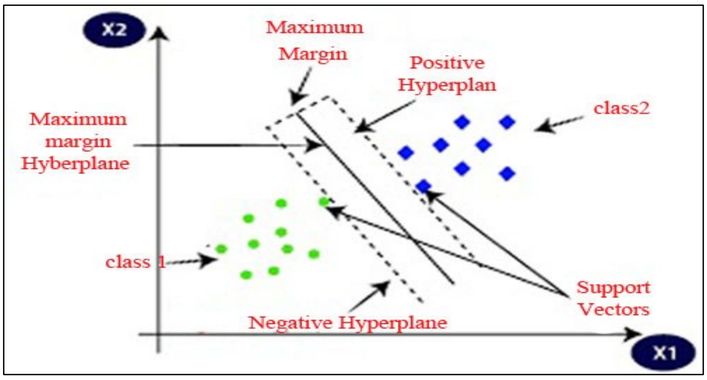


Figure (3): SVM Structure [23].

C. Convolutional neural networks (CNN)- long short term memory (LSTM)

LSTM is a DL technique designed for sentiment analysis, language modeling, text data prediction, and speech analysis. It is regarded as the most distinctive artificial neural network, possessing the ability to analyze long-term connections in textual input [24].

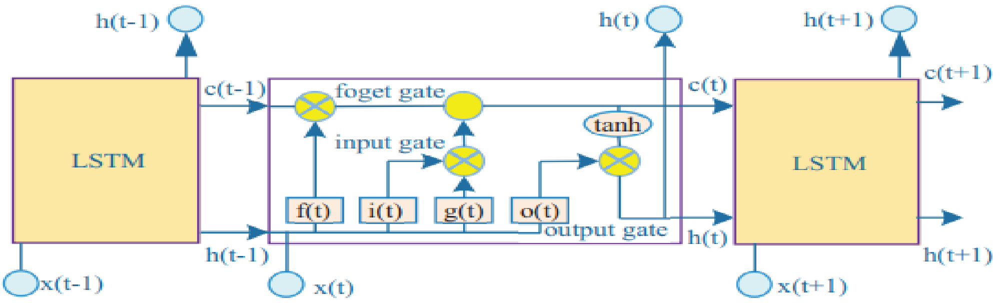


Figure (4): Long short-term memory (LSTM) cell architecture. [25]

D. Gated Recurrent Unit (GRU)

The GRU architecture is a more advanced type of RNN that addresses the issues of gradient disappearance and explosion that occur as network layers and iterations increase in traditional recurrent neural networks. GRU is a variant of LSTM that has fewer gates and a simpler architecture. It employs an update gate and a reset gate to update the hidden state information and uses the sigmoid function to provide between 0 and 1 for retention levels. GRU works well for long-term dependencies as it selectively updates and forgets the data. Figure 5 Structure of GRU Unit. Assuming xt is the input and ht is the output of the hidden layer, GRU calculates ht as follows [26] [27].

$$z_{t=\sigma}\, W^{z} x_t + U^{Z} h_{t-1}$$

$$r_{t=\sigma}\, W^{r} x_t + U^{r} h_{t-1}$$

$$\tilde{h_t} = tanh r_t \circ U h_{t-1} + W x_t$$

$$h_{t=1-} z_t \circ \tilde{h_t} + z_t \circ h_{t-1}$$

The update and reset gates are zt and rt, while ht~ is the total of the previous hidden layer's input xt and output ht−1. σ is a sigmoid function, tanh is hyperbolic tangent, Uz, Wz, Ur, Wr, U, and W are training parameter matrices, and zt°ht−1 is the composite relation between zt and ht−1[27]. GRU LSTMs are popular. GRUs calculate less than LSTMs with equal accuracy, and the update gate helps the model decide what information to convey based on previous time step data. The reset gate function helps prevent gradient fading by determining how much past information to forget. GRU has reset and update gates [28].

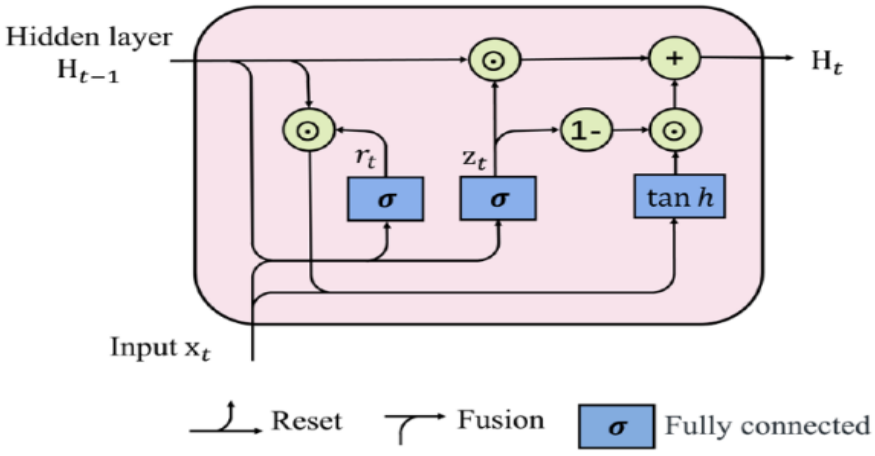
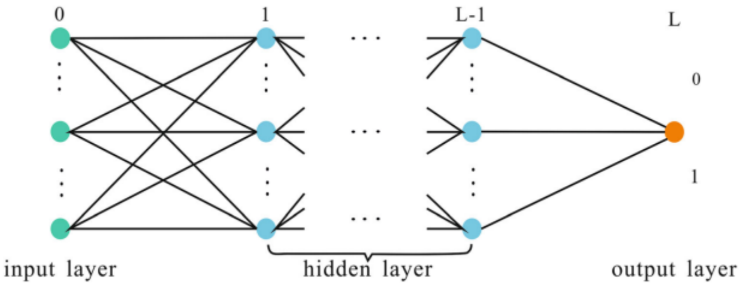


Figure (5): The GRU's unit structure [27]

GRU handles temporal data well: weather, wind speed, and wind power predictions. The GRU can learn long-term dependencies and patterns in time series because to its unique architecture. Decomposition techniques were used to anticipate wind power subsequences with the GRU model. In reference, the GRU has higher prediction accuracy, faster training, and lower noise sensitivity [27].

E. Deep Neural Network (DNN)

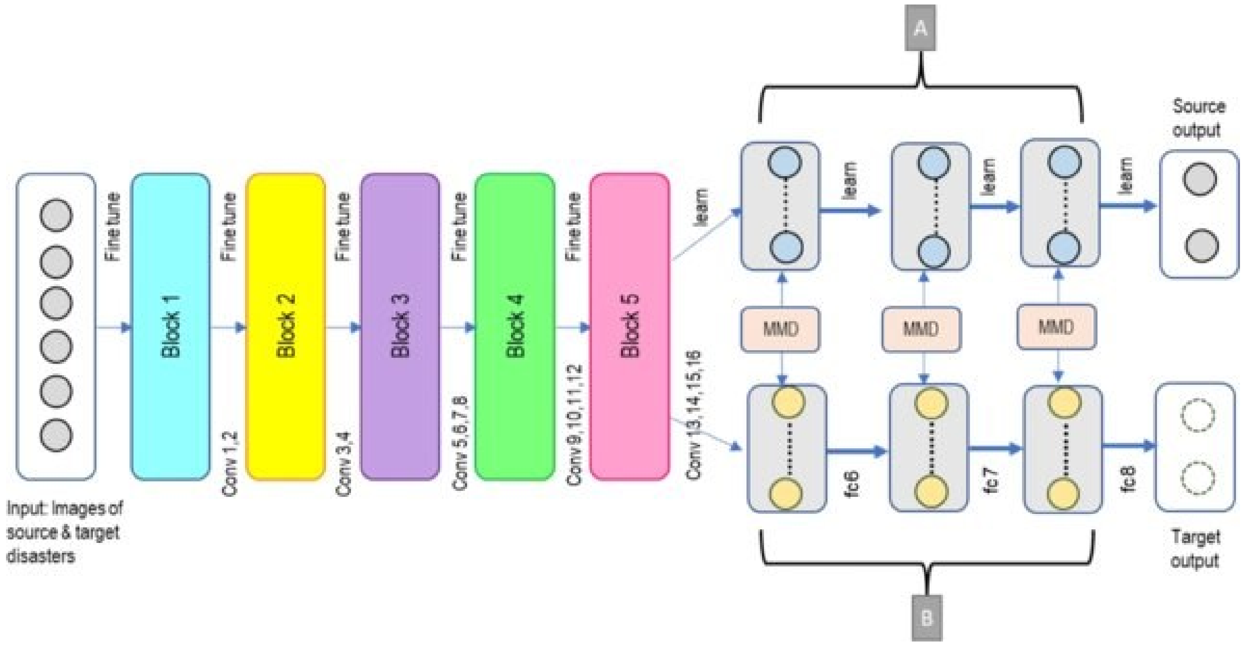
DNNs, also mentioned to as DL, are currently commonly used for numerous AI (AI) applications through image recognition to speech recognition the applications number that use DNNs has increased[29], as in figure 6.



Figure(6): The basic structure of DNN [30]

F. Visual Geometry Group (VGG19)

VGG is a creative object recognition model which supports up to 19 layers. ImageNet datasets has used to pre train VGG. VGG is one of the most used architectures of image recognition because it is still able to superiority other invisible datasets. VGG Net has multiple Variables including VGG-16 and VGG-19,these Variables only vary in the total layers number in the neural network[22], as infigure 7.



Figure(7): The structure of VGG19 [31]

5. Evaluation Criteria

By comparing the expected brain cancer marker with the actual value available with the data set, the brain cancer prediction performance is evaluated using different evaluation criteria. Accuracy, sensitivity, specificity, Precision, scale F and average G are among the evaluation scales, which are defined as follows:

$$\text{Sensitivity [Sen]} = \frac{\text{TP}}{\text{TP+FN}} \qquad (1)$$

$$\text{precision}[Pr] = \frac{\text{TP}}{\text{TP+FP}} \qquad (2)$$

$$F - \text{Measure} = \frac{2\times \text{sensitivity} \times \text{precision}}{\text{sensitivity} + \text{precision}} \, (3)$$

$$Accracy[Ac] = \frac{\text{TP + TN}}{\text{TP + TN + FP + FN}} \, (4)$$

Where: True Positive [TP], False Positive [FP], True Negative [TN] and True Negative [TN].

the F1-score to measure the confusion matrix more accurately. Regardless the successful classifications high rate. Finally, depending on equations (1), (2) (30 and (4) which are defined as:

Accuracy

Accuracy indicated to the ratio of correctly classified samples to the total number of test samples utilized during training and testing.

Recall

The recall or sensitivity is calculated by dividing the true positive) TP)on the summation of true positive)

TP)and false negative (FN), which means (TP + FN) Based on equation (1).

F1-score

F1-score means how new face samples many that pre-trained FER models can successfully classify and minimize the misplacement number [47]. The F1-score equation is expressed as equation (3) [32].

6. Results and Discussion

Six DL models were used with the two datasets, the first of which was (7000)images, and the second of which was (3000)images. The data was divided into 80% for training and 20% for testing. The accuracy results indicators indicated that there was a difference in values between the models. Some models achieved excellent results in detecting brain tumors, and some failed somewhat. Table 2 shows the accuracy factor when applying CNN, CNN-SVM, CNN-LSTM, CNN-GRU, DNN, and VGG19 models on dataset 1 and dataset 2.

Table 2: Comparison of six models in accuracy factor

Data set	**CNN**	**CNN-SVM**	**CNN-LSTM**	**CNN-GRU**	**DNN**	**VGG19**
Dataset1 [7000] image	99%	87%	92%	99%	61%	94%
Dataset2 [3000] image	99%	95%	97%	60%	62%	96%

Figure 3. Shows a plot of the accuracy results of the experiment. The results show that the CNN method performs better than the rest of the models when applied to both data sets. The first model (CNN) when applied to the both datasets performed excellently with very high precision, recall, and F1-scores across all

Table 3: Comparison between dataset1 and dataset2 in precision recall and f1-score factors

state	**Data set1 [7000] image**			**Data set2 [3000] image**		
	precision	**recall**	**f1-score**	**Precision**	**recall**	**f1-score**
Glioma	99%	100%	99%	100%	98%	99%
Meningioma	99%	96%	97%	100%	97%	98%
Notumor	98%	100%	99%	96%	100%	98%
Pituitary	99%	99%	99%	98%	100%	99%

Figures 4, 5, 6 illustrate the confusion matrix, validation accuracy, and validation loss when applying the CNN model to the dataset1 and dataset2.

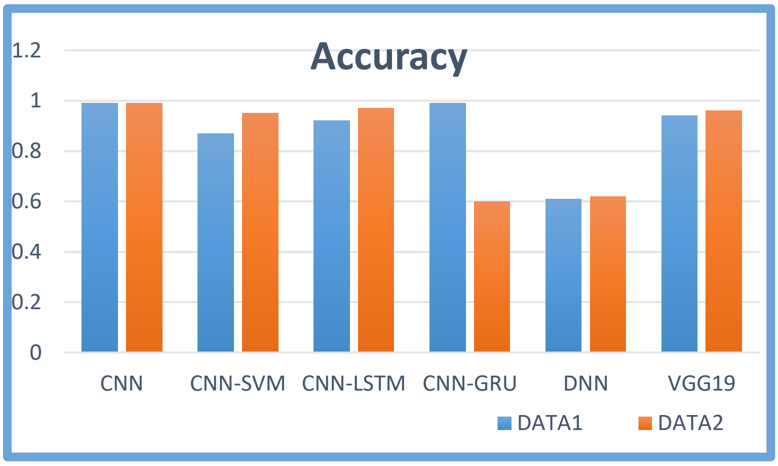
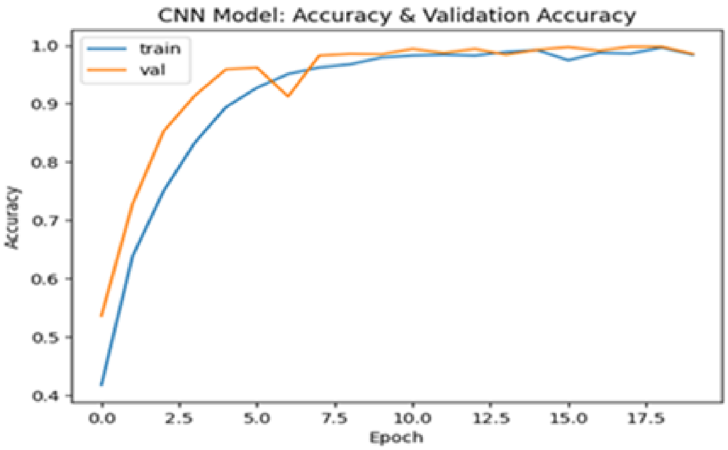
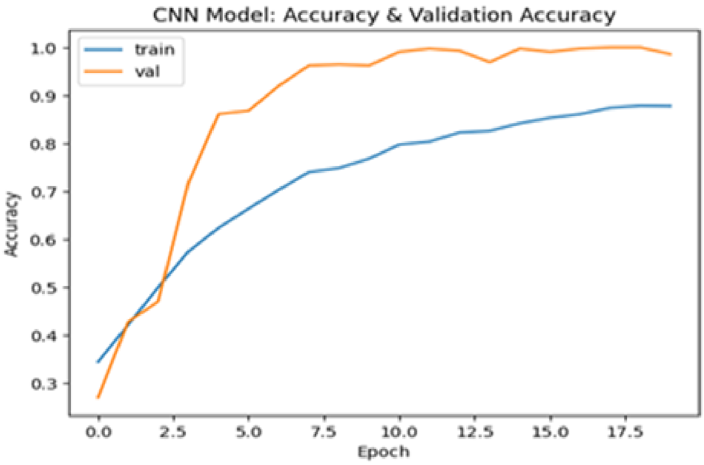


Figure 4: Best accuracy results among models

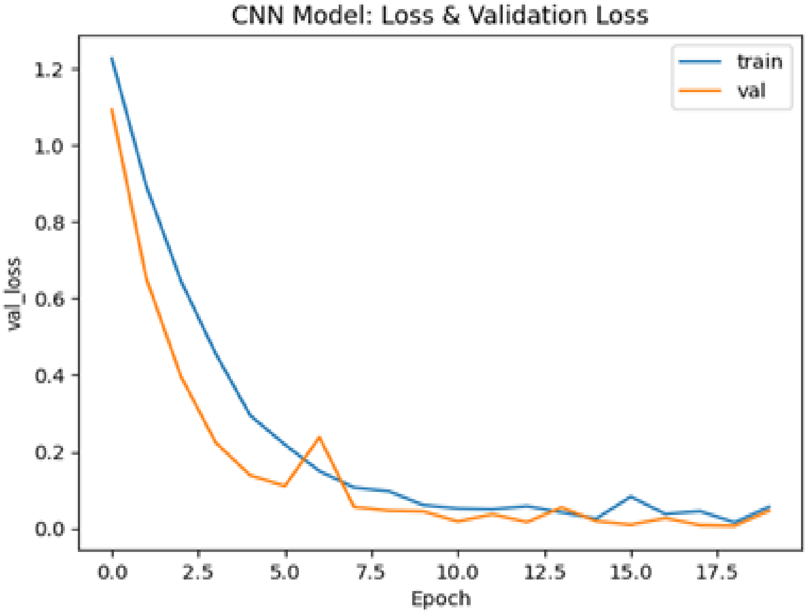


(a) Dataset1

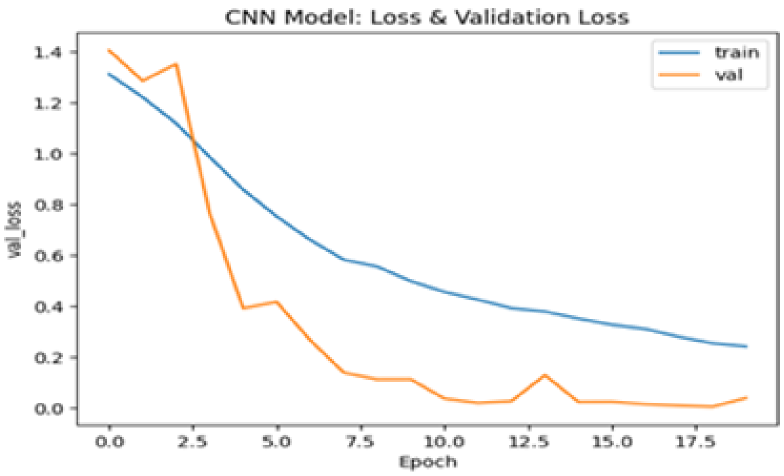


(b) Dataset2

Figure 5: Validation accuracy for (a) Dataset1 (b) Dataset2



(a) Dataset1



(b) Dataset2

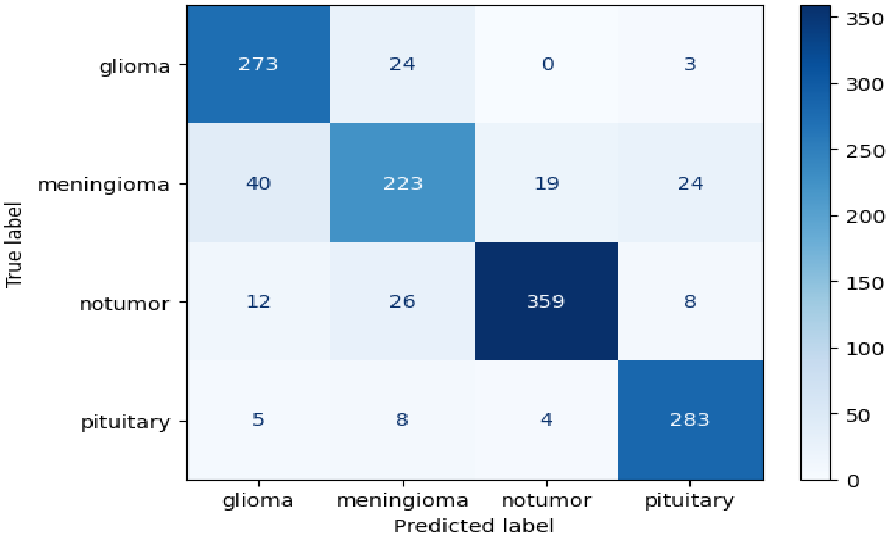
Figure 6: Validation loss for (a) dataset1 (b) dataset2

The CNN-SVM model has good performance with precision and recall ranging from 0.73 to 0.94 across classes for data set1 and high precision and recall for most classes (0.89-1.00) for data set2. Table 4 shows the comparison between dataset 1 and dataset 2 in terms of recall factors, precision, and f1 score when applying the CNN-SVM model.

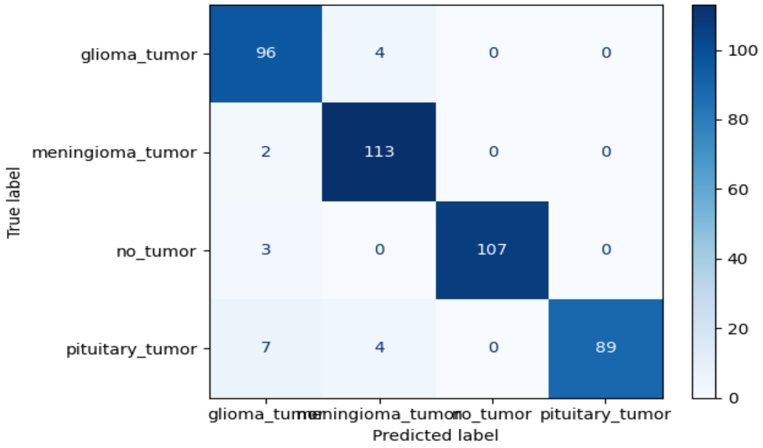
Table 4: comparison between dataset1 and dataset2 in precision recall and f1-score factors

state	**Data set1 [7000] image**			**Data set2 [3000] image**		
	precision	**recall**	**f1-score**	**Precision**	**recall**	**f1-score**
Glioma	83%	91%	87%	89%	96%	92%
Meningioma	79%	73%	76%	93%	98%	96%
Notumor	94%	89%	91%	100%	97%	99%
Pituitary	89%	94%	92%	100%	89%	94%

Figure 7. Illustrate the confusion matrix when applying the CNN-SVM model to the dataset1 and dataset2.



(a) Dataset1



(b) Dataset2

Figure 7: Confusion matrix for (a) dataset1 (b) dataset2

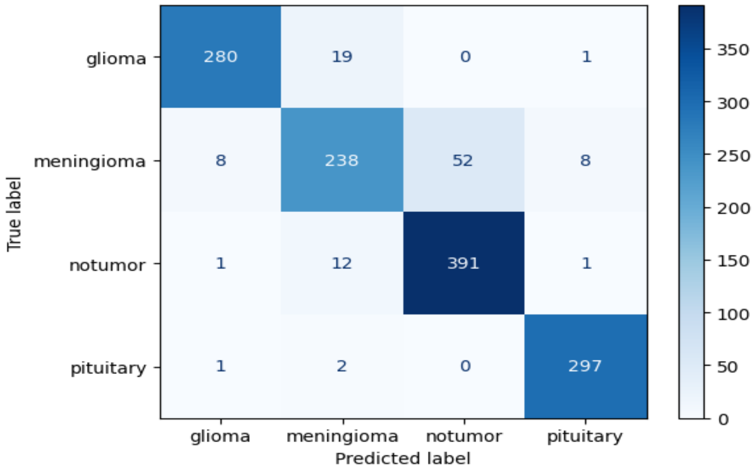
Note that loss and accuracy cannot be plotted directly in the case of using the CNN-SVM algorithm because it relies on using CNN only as a feature extractor from images. In other words, images are passed through a CNN to extract specific features which are then used to train an SVM model. When an SVM is trained using the features extracted from the CNN, the usual training information such as loss and accuracy for each epoch is not used as in training deep neural networks [CNN or DNN]. Instead, the SVM is trained using the extracted features in a single step without having to calculate the loss and accuracy at each stage.

CNN-LSTM achieved high precision and recall (0.78-0.99) across most classes for data set1. When applying the data set2, the results were respectively high precision, recall, and F1 score across all classes (0.93-1.00). Table 5 shows the comparison between dataset 1 and dataset 2 in terms of recall factors, precision, and f1 score when applying the CNN-LSTM model.

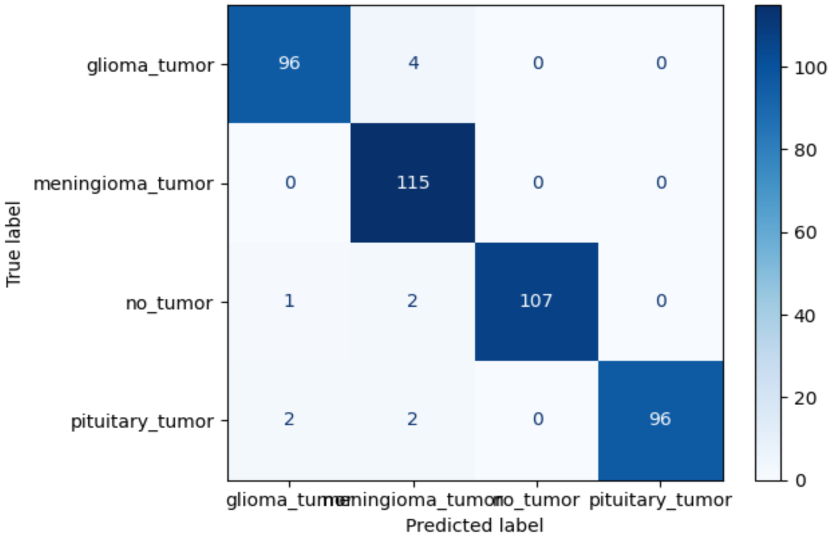
Table 5: Comparison between dataset1 and dataset2 in precision recall and f1-score factors

state	**Data set1 [7000] image**			**Data set2 [3000] image**		
	precision	**recall**	**f1-score**	**Precision**	**recall**	**f1-score**
Glioma	97%	93%	95%	97%	96%	96%
Meningioma	88%	78%	82%	93%	100%	97%
Notumor	88%	97%	92%	100%	97%	99%
Pituitary	97%	99%	98%	100%	96%	98%

Figures 8, 9, 10, illustrate the confusion matrix, validation accuracy, and validation loss when applying the CNN-LSTM model to the dataset1 and dataset2.

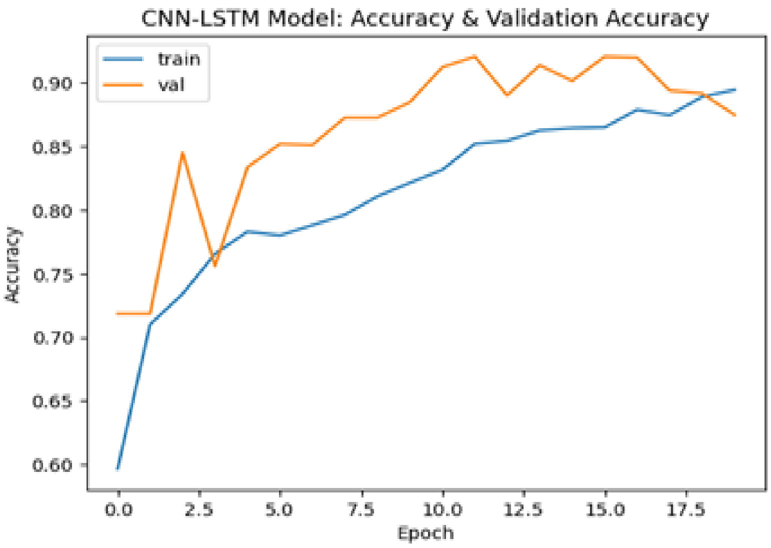


(a) Dataset1

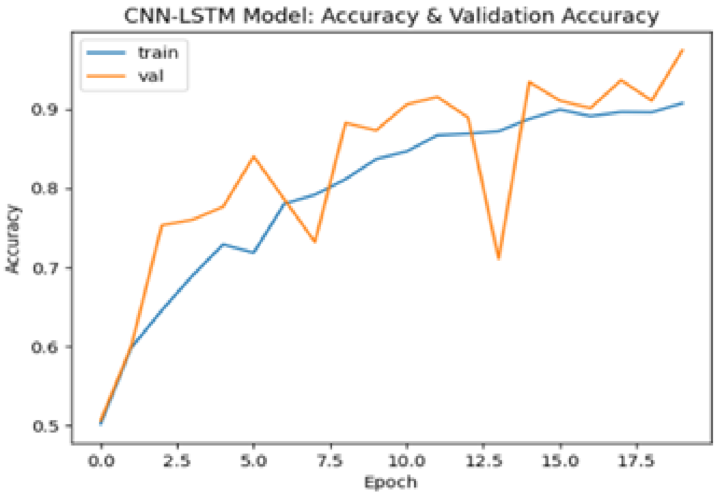


(b) Dataset2

Figure 8: Confusion matrix for [a] dataset1 [b] dataset2

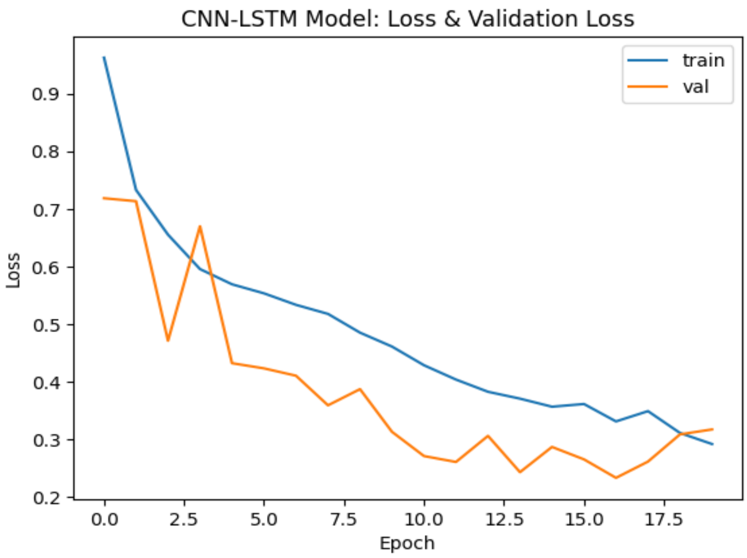


(a) Dataset1

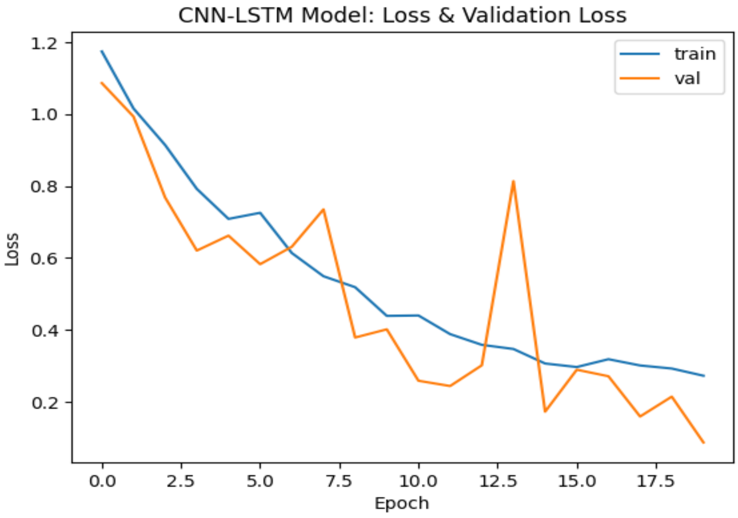


(b) Dataset2

Figure 9: Validation accuracy for [a] dataset1 [b] dataset2



(a) Dataset1



(b) Dataset2

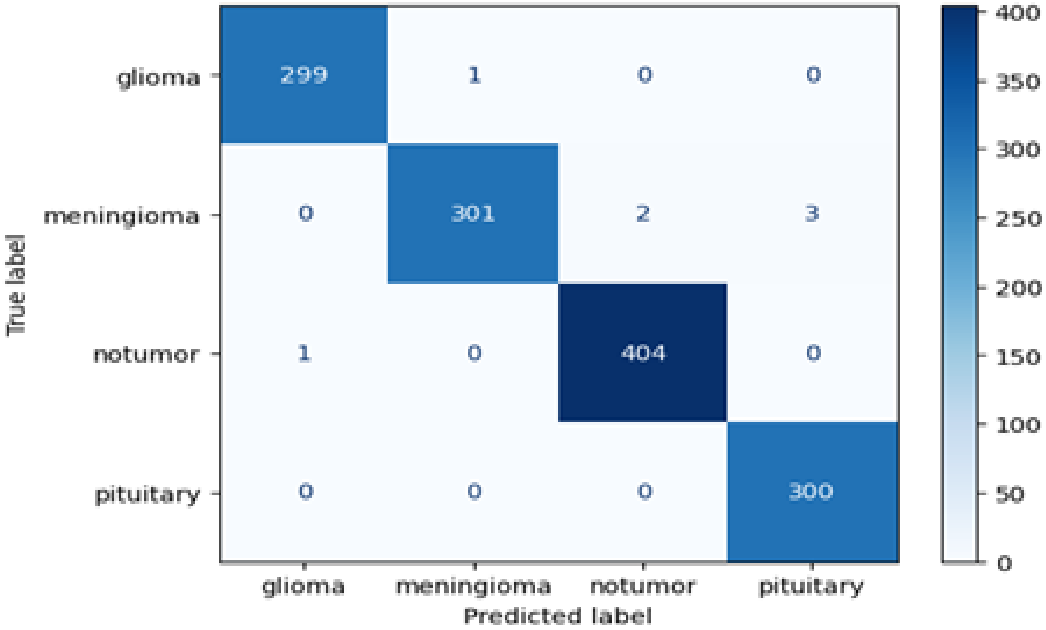
Figure 10: Validation loss for [a] dataset1 [b] dataset2

CNN-GRU achieved near-perfect precision, recall, and F1-scores [0.98-1.00] for all classes of data set1 while low precision and recall across all classes of data set 2. Table 6 shows the comparison between dataset 1 and data set 2 in terms of recall factors, precision, and f1 score when applying the CNN-GRU model.

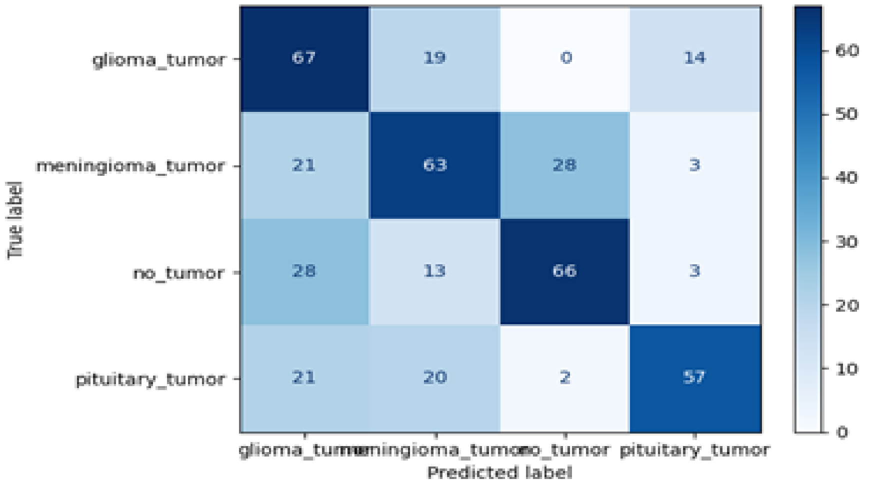
Table 6: Comparison between dataset1 and dataset2 in precision recall and f1-score factors

state	Data set1 [7000] image			Data set2 [3000] image		
	precision	recall	f1-score	Precision	Recall	f1-score
Glioma	100%	100%	100%	49%	67%	57%
Meningioma	100%	98%	99%	55%	55%	55%
Notumor	100%	100%	100%	69%	60%	64%
Pituitary	99%	100%	100%	74%	57%	64%

Figures 11, 12, 13, illustrate the confusion matrix, validation accuracy, and validation loss when applying the CNN-GRU model to the dataset1 and dataset2.

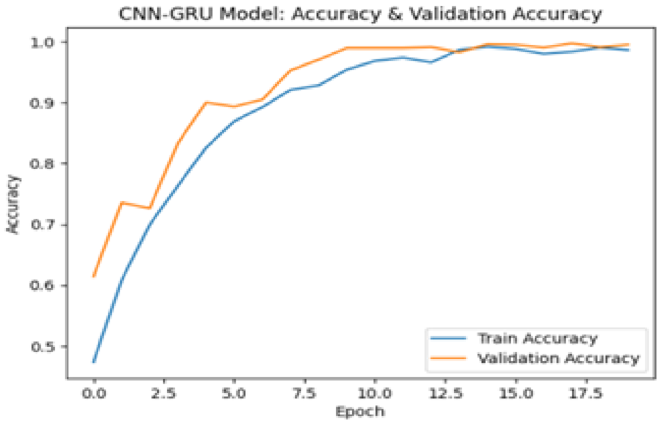


(a) Dataset1

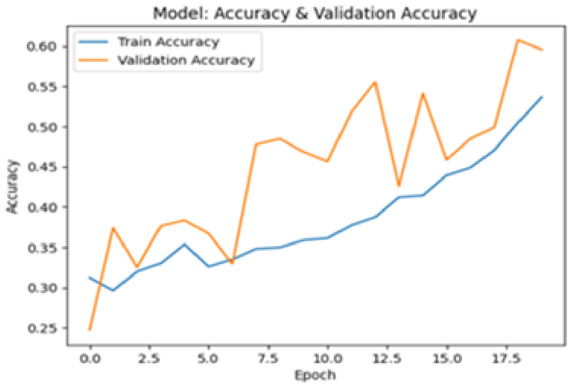


(b) Dataset2

Figure 11: Confusion matrix for [a] dataset1 [b] dataset2

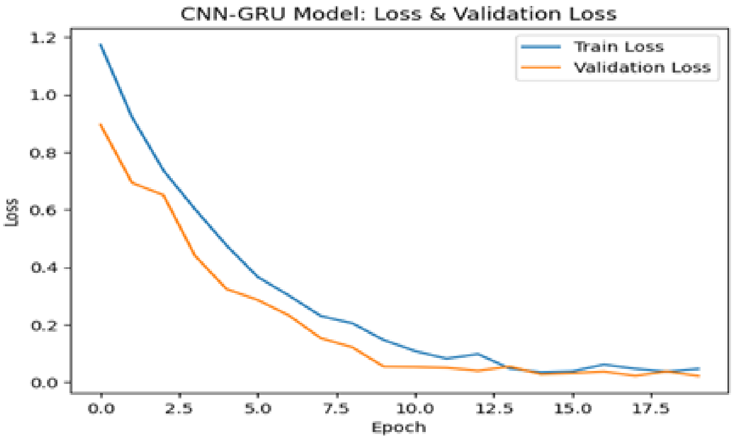


(a) Dataset1

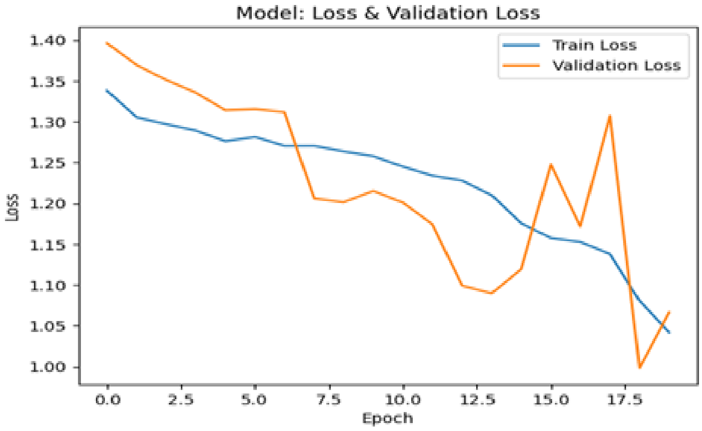


(b) Dataset2

Figure 12: Validation accuracy for [a] dataset1 [b] dataset2



(a) Dataset1



(b) Dataset2

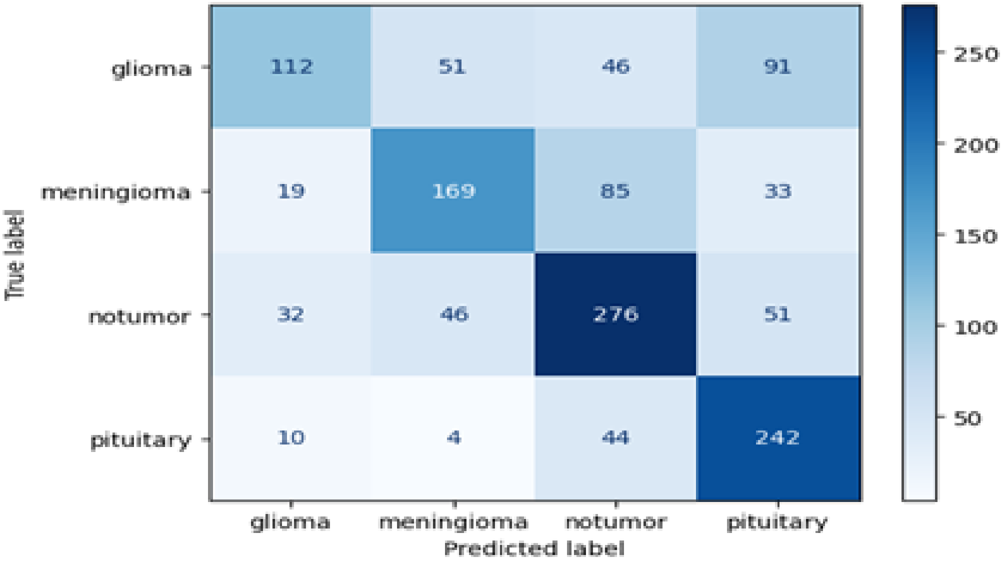
Figure 13: Validation loss for [a] dataset1 [b] dataset2

DNN achieved Inconsistent precision and recall [0.37-0.81] across classes of both data sets. Table 7 shows the comparison between dataset 1 and dataset 2 in terms of recall factors, precision, and f1 score when applying the DNN model.

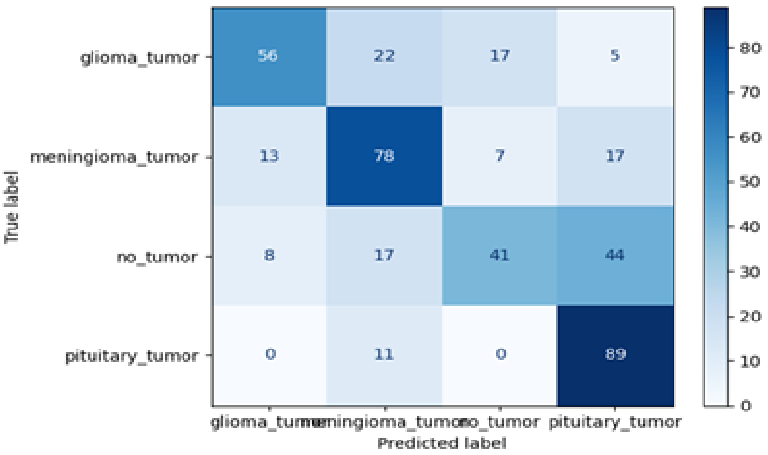
Table 7: Comparison between dataset1 and dataset2 in precision recall and f1-score factors.

state	**Data set1 [7000] image**			**Data set2 [3000] image**		
	precision	**recall**	**f1-score**	**Precision**	**recall**	**f1-score**
Glioma	65%	37%	47%	73%	56%	63%
Meningioma	63%	55%	59%	61%	68%	64%
Notumor	61%	68%	64%	63%	37%	47%
Pituitary	58%	81%	68%	57%	89%	70%

Figures 14, 15, 16, illustrate the confusion matrix, validation accuracy, and validation loss when applying the DNN model to the dataset1 and dataset2.

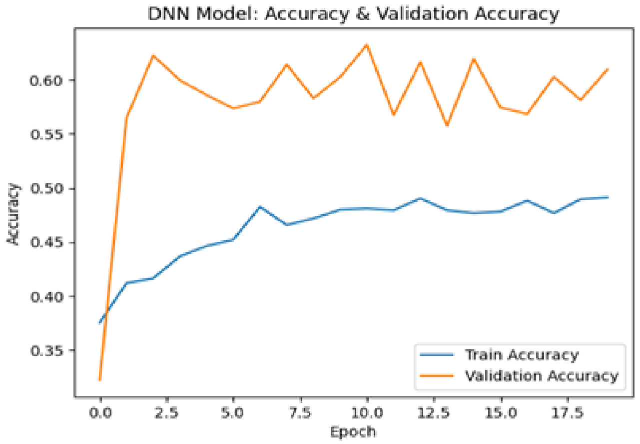


(a) Dataset1

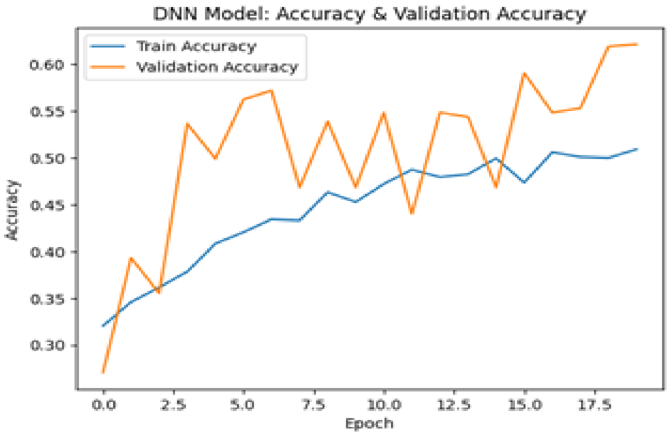


(b) Dataset2

Figure 14: Confusion matrix for (a) dataset1 (b) dataset2

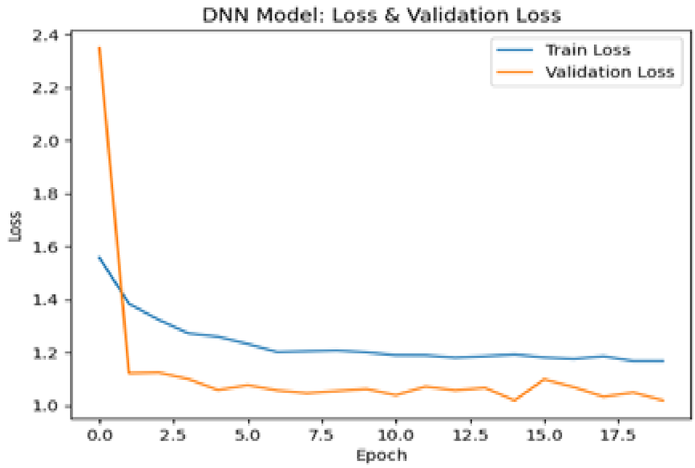


(a) Dataset1

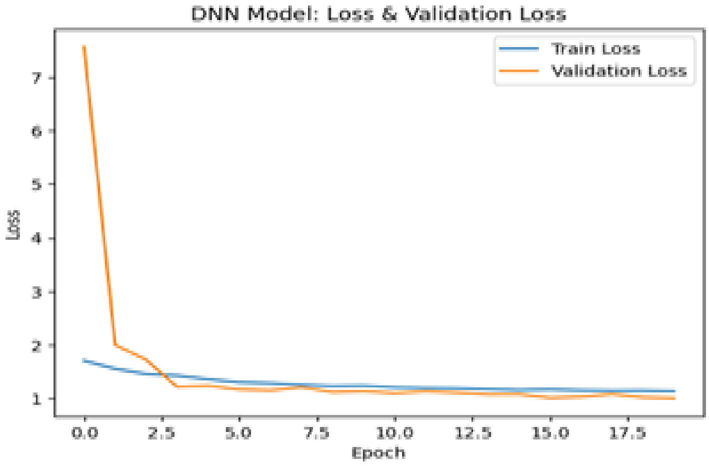


(b) Dataset2

Figure 15: Validation accuracy for (a) dataset1 (b) dataset2



(a) Dataset1



(b) Dataset2

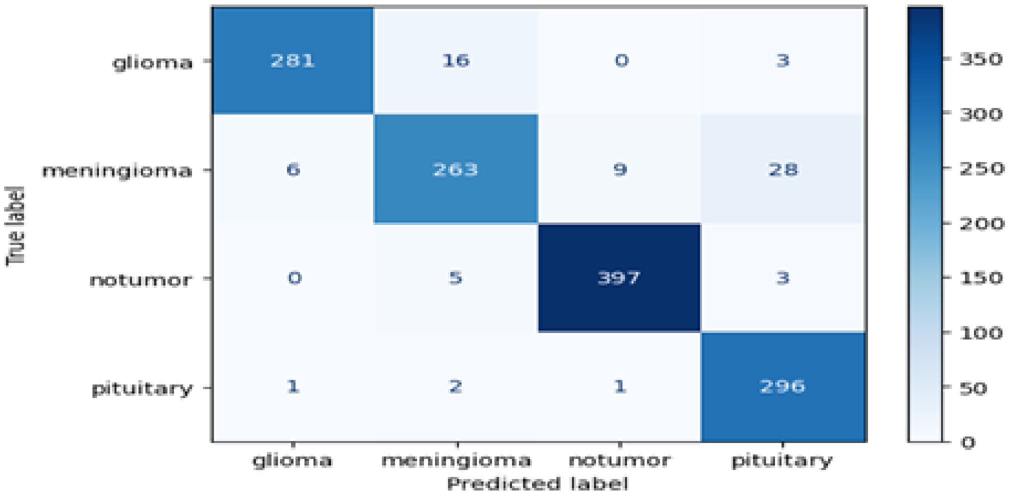
Figure 16: Validation loss for (a) dataset1 (b) dataset2VGG19 achieved high precision, recall, and F1-scores (0.86-0.99) across all classes for both data sets. Table 8 shows the comparison between dataset 1 and dataset 2 in terms of recall factors, precision, and f1 score when applying the VGG19 model.

Table 8: Comparison between dataset1 and dataset2 in precision recall and f1-score factors

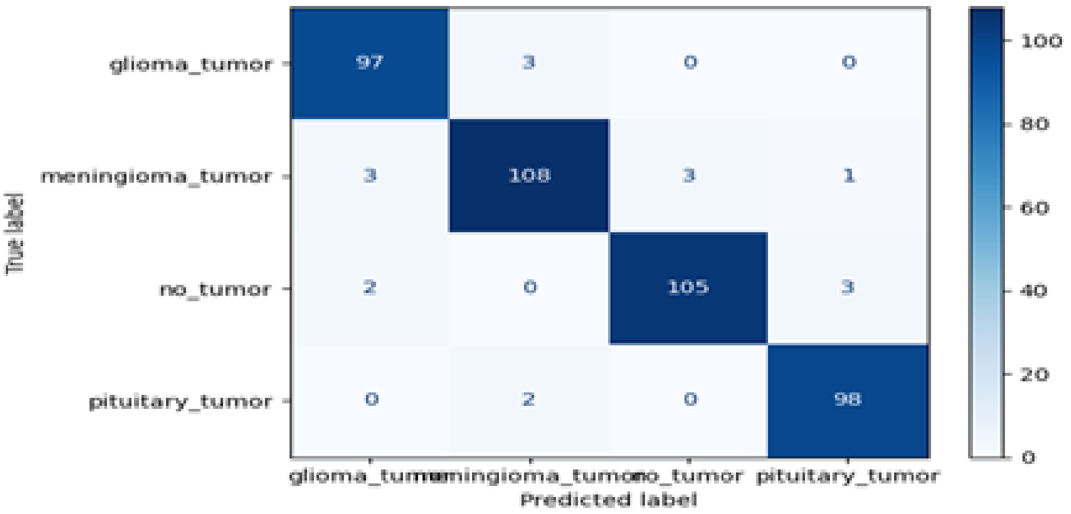
state	**Data set1 [7000] image**			**Data set2 [3000] image**		
	precision	**recall**	**f1-score**	**Precision**	**recall**	**f1-score**
Glioma	98%	94%	96%	95%	97%	96%
Meningioma	92%	86%	89%	96%	94%	95%

Notumor	98%	98%	98%	97%	95%	96%
Pituitary	90%	99%	94%	96%	98%	97%

Figures 17, 18, 19, illustrate the confusion matrix, validation accuracy, and validation loss when applying the VGG19 model to the dataset1 and dataset2.

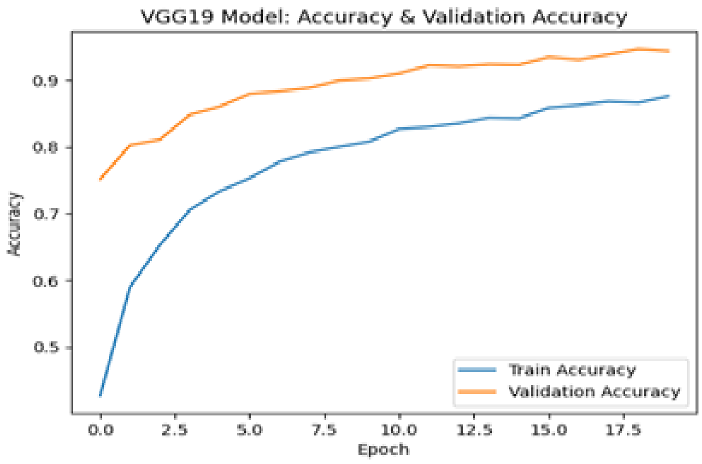


(a) Dataset1

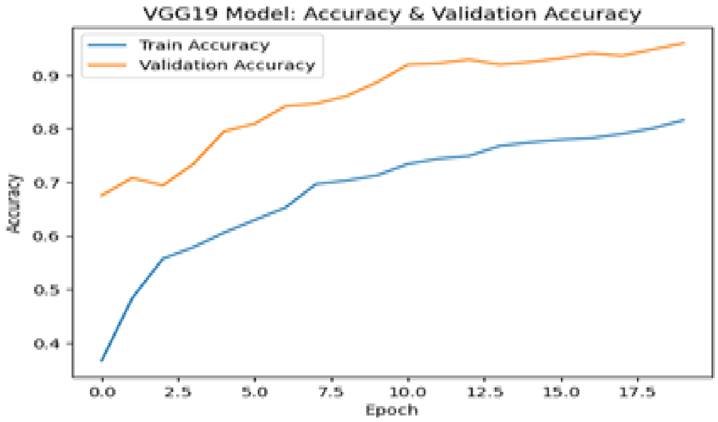


(b) Dataset2

Figure 17: Confusion matrix for (a) dataset1 (b) dataset2

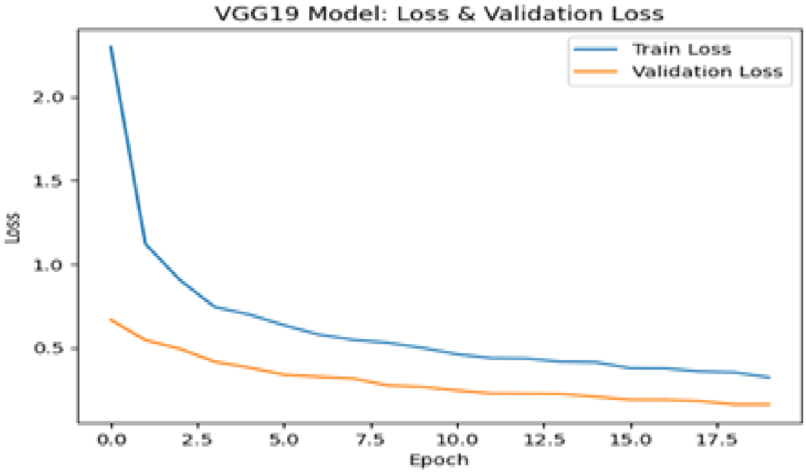


(a) Dataset1

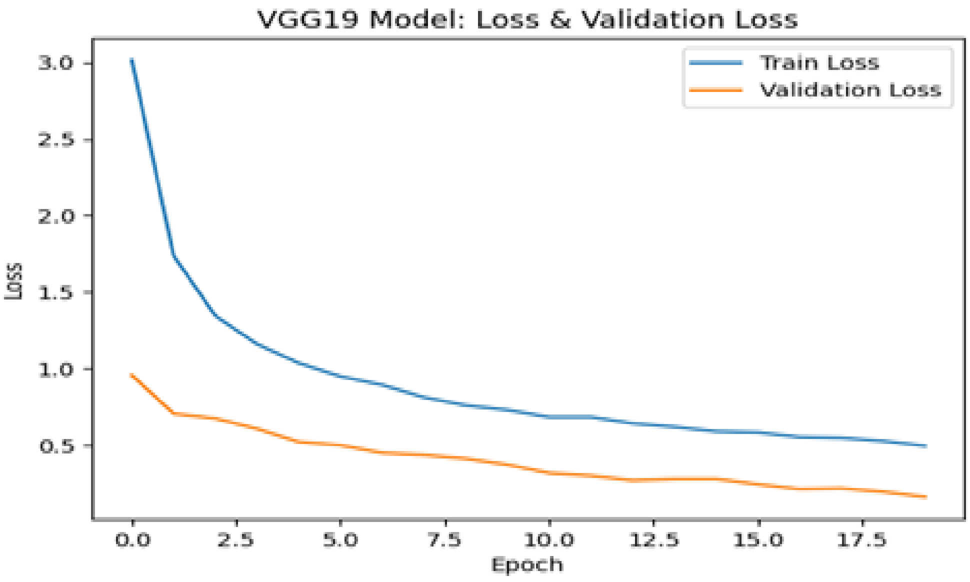


(b) Dataset2

Figure 18: Validation accuracy for (a) dataset1 (b) dataset2



(a) Dataset1



(b) Dataset2

Figure 19: Validation loss for (a) dataset1 (b) dataset2

The analysis of brain tumor detection models reveals that CNN and VGG19 are the top performers, achieving high accuracy and consistent metrics across datasets. CNN excels due to its ability to learn image features effectively, while VGG19 benefits from its deep architecture. CNN-SVM and CNN-LSTM show good results but with some variability depending on the dataset, with CNN-SVM performing better on Dataset 2 and CNN-LSTM showing improvement in Dataset 2 over Dataset 1. CNN-GRU performs exceptionally well on Dataset 1 but struggles with generalization on Dataset 2. DNN models lag behind, struggling with accuracy and validation metrics due to the absence of convolutional layers. Overall, convolutional approaches are preferred for brain tumor detection, though model selection should be tailored to specific dataset characteristics.

7. CONCLUSIONS

DL has shown great potential in aiding brain cancer diagnosis through image analysis. By training DL algorithms on large datasets of brain tumor images, researchers have achieved highly accurate results in identifying and classifying different types of brain tumors. In addition, DL algorithms can help doctors identify and analyze tumor characteristics, such as size, shape, and volume. This helps in treatment planning, as doctors can accurately assess tumor aggressiveness, response to treatments, and monitor disease progression. However, while DL has shown promise in brain cancer diagnosis, there are still challenges that need to be addressed. The need for large and diverse datasets, potential biases in the data, and the interpretability of DL models are some areas that require attention to ensure reliable and robust diagnostic tools. The research has compared six DL models for the task of brain tumor detection using two different datasets with evaluation criteria such as: accuracy, recall, and F1-score. According to the results, CNN and VGG19 are the best models that present high accuracy and robust metrics constantly. CNN showed great performance in both datasets. It achieved very high precision, recall, and F1-score scores for all classes of tumors with a few misclassifications. Its high accuracy is because of its capability in learning image features. VGG19 outperformed with high precision, recall, and F1-scores on both datasets due to its deep architecture that further enhances feature extraction and classification. CNN-SVM performed well with good results for dataset2 compared to dataset1, where the model had high precision and recall for all classes for most of the classes in dataset2, though weaker on Dataset 1. CNN-LSTM performed with high precision and recall, especially on dataset2, and is a candidate for capturing temporal dependencies in data, though the results on dataset 1 were not consistent. CNN-GRU gained near-perfect results on Dataset 1 but was hard to generalize on dataset2, which means strong performances are characteristics of some specific dataset characteristics. Finally, DNN was weak among models and showed inconsistent precision and recall but lower accuracy in general. It probably does not have theconvolutional layers, which are basic for useful image analyses. Finally, those of convolution, such as CNN and VGG19, are generally recommended for the detection of brain tumors by performance. The model type shall depend on the nature of the dataset. In contrast, standalone deep neural networks without convolutional layers have struggled with this image-based task. However, despite their promise, there are still some challenges and limitations with DL for brain cancer diagnosis. The availability of high-quality, well-annotated datasets is critical for training accurate models. Additionally, rigorous validation and testing of these models on diverse patient populations is needed to ensure their robustness and generalizability. Integrating DL algorithms into clinical workflows and obtaining regulatory approval for their use in healthcare settings is a complex process that requires careful consideration of ethical and legal implications. Furthermore, despite the promising results, DL models should not replace human experts. Rather, they should be used as complementary tools to assist radiologists and oncologists in making more accurate and efficient diagnoses.

In conclusion, DL-based diagnosis of brain cancer in medical images has great potential to improve accuracy and efficiency. With further advancements and collaboration between medical professionals and AI experts, these models could become valuable tools in the fight against brain cancer.

ACKNOWLEDGEMENTS

I would like to extend my thanks and appreciation to AL-yarmouk Journal for its good treatment and accuracy in implementing the publishing instructions.

CONFLICT OF INTEREST

The author declares that the publishing of this article does not include any conflicts of interest. Furthermore, the author has strictly adhered to ethical problems such as plagiarism, informed consent, misconduct, data fabrication and falsification, multiple publishing and submission, and redundancy.

REFERENCES

[1] Mathivanan, S. K., Sonaimuthu, S., Murugesan, S., Rajadurai, H., Shivahare, B. D., & Shah, M. A. [2024]. Employing deep learning and transfer learning for accurate brain tumor detection. Scientific Reports, 14[1], 7232.https://doi.org/10.1038/s41598-024-57970-7

[2] Rasool, N., & Bhat, J. I. [2024]. Brain tumour detection using machine and deep learning: a systematic review. *Multimedia Tools and Applications*, 1-54. https://doi.org/10.1007/s11042-024-19333-2

[3] Bienkowski, M., Furtner, J., & Hainfellner, J. A. [2018]. Clinical neuropathology of brain tumors. *Handbook of clinical neurology*, *145*, 477-534. https://doi.org/10.1016/B978-0-12-802395-2.00032-8

[4] Rasool, N., & Bhat, J. I. [2023, March]. Glioma brain tumor segmentation using deep learning: A review. In *2023 10th International Conference on Computing for Sustainable Global Development [INDIACom]* [pp. 484-489]. IEEE.. https://ieeexplore.ieee.org/abstract/document/10112331

[5] Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., ... & Canaider, S. [2013]. An estimation of the number of cells in the human body. *Annals of human biology*, *40*[6], 463-471.https://doi.org/10.3109/03014460.2013.807878

[6] Arunkumar, N., Ramkumar, K., Venkatraman, V., Abdulhay, E., Fernandes, S. L., Kadry, S., & Segal, S. [2017]. Classification of focal and non focal EEG using entropies. *Pattern Recognition Letters*, *94*, 112-

117.https://doi.org/10.1016/j.patrec.2017.05.007

[7] Elazab, A., Wang, C., Gardezi, S. J. S., Bai, H., Hu, Q., Wang, T., ... & Lei, B. [2020]. GP-GAN: Brain tumor growth prediction using stacked 3D generative adversarial networks from longitudinal MR Images. *Neural Networks*, *132*, 321-332. https://doi.org/10.1016/j.neunet.2020.09.004

[8] Zhang, F., Breger, A., Cho, K. I. K., Ning, L., Westin, C. F., O'Donnell, L. J., & Pasternak, O. [2021]. Deep learning based segmentation of brain tissue from diffusion MRI. *Neuroimage*, *233*, 117934. https://doi.org/10.1016/j.neuroimage.2021.117934

[9] Shoeibi, A., Khodatars, M., Jafari, M., Ghassemi, N., Moridian, P., Alizadehsani, R., ... & Gorriz, J. M. [2023]. Diagnosis of brain diseases in fusion of neuroimaging modalities using deep learning: A review. *Information Fusion*, *93*, 85-117.https://doi.org/10.1016/j.inffus.2022.12.010

[10] Bourennane, M., & Naimi, H. [2023]. An efficient hybrid model of CNNs and different kernels of SVM for brain tumor classification. *The Journal of Engineering and Exact Sciences*, *9*[8], 16739-01e.https://doi.org/10.18540/jcecvl9iss8pp16739-01e

[11] Mallick, Sh., Mishra. S. P., An Automatic Detection of Brain Tumor using CNN & VGG19, International Journal on Recent and Innovation Trends in Computing and Communication, 2023, 11[11s]:99-106, , https://doi.org/10.17762/ijritcc.v11i11s.8075

[12] Gomaa, M. M., Zain elabdeen, A. G., Elnashar, A., & Zaki, A. M. [2024]. Brain tumor X-ray images enhancement and classification using anisotropic diffusion filter and transfer learning models. *International Journal of Information Technology*, 1-9.https://doi.org/10.1007/s41870-024-01830-0

[13] Barode, D. M., Awhad, R. S., Dhangar, V. D., & Kawathekar, S. S. Comparative Analysis of SVM and CNN Techniques for Brain Tumor Detection. https://doi.org/10.35940/ijitee.G9908.13070624

[14] Srinivasan, S., Francis, D., Mathivanan, S. K., Rajadurai, H., Shivahare, B. D., & Shah, M. A. [2024]. A hybrid deep CNN model for brain tumor image multi-classification. *BMC Medical Imaging*, *24*[1], 21.https://doi.org/10.1186/s12880-024-01195-7

[15] Tiwari, A., Kumar, D., Sidh, H., Sahare, P., Diwan, T., & Satpute, V.

[2024]. Deep learning based 3D multimodal CNN for brain tumor segmentation and detection. *Iran Journal of Computer Science*, 1-17. https://doi.org/10.1007/s42044-024-00201-z

[16] Shanjida, S., Islam, M. S., & Mohiuddin, M. [2024, May]. Hybrid model-based Brain Tumor Detection and Classification using Deep CNN-SVM. In *2024 6th International Conference on Electrical Engineering and Information & Communication Technology [ICEEICT]* [pp. 1467-1472]. IEEE.https://ieeexplore.ieee.org/abstract/document/10534376

[17] Ghazvini, M. M., Dehlaghi, V., Papi, A., & Mansoory, M. S. Diagnosis and Classification of Brain Tumors from MRI Images Using the SVM Algorithm. *Journal of Clinical Research in Paramedical Sciences*, *13*[1]. https://doi.org/10.5812/jcrps-148703

[18] https://www.kaggle.com/datasets/adityakomaravolu/brain-tumor-mri-images.

[19] https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri

[20] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. [1998]. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*[11], 2278-2324. https://doi.org/10.1109/5.726791

[21] X. Kang, B. Song, and F. Sun, **"A deep similarity metric method based on incomplete data for traffic anomaly detection in IoT",** Applied Sciences (Switzerland), vol. 9, no. 1. 2019, doi: 10.3390/app9010135

[22] Ashraf, N., Nayel, H., Aldawsari, M., Shashirekha, H., & Elshishtawy, T. [2024, August]. BFCI at AraFinNLP2024: Support Vector Machines for Arabic Financial Text Classification. In *Proceedings of The Second Arabic Natural Language Processing Conference* [pp. 446-449]. https://doi.org/10.18653/v1/2024.arabicnlp-1.42

[23] Awad, M., Motai, Y., Näppi, J., & Yoshida, H. (2010). A clinical decision support framework for incremental polyps classification in virtual colonoscopy. *Algorithms*, *3*(1), 1-20.

[24] Heikal, M., Torki, M., & El-Makky, N. (2018). Sentiment analysis of Arabic tweets using deep learning. *Procedia Computer Science*, *142*, 114-122.

[25] Kong, Z., Cui, Y., Xia, Z., & Lv, H. (2019). Convolution and long short-term memory hybrid deep neural networks for remaining useful life prognostics. *Applied Sciences*, *9*(19), 4156.

[26] Zhu, X., Liu, R., Chen, Y., Gao, X., Wang, Y., & Xu, Z. [2021]. Wind

speed behaviors feather analysis and its utilization on wind speed prediction using 3D-CNN. *Energy*, *236*, 121523. https://doi.org/10.1016/j.energy.2021.121523

[27] Huang, X., Zhang, Y., Liu, J., Zhang, X., & Liu, S. [2023]. A short-term wind power forecasting model based on 3d convolutional neural network–gated recurrent unit. *Sustainability*, *15*[19], 14171. https://doi.org/10.3390/su151914171

[28] Adam, A. Z. R., & Setiawan, E. B. [2023]. Social media sentiment analysis using convolutional neural network [CNN] dan gated recurrent unit [GRU]. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika [JITEKI]*, *9*[1], 119-131.https://doi.org/10.12928/jiteki.v9i1.23696

[29] Albashish, D., Al-Sayyed, R., Abdullah, A., Ryalat, M. H., & Almansour, N. A. [2021, July]. Deep CNN model based on VGG16 for breast cancer classification. In *2021 International conference on information technology [ICIT]* [pp. 805-810]. IEEE..https://doi.org/10.1109/ICIT52682.2021.9491631

[30] Zhang, Y., Xie, Y., Zhang, Y., Qiu, J., & Wu, S. (2021). The adoption of deep neural network (DNN) to the prediction of soil liquefaction based on shear wave velocity. *Bulletin of Engineering Geology and the Environment*, *80*, 5053-5060.

[31] Khattar, A., & Quadri, S. M. K. (2022). Generalization of convolutional network to domain adaptation network for classification of disaster images on twitter. *Multimedia Tools and Applications*, *81*(21), 30437-30464.

[32] Kisvari, A., Lin, Z., & Liu, X. [2021]. Wind power forecasting–A data-driven method along with gated recurrent neural network. *Renewable Energy*, *163*, 1895-1909.https://doi.org/10.1016/j.renene.2020.10.123.