International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 10, Issue 1 (2024)
https://doi.org/10.31436/ijpcc.v10i1.460

Utilising VGG-16 of Convolutional Neural Network for Medical Image Classification

Amelia Ritahani Ismail*, Syed Qamrun Nisa, Shahida Adila Shaharuddin, Syahmi Irdina Masni, Syaza Athirah Suharudin Amin
Department of Computer Science, Kulliyyah of Information and Communication Technology
International Islamic University Malaysia

*Corresponding author: amelia@iium.edu.my
(Received: 14th December 2023; Accepted: 30th December 2023; Published on-line: 28th January 2024)

Abstract—Medical image classification, which involves accurately classifying anomalies or abnormalities within images, is an important area of attention in healthcare domain. It requires a fast and exact classification to ensure appropriate and timely treatment to the patients. This paper introduces a model based on Convolutional Neural Network (CNN) that utilises the VGG16 architecture for medical image classification, specifically in brain tumour and Alzheimer dataset. The VGG16 architecture, is known for its remarkable ability to extract important features, that is crucial in medical image classification. To enhance the precision of diagnosis, a detailed experimental setup is conducted, which includes the careful selection and organisation of a collection of medical images that cover different illnesses and anomalies to the dataset. The architecture of the model is then adjusted to achieve optimal performance in for image classification. The results show the model's efficiency in identifying anomalies in medical images especially for brain tumour dataset. The sensitivity, specificity, and F1-score evaluation metrics are presented, emphasising the model's ability to accurately differentiate between various medical image diseases.

Keywords— Deep learning, Convolutional Neural Network (CNN), VGG-16, medical image classification.

I. Introduction

Medical image classification is a crucial component of healthcare, fundamentally transforming how medical practitioners identify and treat different diseases in assisting the patients. Recently, Convolutional Neural Networks (CNNs) have become a prominent method in medical image analysis for automating detection and classification tasks, such as the VGG-16 model that has the ability in improving the precision and dependability of medical image classification. The importance of medical image classification cannot be overstated as it assists radiologists and healthcare professionals in promptly and precisely detecting abnormalities, lesions, tumours, and other medical irregularities in different imaging techniques such as X-rays, MRI (Magnetic Resonance Imaging), CT (Computed Tomography) scans, and ultrasound pictures. Prompt detection is crucial for timely action, improved patient results, and decreased healthcare expenses.

This paper investigates the use of Convolutional Neural Networks (CNNs), namely the VGG-16 architecture, for medical image classification. It begins with the examination of the current literature, highlighting valuable perspectives on the most advanced techniques and emphasising the importance of automated medical image classification in today's healthcare domain. Next, the experimental setup is discussed to guarantee the detailed assessment of the CNN-based approach and VGG-16 in medical image classification. Two datasets, which are the brain tumour and Alzheimer dataset is collected to evaluate the model on balanced and imbalanced dataset. Data augmentation methods, including rotation, scaling, and flipping, are used to increase the size of the dataset and address problems associated with the uneven distribution of classes in the datasets. The outcome of these model is explained in the results section, where a comprehensive evaluation of the model's performance is displayed. The evaluation parameters, such as accuracy, sensitivity, specificity, and F1-score, are utilised to evaluate the performance of CNN and VGG-16 in medical image classification comprehensively.

II. Literature Review

Medical image analysis is an advancing discipline that combines medicine, computer science, and engineering. Its main objective is to develop and utilise computational methods for analysing and interpreting medical images. The images acquired from medical image analysis are like X-ray, MRI (Magnetic Resonance Imaging), CT (Computed Tomography), ultrasound, and PET (Positron Emission Tomography), which offer crucial visual data regarding the

https://doi.org/10.31436/ijpcc.v10i1.460

structure and diseases of the human body. Medical image analysis has become a clinical practice across radiology, oncology, cardiology, and neurology as it facilitates the timely identification of illnesses, evaluation of therapy efficacy, and formulation of surgical strategies if it is needed. Furthermore, it enhances the field of personalised medicine by enabling customised treatment strategies informed by each patient's unique imaging data provided by the medical images. The advancement of medical image detection technology encounters obstacles such as effectively handling substantial amounts of imaging data, guaranteeing precision and dependability, safeguarding patient confidentiality, and seamlessly integrating with current healthcare systems. However, continuous research and innovation in this domain hold the potential to overcome these obstacles and uncover novel opportunities in medical diagnostics and therapy.

A brain tumour is an aggregation of cells, specifically a neoplasm that have undergone unregulated proliferation and multiplication in many regions of the brain, including glial cells, neurons, lymphatic tissue, blood vessels, the pituitary and pineal glands, the skull, or because of metastasis from other organs [1][2][3]. Brain tumours are categorised according to their precise intracranial location and their characteristics, such as benign or malignant, and other relevant factors that are associated with them. Diverse imaging techniques, including known ones such as Magnetic Resonance Imaging (MRI), provide supplementary information regarding brain tumours' genetic makeup, prognosis, and predictive characteristics. Magnetic Resonance Imaging (MRI) is vital in enhancing preoperative diagnosis, predicting tumour grade and patient prognosis, guiding surgical and radiation treatment planning, and evaluating treatment effectiveness. Imaging continues to be an effective and non-invasive tool that positively affects the treatment of patients with brain tumours [1]. A system was suggested by [4] and [5] to automatically diagnose brain tumours using the machine learning and VGG-16 with 23-layer of CNN architecture. Both experimental findings demonstrate that the models able to detect brain tumour as it can achieve up to 97.8% and 100% prediction accuracies.

Convolutional Neural Networks (CNNs) are specialised deep neural networks primarily employed for visual analysis. Convolutional neural networks show great potential in problems involving image recognition and classification due to their ability to gather important features of information from input images independently and adaptively. A Convolutional Neural Network (CNN) comprises many layers, including convolutional layers, pooling layers, and fully connected layers, each with its specific function in processing the input data. Convolutional Neural Networks (CNNs) can be trained to recognise distinct patterns and abnormalities in brain scans, such as MRI pictures, that indicate the existence of brain cancers and other medical images. VGG-16 is a CNNs architecture known for its simplicity. It consists exclusively of 3x3 convolutional layers stacked together, with each layer having a greater depth than the previous one. The model comprises 16 layers with corresponding weights, essential for accurately capturing complicated patterns features in high-level characteristics. VGG16 is particularly effective in image classification due to its proficiency in collecting intricate details and textures in medical images, which is crucial, for instance, in recognising tumour tissues as well as other medical images. [2] provides more evidence by showcasing the wide range of applications of CNNs in medical imaging. Their research investigates the adaptability of CNN architectures, demonstrating their effective adaptation to tackle the challenges faced in brain tumour classification.

The research conducted by [3] on dual convolutional neural network (CNN) models enhances understanding of the potential of hierarchical neural networks in medical diagnostics. By utilising two Convolutional Neural Networks (CNNs) concurrently, this approach provides an extra layer of verification and precision, ensuring the utmost accuracy in the diagnosis. This dual-model technique aims to address a significant challenge in medical imaging, which is the requirement to limit both false positives and false negatives. By implementing this approach, the strategy improves the reliability of diagnostic processes.

[6] Presented the Convolutional Neural Network (CNN) model for identifying tumours in Magnetic Resonance Imaging (MRI). At first, Support Vector Machines (SVM) were combined with Convolutional Neural Networks (CNN) but needed better accuracy. Further improvement was achieved by fine-tuning parameters, such as changing the final layer parameter to SoftMax and employing the AdaMax optimisers. However, the RMSProp optimiser achieved better results, resulting in a remarkable increase in output accuracy. [7] employed a single CNN architecture as a binary classifier to detect various abnormalities individually. The analysis utilised three convolutional layers and two fully linked layers for each irregularity. [8] presents a specialised network architecture designed for accurately identifying numerous anomalies in medical images. The network has a central module for extracting shared characteristics and two supplementary modules for performing classification and segmentation tasks. [9] proposes a method for identifying COVID-19 cases from chest X-rays while also differentiating between regular X-rays and those affected by Viral Pneumonia. The approach utilises Deep CNNs model. In the model, transfer learning is employed to evaluate three pre-trained CNN models, namely EfficientNetB0, VGG16, and InceptionV3 because they offer a favourable combination of

https://doi.org/10.31436/ijpcc.v10i1.460

accuracy and efficiency. The results confirm that it is possible to use advanced deep CNN computer models in healthcare settings to screen and diagnose COVID-19 from chest X-rays.

[10] presents a model that utilises a CNN to identify numerous rib fractures by analysing quality-normalised chest radiographs. The research encompassed several stages, including gathering data from multiple centres, normalising image using the CNN models, evaluating its performance, and comparing it with assessments made by radiologists to validate the results. The results imply that the CNN model exhibited high diagnostic effectiveness, indicating the model ability to enhance the detection of rib fractures in chest radiographs. This can potentially reduce missed diagnoses and lessen the workload of radiologists. However, additional verification of the detection capabilities is necessary and still needed to substantiate the prospective function of CNN in medical image detection.

III. Experimental Setup

The experimental design began with a process of collecting a dataset, followed by data pre-processing, that encompasses the tasks of cleaning, normalising, and transforming the data to ensure its suitability for analysis. Next, data augmentation techniques, including rotations, flips, and shifts, were utilised to increase the variety of the dataset and assist model training. Furthermore, data visualisation was conducted, providing valuable insights into the dataset's attributes, and assisting in identifying trends and abnormalities in the dataset. Afterwards, the VGG16 Model is applied to the dataset. Finally, the analysis of results is conducted thoroughly assessing the model's performance, using metrics such as accuracy, precision, and recall evaluating the model's effectiveness. The Flow chart of the experimental design is shown in Fig. 1.

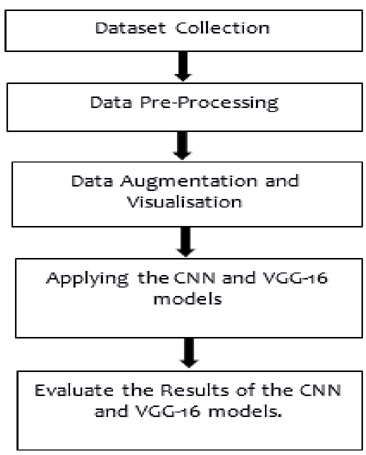

Fig. 1 Flow chart of experimental design.

A. Dataset Description

The Brain Tumour MRI dataset used is obtained from Kaggle, consisting of 7023 pictures of MRI scans of the human brain. Four classes are used to categorise these images: glioma, meningioma, no tumour, and pituitary. A total of 5712 labelled images are included in the training set, which accounts for 80% of the total images. The testing set consists of 1311 images of brain tumours, making up 20% of the total images.

The second dataset used is Alzheimer's Dataset that contains MRI images that have been categorised into four classes. A significant disparity is observed in the number of images for each file in the Alzheimer's dataset, leading to imbalanced data. The accuracy between these two data sets will be compared and the performance will be evaluated accordingly.

B. Data Augmentation and Visualization

Image augmentation is applied to increase the diversity of images of the dataset. This is adjusted by increasing:

- Random Brightness: 80% - 120%
- Random Contrast: 80% - 120%
- Random sharpness: 80% - 120%

Visualisation is crucial in transforming extensive data sets into meaningful visual graphical representations, such as charts and graphs, to provide insights into complex data. The dataset focused solely on images of four brain tumour classes: glioma, meningioma, no tumour (no tumour), and pituitary. A more precise understanding of the distinctions between various types of brain tumours was aimed to be provided by visualising a selection of augmented images and corresponding labels. The subplots of MRI images of the kind of brain tumours with labels are illustrated in Fig. 2.

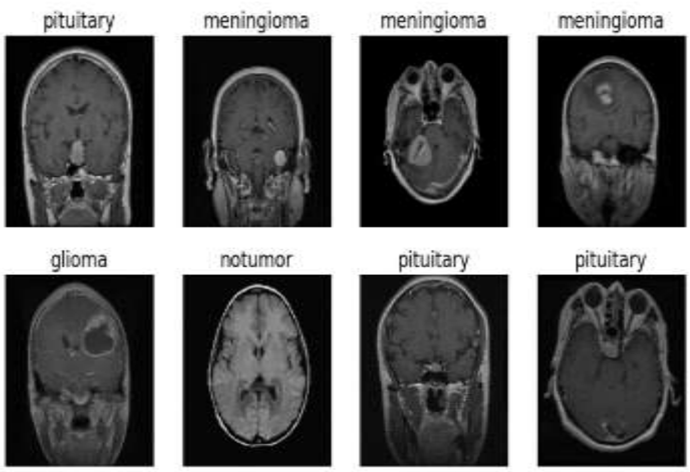

Fig. 2 Subplots of MRI images of type of brain tumours with labels

https://doi.org/10.31436/ijpcc.v10i1.460

IV. MODELLING

A. *CNN Model*

The In this section, a Convolutional Neural Network (CNN) model is constructed. The model consists of 3 convolutional layers with ReLU as its activation function. The middle convolution layer has the highest number of filters which is 64 compared to the other 2 layers which are 32 filters. After every convolution layer, a pooling layer is implemented to reduce the spatial dimensions of the input. Lastly, a process of flattening the 2D-matrix is done. The flattened 1D-array is connected to a fully dense neural network which is set up to perform classification of the label. In the model architecture, each convolution layer is set with a kernel size of 3x3 with a Rectified Linear Unit (ReLU) activation function. It helps capture local and non-linear patterns in the input image. Additionally, a maximum pooling layer is used to reduce the computational complexity. In this model a maximum pooling approach with a 2x2 size matrix pool is used. Furthermore, the model is compiled using the Adam optimiser, which is an algorithm that adjusts network weights during training to minimise the loss function. The model tries to minimise the difference between the predicted label and the actual label. Finally, the accuracy metric is used to see the performance of the model by calculating the percentage of correctly predicted labels.

B. *The VGG-16 Model*

In this experiment, the VGG-16 model as shown in Fig. 3 provided by the Keras library was used. The VGG-16 model is a deep neural network architecture that has been pre-trained on large datasets for image classification tasks. First, the VGG-16 model is imported which contains the parameters. The layers in the base VGG16 model are initially frozen which means the weights are not updated during training, but the last few layers are made trainable to refine the model for our specific brain tumour classification task. A few additional layers, such as flattening, dropout, and dense layers, are added after the VGG16 model to further refine the extracted features and make predictions.

The VGG16 components, is then compiled with Adam optimiser, loss function and evaluation metric. The model is then trained using the datasets and the model trained accuracy can be seen in Fig. 4, which shows accuracy and loss values across epochs. It can be observed that the accuracy steadily increased, and the loss consistently decreased with each epoch. This demonstrates that as the model learned from the training data, its performance improved.

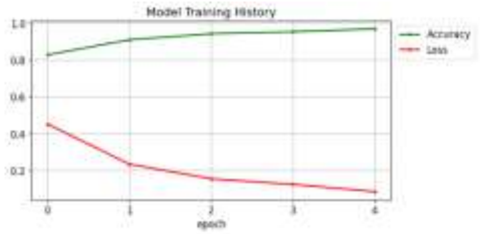

Fig. 4 *Line graph of accuracy of model training history*

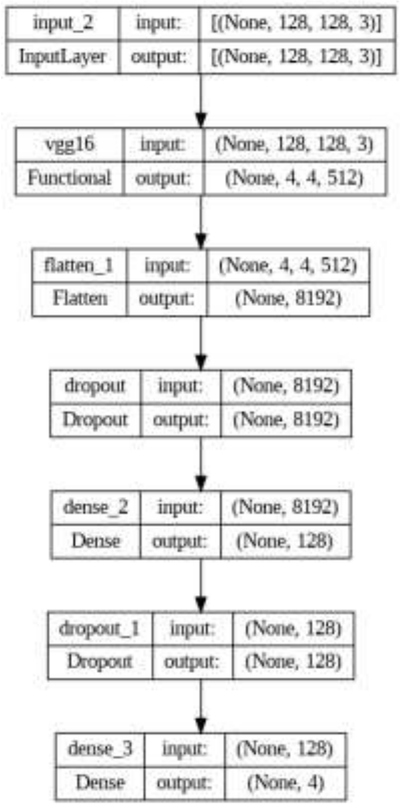

Fig. 3 VGG-16 model.

V. RESULTS AND DISCUSSION

A. *Result of CNN Model for Brain Tumour.*

The CNN model achieved a training accuracy of 96.27% and a test accuracy of 90.47%. This indicates that CNN model alone performs well in classifying brain tumour images. Table I shows the results of train and test accuracy for the CNN model.

TABLE I.
Results of train and test accuracy for CNN model

Train Accuracy	**Test Accuracy**
0.9627	0.9046

https://doi.org/10.31436/ijpcc.v10i1.460

TABLE II presents a comprehensive evaluation of the model's performance on several classifications of brain tumours, including multiple evaluation measures. The model achieved high precision for the 'no tumour' class, with a perfect score 1.00. This suggests that it accurately classified all photos without tumours. Regarding recall, the class has the highest value of 0.98. Subsequently, the 'no tumour' class demonstrated an F1-score of 0.97, signifying a good precision and recall. The CNN model achieved an accuracy of 0.92, indicating that it correctly categorised 92% of the images in the evaluation dataset.

TABLE II.
Classification Report of CNN Model

	Precision	Recall	F1-score	Support
no tumour	1.00	0.94	0.97	405
glioma	0.95	0.83	0.89	300
pitutary	0.94	0.98	0.96	300
meningioma	0.79	0.92	0.85	306
accuracy			0.92	1311
macro avg	0.92	0.92	0.92	1311
weighted avg	0.92	0.92	0.92	1311

B. Result of VGG-16 for Brain Tumour.

The utilisation of the VGG16 architecture leads to an enhancement in the model's accuracy. The model achieved an accuracy of 0.97 in the classification report, suggesting that it correctly categorised 97% of the images in the evaluation dataset. The precision, recall, and F1-scores metrics for each class exhibited enhancement, suggesting an overall improvement in performance. The results indicate that including the VGG-16 architecture improved the model's accuracy in classifying brain tumour images. The categorisation report for the VGG 16 Model is displayed in TABLE III.

TABLE III.
Classification result for VGG-16 model.

	Precision	Recall	F1-score	Support
glioma	0.94	0.96	0.95	300
meningioma	0.96	0.92	0.94	306
notumor	0.98	1.00	0.99	405
pituitary	0.98	0.98	0.98	300
accuracy			0.97	1311
macro avg	0.97	0.96	0.97	1311
weighted avg	0.97	0.97	0.97	1311

C. Result of VGG-16 for Alzheimer's Dataset

The accuracy of the CNN model is further analysed in the Alzheimer's dataset, which is imbalanced and consists of four classes (MildDemented, VeryMildDemented, NonDemented, and ModerateDemented), was 0.62. The model is further augmented with data to enhance its performance and generalisability. Augmentation techniques such as random rotations, flips, and zooms were applied to the images. Upon comparing the accuracy with and without data augmentation, it was seen that the accuracy drastically dropped to 0.59 in the absence of augmentation. The significant decrease in accuracy underscores the necessity of employing data augmentation to enhance the model's performance on the Alzheimer's dataset. Data augmentation enhances the variety and volume of the training dataset, enabling the model to acquire more resilient characteristics and generalise more efficiently to previously unobserved images. While achieving an accuracy of 0.62 signifies a certain level of competence, there is still potential for further improvement. The accuracy limits may have been influenced by factors such as class imbalance and variations in tumour characteristics within each class. TABLE IV and TABLE V display the accuracy of an augmented and non-augmented dataset for Alzheimer's disease, respectively.

TABLE IV.
The accuracy for an augmented Alzheimer dataset

	Precision	Recall	F1-score	Support
VeryMildDemented	0.57	0.59	0.58	448
MildDemented	0.60	0.18	0.28	179
NonDemented	0.65	0.77	0.71	640
ModerateDemented	1.00	0.17	0.29	12
accuracy			0.62	1279
macro avg	0.71	0.43	0.46	1279
weighted avg	0.62	0.62	0.60	1279

TABLE V.
The accuracy for a non-augmented Alzheimer dataset

	Precision	Recall	F1-score	Support
VeryMildDemented	0.52	0.69	0.59	448
MildDemented	0.65	0.12	0.21	179
NonDemented	0.65	0.67	0.66	640
ModerateDemented	0.00	0.00	0.00	12
accuracy			0.59	1279
macro avg	0.45	0.37	0.36	1279
weighted avg	0.60	0.59	0.57	1279

VI. Conclusions

To summarise, the VGG-16 model has produced encouraging results in predicting brain tumours. The Brain Tumour MRI dataset demonstrated markedly superior accuracy compared to the Alzheimer's dataset, which is imbalanced. This highlights the significant influence of dataset attributes and complexity on the model's performance. The Brain Tumour MRI dataset exhibited prominent and well-defined characteristics, which greatly facilitated precise categorisation by the models. Increasing the training data significantly contributed to improving the accuracy. The models were exposed to more data by applying various transformations to the training images, such as adjusting brightness, contrast, and sharpness. This enhanced their capacity to make generalisations and reduced the likelihood of overfitting.

Acknowledgment

The authors would like to extend an appreciation to all those that participated in this research.

Conflict of Interest

The authors declare that there is no conflict of interest

References

[1] M. C. Mabray, R. F. Barajas, & S. Cha, "Modern Brain Tumor Imaging". Brain Tumor Research and Treatment, 2015. https://synapse.koreamed.org/articles/1059671

[2] S. Roy, S. Nag, I. K. Maitra, & S. K. Bandyopadhyay, "A review on Automated Brain Tumor Detection and segmentation from MRI of brain". NASA/ADS, 2013. https://ui.adsabs.harvard.edu/abs/2013arXiv1312.6150R/abstract

[3] M. Toğaçar, B. Ergen, & Z. Cömert, "BrainMRNet: Brain tumor detection using magnetic resonance images with a novel convolutional neural network model". *Medical Hypotheses*, *134*, 109531, 2020. https://doi.org/10.1016/j.mehy.2019.109531

[4] K. Sharma, A. Kaur, & S. Gujral, "Brain Tumor Detection based on Machine Learning Algorithms". International Journal of Computer Applications, 103(1), 7–11, 2014. https://doi.org/10.5120/18036-6883

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

[6] A. Chattopadhyay, & M. Maitra, "MRI-based brain tumour image detection using CNN based deep learning method". *Neuroscience informatics*, 2(4), 100060, 2022.

[7] A. K. Sekuboyina, S. T. Devarakonda and C. S. Seelamantula, "A convolutional neural network approach for abnormality detection in Wireless Capsule Endoscopy," IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, Australia, 2017, pp. 1057-1060, doi: 10.1109/ISBI.2017.7950698.

[8] M. Hajabdollahi, R. Esfandiarpoor, E. Sabeti, N. Karimi, S.M.R. Soroushmehr, S. Samavi, "Multiple abnormality detection for automatic medical image diagnosis using bifurcated convolutional neural network". *Biomedical Signal Processing and Control*, 57, 101792, 2020. https://doi.org/10.1016/j.bspc.2019.101792.

[9] L. Gaur, U. Bhatia, N. Z. Jhanjhi, G. Muhammad, & M. Masud, "Medical image-based detection of COVID-19 using deep convolution neural networks". *Multimedia systems*, 29(3), 1729-1738, 2023. https://doi.org/10.1007/s00530-021-00794-6

[10] J. Wu, N. Liu, X. Li, Q. Fan, Z. Li, J. Shang, F. Wang, B. Chen, Y. Shen, P. Cao, & Z. Liu, "Convolutional neural network for detecting rib fractures on chest radiographs: a feasibility study". *BMC Medical Imaging*, 23(1), pp.1-12, 2023. https://doi.org/10.1186/s12880-023-00975-x