Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 12, No. 3, September 2024, pp. 926~939
ISSN: 2089-3272, DOI: 10.52549/ijeei.v12i3.5977

Enhancing Confidence in Brain Tumor Classification Models With Grad-CAM and Grad-CAM++

Hoang-Tu Vo, Nhon Nguyen Thien , Kheo Chau Mui, Phuc Pham Tien
Information Technology Department, FPT University, Can Tho 94000, Vietnam

Article Info

Article history:

Received Oct 18, 2024
Revised Nov 7, 2024
Accepted Nov 23, 2024

Keywords:

Brain tumors
Explainable AI
XAI
Grad-CAM
Grad-CAM++
Transfer Learning
Pre-trained models

ABSTRACT

Brain tumors are a terrible and dangerous health problem, often posing a significant threat to individuals due to their high probability of death. Detecting these tumors at an early stage is crucial, as it not only increases the chances of successful treatment but also plays a pivotal role in reducing total healthcare costs. Early detection allows medical professionals to take action quickly, enabling a more targeted and effective treatment approach. Numerous studies are currently employing Machine Learning (ML) and Deep Learning (DL) to classify brain tumors, promising improved accuracy and efficiency in tumor identification for potential breakthroughs in medical diagnosis. However, a significant challenge lies in these models being "black box" as their complex inner workings are not easily understood by humans. Explainable Artificial Intelligence (XAI) refers to the capability of an artificial intelligence (AI) system to provide understandable and interpretable explanations for its decisions or predictions. In this study, we propose a classification model based on various network architectures, namely DenseNet201, DenseNet169, Xception, MobileNetV2 and ResNet50. We then used Grad-CAM and Grad-CAM++ to interpret the model's results, evaluating its ability to distinguish important features in Magnetic resonance imaging (MRI) images of brain tumors during the decision-making process. The integration of Grad-CAM and Grad-CAM++ enhances the interpretability of the brain tumor classification model, providing valuable evidence of its effectiveness by focusing on crucial features in MRI images of brain tumors during decision-making. Research results contribute to the development of systems that support early diagnosis of tumors. This contribution is pivotal as it not only enhances the model's transparency but also validates its effectiveness in accurately identifying brain tumors.

Corresponding Author:

Hoang-Tu Vo
Information Technology Department, FPT University, Can Tho 94000, Vietnam.
Email: tuvh6@fe.edu.vn

1. INTRODUCTION

Brain tumors are a serious health issue, putting people at risk of dying. Finding these tumors early is super important because it not only helps treat them better but also saves money on healthcare. Brain tumors pose a serious risk to individuals, being a common and potentially life-threatening medical condition linked to higher chances of death [1], [2]. The increasing common of this dangerous disease emphasizes the critical need for detailed examination and advanced medical treatment. Establishing an early detection system is crucial, but the current diagnostic process is time-consuming. The extended time taken for diagnosis presents a significant challenge, mainly due to the sequence of diagnostic procedures that patients undergo to confirm the presence of a brain tumor. These complex measures contribute to delays in starting necessary medical operations on time. Brain tumors are commonly categorized into two Categories: benign and malignant. This classification serves as a fundamental framework for understanding the varying degrees of severity associated with these tumors. The ability to differentiate between benign tumors, usually non-cancerous, and malignant tumors, which have the potential to cause cancer, is crucial for devising effective treatment strategies.

Several studies in the scientific literature emphasize the significance of leveraging artificial intelligence (AI) methodologies for the early diagnosis of brain tumors. A study conducted by Havaei et al. (2017) employed deep learning algorithms on MRI scans to classify brain tumors with remarkable accuracy, demonstrating the potential of AI in improving diagnostic capabilities [3]. Additionally, the work of Maqsood et al. (2022) explored the use of a combination of a deep neural network architecture and a multiclass support vector machine (SVM) for the classification of brain tumors, showcasing the capacity of these technologies to enhance diagnostic precision [4].

The application of ML in brain tumor classification has been further validated by research conducted by Akkus et al. (2017), which demonstrated the efficacy of a DL algorithm for brain segmentation on MRI images [5]. These findings collectively highlight the increasing number of evidence supporting the potential breakthroughs in medical diagnosis facilitated by ML and DL approaches in the realm of brain tumor detection.

However, the significant challenge associated with many advanced machine learning models, including neural networks and deep learning algorithms, lies in their inherent "black box" nature. This term refers to the opacity and complexity of the internal mechanisms of these models, making it difficult for humans to understand how these models make predictions or decisions. Unlike traditional, logistic regression [6], decision trees (DT) [7], k-nearest neighbors (KNN) [8] model where the decision-making process is transparent and interpretable, "Black box" models operate with complex layers of interconnected neurons, often in high-dimensional space. The absence of clear interpretability gives rise to apprehensions, particularly in vital fields like healthcare or finance, where understanding the reasoning behind a model's decision is crucial for trust.

Scientists are currently investigating various approaches to enhance the interpretability of these models, such as developing explainability techniques, creating simpler interpretable models, or integrating transparency features into complex architectures. Addressing the "black box" challenge is essential for improving user confidence in deploying these advanced models.

Despite the promising advances of ML and DL methods in promoting brain tumor detection. An important obstacle remains in the form of the fundamental "black box" characteristics of these complex models. The opacity and complexity of the internal mechanisms of these models prevent our ability to understand decision-making processes, giving rise to significant concerns, especially in critical areas such as healthcare.

In this context, this study endeavors to address the interpretability challenge associated with advanced machine learning models, focusing on their application in brain tumor classification. Our objective is to propose a classification model based on various network architectures - DenseNet201, DenseNet169, Xception, MobileNetV2 and ResNet50 and employ XAI techniques to enhance the transparency of the decision-making process. we seek to provide an in-depth understanding of the important features in MRI scans, aiming to bridge the gap between model complexity and interpretability of results.

The main contributions of this paper are as follows:

- Proposing a Brain Tumor Classification Model: We introduce a classification model based on various network architectures, including DenseNet201, DenseNet169, Xception, MobileNetV2, and ResNet50.
- Leveraging Grad-CAM and Grad-CAM++: We employ these visualization techniques to interpret the model’s results, focusing on crucial features in Magnetic Resonance Imaging (MRI) scans of brain tumors during decision-making.
- Enhancing Interpretability: By integrating Grad-CAM and Grad-CAM++, we enhance the model’s interpretability, bridging the gap between accuracy and transparency in brain tumor diagnosis.

The structure of the paper is as follows: In Part 2, we present a thorough literature review. Part 3 outlines the classification method and explains the model results, including details about the proposed deep learning methodology, the dataset, data preparation, evaluation metrics for the model, and Classification process and explain the results of the model. Moving on to Part 4, we delve into the experimental system and present the final results. Finally, Part 5 summarizes our study’s findings, offers concluding remarks, and delineates possible directions for future investigation.

2. RELATED WORK

The use of ML and DL algorithms for medical image analysis has gained popularity in recent years, revolutionizing the field of diagnostic imaging. With the advent of complex algorithms and the availability of massive datasets, researchers have been able to develop advanced models that automate and enhance medical image interpretation. These studies span a variety of medical imaging modalities, including but not limited to

MRI, computed tomography (CT), ultrasound, and X-ray. ML and DL models have demonstrated remarkable success in tasks such as lesion detection, image segmentation, and disease classification.

Pathak, Yadunath, et al. in the article [9] employs a deep transfer learning (DTL) method for COVID-19 patient classification and addresses challenges in noisy and imbalanced datasets. Experimental findings demonstrate the efficacy of the proposed DTL model, revealing its efficiency in comparison to other supervised learning models. The authors in This study [10] aims to employ pre-trained DL architectures for automated detection and diagnosis of COVID-19 in chest CT scans. The proposed DenseNet201 architecture, utilizing learned weights from the ImageNet dataset and a convolutional neural architecture, demonstrates superior performance in classifying COVID-19 patients compared to other competitive approaches, as revealed through extensive experimental evaluations.

In this paper [11] introduces a CNN model with fewer learning parameters, consisting of five main convolutional layers, and proposes a multi-stage fusion functionality to enhance the efficiency of COVID-19 screening. Experimental results, utilizing publicly available LUS photographs and video datasets, demonstrate that the proposed fusion technique achieves significantly higher performance metrics with 92.5% precision, 91.8% accuracy, and 93.2% retrieval compared to cutting-edge CNN versions. To detect positive COVID-19 cases early and prevent further spread.

Karhan, Zehra, and A. K. A. L. Fuat. in the study [12] utilizes the ResNet50 CNN to classify chest X-ray images alongside RT-PCR tests, demonstrating encouraging results for computer-aided pathology analysis, particularly when there are limitations on resources and PCR testing.

This study [13] introduces an automatic brain tumor detection method employing CNN with a deeper architecture designed using small kernels and small neuron weights, achieving a high accuracy rate of 97.5%, demonstrating superior performance compared to other cutting-edge methods while maintaining low complexity.

Pugalenthi, R., et al. in this study [14] presents an automated method for distinguishing between cancerous and non-cancerous brain MRI. The approach involves lesion segmentation using various techniques, followed by the selection of a feature set based on shape, texture, and intensity, and employs a SVM classifier, achieving an average accuracy of 97.1% specificity across three benchmark datasets (Harvard, RIDER, and Local), demonstrating superior performance and faster processing compared to existing methods.

This paper [15] focuses on early diagnosis of brain tumors through multi-classification using CNN. Three distinct CNN models are proposed, achieving brain tumor diagnosis with 99.33% accuracy, 92.66% accuracy for classifying five tumor types, and 98.14% accuracy for categorizing tumors into three grades; all important hyper-parameters are optimized using a grid search algorithm. The proposed CNN models, tuned by the grid search optimizer, are compared with other cutting-edge models, demonstrating satisfactory classification results on large clinical datasets and potential utility in assisting physicians and radiologists in brain tumor screening.

This article [16] introduces a new CNN model that can categorize three different kinds of brain cancers, demonstrating simplicity compared to existing pre-trained networks. The network was evaluated using four different approaches, and the highest accuracy of 96.56% was achieved with a record-wise cross-validation method on an augmented dataset, showcasing improved generalization capabilities.

In this article [17], Aamir, Muhammad, et al. presents an automated technique for brain tumor detection using MRI, involving pre-processing of images, extraction of features with two pre-trained DL models, and subsequent classification using a hybrid feature vector generated through partial least squares (PLS) method, achieving a high classification accuracy of 98.95%, outperforming existing approaches.

In this proposed framework [18], transfer learning is employed with various pre-trained deep CNNs to extract features from brain MR pictures, and the top-performing deep features are selected and combined into an ensemble. This ensemble, when utilized with machine learning classifiers, particularly the Support Vector Machine (SVM) utilizing a radial basis function kernel, demonstrates significant performance improvement in brain tumor classification, especially for large datasets, as confirmed by experimental results using three openly accessible brain MRI datasets.

The research in [19] introduces a fully automated technique for segmenting brain tumors utilizing U-Net based deep convolutional networks, which is evaluated on the Multimodal BRATS 2015 datasets comprising 220 high-grade and 54 low-grade tumor cases, demonstrating promising and efficient segmentation through cross-validation.

This study introduces a classification model leveraging DenseNet201, DenseNet169, Xception, MobileNetV2, and ResNet50 architectures. Grad-CAM and Grad-CAM++ techniques are employed to interpret the model's decisions, evaluating its ability to identify significant features in MRI images of brain tumors during the decision-making process.

While the reviewed studies achieve impressive results in various medical imaging tasks, they primarily focus on model accuracy and lack in-depth explanations for the models' decision-making process.

This makes it difficult for diagnosticians to understand why models make such decisions. Table 1 shows the performance comparison of deep learning models for medical imaging tasks.

Our research focused on developing a deep learning model to aid in the early diagnosis of brain tumors using MRI scans. While achieving high accuracy in tumor detection is essential, we aimed to go a step further. By employing Grad-CAM and Grad-CAM++ techniques, we investigated how the model interprets the MRI images during its decision-making process. This allowed us to examine the model's performance to distinguish the critical features within the scans that differentiate brain tumors from healthy tissue. This in-depth analysis provides valuable insights beyond just a classification result. It can not only increase trust in the model's predictions but also potentially guide medical professionals in interpreting the MRI scans themselves.

Table 1. Performance Comparison of DL Models for Medical Imaging Tasks

Paper	Medical Imaging Modality	Task	Model	Key Finding
[9]	Chest X-ray	COVID-19 classification	Deep Transfer Learning (DTL)	Effective for noisy and imbalanced datasets
[10]	Chest CT scans	COVID-19 detection and diagnosis	Pre-trained DenseNet201	Superior performance compared to other methods
[11]	Lung ultrasound (LUS)	COVID-19 screening	CNN with multi-layer fusion	High accuracy (precision, recall, retrieval)
[12]	Chest X-ray	COVID-19 classification	ResNet50 CNN	Promising for resource-limited settings
[13]	Brain MRI	Brain tumor detection	Deep CNN with small kernels and weights	High accuracy, low complexity
[14]	Brain MRI	Brain tumor classification	SVM with various segmentation techniques	High accuracy and speed
[15]	Brain MRI	Brain tumor classification	Three CNN models	High accuracy for diagnosis and grading
[16]	Brain MRI	Brain tumor classification	New CNN model	High accuracy with data augmentation
[17]	Brain MRI	Brain tumor detection	Hybrid feature extraction with PLS	High classification accuracy
[18]	Brain MRI	Brain tumor classification	Ensemble of pre-trained CNN features with SVM	Improved performance for large datasets

3. MATERIALS AND METHODS

3.1. The process of gathering and preparing data

The dataset utilized in this research originates from a publicly accessible MRI dataset [20]. It comprises a total of 4,600 MRI brain images, including 2,513 images indicating cancer and 2,087 images representing healthy conditions. Each MRI image is labeled accordingly, facilitating the supervised learning process. Figure 1 displays sample brain MRI images from this dataset, while Figure 2 visually illustrates the distribution of the dataset, highlighting the proportion of cancerous versus healthy images. We divided this dataset into three different categories using a training, validation, and testing ratio of 80:10:10. Before training the model, the data processing process involves using the ImageDataGenerator class from the Keras library with specified augmentation parameters. The images were rescaled to ensure that pixel values were normalized to the range [0, 1]. Additionally, due to the varying sizes of the data, all images were resized to 224 × 224 pixels. Augmentation techniques such as rotation with a range of 0.2 radians, horizontal and vertical shifts with ranges of 0.05, constant fill mode, horizontal and vertical flips, and a zoom range of 0.2 are applied to enhance the diversity of the training dataset.

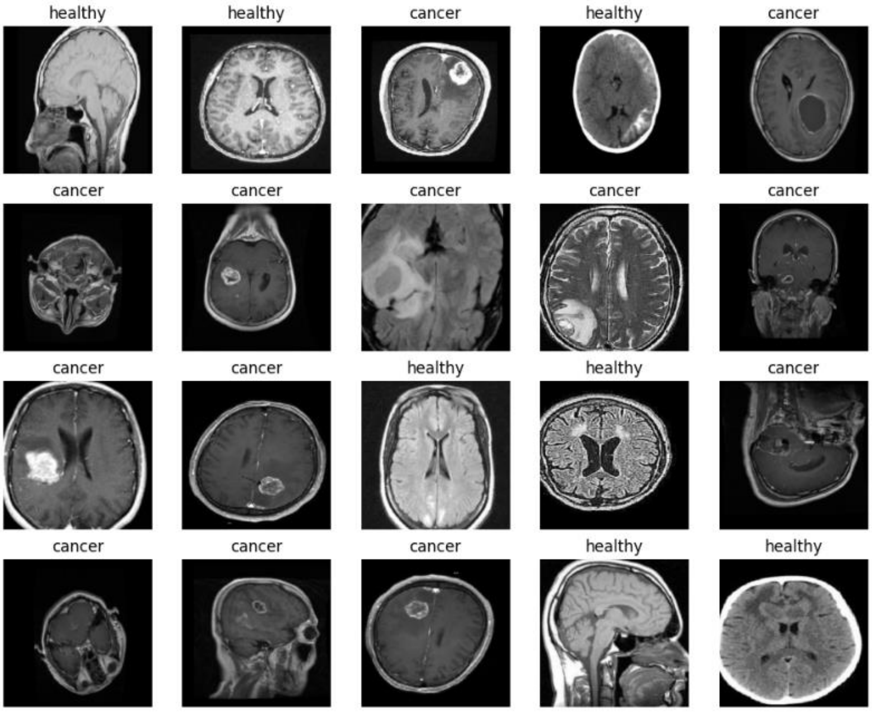

Figure 1. Sample images from the Brain MRI dataset.

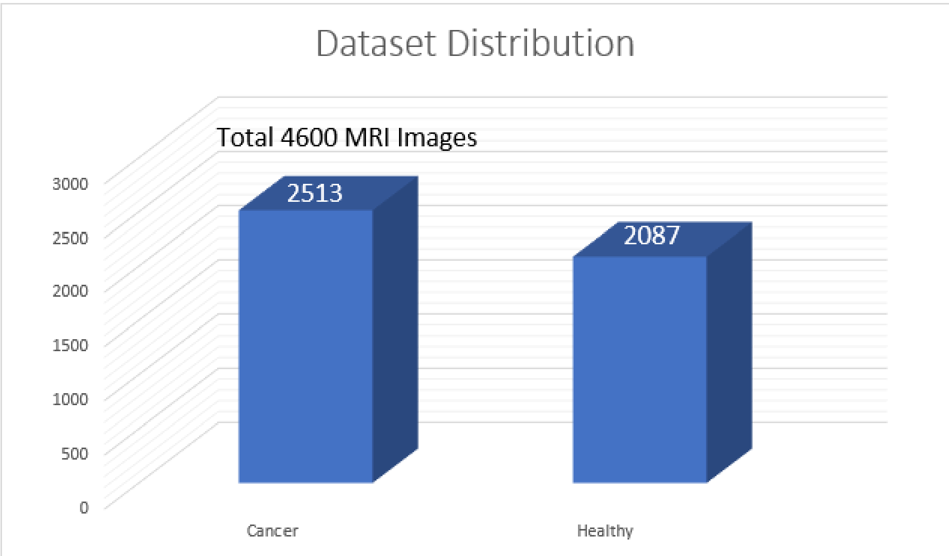

Figure 2. A distribution of datasets.

3.2. Overall Methodology

Figure 3 presents a diagram that illustrates the proposed deep learning methodology. This methodology is used for training a brain tumor classification model.

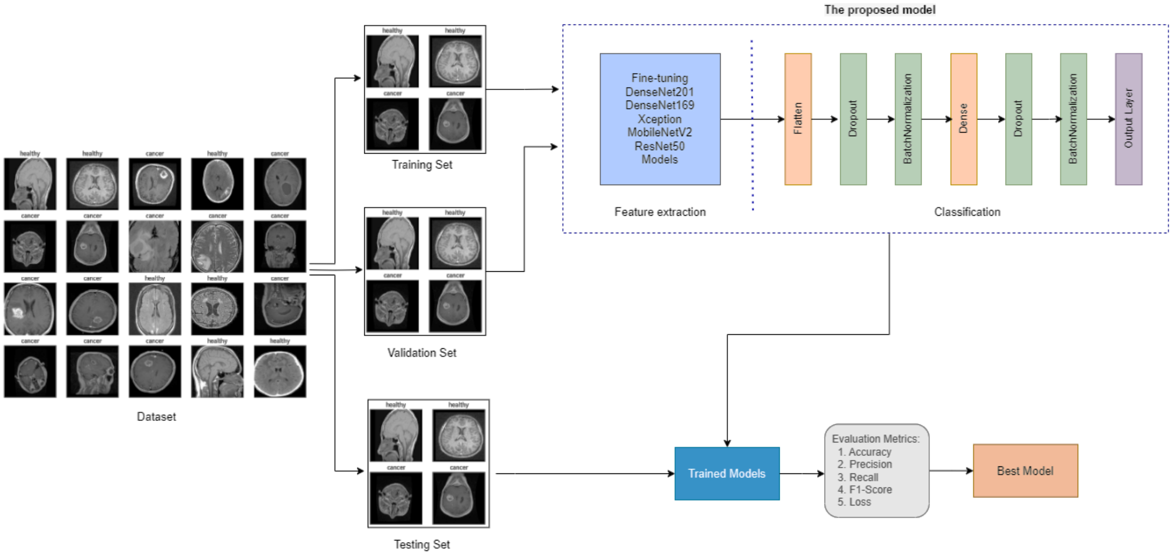

Figure 3. Diagram illustrating the proposed deep learning methodology for training a brain tumor classification model.

The research typically consists of two main Phases:

Phase 1: Train brain tumor classification transfer learning models in MRI images

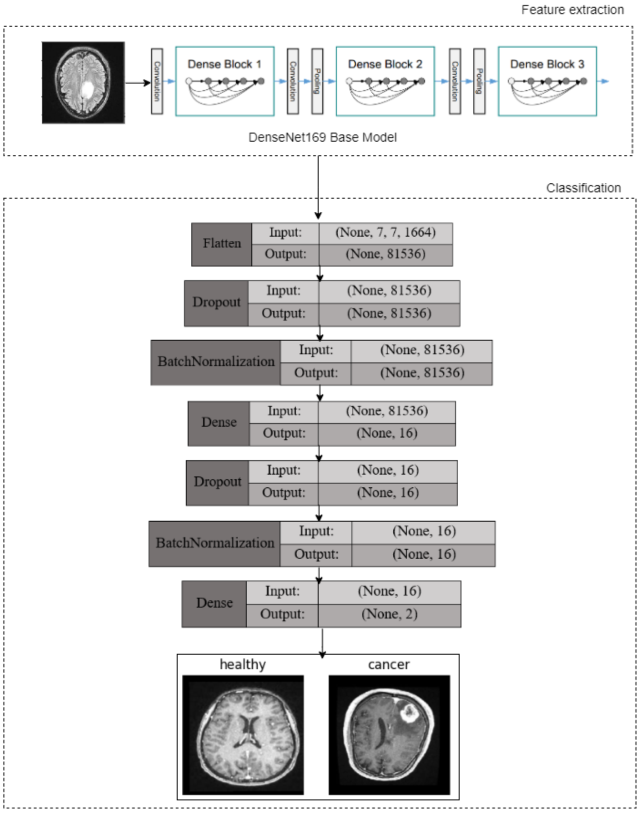

Figure 4. The proposed DenseNet169-based brain tumor classification model.

Transfer learning (TL) is a ML method where a model initially trained for one task is then adjusted to suit a second (related task). In the context of neural networks, TL involves using a pre-trained model, often trained on a large dataset for a particular task, as a starting point for a new model on a different but related task. Instead of training a new model from scratch, TL allows us to leverage these learned features and fine-tune the model on a smaller dataset for a specific task. For the classification model, five pre-trained models have been used.

Pre-trained models include DenseNet201 [21], DenseNet169 [21], Xception [22], MobileNetV2 [23] and ResNet50 [24]. Figure 3 presents a diagram illustrating the proposed deep learning methodology for training a brain tumor classification model. Figure 4 provides an overview of the brain tumor classification model suggested in the study.

The classification model suggested in this paper consists of:

- Using the architecture of pre-trained models as their base model, initialized with pre-trained weights from ImageNet and configured to exclude the top layer while setting the input shape to (224, 224, 3).
- The output of the last layer of the base model is then passed through a flattening layer.
- Followed by a dropout layer and subsequent batch normalization.
- A dense layer with ReLU activation is added, followed by another dropout layer and additional batch normalization.
- The final layer consists of two neurons with softmax activation, suitable for binary classification tasks.

Phase 2: Using Grad-CAM and Grad-CAM++ to Observe and Interpret Model Predictions

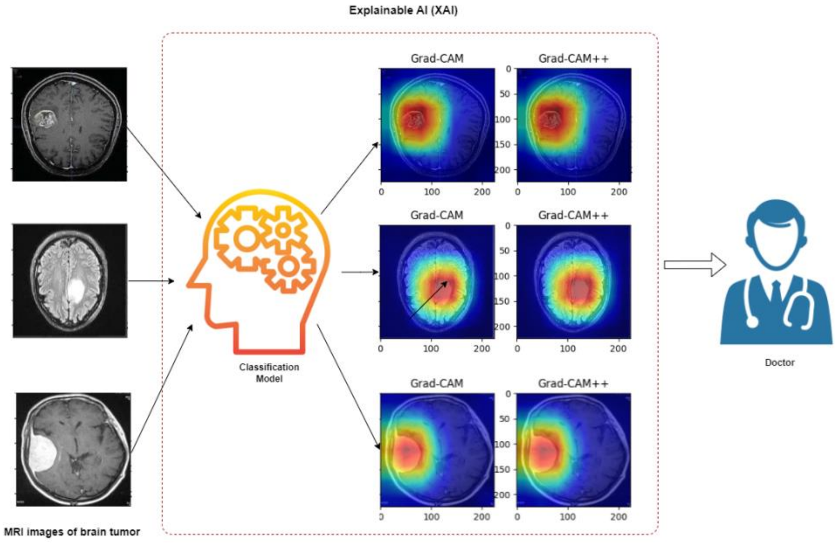

Figure 5. Classification process and explain the model's results

In addition to proposing brain tumor classification models on MRI images, the purpose of this research is to highlight the image regions that are crucial and significantly affect the predictions of the model.

In this research, we used Grad-CAM and Grad-CAM++ approaches to achieve this. These techniques allow for the observing of critical image regions that are very significant in the decision-making procedure of the model, offering insights into the areas that the model concentrates on when making predictions.

In particular, Grad-CAM generates a "heatmap" that emphasizes important areas in the MRI picture. This heatmap clearly indicates the areas of the picture that the model is interested in during the prediction phase. High values on this heatmap often indicate important areas that are valuable to the task of classifying and making decisions.

The study team also employed Grad-CAM++ as an additional technique to provide a comparative analysis with Grad-CAM. This method intends to provide a deeper understanding of the specific visual components that attract the model's attention during the prediction phase. The classification process and an explanation of the model's results are displayed in Figure 5.

3.4. Performance Evaluation Measures

In the realm of machine learning and data science research, evaluating model performance is crucial. Several metrics serve as yardsticks to assess the effectiveness of models. Accuracy measures the proportion of correctly predicted instances out of the total. Precision quantifies the true positive rate among the predicted positive instances, while Recall (also known as sensitivity) gauges the true positive rate among actual positive instances. The f1-score, which balances precision and recall, provides a comprehensive view of model performance. Additionally, loss functions play a pivotal role during model training, guiding optimization by quantifying the discrepancy between predicted and actual values.

In which, FP represent False Positive (incorrectly identifies a tumor when none is present), TN denote True Negative (correctly identifies absence of a tumor), TP signify True Positive (correctly detects a tumor), and FN indicate False Negative (Fails to detect a tumor that is present.). The variable k represents the number of classes, while y corresponds to the actual value, ŷ is prediction value.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recal = \frac{TP}{TP + FN} \quad (5)$$

$$F_1 - Score = \frac{Precision * Recall}{Precision + Recall} \quad (6)$$

$$Loss = -\sum_{j=1}^{k} y_i \log(\hat{y}_i) \quad (7)$$

4. RESULTS AND DISCUSSION

4.1. Environmental settings

The experiments were carried out on the Kaggle platform to obtain results. A Tesla T4x2 GPU with 30GB of memory was utilized, and the system had 29GB of RAM.

4.2. Experiments

4.2.1 Experiments 01: Train brain tumor classification transfer learning models in MRI images.

Table 2. Hyperparameters Of The Proposed Deep Learning Methodology For Training A Brain Tumor Classification Model

Hyperparameters	**Value**
Batch Size	32
Number of Epochs	100
Optimizer	Adam
Loss Function	Binary Crossentropy
Activation Function in Hidden Layer	Relu
Activation Function in Output Layer	Softmax
EarlyStopping Monitor	Val Loss
EarlyStopping Patience	10
EarlyStopping Mode	Min
Learning Rate	0.0001

In this experience, we train TL models for brain tumor classification using MRI images. The hyperparameters of the proposed DL methodology for training a brain tumor classification model are displayed in the Table 2. The results of our experiments are summarized in Table 3.

Table 3. Table of results for fine-tuning models on the testing set for brain tumor classification using MRI images.

Model	Loss	Precision(%)	Recall(%)	F1-Score(%)	Accuracy(%)	Model size mb
DenseNet201	0.0325	99.13	99.13	99.13	99.13	97
DenseNet169	**0.0160**	**99.67**	**99.67**	**99.67**	**99.67**	171
Xception	0.0248	99.46	99.46	99.46	99.46	273
MobileNetV2	0.0832	98.04	98.04	98.04	98.04	41
ResNet50	0.0365	99.24	99.24	99.24	99.24	306

The table 3 presents the performance outcomes of various classification models, encompassing metrics such as loss, precision, recall, F1-score, accuracy, and model size in megabytes. Notably, each model, including DenseNet201, DenseNet169, Xception, MobileNetV2, and ResNet50, achieves exceptional precision, recall, and F1-score, reflecting their robust classification capabilities. The accuracy values for all models range from 98.04% to 99.67%, demonstrating consistently high performance. Among these models, DenseNet169 stands out with the highest accuracy of 99.67%. The accuracy of the model in this study demonstrates its effectiveness in correctly classifying unseen data instances. It is worth noting that, while accuracy is a crucial measure, other factors such as model size should also be considered when selecting the most suitable model for a given application. The accuracy and loss during training and validation for the classification model based on DenseNet169 are visualized in Figure 6 and Figure 7. Comparison results for fine-tuning models on the testing set for brain tumor classification using MRI images are shown in Figure 8.

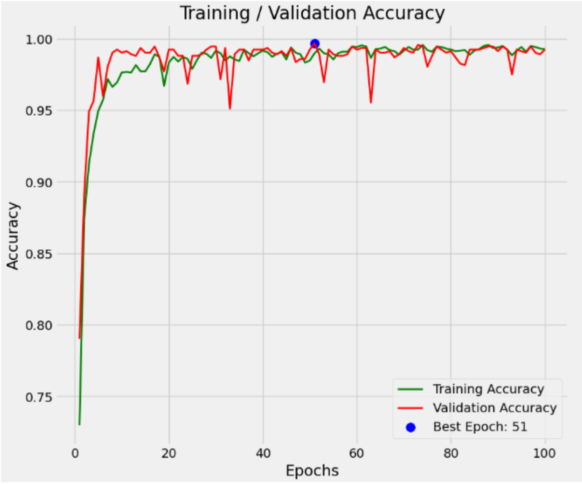

Figure. 6. The training and validation accuracy of the DenseNet169-based classification model

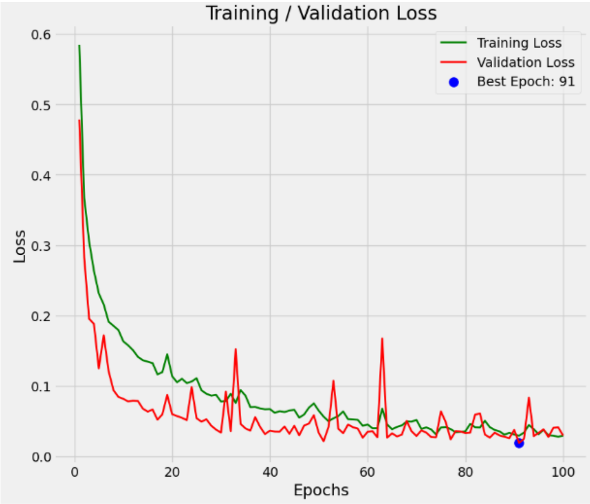

Figure. 7. The training and validation loss of the DenseNet169-based classification model

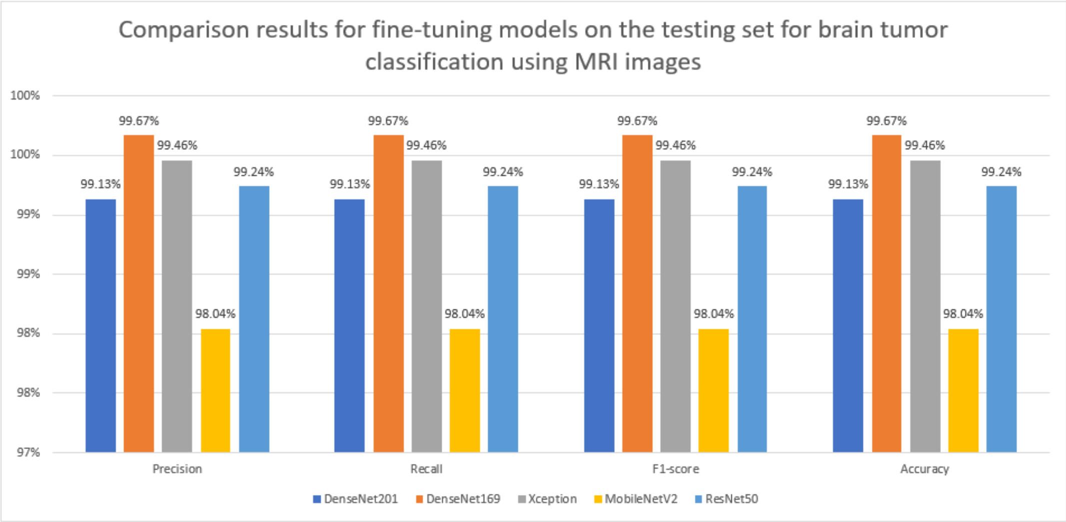

Figure 8. Comparison results for fine-tuning models on the testing set

4.2.2 Experiments 02: Using Grad-CAM and Grad-CAM++ to Observe and Interpret Model Prediction

To re-evaluate the accuracy and reliability of brain tumor MRI image classification models examined in phase 1. The research team employed Grad-CAM and Grad-CAM++ techniques on original images to assess whether the models determining the presence of tumors (cancer) truly focus on features within the image area containing tumors. Based on the results presented in Figure 9, all models correctly determine the presence of a tumor. Both Grad-CAM and Grad-CAM++ demonstrate accurate focus on the image area containing the tumor.

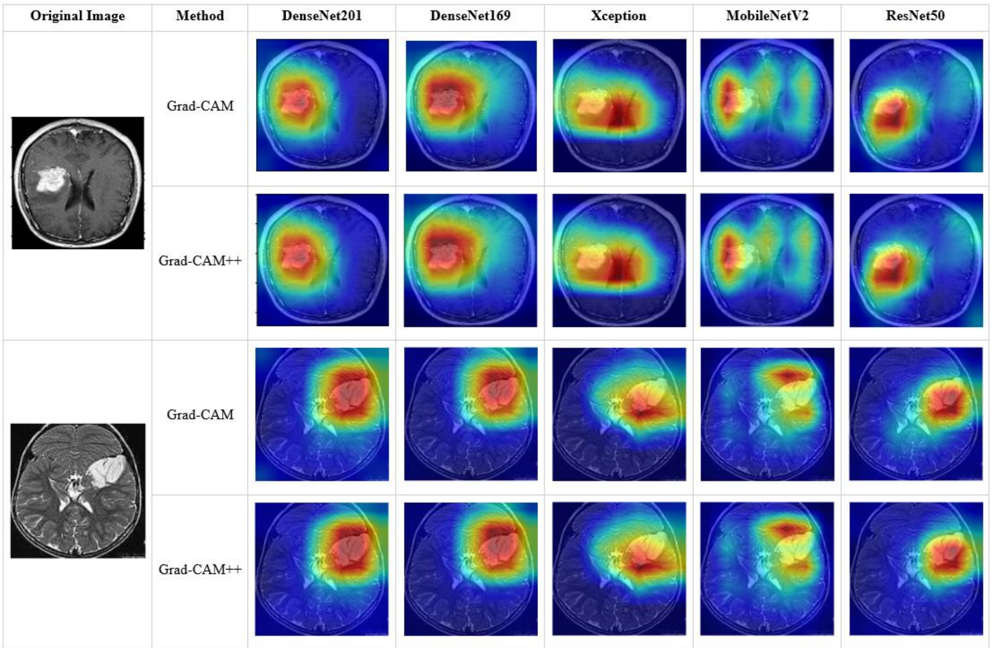

Figure 9. Brain tumor examples were classified using five different models in this study and explained using Grad-CAM and Grad-CAM++

However, the experimental results presented in Figure 10 on other MRI images containing tumors, DenseNet201 and DenseNet169 models provide highly accurate Grad-CAM and Grad-CAM++ results, precisely focusing on the area containing the tumor in the image. While Xception and ResNet50 in phase 1 have relatively high accuracy, the Grad-CAM and Grad-CAM++ results do not entirely concentrate on the region containing the tumor in the image. Notably, MobileNetV2 in phase 1 gave the lowest results among the models, with Grad-CAM and Grad-CAM++ results similarly reflecting this experiment.

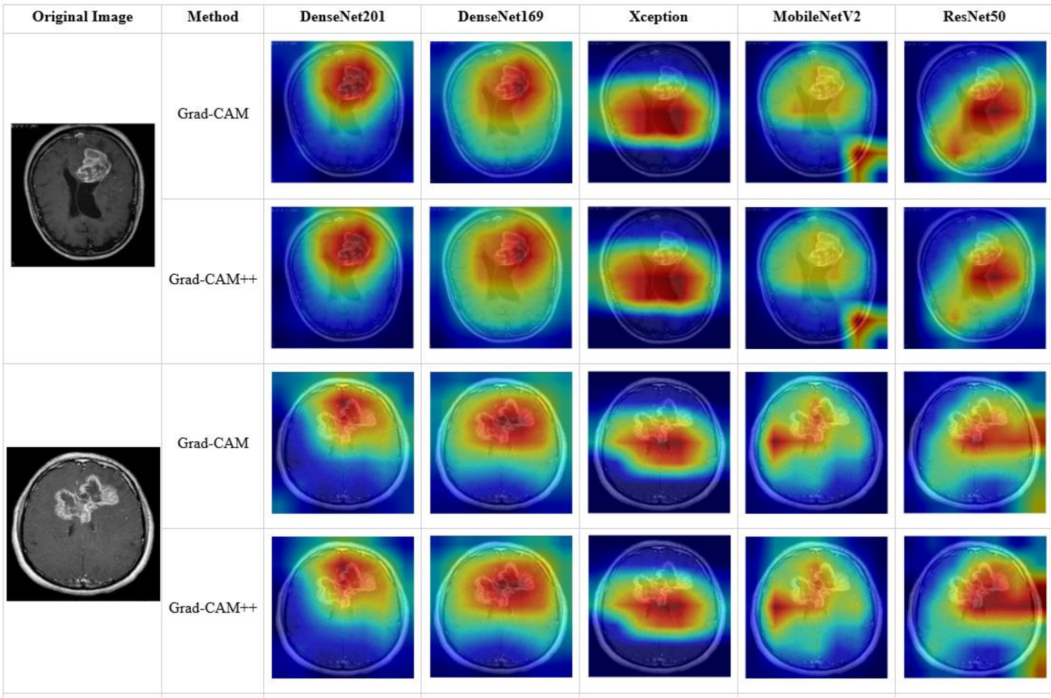

Figure 10. Brain tumor examples were classified using five different models in this study and explained using Grad-CAM and Grad-CAM++

Figure 11 presents samples of Brain Tumors classified using the DenseNet169-based model and explained using Grad-CAM and Grad-CAM++.

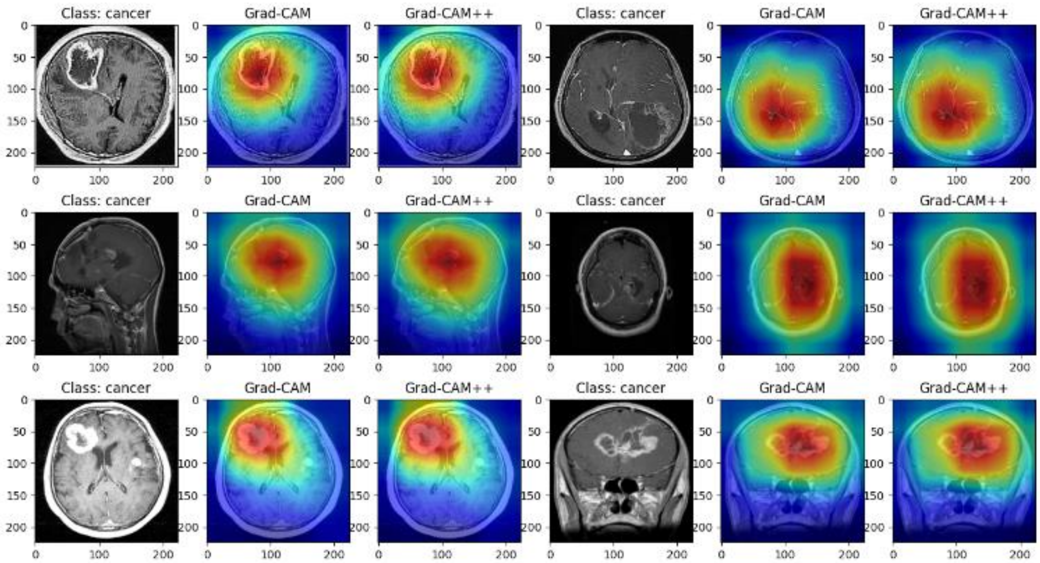

Figure 11. Examples of Brain Tumors classified using the DenseNet169-based model and explained using Grad-CAM and Grad-CAM++

Table 4 presents a comparison of accuracy and interpretability across several studies on the brain tumor classification problem. The table includes five previously published studies, referenced as [13], [14], [15], [16], and [18], along with the proposed model. The accuracy of these models ranges from 94.00% to 99.33%, with the proposed model achieving the highest accuracy of 99.67%. Furthermore, the proposed model incorporates Grad-CAM and Grad-CAM++ methods to enhance interpretability. This demonstrates that the proposed model outperforms the others, not only in terms of accuracy but also by offering better interpretability through the use of these techniques.

Table 4. Table of Comparison of Studies for Brain Tumor Classification with Proposed Model Performance.

Model	Accuracy(%)	Grad-CAM	Grad-CAM++
[13]	97.50		
[14]	94.00		
[15]	99.33		
[16]	96.56		
[18]	98.67		
Our Proposed	**99.67**	✓	✓

5. CONCLUSION AND FUTURE WORK

In conclusion, our two-phase experiment aimed to advance the understanding and application of brain tumor classification models using MRI images. In Experiment 1, we successfully trained and evaluated transfer learning models, including DenseNet201, DenseNet169, Xception, MobileNetV2, and ResNet50, for brain tumor classification. The comprehensive performance analysis, detailed in Table II, highlighted exceptional precision, recall, and F1-score across all models, with DenseNet169 leading in accuracy at 99.67%.

Additionally, Experiment 2 employed Grad-CAM and Grad-CAM++ techniques to visualize and interpret model predictions, aiming to reevaluate the accuracy and reliability of brain tumor MRI image classification models examined in phase 1. While all models demonstrated accurate tumor detection, DenseNet201 and DenseNet169 models exhibited highly accurate Grad-CAM and Grad-CAM++ results. Notably, the finding in this work offer valuable insights into the interpretability of the models, providing an understanding of how models make decisions and contributing to the advancement of brain tumor classification methodologies.

The main contributions of this research include: Development of a robust classification model for brain tumor detection using state-of-the-art deep learning architectures. Integration of interpretability techniques to enhance the transparency and understanding of model predictions. Demonstration of the potential deep learning in advancing medical diagnostics and improving patient outcomes.

Future research in the domain of brain tumor classification using MRI images should aim to enhance the interpretability of models beyond the employed Grad-CAM and Grad-CAM++ techniques. Investigating novel methods for interpreting model decisions could provide more comprehensive insights into the complexity of brain tumor classification.

REFERENCES

[1] Cancer, available online: https://www.who.int/news-room/fact-sheets/detail/cancer.

[2] Sarah Ali Abdelaziz Ismael, Ammar Mohammed, and Hesham Hefny. An enhanced deep learning approach for brain cancer mri images classification using residual networks. *Artificial intelligence in medicine*, 102:101779, 2020.

[3] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31, 2017.

[4] Maqsood, Sarmad, Robertas Damaševičius, and Rytis Maskeliūnas. Multi-modal brain tumor detection using deep neural network and multiclass SVM. *Medicina* 58.8 (2022): 1090.

[5] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging*, 30:449–459, 2017.

[6] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. Applied logistic regression, volume 398. *John Wiley & Sons*, 2013.

[7] Yacine Izza, Alexey Ignatiev, and Joao Marques-Silva. On explaining decision trees. arXiv preprint arXiv:2010.11034, 2020.

[8] Leif E Peterson. K-nearest neighbor. Scholarpedia, 4(2):1883, 2009.

[9] Yadunath Pathak, Prashant Kumar Shukla, Akhilesh Tiwari, Shalini Stalin, and Saurabh Singh. Deep transfer learning based classification model for covid-19 disease. *Irbm*, 43(2):87–92, 2022.

[10] Aayush Jaiswal, Neha Gianchandani, Dilbag Singh, Vijay Kumar, and Manjit Kaur. Classification of the covid-19 infected patients using densenet201 based deep transfer learning. *Journal of Biomolecular Structure and Dynamics*, 39(15):5682–5689, 2021.

[11] Ghulam Muhammad and M Shamim Hossain. Covid-19 and non-covid-19 classification using multi-layers fusion from lung ultrasound images. *Information Fusion*, 72:80–88, 2021.

[12] Zehra Karhan and AKAL Fuat. Covid-19 classification using deep learning in chest x-ray images. *In 2020 Medical Technologies Congress (TIPTEKNO)*, pages 1–4. IEEE, 2020.

[13] J Seetha and S Selvakumar Raja. Brain tumor classification using convolutional neural networks. *Biomedical & Pharmacology Journal*, 11(3):1457, 2018.

[14] R Pugalenthi, MP Rajakumar, J Ramya, and V Rajinikanth. Evaluation and classification of the brain tumor mri using machine learning technique. *Journal of Control Engineering and Applied Informatics*, 21(4):12–21, 2019.

[15] Emrah Irmak. Multi-classification of brain tumor mri images using deep convolutional neural network with fully optimized framework. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, 45(3):1015–1036, 2021.

[16] Milica M Badza and Marko C Barjaktarovic. Classification of brain tumors from mri images using a convolutional neural network. *Applied Sciences*, 10(6):1999, 2020.

[17] Muhammad Aamir, Ziaur Rahman, Zaheer Ahmed Dayo, Waheed Ahmed Abro, M Irfan Uddin, Inayat Khan, Ali Shariq Imran, Zafar Ali, Muhammad Ishfaq, Yurong Guan, et al. A deep learning approach for brain tumor classification using mri images. *Computers and Electrical Engineering*, 101:108105, 2022.

[18] Jaeyong Kang, Zahid Ullah, and Jeonghwan Gwak. Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers. *Sensors*, 21(6):2222, 2021.

[19] Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. *In Medical Image Understanding and Analysis: 21st Annual Conference*, MIUA 2017, Edinburgh, UK, July 11–13, 2017, Proceedings 21, pages 506–517. Springer, 2017.

[20] Brian tumor dataset, available online: https://www.kaggle.com/datasets/preetviradiya/brian-tumor-dataset.

[21] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[22] Francois Chollet. Xception: Deep learning with depthwise separable convolutions. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[23] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

BIOGRAPHY OF AUTHORS

HoangTu-Vo holds a Bachelor of Information Systems from Can Tho University, Vietnam, in 2011. In 2013, He graduated with a master's degree in Information Systems from Can Tho University in Vietnam. Currently, He is working as a lecturer in Information Technology Department at FPT University, Can Tho campus in Vietnam. His research interests include Machine learning, Deep learning, Image processing, and Computer vision. He can be contacted at email: tuvh6@fe.edu.vn

Nhon Nguyen Thien holds a Bachelor of Information Systems from Can Tho University, Vietnam, in 2013. In 2017, He graduated with a master's degree in Information Systems from Can Tho University in Vietnam. Currently, He is working as a lecturer in Information Technology Department at FPT University, Can Tho campus in Vietnam. His research areas of interest include Data science, Machine learning, Deep learning, and Web application development. He can be contacted at email: nhonnt9@fe.edu.vn

Kheo Chau Mui holds a Bachelor of Information Systems from Can Tho University, Vietnam, in 2010. In 2020, She graduated with a master's degree in Computer Science from Can Tho University in Vietnam. Currently, She is working as a lecturer in Information Technology Department at FPT University, Can Tho campus in Vietnam. Her research areas of interest include Image processing, Image classification, Data science, Object detection, Deep learning, and Machine learning. She can be contacted at email: kheocm@fe.edu.vn

Phuc Pham Tien earned his Bachelor's degree in Information Systems from Can Tho University, Vietnam, in 2003. In 2010, he completed a Master's degree in Information Technology at the same university. Currently, he is a lecturer in Information Technology Department at FPT University, Can Tho campus, Vietnam. His research interests include machine learning and deep learning. He can be reached at phucpt10@fe.edu.vn