International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 6, December 2024, pp. 6360~6372
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i6.pp6360-6372

Deep learning for magnetic resonance imaging brain tumor detection: evaluating ResNet, EfficientNet, and VGG-19

Fatima Muftic[1], Merjem Kadunic[1], Almina Musinbegovic[1], Ali Abd Almisreb[1,2], Hajar Jaafar[3]
[1]Faculty of Engineering and Natural Sciences, International University of Sarajevo, Sarajevo, Bosnia and Herzegovina
[2]Department of Cybersecurity, International Information Technology University, Almaty, Kazakhstan
[3]School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA Terengganu Branch, Terengganu, Malaysia

Article Info

Article history:

Received May 2, 2024
Revised Jul 25, 2024
Accepted Aug 6, 2024

Keywords:

Brain tumor
Convolutional neural networks
EfficientNetB3
ResNet50
VGG-19

ABSTRACT

This paper investigates the application of convolutional neural networks (CNNs) for the early detection of brain tumors to enhance diagnostic accuracy. Brain tumors present a significant global health challenge, and early detection is vital for successful treatments and patient outcomes. The study includes a comprehensive literature review of recent advancements in brain tumor detection techniques. The main focus is on the development and evaluation of CNN models, including EfficientNetB3, residual networks-50 (ResNet50) and visual geometry group-19 (VGG-19), for binary image classification using magnetic resonance imaging (MRI) scans. These models demonstrate promising results in terms of accuracy, precision, and recall metrics. However, challenges related to overfitting and limited dataset size are acknowledged. The study highlights the potential of artificial intelligence (AI) in improving brain tumor detection and emphasizes the need for further research and validation in real-world clinical settings. EfficientNetB3 reached 99.44% training accuracy but showed potential overfitting with validation accuracy dropping to 89.47%. ResNet50 steadily improved to 83.62% training accuracy and 89.47% validation accuracy. VGG19 struggled, with only 62% accuracy.

Corresponding Author:

Hajar Jaafar
School of Electrical Engineering, Coolege of Engineering, Universiti Teknologi MARA
Terengganu Branch, 23000, Dungun, Terengganu, Malaysia
Email: hajarj422@uitm.edu.my

1. INTRODUCTION

The human brain is a command center and an essential organ of the human nervous system responsible for accomplishing daily life activities. The brain collects stimuli or signals from the body's sensory organs, handles processing, and directs the ultimate decisions and output information to the muscles. A brain tumor is one of the most severe situations related to the human brain, where a group of abnormal brain cells grow in an undisciplined manner [1]. They can be divided into two main types: primary and secondary metastatic. The primary brain tumors are generally non-cancerous and originate from human brain cells. In contrast, secondary metastatic tumors spread to the brain with blood flow from other body parts [2].

Various advancements in the field of computer-aided diagnosis of brain tumors have been developed during the previous decade. These approaches are always available to assist radiologists who are unsure about the type of tumor they are looking at or wish to investigate further. To identify tumors, doctors utilize magnetic resonance imaging (MRI) and computed tomography scan (CT-scan) [3].

Artificial intelligence (AI) techniques play a significant and influential role in the medical domain by developing complex algorithms that can automatically analyze and interpret large amounts of medical data. This data includes various types of medical imaging, not only MRI and CT-scan, but X-ray, ultrasound, and other types of medical data, such as genomics and electronic health records. The most famous of these techniques is deep learning [4]. Deep learning algorithms extract features from medical datasets, learn new patterns, help diagnose diseases, and help physicians and healthcare specialists make the right and accurate decisions [5].

Early detection of brain tumors can play an indispensable role in improving the treatment possibilities, and a higher gain of survival possibility can be accomplished [6]. The main aim of this work is to contribute to the understanding and advancement of brain tumor detection and classification using machine learning techniques. By exploring the background and characteristics of brain tumors, along with the principles of convolutional neural networks (CNNs) and transfer learning, this research aims to provide valuable insights into the application of these technologies in the medical field.

This research underscores the potential of artificial intelligence in enhancing the early detection of brain tumors, which is crucial for successful treatment and patient outcomes. By employing CNNs like EfficientNetB3, residual networks-50 (ResNet50), and visual geometry group-19 (VGG-19), the study demonstrates significant improvements in diagnostic accuracy using MRI scans. Despite challenges such as overfitting and limited dataset sizes, the promising results highlight the importance of further research and validation. This study paves the way for more reliable AI applications in healthcare, aiming to improve diagnostic techniques and ultimately patient care.

In the upcoming section, the background of brain tumors will be delved into, examining its various types, potential causes, and available treatment approaches. Additionally, the fascinating realm of convolutional neural networks and transfer learning will be explored, unraveling its underlying principles and showcasing its diverse applications in different domains. In section 3, the methodology employed will be dug to compare EfficientNetB3, ResNet50, and VGG19 models, covering data preprocessing, training procedures, and the selection of validation data, along with the metrics used to evaluate model performance. Section 4 will present the results obtained from each model, represented through informative graphs displaying performance metrics. As section 5 progresses, an overarching discussion about the results, analyzing various factors that may have influenced the outcomes and identifying areas for future analysis and improvement will be engaged in. Finally, a comprehensive conclusion will be provided, summarizing the research conducted in this study and offering valuable suggestions for future works and research endeavors.

2. BACKGROUND STUDY

2.1. What is a brain tumor?

Brain tumors are a heterogenous group of common intracranial tumors that cause significant mortality and morbidity. Malignant brain tumors are among the most aggressive and deadly neoplasms in people of all ages, with mortality rates of 5.4/100,000 men and 3.6/100,000 women per year being reported between 2014 and 2018 [7]. According to the 2021 World Health Organization (WHO) classification of tumors of the central nervous system, brain tumors are classified into four grades (I to IV) of increasingly aggressive malignancy and worsening prognosis. Indeed, in clinical practice, tumor type and grade influence treatment choices. Within WHO Grade IV tumors, glioblastoma is the most aggressive primary brain tumor, with a median survival after diagnosis of just 12–15 months [8]. The title, low-grade tumor (I and II), is appropriate for tumors with excellent and good prognosis in contrast to high-grade tumors (III and IV), which tend to be malignant, thereby leading to serious complications. Primary brain tumor categories owing to the cell type from which they originate are gliomas; having their association with glial cells, meningiomas; abnormal growth of meninges, ependymomas; originating from cells (*ependymocytes*) lining the cerebrospinal fluid (CSF)-filled ventricles, astrocytoma; developing from star-shaped glial cells (*astrocytes*) and so on [9].

Diagnosis usually involves imaging of the brain, such as a magnetic resonance imaging (MRI) or computed tomography (CT) scan, and a biopsy, which involves taking a sample of the affected tissue for analysis. Treatments for brain tumors may include surgical removal of the tumor, radiation therapy, chemotherapy, or targeted drug therapy [10]. The conventional method of detecting brain tumors includes a doctor or a radiologist examining MR image for anomalies and making decisions. It is strongly dependent on a doctor's medical expertise; disparities in experience levels and the nature of images create extra complexity for diagnosing with naked human eyes [11].

It is challenging for a doctor to interpret these images in a limited period since it normally comes with several abnormalities or noisy data. As the volume of information increases, assessing a massive amount of information gets even more challenging. The manual detection of a brain tumor becomes more time-consuming and costly. Artificial intelligence plays an essential role in identifying and diagnosing brain tumors. The discipline of brain tumor surgery is an excellent choice for additional AI integration due to its complicated and elaborated processes. Deep learning, particularly neural networks, gains substantial importance when it obtains promising results. CNNs in particular are remarkable for learning features and providing unlimited precision [12].

2.2. Convolution neural networks

CNNs are a special type of neural networks developed specifically for image processing and analysis. This network architecture has proven to be extremely successful in areas such as object recognition, image classification, and image segmentation. CNN consists of convolution layers, which filter images to extract features and shapes. These layers use convolutional kernels, which are small weight matrices, to apply various filtering operations to the input images. Convolution layers are followed by subsampling (pooling) layers, which reduce the dimensionality of the data while retaining important features. This is followed by densely connected layers, which are used for classification and decision-making based on learned features [13]. CNN is different in terms of the number of layers employed, the size and number of images, as well as the type of activation functions employed. In the CNNs, the parameters are chosen experimentally based on trial and error. In other words, each CNN consists of several layers, the main layers of which are the Convolutional layer and the Sub-sampling layer [14]. Recently, supervised deep CNN have attracted lots of interest. Compared to conventional supervised machine learning methods, these deep learning-based methods are not dependent on hand-crafted features but automatically learn a hierarchy of increasingly complex features directly from data [15].

CNNs are the most widely used machine learning algorithm for visual learning and brain tumor recognition. Brain tumor classification is possible with the help of the fully automated CNN model to make fast and accurate decisions [16]. Through training and optimization of the CNN model, it is possible to teach the network to recognize the characteristic features of tumors and distinguish them from normal brain tissue. There are many different CNN models, such as AlexNet, GoogleNet, VGGNet, EfficientNet, and DenseNet [17]. During this research, EfficientNet, VGG, and ResNet would be implemented and compared to each other to determine the most accurate model corresponding to the given data from brain MRI images for brain tumor detection dataset.

2.3. Transfer learning

To achieve satisfactory results, the application of deeply pre-trained convolutional neural networks based on transfer learning in medical imaging requires appropriate adjustment of both hyper-parameters and learning parameters of the models. Training the networks with transfer learning is usually much faster and easier than training the networks with randomly initialized weights [18]. The essence of transfer learning can support interdisciplinary research to get better performance. This paper presents a comprehensive review of research focusing on transfer learning (TL) in brain tumor detection. TL is applied to build up more complicated and accurate convolutional neural networks to cope with problems like data scarcity or computational inefficiency, which is the subfield of deep learning (DL). Due to data privacy or lack of data, transfer learning has been widely employed in the medical field to help improve the performance of learning [19].

Several relevant papers [20], [21] focusing on one model or multiple models' comparison with TL are presented. This paper referred to some reviews about the application of TL and other tumor detection tasks but mainly focused on evaluating different models classifying MR images. Deep learning models like ResNet, EfficientNet, and VGG19 have shown great promises in various domains, including medical imaging. Transfer learning complements these models by leveraging their pre-trained knowledge and adapting them to specific tasks, such as MRI brain tumor detection. This approach enables accurate and efficient tumor detection even with limited labeled data, making it a valuable technique in medical applications.

3. METHOD

3.1. Data preprocessing

In the context of this research study, the main aim was to leverage artificial intelligence for the detection of brain tumors using magnetic resonance imaging (MRI) scans. The dataset, sourced from Kaggle, consisted of two categories of brain MRI images: 155 images labeled 'yes', indicating the presence of a brain tumor, and 98 images labeled 'no', indicating the absence of a brain tumor. Example images are shown in Figures 1 and 2.

Given the heterogeneity in the size of the original images, the initial step in data preprocessing was to investigate the dimensions of the images. Dimensions are represented in Figure 3. Statistical metrics, such as mean and median, were employed to better understand the distribution of image dimensions in both categories. Consequently, all images were resized to a uniform dimension of 300×300 pixels for consistency, ensuring the AI model's input layer receives data of uniform size.

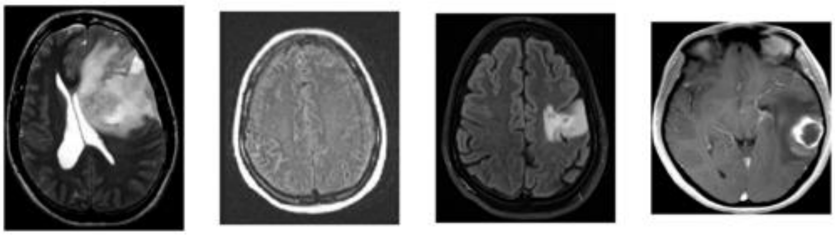

Figure 1. The images included in this figure are labeled ‘yes’ and depict the presence of a brain rumor

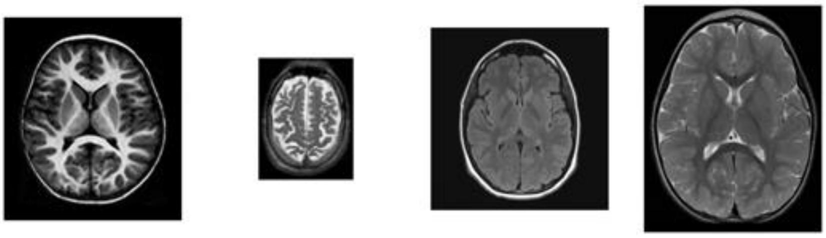

Figure 2. The images included in this figure are labeled ‘no’ and depict the absence of a brain rumor

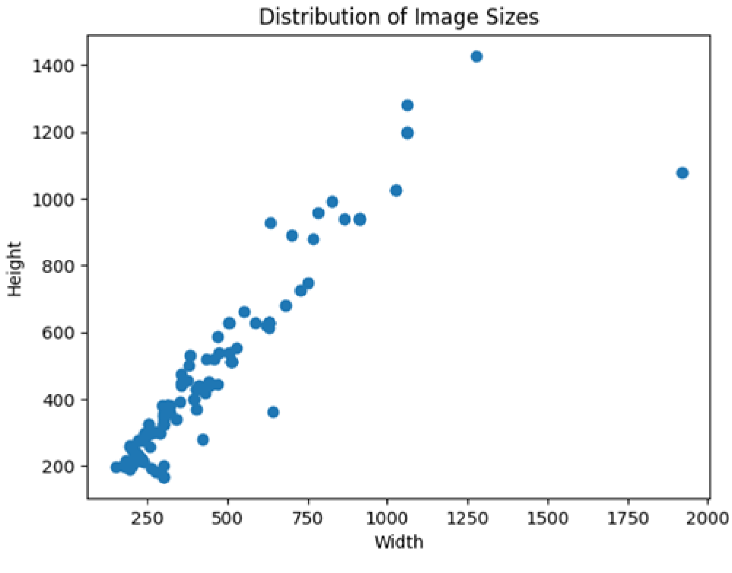

Figure 3. Distribution of image sizes in dataset

Furthermore, to ensure the robustness of the model in handling images of different orientations and positions, data augmentation techniques were applied. These included random rotations, shifts, zooms, and horizontal flips. A sample of augmented images from each category was examined visually to verify the appropriateness of transformations applied. The pixel intensities in the images were then converted from an integer format to a floating-point format and normalized them to a range [0,1]. Normalization is a common practice in deep learning for better convergence of the model.

In preparation for the training phase, the dataset was partitioned into training (70%), validation (15%), and test (15%) sets. This enables optimal parameter learning during training, hyperparameter tuning during validation, and an unbiased evaluation of the model's performance during testing. Given the class imbalance observed in the dataset, class weights were computed. The use of class weights ensures that the model does not display a bias towards the majority classes, thus aiding in improved predictive performance. To summarize, the data preprocessing phase was critical in preparing the dataset for the subsequent phases of model development and evaluation. The approaches adopted were aimed at ensuring model robustness, performance, and fairness, thus aligning with the overarching goal of the study-the effective detection of brain tumors through artificial intelligence.

3.2. EfficientNetB3 model

This research utilizes the transfer learning technique to build a binary classification model using the pre-trained EfficientNetB3 [22], [23] model. This methodology section details the step-by-step process of developing the model. We begin with model setup, model training and evaluation and lastly model improvement.

3.2.1. Model setup

We employed the EfficientNetB3 model pretrained on ImageNet dataset and utilized its transfer learning capabilities. The model's architecture was modified to suit the binary classification task at hand. A GlobalAveragePooling2D layer was added to reduce the spatial dimensions of the output from the EfficientNetB3.

Then, a Dense layer with 1,024 neurons and rectified linear unit (ReLU) activation was added to help the model learn more complex patterns. Finally, a single neuron with sigmoid activation was added to classify the inputs into the two categories. The layers of the base model were frozen to preserve the pre-trained weights. Then, the model was compiled using Adam optimizer and binary cross entropy as the loss function since this is a binary classification problem.

3.2.2. Model training and evaluation

The model was trained on training data for 10 epochs with a batch size of 32. The training process was monitored using accuracy and loss on the validation set. After each epoch, the model's accuracy and loss for the training and validation set were plotted to visually monitor the model's performance.

3.2.3. Model improvement

To enhance the performance of the model, early stopping and learning rate reduction were implemented on the plateau. Early stopping prevents overfitting by halting the training process if the model's performance on the validation set does not improve for a certain number of epochs. On the other hand, ReduceLROnPlateau reduces the learning rate when the validation loss does not improve, which helps the model to fine-tune and reach a better minimum.

Furthermore, additional metrics-precision and recall-were incorporated into the model for better understanding of its performance. The improved model was trained again for 50 epochs with the same batch size. Its performance was again evaluated using the same evaluation criteria as before.

3.3. ResNet50 model

ResNet50 [24] is a deep neural network with 50 layers, trained on over a million images from the ImageNet database. It utilizes “skip connections” or “shortcuts” to address the vanishing gradient problem, which often hinders the training of very deep networks. These skip connections allow gradients to flow more easily through the network during backpropagation, facilitating the training of deeper models. By incorporating these shortcuts, ResNet50 enables the effective learning of complex features in high-dimensional data. This architecture significantly improves the performance of deep learning models by making them more robust and easier to train.

3.3.1. Model setup

The ResNet50 model is loaded with pre-trained weights from ImageNet, with the top layers excluded. This model serves as a feature extractor, with its output fed into a global average pooling 2D layer to reduce dimensionality. The pooled output is then passed through a fully connected Dense layer with 1,024 nodes using a ReLU activation function. The final layer is a single node Dense layer using a sigmoid activation function to provide binary classification outputs.

3.3.2. Model training and evaluation

The layers of the base model (ResNet50) are frozen during training, which means that only the weights of the added Dense layers are updated during the training process. This approach allows leveraging the pre-existing feature detection capabilities of the ResNet50 model while training existed model to classify images specific to the dataset.

The model is compiled with an Adam optimizer using a learning rate of 0.0001. Binary cross-entropy is used as the loss function, appropriate for the binary classification problem. The model metrics tracked during training are accuracy, precision, and recall.

The model is trained with early stopping and learning rate reduction strategies. Early stopping is implemented to prevent overfitting by halting the training process if the validation loss does not decrease for five consecutive epochs. On the other hand, if the validation loss does not improve for two epochs, the learning rate is reduced by a factor of 0.1, down to a minimum of 1e-7.

The model is trained over 50 epochs with a batch size of 32. Class weights are provided to the model during training to handle any potential class imbalance in the dataset. The training and validation datasets are supplied during training.

Once the model has been trained, it is used to predict the labels of the images in the test dataset. These predictions are compared with the actual labels to evaluate the model's performance. The metrics tracked during training (accuracy, precision, and recall) can be used to assess the model's ability in correctly classifying images.

3.3.3. Model improvement and two-stage training

To optimize the performance of the model, a two-stage training process was introduced. In the second stage, the model was fine-tuned by unfreezing and training the last two blocks of the ResNet50 model. After this initial training phase, the optimizer was switched to stochastic gradient descent (SGD) optimizer with a learning rate of 0.0001 and a momentum of 0.9 for the fine-tuning stage.

Each stage of training included 20 and 30 epochs, respectively, with a batch size of 32. Class weighting was applied during training to handle any imbalance in the dataset. Additionally, early stopping and learning rate reduction strategies were employed to prevent overfitting and dynamically adjust the learning rate.

Following training, the model's performance was evaluated based on binary accuracy, precision, and recall, with predictions generated on the test data. This enhanced methodology introduces several improvements over the initial model:

a. Selective layer training: The ResNet50 layers are selectively unfrozen and fine-tuned in the second stage of training, leading to a better task-specific adaptation.
b. Optimizer adjustment: Adam optimizer in the initial training stage was switched to the SGD optimizer in the fine-tuning stage, providing more stability when fine-tuning the pre-trained features.
c. Layer-wise learning: Deeper layers in the ResNet50 model are gradually trained, preserving useful pre-trained features while adjusting the task-specific ones.

3.4. VGG19 model

visual geometry group-19 (VGG19) [25], a well-known CNN model, was created by the VGG at the University of Oxford. The ImageNet large scale visual recognition challenge (ILSVRC) in 2014 saw this deep neural network architecture perform at the cutting edge of performance. VGG19 model's name is given because of the model's structure, a 19 layers comprising 16 convolutional layers and 3 fully linked layers. In the convolutional layers, 3×3 filters with a stride of one are used, while 2×2 filters with a stride of two are used in the max-pooling layers. At the network's outermost layer, classification is performed using its completely connected layers.

3.4.1. Model setup

As in each model above, VGG19 is loaded with pre-trained weights from ImageNet. Training, test, and validation subnets were transformed to a NumPy array for the image data generator specific to this model. An image data generator is used to create training and validation values for the process of fitting the model. The main difference between this model and the models above is that this is a sequential model where all layers in the model are arranged in sequence. To compile the model and see details or improvements, the following parameters are used: Adam optimizer, binary cross-entropy loss, and followed metrics like accuracy, precision, and recall. Additionally, except for all VGG19 model layers, Flatten, Dropout, and Dense layers are used to improve the model.

3.4.2. Model training and evaluation

Only the weights of the additional Dense layers are updated during training since the layers of VGG base model are locked. Using this method, the model is trained to classify photos specific to the dataset while utilizing the VGG model's built-in feature detection capabilities. The Adam optimizer is used to build the model, using a learning rate of 0.0001. The loss function, binary cross-entropy, is acceptable for binary classification problems. Accuracy, precision, and recall are the model metrics that are monitored throughout training. For training models, except the learning rate, the callback function was created to stop the execution if the accuracy reaches 0.9 or 90%. If the learning rate is decreased by a factor of 0.1, down to a minimum of 1e-7, the validation loss does not improve for two epochs.

The model is trained using a batch size of 32 across 50 epochs. During training, class weights are given to the model to address any potential class imbalance in the dataset. Throughout training too, both training and validation datasets are provided. Model is used to predict the labels of the images in the test dataset once it has been trained. To assess the model's effectiveness, these predictions are contrasted with the actual labels. The model's aptitude for accurately classifying images can be evaluated using the parameters monitored during training (accuracy, precision, and recall). The model improvement is implemented in the same way as in ResNet50 model with selective layer training, SGD optimizer adjustment and layer-wise learning.

4. RESULT AND DISCUSSION

4.1. EfficientNetB3

A CNN was constructed based on the EfficientNetB3 architecture for binary image classification. The model was pretrained on the ImageNet dataset and fine-tuned for the specific task. Initially, we trained the model for ten epochs, after which it achieved promising results.

The initial training phase showed consistent improvement over each epoch. Specifically, the training accuracy increased from 58.19% to 94.35%, while the validation accuracy grew from 60.53% to 92.11% within ten epochs. Loss values correspondingly decreased, highlighting the model's improved capacity to predict class labels correctly. The confusion matrix presented an overall good performance, showing that the model correctly classified most of the validation images. Initial model results are plotted in Figure 4.

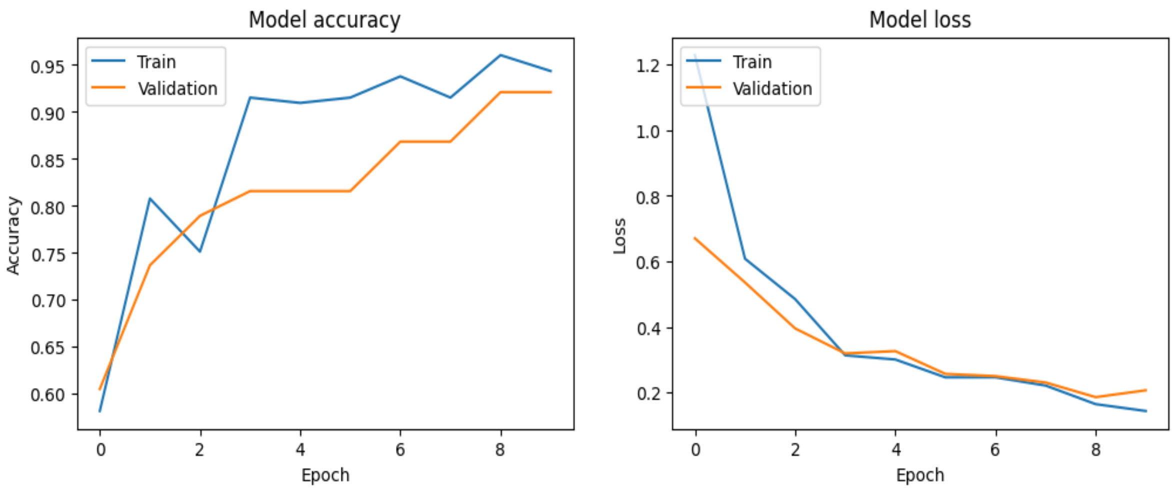

Figure 4. EfficientNetB3 initial model metrics

To further enhance model performance, early stopping and learning rate reduction methods were incorporated into the training process. The initial model reached a validation accuracy of 92.11% after 10 epochs, while the improved model reached the same accuracy after 8 epochs, suggesting faster convergence. The improved model trained for 22 epochs before early stopping was triggered. At this stage, the model attained a training accuracy of 99.44%, with a precision and a recall of 100% and 99.09% respectively, which indicates a near-perfect fit to the training data.

The validation accuracy was 89.47% at this point, slightly lower than the peak validation accuracy of 94.74% observed in previous epochs. This decrease in validation accuracy, coupled with near-perfect training metrics, suggests that the model might be starting to overfit the training data. Precision and recall on the validation data were 89.29% and 96.15%, respectively, revealing that the model was still effectively identifying the positive class, despite the slight drop in validation accuracy. Improved model results are plotted in Figure 5.

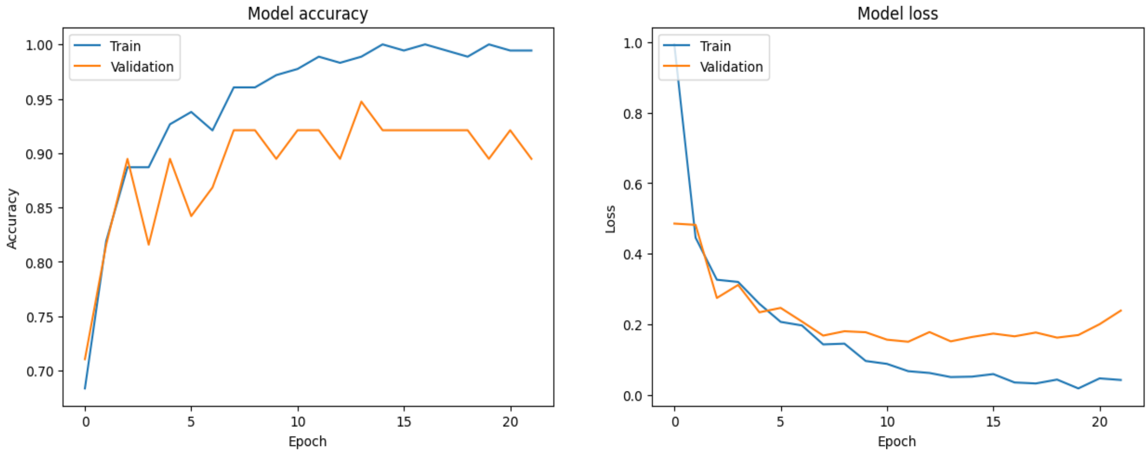

Figure 5. EfficientNetB3 improved model metrics

4.2. ResNet50

Our initial model's performance notably improved over the 50 epochs of training. The training loss started at 79.78% and ended at 56.94%, while validation loss decreased from 59.48% to 44.97%. Initial accuracy, precision, and recall on the training set were all 62.15%, and on the validation set 68.42%. After training, these metrics increased to 75.71% (training) and 81.58% (validation) for accuracy, 73.76% (training) and 85.19% (validation) for precision and remained high at 94.55% (training) and 88.46% (validation) for recall.

Notably, the model generalized better to the validation set from the 20th epoch, evidenced by the narrowing gap between training and validation metrics. The ReduceLROnPlateau callback reduced the learning rate after the 30th epoch, leading to a noticeable increase in validation accuracy. In summary, the model shows improved performance, with significant enhancements from the 20^{th} epoch onwards. Initial model results are plotted in Figure 6.

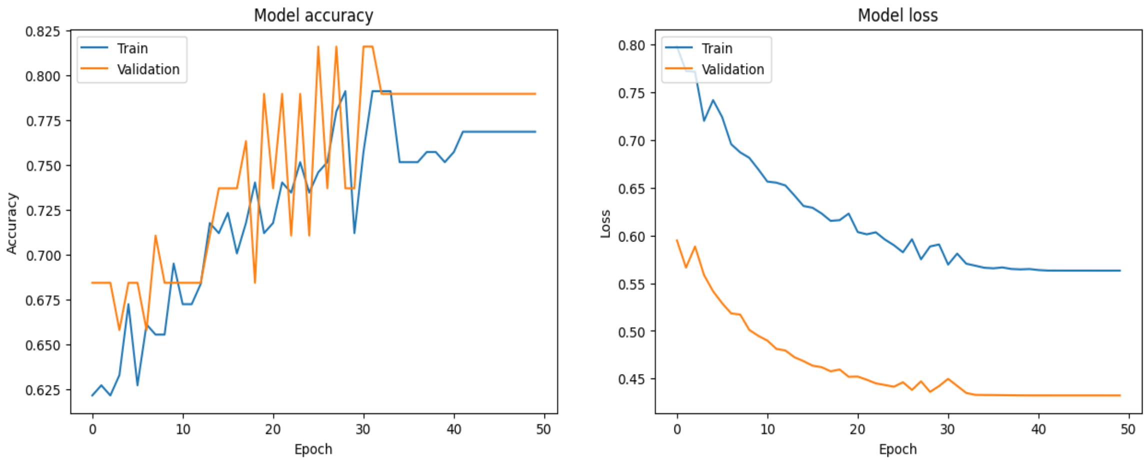

Figure 6. ResNet50 initial model metrics

The improved model utilized two stage training. In the first stage it was trained for 20 epochs and evaluated on a validation dataset. Starting from the first epoch, the model exhibited promising signs of learning with an initial training loss of 90.54%, accuracy of 57.63%, precision of 62.07%, and recall of 81.82%. The validation metrics at this stage were also encouraging with a validation loss of 68.88%, accuracy of 68.42%, precision of 68.42%, and a perfect recall of 100%.

Over the training process, the model showed consistent improvement in all areas. By the final epoch, the training loss was reduced to 67.40%, and the accuracy, precision, and recall were 76.84%, 75.18%, and 93.64% respectively. The validation metrics at the 20th epoch were also significantly improved with a validation loss of 46.52%, accuracy of 73.68%, precision of 76.67%, and recall of 88.46%. The second stage of the model incorporated a similar ResNet-style architecture as the first part, with convolutions, batch normalizations, and ReLU activations arranged in layers. The model displayed notable early stopping at the 11th epoch.

Over the training period, the model demonstrated improving performance on both training and validation sets. In terms of accuracy, the model's performance increased from an initial 63.84% in epoch 1 to a high of 83.62% in epoch 11 on the training set. Similarly, precision increased from an initial 66.91% to 86.49% in epoch 11 on the training set, while recall demonstrated a slight decrease from an initial 82.73% to 87.27% in epoch 11. This suggests that the model was becoming increasingly precise in its predictions while maintaining a relatively high true positive rate.

On the validation set, the model's performance was generally higher, demonstrating its generalization ability. The model's accuracy on the validation set increased from an initial of 78.95% in epoch 1 to a high of 89.47% in epoch 12. Precision on the validation set reached 95.83%, indicating a high rate of true positive predictions, while recall reached 88.46%, indicating the model's ability to correctly identify a high proportion of positive cases. However, the learning rate was gradually reduced as a result of the ReduceLROnPlateau function which was triggered due to lack of improvement in the validation loss. After the 11th epoch, the learning rate was reduced to 9.99999883788405e-07, suggesting that the model might not be able to further learn from the training data beyond this point. Notably, the model's performance demonstrated a balance between precision and recall. This is beneficial as it indicates the model was not overly biased towards either predicting the positive class or avoiding false negatives. The second stage of improved model results are plotted in Figure 7.

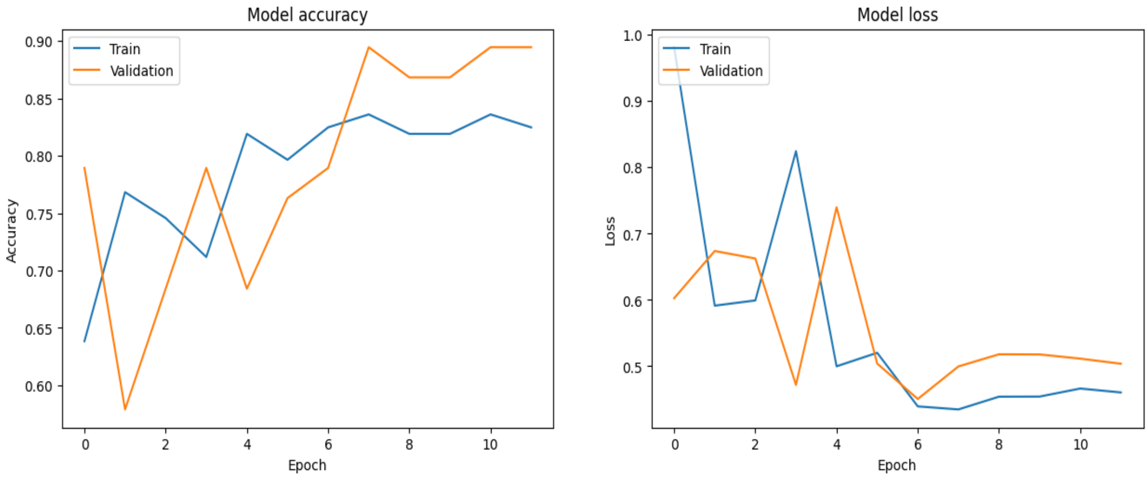

Figure 7. ResNet50 improved model second stage metrics

4.3. VGG19 model

In all 50 epochs, the overall model performance did not noticeably change. Accuracy improved from 57% to 62%, which does not result in obvious improvement. This is because the loss was varying between 86% and 97%. The recall is mostly 100%, which is large enough to be real. The training data was not modified well enough for this specific model. As mentioned above, the callback function was used to stop iterations if the accuracy is 90% as the best case of the model and the learning rate did not go under the limit, which presents the worst case of the model. So, the model was performed in epochs till the end, but the metrics data are not satisfying. In Figure 8, the validation and training data are compared with their accuracy and loss metrics.

After analyzing the performance data of the first model, attempts were made to improve its performance. Differences and improvement were visible, but accuracy was still varying from 57% to 62%, and the surprise decrease in the learning rate stopped the model performance. According to research, this is a specific model which is hard to implement, and a lot of experience and calculations should be included in it. After model improvement with given training and validation data, which were used for all three models, Figure 9 shows results of accuracy and loss metrics in 5 epochs, because performance was stopped when the learning rate decreased under 1e-7.

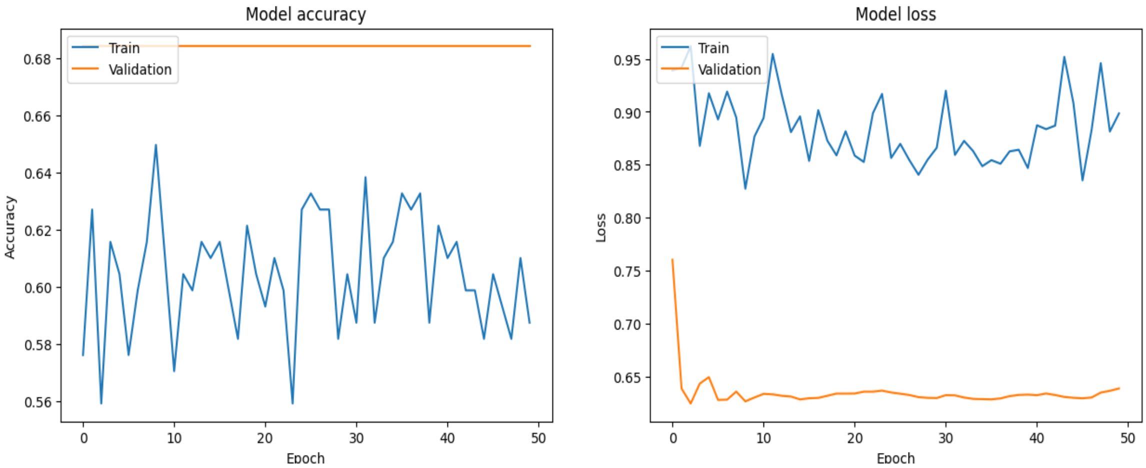

Figure 8. VGG19 initial model metrics

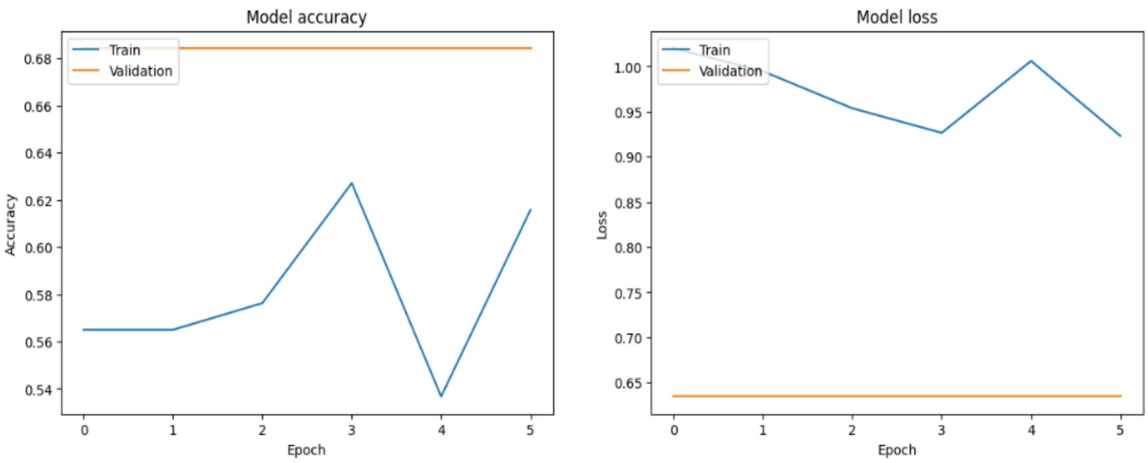

Figure 9. VGG19 improved model metrics

4.4. Discussion

This research study aimed to build an artificial intelligence model that could efficiently detect the presence or absence of brain tumors using MRI scans. The presented results describe the performance of three different models, EfficientNetB3, ResNet50 and VGG19, for binary image classification, each pre-trained on the ImageNet dataset. Pre-trained models were used to extract features and the training process was used to teach models to recognize patterns specific to the dataset.

The EfficientNetB3 model was showing promising results after 10 epochs of training. The training accuracy increased to 94.35% and validation accuracy to 92.11%, all while decreasing loss values. The highest accuracy was obtained after 22 epochs, where the EfficientNetB3 reached 99.44% training accuracy, with precision and recall of 100% and 99.09%, respectively. However, despite the near-perfect training metrics, the validation accuracy of the improved model slightly decreased. This decrease, coupled with the high training metrics, suggests the model might be starting to overfit to the training data. Nonetheless, the results of the precision and recall on the validation data indicated that the model was still effectively identifying the positive class.

The ResNet50 model also showed improved performance over the course of 50 epochs of training. The accuracy, precision, and recall metrics increased for both the training and validation sets, demonstrating the model's ability to generalize better to the validation set as training progressed. Trained with a two-stage process, it achieved a training accuracy of 83.62% and a validation accuracy of 89.47% after 11 epochs in the second stage. The model's recall slightly decreased but remained relatively high. However, the learning rate was gradually reduced due to the lack of improvement in the validation loss. After the 11th epoch, the learning rate was reduced, suggesting that the model might have reached its learning capacity.

On the other hand, the initial VGG19 model's performance remained relatively unchanged over 50 epochs, with accuracy improving marginally from 57% to 62%. The varying loss values between 86% and 97% indicate inconsistency in the model's learning. While the recall of 100% suggests effective identification of positive cases, it also raises concerns about potential overfitting. The fluctuating loss and the unexpected termination due to a decrease in the learning rate indicate that the model's performance was unsatisfactory. The complexity of the model requires further exploration and expertise to achieve better results.

The models showed a reasonable ability to identify positive cases (tumors present) and avoid false negatives, as suggested by the precision and recall metrics. However, all three models displayed limitations. One notable limitation was their tendency to overfit, evidenced by a decline in validation accuracy while training accuracy remained exceptionally high. This observation highlights the models' increased specificity to the training data, potentially limiting their ability to generalize to new, unseen data. Additionally, the dataset used in the models' training was relatively small and imbalanced. Despite efforts to minimize bias by calculating and incorporating class weights during the training process, employing larger and more balanced datasets could enhance the models' resilience and performance to a greater extent. Further analysis and comparisons between these models could provide insights into their relative strengths and weaknesses for specific use cases.

5. CONCLUSION

This study sheds light on the severity of brain tumors and provides a concise literature review of recent studies conducted in 2023 on brain tumor detection. By exploring the application of artificial intelligence and deep learning techniques, the study aims to contribute to the growing body of research focused on improving the accuracy and efficiency of brain tumor detection methods. Moreover, the potential of transfer learning is demonstrated in this study in detecting brain tumors using MRI scans. The models EfficientNetB3, ResNet50, and VGG19 were fine-tuned on the brain MRI scans dataset, and they showed promising results, achieving high accuracy, precision, and recall metrics. Data augmentation techniques, early stopping, learning rate reduction, and two-stage training were valuable strategies for improving model performance and preventing overfitting. The EfficientNetB3 model demonstrated a validation accuracy of 92.11% after incorporating early stopping and learning rate reduction. On the other hand, the ResNet50 model achieved a validation accuracy of 83.68% through selective layer training and optimizer adjustment in a two-stage training process. The training data may not be modified adequately for VGG19, and the surprise decrease in the learning rate prematurely halted the model's performance. The model exhibited minimal improvement throughout the 50 epochs, with accuracy only slightly increasing from 57% to 62%. Additionally, the challenging nature of this particular model requires extensive expertise and calculations. Despite the limitations of potential overfitting, these results underscore the potential of artificial intelligence and deep learning in improving the detection of brain tumors, leading to early diagnosis and better treatment planning. However, there is room for further improvement and validation by utilizing more extensive and more diverse datasets. To fully leverage the capabilities of artificial intelligence in brain tumor detection, future research should focus on addressing challenges related to interpretability, generalization to different populations, and integration with existing medical practices. Also, it would be valuable to compare the performance of other state-of-the-art models on the same task. Furthermore, deploying and testing the models in real-world clinical settings would be the ultimate test of their utility in practical applications. By continuing to advance in this field, we can unlock the full potential of artificial intelligence in improving healthcare outcomes for patients with brain tumors.

ACKNOWLEDGEMENTS

Authors thank Faculty of Engineering and Natural Sciences, International University of Sarajevo, Bosnia and Herzegovina for support in this research and School of Electrical, College of Engineering, UiTM Terengganu Branch for financial support under grant 600-RMC/KEPU 5/3 (004/2023).

REFERENCES

[1] L. Zhao and K. Jia, “Multiscale CNNs for brain tumor segmentation and diagnosis,” *Computational and Mathematical Methods in Medicine*, vol. 2016, pp. 1–7, 2016, doi: 10.1155/2016/8356294.

[2] A. Raza *et al.*, “A hybrid deep learning-based approach for brain tumor classification,” *Electronics*, vol. 11, no. 7, Apr. 2022, doi: 10.3390/electronics11071146.

[3] T. Shelatkar, D. Urvashi, M. Shorfuzzaman, A. Alsufyani, and K. Lakshmanna, “Diagnosis of brain tumor using light weight deep learning model with fine-tuning approach,” *Computational and Mathematical Methods in Medicine*, vol. 2022, pp. 1–9, Jul. 2022, doi: 10.1155/2022/2858845.

[4] M. M. Mijwil, R. Doshi, K. K. Hiran, O. J. Unogwu, and I. Bala, “MobileNetV1-based deep learning model for accurate brain tumor classification,” *Mesopotamian Journal of Computer Science*, pp. 32–41, Mar. 2023, doi: 10.58496/MJCSC/2023/005.
[5] G. Battineni, G. G. Sagaro, N. Chinatalapudi, and F. Amenta, “Applications of machine learning predictive models in the chronic disease diagnosis,” *Journal of Personalized Medicine*, vol. 10, no. 2, Mar. 2020, doi: 10.3390/jpm10020021.
[6] T. Hossain, F. S. Shishir, M. Ashraf, M. D. A. Al Nasim, and F. Muhammad Shah, “Brain tumor detection using convolutional neural network,” in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, May 2019, pp. 1–6, doi: 10.1109/ICASERT.2019.8934561.
[7] Y. Xie *et al.*, “Convolutional neural network techniques for brain tumor classification (from 2015 to 2022): review, challenges, and future perspectives,” *Diagnostics*, vol. 12, no. 8, Jul. 2022, doi: 10.3390/diagnostics12081850.
[8] D. N. Louis *et al.*, “The 2021 WHO classification of tumors of the central nervous system: a summary,” *Neuro-Oncology*, vol. 23, no. 8, pp. 1231–1251, Aug. 2021, doi: 10.1093/neuonc/noab106.
[9] A. A. K. Abolanle *et al.*, “Brain tumor: an overview of the basic clinical manifestations and treatment,” *Global Journal of Cancer Therapy*, vol. 6, pp. 38–41, 2020.
[10] M. I. Mahmud, M. Mamun, and A. Abdelgawad, “A deep analysis of brain tumor detection from MR images using deep learning networks,” *Algorithms*, vol. 16, no. 4, Mar. 2023, doi: 10.3390/a16040176.
[11] G. Raut, A. Raut, J. Bhagade, J. Bhagade, and S. Gavhane, “Deep learning approach for brain tumor detection and segmentation,” in *2020 International Conference on Convergence to Digital World - Quo Vadis, ICCDW 2020*, 2020, pp. 1–5, doi: 10.1109/ICCDW45521.2020.9318681.
[12] H. A. Shah, F. Saeed, S. Yun, J. H. Park, A. Paul, and J. M. Kang, “A robust approach for brain tumor detection in magnetic resonance images using finetuned EfficientNet,” *IEEE Access*, vol. 10, pp. 65426–65438, 2022, doi: 10.1109/ACCESS.2022.3184113.
[13] S. Solanki, U. P. Singh, S. S. Chouhan, and S. Jain, “Brain tumor detection and classification using intelligence techniques: an overview,” *IEEE Access*, vol. 11, pp. 12870–12886, 2023, doi: 10.1109/ACCESS.2023.3242666.
[14] M. Siar and M. Teshnehlab, “Brain tumor detection using deep neural network and machine learning algorithm,” in *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, Oct. 2019, pp. 363–368, doi: 10.1109/ICCKE48569.2019.8964846.
[15] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, “Automatic brain tumor detection and segmentation using U-net based fully convolutional networks,” *Communications in Computer and Information Science*, vol. 723, pp. 506–517, 2017, doi: 10.1007/978-3-319-60964-5_44.
[16] D. R. Nayak, N. Padhy, P. K. Mallick, M. Zymbler, and S. Kumar, “Brain tumor classification using Dense Efficient-Net,” *Axioms*, vol. 11, no. 1, Jan. 2022, doi: 10.3390/axioms11010034.
[17] S. Saeedi, S. Rezayi, H. Keshavarz, and S. R. Niakan Kalhori, “MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques,” *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, Jan. 2023, doi: 10.1186/s12911-023-02114-6.
[18] R. Chelghoum, A. Ikhlef, A. Hameurlaine, and S. Jacquir, “Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images,” in *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 2020, pp. 189–200.
[19] Z. Kuang, “Transfer learning in brain tumor detection: from AlexNet to Hyb-DCNN-ResNet,” *Highlights in Science, Engineering and Technology*, vol. 4, pp. 313–324, Jul. 2022, doi: 10.54097/hset.v4i.919.
[20] Rroan M, “A review of brain tumor segmentation of MRI image using machine learning algorithm,” *Journal of Applied Engineering and Science*, vol. 1, no. 1, pp. 6–11, 2023.
[21] M. Aggarwal, A. K. Tiwari, M. P. Sarathi, and A. Bijalwan, “An early detection and segmentation of brain tumor using deep neural network,” *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, Apr. 2023, doi: 10.1186/s12911-023-02174-8.
[22] S. Athisayamani, R. S. Antonyswamy, V. Sarveshwaran, M. Almeshari, Y. Alzamil, and V. Ravi, “Feature extraction using a residual deep convolutional neural network (ResNet-152) and optimized feature dimension reduction for MRI brain tumor classification,” *Diagnostics*, vol. 13, no. 4, Feb. 2023, doi: 10.3390/diagnostics13040668.
[23] M. Tan and Q. Le, “EfficientNet: rethinking model scaling for convolutional neural networks,” *International conference on machine learning*, pp. 6105–6114, 2019.
[24] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[25] S. Karen and Z. Andrew, “Very deep convolutional networks for large-scale image recognition,” in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015, pp. 1–14.

BIOGRAPHIES OF AUTHORS

Fatima Muftic graduated with a degree in computer science and engineering from the International University of Sarajevo and is currently pursuing her master's degree. She has accumulated valuable experience as an IT business analyst/product owner at extreme software group, focusing on Agile methodologies to enhance software development efficiency. Her true passion lies at the confluence of artificial intelligence and healthcare. Fatima is committed to researching and applying AI in medical contexts, aiming to make healthcare more accessible and effective. She can be contacted at email: fmuftic@student.ius.edu.ba.

Merjem Kadunic is an electrical engineer from Bugojno, Bosnia and Herzegovina. Merjem received her Bachelor of Science degree in electrical engineering with a specialization in telecommunications from the Faculty of Electrical Engineering at the University of Sarajevo in Sarajevo. She is currently pursuing a master's degree in electrical and electronics engineering from the International University of Sarajevo in Sarajevo. During her studies, she is attending advanced courses in AI and software engineering to develop further her knowledge and expertise in these exciting and dynamic fields. She can be contacted at email: hajar_3112@yahoo.com.my

Almina Musinbegovic holds a dual bachelor's and master's degree in computer science and engineering from the International University of Sarajevo, Bosnia. Her research interests lie in the application of artificial intelligence within the healthcare sector. Currently, she works as a Software Developer at Waltercode Company. She can be contacted at email: 220302354@student.ius.edu.ba.

Ali Abd Almisreb is currently an associate professor at the Faculty of Computer Sciences and Engineering, Director of Graduate Council and Editor in Chief at International University of Sarajevo. He received a master's degree in computer science and doctorate doctor's degree in electrical engineering/computer engineering from Universiti Teknologi MARA, Malaysia. His major interests include deep learning, machine learning, computer vision voice recognition and quantum computing. He can be contacted at email: alimes96@yahoo.com.

Hajar Jaafar is senior lecturer of radio-frequency and microwave, IoT system at Univerisiti Teknologi MARA, Malaysia and fellow researcher at Antenna Research Center in Univerisiti Teknologi MARA, Malaysia since 2016. She was awarded in PhD in electrical engineering from Universiti Teknologi MARA in February 2016. She has published more papers including high impact journal and international conference papers. She was also a reviewer of IEEE Antenna Wireless Propagation Letter (AWPL), journals and international conference paper. She has been awarded numerous research grant proposal. She has wide experience supervising postgraduate students related to radio-frequency, microwave fields and IoT system. She also as an active member in graduate engineer, Board of Engineers Malaysia (BEM) (No: 97896A), The Institution of Engineers, Malaysia (IEM) (No: 36254), Member (M) of IEEE (No: 92628020), IEEE Antennas and Propagation Society, International Association of Engineers (IAENG) (No: 193643), Malaysia Board of Technologists (MBOT) (No: PT20090126). She can be contacted at email: hajarj422@uitm.edu.my.