Degree Program in Computer Engineering/Engineering and Economics

First Cycle 15 credits

Image-classification for Brain Tumor using Pre-trained Convolutional Neural Network

AHMAD OSMAN

BUSHRA ALSABBAGH

STOCKHOLM, SVERIGE 2023

Image-classification for Brain Tumor using Pre-trained Convolutional Neural Network

Bildklassificering för hjärntumör med hjälp av förtränat konvolutionellt neuralt nätverk

AHMAD OSMAN

BUSHRA ALSABBAGH

Examensarbete inom datateknik/teknik och ekonomi
Grundnivå, 15 hp
Handledare på KTH: Fredrik Kilander
Examinator: Nicklas Branefelt
TRITA-CBH-GRU-2023:105

KTH
Skolan för kemi, bioteknologi och hälsa
141 52 Huddinge, Sverige

Abstract

Brain tumor is a disease characterized by uncontrolled growth of abnormal cells in the brain. The brain is responsible for regulating the functions of all other organs, hence, any atypical growth of cells in the brain can have severe implications for its functions. The number of global mortality in 2020 led by cancerous brains was estimated at 251,329. However, early detection of brain cancer is critical for prompt treatment and improving patient's quality of life as well as survival rates. Manual medical image classification in diagnosing diseases has been shown to be extremely time-consuming and labor-intensive. Convolutional Neural Networks (CNNs) has proven to be a leading algorithm in image classification outperforming humans. This paper compares five CNN architectures namely: VGG-16, VGG-19, AlexNet, EffecientNetB7, and ResNet-50 in terms of performance and accuracy using transfer learning. In addition, the authors discussed in this paper the economic impact of CNN, as an AI approach, on the healthcare sector. The models' performance is demonstrated using functions for loss and accuracy rates as well as using the confusion matrix. The conducted experiment resulted in VGG-19 achieving best performance with 97% accuracy, while EffecientNetB7 achieved worst performance with 93% accuracy.

Keywords

Brain tumor, Deep learning, Convolutional Neural Network (CNN), diagnosis, Image classification, pre-trained models, dataset, economic impact.

Sammanfattning

Hjärntumör är en sjukdom som kännetecknas av okontrollerad tillväxt av onormala celler i hjärnan. Hjärnan är ansvarig för att styra funktionerna hos alla andra organ, därför kan all onormala tillväxt av celler i hjärnan ha allvarliga konsekvenser för dess funktioner. Antalet globala dödligheten ledda av hjärncancer har uppskattats till 251329 under 2020. Tidig upptäckt av hjärncancer är dock avgörande för snabb behandling och för att förbättra patienternas livskvalitet och överlevnadssannolikhet. Manuell medicinsk bildklassificering vid diagnostisering av sjukdomar har visat sig vara extremt tidskrävande och arbetskrävande. Convolutional Neural Network (CNN) är en ledande algoritm för bildklassificering som har överträffat människor. Denna studie jämför fem CNN-arkitekturer, nämligen VGG-16, VGG-19, AlexNet, EffecientNetB7, och ResNet-50 i form av prestanda och noggrannhet. Dessutom diskuterar författarna i studien CNN:s ekonomiska inverkan på sjukvårdssektorn. Modellens prestanda demonstrerades med hjälp av funktioner om förlust och noggrannhets värden samt med hjälp av en Confusion matris. Resultatet av det utförda experimentet har visat att VGG-19 har uppnått bästa prestanda med 97% noggrannhet, medan EffecientNetB7 har uppnått värsta prestanda med 93% noggrannhet.

Nyckelord

Cancer, Hjärntumör, Artificiell intelligens (AI), djupinlärning, konvolutionellt neuralt nätverk (CNN), Diagnostik, Bildklassificering, förtränade modeller, dataset.

Contents

1 Introduction

Cancer refers to a wide range of diseases characterized by uncontrolled growth of abnormal cells that may develop in various organs or tissue throughout the body [1]. Cancer ranks as the second leading cause of death worldwide after heart disease, accounting for 21% of global mortality. Nonetheless, cancer is considered the leading cause of mortality among women aged 40 to 79 years and men aged 60 to 79 years [2].

More specifically, the number of deaths led by cancerous brain in the United States in 2023 is estimated at 18,990. While global mortality in 2020 led by cancerous brain was estimated at 251,329 [3]. As a matter of fact, the brain is an essential organ as it regulates the functions of other organs in the human body, including responses, sensation, movement, and memory [4]. Moreover, all information and instructions are distributed by the central nervous system via the brain. Therefore, any abnormal cell behaviour in the brain will have serious consequences depending on the damaged part of the brain and may also affect other organs in the body [5].

As a result, early detection of brain cancer is critical for prompt treatment and improving patient quality of life as well as survival rates. Delays in diagnosis, on the other hand, may have a negative impact on patients' survival and quality of life. Consequently, emphasizing the importance of technological advancements in healthcare for early and accurate diagnosis [6] and [7].

Artificial Intelligence (AI) applications, such as computer-aided diagnostics, have shown considerable benefits in assisting clinical decision-making in the healthcare sector [8]. In specific, the use of AI in the diagnosis of brain cancers via brain scans has shown promising results. A study has reported AI's ability to accurately identify 98% of brain cancers [43]. For instance, machine learning algorithms in AI can analyze medical images, including brain scans, and potentially extract hidden features that may not be detectable by human experts [9].

Over time, the science of Computer Vision and Deep Learning has seen significant breakthroughs and advancements, with a focus on one specific algorithm, namely the Convolutional Neural Network (CNN) [10]. This Deep Learning method, CNN, is commonly utilized in visual image analysis to classify and detect images [11].

1.1 Problem statement

Manual medical image classification has been shown to be a time-consuming task [12]. As a result, many neural networks such as Convolutional neural networks (CNNs) were developed in order to assist doctors in several medical image processing, for example, tumor detection and classification. Furthermore, CNNs are gaining popularity as a leading algorithm for image classification [15]. There are common CNN architectures known for their high accuracy rates. The problem addressed in this study was that it was not well known which of the five CNN models VGG-16, VGG-19, ResNet-50, EfficientNetB7, and AlexNet performed best on the problem of brain tumor classification.

1.2 Goal

The purpose of this study was to investigate five different CNN models namely the VGG-16, VGG-19, ResNet-50, AlexNet, and EfficientNetB7 in classifying brain tumor images. Additionally, the study analyzed how image classification, as an AI approach, affects resources use and healthcare costs. The goal of this study was to help improve the image classification field regarding brain tumors. Hence, providing better treatment at low costs for patients.

1.3 Limitations

This study was limited by comparing five pre-trained CNN models, the limited dataset of brain tumor images, and the graphics processing unit GPU 1080TI. In the training phase, the authors employed a CPU processor. However, training each model was extremely time-consuming. To address the problem, the authors used another computer with a GPU processor to reduce the training time. While training the EfficientNetB7 model, the GPU processor could not process large amounts of data. As a consequence, the authors reduced the value of a hyperparameter, namely batch size, so that the data could be processed by the GPU. However, a smaller batch size resulted in worse performance for this tested model.

2 Theory and background

2.1 AI impact on healthcare sector

The applications of AI in the healthcare field are being increasingly and frequently utilized, which had a positive impact on improving the healthcare industry with its high performance on a daily basis. According to [14], the impact of AI in the healthcare field has empowered medical professionals to make decisions relying on high information accuracy, resulting in time and cost savings. As a result, AI has been introduced into several fields in the healthcare sector. In [15], the study states that AI has been applied in many medical domains including diagnostic, clinical, surgical, rehabilitative, and predictive practice.

There are many advantages of applying AI in healthcare whether for patients or medical practitioners. [16] analyzes the benefits of AI as it helps with reducing human mistakes. For instance, medical errors and misdiagnosis were the third leading cause of death in the United States in 2016, according to the Centers for Disease Control and Prevention. AI also assists with accurate and fast diagnoses. Accurate medical diagnosis is one of the most crucial aspects of patient treatment as it increases the probability of successful recovery. Additionally, AI helps minimize time spent in the healthcare sector, thus taking on many tasks that were previously carried out by doctors. For example, AI assists in test results of MRI, CT scans, ultrasound, etc. As a result, AI can lead to better patient treatment and higher quality of healthcare [16].

2.2 Deep learning and image classification

Deep learning is an AI method that makes computers able to process data and automate processes that typically involve cognitive ability, such as image classification. Deep learning models are capable of identifying complicated patterns in images, text, audio, and other data. It also has various applications in automotive, manufacturing, electronics, medical research, and other industries. In the healthcare industry, deep learning is helping professionals and researchers analyze images and allows doctors to diagnose diseases more accurately providing improved performance and resulting in better medical outcomes [17]. Figure 2.1 illustrates that deep learning models including CNN, Deep Neural Network (DNN), and Recurrent Neural Network (RNN) helps in analyzing data and using the analysis results in many applications like disease classification and disease prediction.

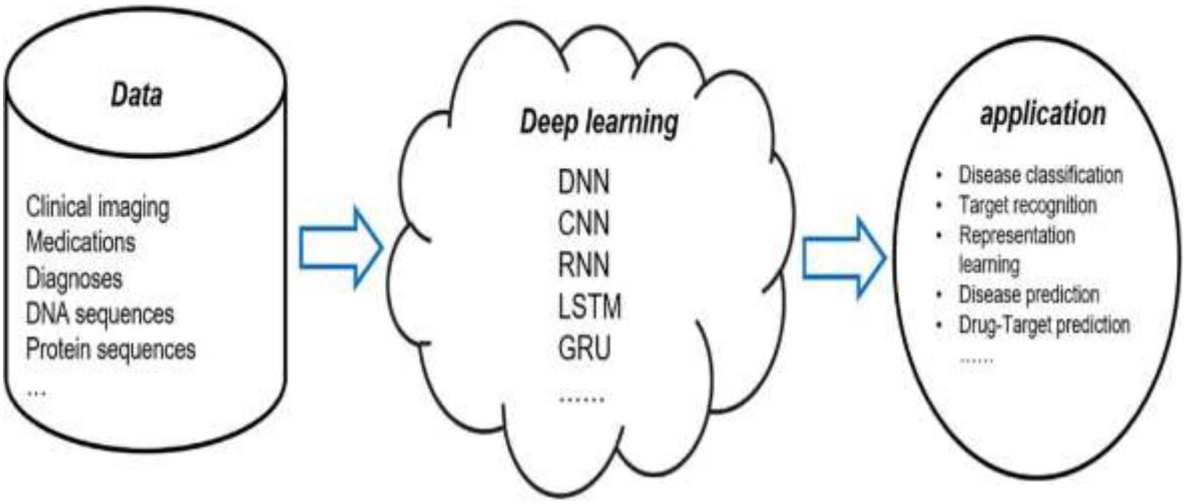

Figure 2.1: Applications of deep learning models in computational medicine, adopted from [18]. Reproduced with permission from Frontiers.

Deep learning networks learn by analyzing complex structures in the input data they receive. Figure 2.2 shows how computational models are composed of multiple processing layers. The input layer where it receives data, several hidden layers where they perform computations on the input, and the output layer where it returns the output [19].

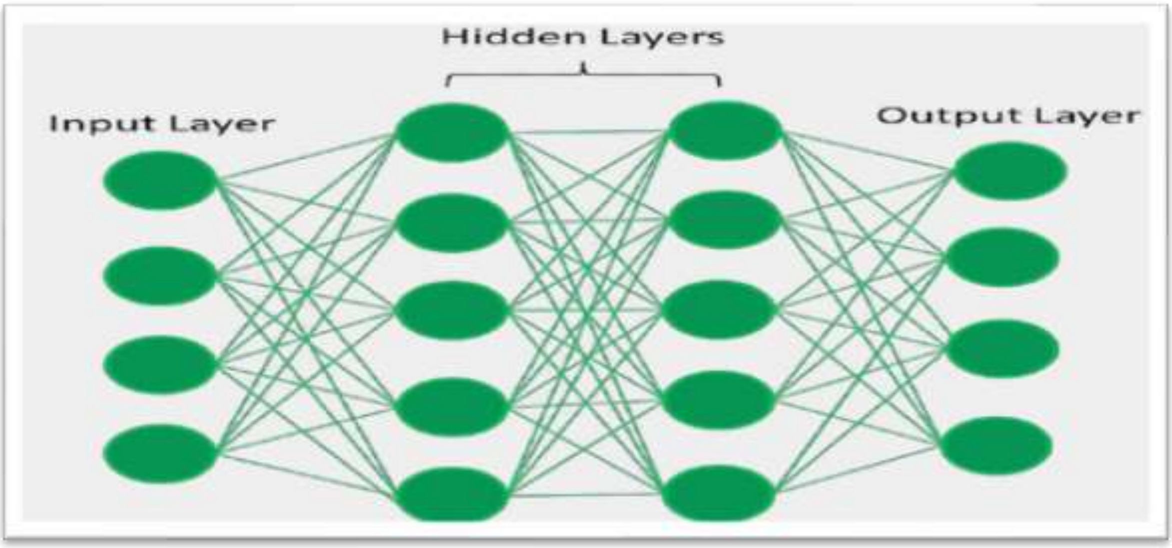

Figure 2.2: A fully connected deep neural network consists of an input layer, multiple hidden layers, and output layer, adopted from [20]. Reproduced with permission from GeeksforGeeks.

One of the most important applications of deep learning is image classification. By using this approach in the healthcare sector, disease diagnosis equipment is becoming more sophisticated and advanced. Simultaneously, image classification technology has the potential to increase the efficiency of diagnosing difficult diseases through the analysis of a significant quantity of picture data and through pathological examination of the symptoms of various illnesses. This will give more time for the recovery of patients as a consequence [21].

2.3 Convolutional Neural Network (CNN)

2.3.1 CNN structure

A CNN, or convolutional neural network, is one of the most popular and established deep learning algorithms that process pixel data such as images [23]. CNN has succeeded in detecting patterns in images, such as lines, circles, or faces. Moreover, CNN has multiple convolutional layers layered on top of one another, each with the ability to identify increasingly complex forms. Layers in CNN are in three dimensions namely height, width, and depth. Moreover, CNN is a feedforward neural network that usually has three layers namely, convolutional, pooling, and a fully connected layer [22].

The first layer and key block of CNN is the convolutional layer where the main computation is assigned. This layer computes the dot product of two matrices, one of which is the input image, which may have three (RGB) channels, and the other matrix is the filter and is known as a kernel. The kernel is smaller in space than an image but more detailed. The kernel slides over the picture's height and breadth during the forward pass, providing an image representation of that receptive region. This results in a two-dimensional representation of the image known as an activation map, which provides the kernel's reaction at each spatial point of the image [24].

The next layer in CNN structure is the pooling layer, it substitutes network output at certain points by calculating a summary statistic of surrounding outputs. This helps to reduce the amount of computations and weights necessary due to the reduced spatial size of the representation. The pooling action is performed on each slice of the representation separately. In addition, the pooling layer minimizes the amount of network parameters and calculations which increases network efficiency and prevents overlearning [24].

The resulting two-dimensional arrays from the previous layer (pooling layer) have to be converted into a one-dimensional array in order to feed the following layer, namely the fully connected (FC) layer. This phase is conducted on another layer called the flattening layer [25]. Therefore it is possible to use three or four convolutional layers. On the other hand, neurons in the fully connected layer have a complete connection with all neurons in the adjacent (preceding and following) layers. As a result, it may be calculated using a matrix multiplication followed by a bias effect. The FC layer aids in mapping the representation between input and output [24]. Figure 2.3 illustrates how the CNN procedure steps starting with receiving an input image, then making the computations in the convolutional layer, after that minimizing the spatial size of the representation, then flattening the output from the pooling layer, and finally classifying the image in the fully connected layer.

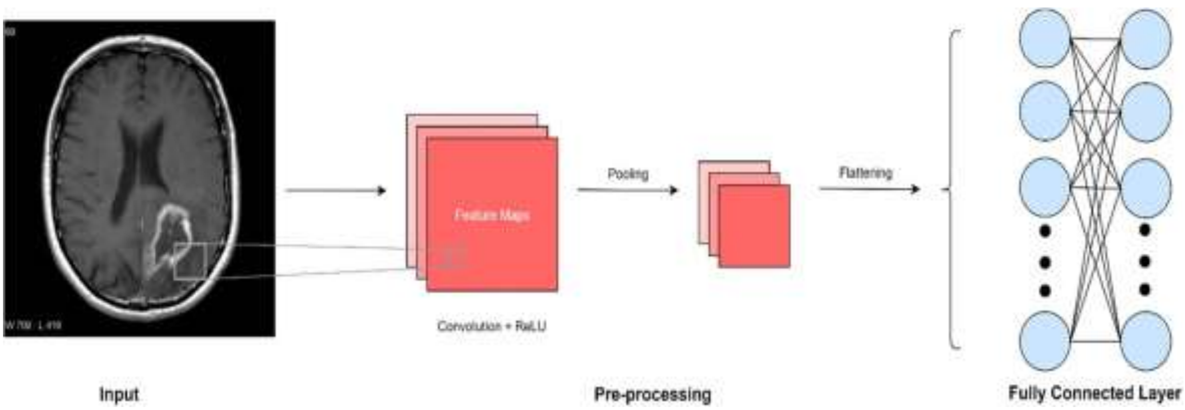

Figure 2.3: The process of classifying an image by CNN. Adopted from [26]. Reproduced with permission from Insights into imaging.

2.3.2 CNN Impact on the healthcare sector

As CNN has achieved high performance in disease detection, the researchers in [27] attempted to improve a CNN model for skin image segmentation. The study aimed to create a skin disease image dataset that could be used for Computer-aided detection (CAD) of multiple skin disease classifications. The outcomes showed that atopic dermatitis can be distinguished from malignant diseases and complications such as mycosis fungoides, impetigo, and herpesvirus infection with about 90% sensitivity and specificity. The efficacy of skin disease categorization in the CNN-segmented picture dataset was very close to that of the manually cropped image dataset and greater than the original image dataset.

Heart disease is one of the leading causes of death. According to the World Health Organization, 17.9 million people die each year as a result of heart disease. The implementation of an effective algorithm can assist in recognizing the existence of cardiac disease before it reveals itself. A study [28] focused on the utilization of state-of-the-art UCI repositories in order to early detect heart diseases. The data were preprocessed before training. On Google Collab, the preprocessed data were combined with the CNN algorithm to predict cardiac illnesses. The suggested system was assessed in terms of performance measures and it obtained 91.71% accuracy, 88.88% precision, and 82.75% recall.

In the human digestive system, gastrointestinal (GI) illnesses are prevalent. Stomach cancer, esophageal cancer, and colorectal cancer are the most frequent occurrences causing death in GI diseases. A study [29] aimed to enhance classification performance for detecting GI disorders. The classification accuracy results for CNN models ResNet-50 and GoogLeNet were 90.28% and 91.38% respectively.

2.4 Economic aspects

As mentioned before, AI has the potential to enhance healthcare quality such as productivity and efficiency. It may help doctors on a daily basis by allowing them to spend more effort and time taking care of patients. It may additionally expedite the release of life-saving cures [30].

AI is bringing social and economic benefits to the healthcare sector helping in resolving key issues. For example, AI is helping medical practitioners in diagnosing diseases and treatments. This economic and social effect can be measured in terms of health conditions, financial capacity, and healthcare professionals' (HCPs') time spent. Nonetheless, the impact of AI on Europe's healthcare systems may be evaluated by measuring the number of lives saved, financial savings, and hours saved up for HCPs [31].

Evaluating the socio-economic effect of AI on European health systems is critical to moving the present debate on the role AI is able to play in health forward. It is estimated that 380,000 to 403,000 lives might be spared each year. Additionally, €170.9 to 212.4 billion can be saved yearly as a result of AI's impact on financial resources [31].

AI applications, for instance, could be utilized throughout the patient's treatment such as wearables, imaging, and laboratory applications. Among AI applications, wearable applications may have a major effect, with the potential savings of up to 313,000 lives and €50.6 billion. Followed by AI applications in monitoring, saved up to 42,000 lives and €45.7 billion, subsequently, AI applications in imaging saving up to 41,000 lives. Finally, AI applications could save between 1,659 and 1,944 million hours per year. This effect is driven by AI applications in virtual health aid (VHA), which may free up to 1,154 million hours per year. Thus, enabling HCPs to devote significant time towards activities of higher value [31].

In [31], the study points out that AI application in the healthcare industry may save €7 billion by early detecting five-year mortality rates for patients at risk of coronary artery diseases (CAD). Additionally, the study shows that the image classification approach in AI can also save 1,900 lives annually by assisting radiologists in diagnosing diseases.

AI is predicted to analyze approximately 250 million images per day for around $1,000, implying possible savings of billions of dollars. Aside from the efficiency and workflow benefits that AI-assisted image interpretation and clinician support can provide, there is an opportunity to decrease the number of employees in the back-office, administrative jobs, scheduling of operating rooms and clinic appointments, and staffing [34].

Cancer has substantial financial expenses affecting the individual and society as an entirety. According to the Agency for Healthcare Research and Quality (AHRQ), the direct medical expenses for cancer in the United States in 2015 were $80.2 billion [32]. For instance, utilizing AI for cancer screening will potentially save society 290 million dollars annually [33].

2.5 Previous work

Patients with brain tumors have a relatively short life expectancy in their most severe form. Therefore, arranging treatment is an important stage in improving patients' standard of life. Brain tumors are classified into two types: low-grade tumors and high-grade tumors. A low-grade brain tumor is not malignant and does not spread to other sections of the brain. As well, a high-grade tumor is referred to as a malignant tumor, and it spreads fast with unclear limits to other parts of the body, it may also result in instant death [35]. The classification of Convolutional Neural Networks (CNN) is proposed for automated brain tumor detection. When compared to all other state-of-the-art methods, experimental results show that CNN archives rate 97.5% accuracy with low complexity [36].

In [37], an experiment was carried out to compare the performance of pre-trained models VGG-19, VGG-16, ResNet-50, and Inception-V3 in classifying brain MRI image, the dataset used for training consisted of 98 unaffected images and 155 affected images. The result of the experiment showed that the VGG-16 model achieved 99% accuracy, the ResNet-50 model achieved 97.92% accuracy, the Inception-V3 model achieved 81.25% accuracy, and the VGG-19 model achieved best performance with 99.48% accuracy.

In [38] researchers conducted an experiment categorizing normal and abnormal brain MRI images, employing the ResNet-34 pre-trained CNN model, a transfer learning technique, and Data Augmentation. The researchers trained the CNN models using a dataset of 613 images for training and testing. There were 27 normal photos and 513 abnormal ones. In the axial plane, all images were 256x256 pixels in size. The study received optimal testing accuracy of 100%.

In [39] an experiment was performed to classify pituitary, glioma, and meningioma brain cancers using three different pre-trained CNN models (VGG-16, AlexNet, and GoogleNet). During this Transfer learning approach, VGG-16 achieved the highest classification and detection accuracy of 98.69%.

Additionally, an experiment [40] aimed to use CNN networks and CNN layers to classify brain tumor images. The Resnet50 architecture was employed, and a hybrid model was provided. The last 5 layers of the Resnet50 architecture were deleted, and 10 layers were inserted into their replacement place, and its accuracy was compared with different pre-trained models including Alexnet, Resnet-50, Inception-V3, GoogleNet, and Densenet201. The researchers trained the CNN models using a dataset made up of two folders. The first folder contained 98 not infected images, whereas the other folder had 155 infected images. The modified ResNet-50 model performed well, attaining 97.01% accuracy.

Another study [41] used Image Processing and Data Augmentation methods on a limited dataset of 253 brain MRI images. Researchers trained the dataset with a basic 8-layer CNN model and used transfer learning to compare the scratched CNN model accuracy to pre-trained VGG-16, ResNet-50, and Inception-V3 models. There were 155 images of malignant cancer and 98 images of benign non-cancerous tumors in the dataset. The dataset was divided into three sections for training, validation, and testing. Model learning used the training data, while the validation data was used for model evaluation and parameter adjustment. The test data was used to evaluate the model in the end. The suggested model performed 96% correctly on training data and 89% correctly on validation dataset. They trained pre-trained VGG-16, ResNet-50, and Inception-V3 CNN models on the same dataset using the transfer learning technique to evaluate the correctness of their CNN model. VGG-16 demonstrated 90% accuracy on training data and an accuracy of 87% on validation data, ResNet-50 showed 92% accuracy on training data and 87% accuracy on validation data, and Inception-V3 demonstrated 93% accuracy on training data and 83% accuracy on validation data.

Using AI Algorithms, CNN, and Deep Learning, a study [42] intended to improve the quality and efficiency of MRI scanners in categorizing brain tumors and identifying their kinds. They used five pre-trained models to train a brain tumor dataset: Xception, ResNet-50, Inception-V3, VGG-16, and MbileNet. The authors trained their CNN models on a dataset of 4480 images, 2880 for training, and 2880 for validation. There were 520 images for training and 200 images for validation for each kind of brain tumor. The final trained models were tested using 800 unseen photos (200 images for each kind). The ResNet-50 model achieved 98.50% accuracy, the Inception-V3 model achieved 98.00% accuracy, the VGG-16 model achieved 97.50% accuracy, the MobileNet model achieved 97.25% accuracy, and the Xception model achieved 97.25% accuracy.

3 Method

This chapter gives an overview of the methods used to obtain the results achieved in this study, mainly to determine which pre-trained CNN model has the best performance in classifying brain tumor images. In order to conduct the experiment in this study, many measures were implemented. Starting with collecting the data to train models. The authors used transfer learning which allows us to use previous knowledge in creating precise models in a shorter period of time instead of creating and training the entire model from scratch. In addition, various techniques were employed to improve the models' performance in brain tumor classification, such as using data augmentation and transformation. This helped preprocess the images and expand the training dataset size by making copies of the existing data with minor differences. The authors also used several hyperparameters in order to define the CNN model’s structure and the model’s performance to get the best accuracy. Hyperparameters that were used in this study were batch size, epochs, learning rate, and optimizer. Furthermore, the authors used an evaluating matrix, namely confusion matrix, to evaluate the performance of CNN models. The authors measured the economic impact of CNN, as an AI approach, by analyzing the value added, i.e. time, lives, and cost saving by implementing CNN in the healthcare sector. The analysis was performed to determine the possible benefits of investment in innovative technology.

3.1 Data collection

The primary goal of data collection is to carry out this study’s experiment evaluating the performance of CNN models. The dataset was imported from the Kaggle website [44], comprising 3861 images. The dataset was divided into four categories, namely no, yes, pred, and Br35H-Mask-RCNN. The yes category contains 1500 images that are affected by brain tumor. While the no category has 1500 images that are not affected by brain tumor. The authors divided images in the yes and no categories, where 1000 images were taken from each category to train the models. While the remaining 500 images from each category were used in the testing phase. The folders pred and Br35H-Mask-RCNN were not used in this study.

3.2 Transfer learning and CNN architectures

Training a model from scratch requires a lot of data and consumes a lot of time, which is why we decided to use transfer learning. The transfer learning approach allows for using previous knowledge in creating precise models in a shorter period of time instead of creating the entire model from the start. This approach as a deep learning method implies using pre-trained models, these models acquired knowledge in solving other challenges by being trained on extensive datasets [45]. Because training such models is computationally expensive and the dataset used in this study was limited, the authors used the pre-trained models. Some of the most popular pre-trained models in CNN are briefly presented.

Visual geometric group (VGG) is one of CNN pre-trained models achieving high accuracy in image classification. VGG architecture can consist of 16 or 19 layers as in VGG-16 and VGG-19 respectively. At ImageNet Large Scale Visual Recognition Competition (ILSVRC)-2014, VGG reached about 92.7% test accuracy using the ImageNet, a dataset that contains more than 14 million images from almost 1000 classes [46].

Another CNN architecture is ResNet that won the ILSVRC-2015. This architecture delivers a high accuracy at a minimal computational expense. Other CNN models that are worth mentioning are

ZFNet which also won the competition in 2013, AlexNet which won the competition in 2012, and EfficientNet which won ILSVRC-2019 with high accuracy [47, 48].

Based on the results in the competitions, the above mentioned models achieved high efficiency in image classification. The pre-trained CNN models that were investigated in this study were (VGG-16, VGG-19, ResNet-50, AlexNet, and EfficentNetB7).

To perform this study with the limited dataset, the authors used the extraction features in transfer learning. These features imply using a pre-trained model with its useful features as a start point in a new study. The authors imported pre-trained models from the Touchvision library in a popular open-source framework PyTorch [49]. Then number of changes were made in these models, for instance, the number of output classes was changed from 1000 to 2 to be able to train the models with the dataset used in this experiment. The 2 classes represent the unaffected brain tumor images and the affected with brain tumor images.

3.3 Data augmentation

Data augmentation is used in CNNs to expand the training dataset size by making copies of the existing data with minor differences. These copies are made by making a series of image transformations involving rotation, scaling, flipping, and cropping. The dataset used to carry out this experiment is large enough, however a larger dataset will result in more accurate predictions. In this study the *torchvision.transforms* module [49] was used, which offers a number of transformations to make various sorts of image data modifications. The authors used multiple data augmentation and transformation in order to train CNN models applying RandomHorizontalFlip, Normalization, and RandomResizedCrop transformers from the module. For instance, RandomHorizontalFlip flips the images horizontally, RandomResizedCrop crops a random part of the image and resizes it, and normalization normalizes an image with mean and standard deviation to guarantee similar distribution across all inputs. The authors used the values [0.485, 0.456, 0.406] and [0.229, 0.224, 0.225] for mean and standard deviation (std) respectively. Those values were provided by PyTorch as default values and can be assigned manually by authors. Moreover, each model had a different default input image size, for example, the VGG-19 input image size was 224x224.

3.4 Hyperparameters

Hyperparameters are variables that can be assigned values before the model training process begins. These variables define the CNN model's structure and the model's performance, in order to get best performance and accuracy for the CNNs models [50]. Multiple hyperparameters were used in this study, namely batch size, epochs, learning rate, and optimizer. The authors chose to use the epoch hyperparameter because it specifies the number of times the neural network works through the entire training dataset, which implies how much the network will learn. While batch size hyperparameter was used in the study to define the number of images worked through by the neural network every iteration. The optimizer hyperparameter was used to modify the parameters and weights of the network. On the other hand, the learning rate was used to determine how quickly the model adapts to a problem.

Moreover, as for the epoch hyperparameter, the authors assigned different values observing if the performance of each model improved as the assigned value increased. For example, the author assigned the values 10 epochs, then 30, and finally 50 resulting in the value 50 giving the best accuracy compared to other assigned values. However, the more the model is trained, the more likelihood of overfitting increases. This means that a model attempts to intensively train and learn for a specific dataset, decreasing the model's efficiency and accuracy. Consequently, the model cannot predict new samples that are not part of the training data. On the other hand, underfitting occurs when a CNN model is unable to learn properly from the training data, reducing accuracy and resulting incorrect predictions [51]. Notably, the authors tested 75 number of epochs on one model, ResNet-50, to see if the accuracy will even be higher or if the model will be overfitted resulting in worse accuracy. It was observed that the accuracy rate was worse than the rate in 50 epochs and the loss rate was higher.

The other hyperparameter used in the study was learning rate that determines how fast the model learns as mentioned above. The learning rate indicates the amount of weights changed during training. The value of learning rate is usually small, between 0.0 and 1.0. Using a high learning rate might lead the model to learn faster with a minimal set of weights, while using a low learning rate may lead the model to learn slowly but with an optimal set of weights [52]. In this study, the authors assigned the learning rate to 0.001, which is the default assigned value by PyTorch.

Since the batch size has no default value, many batch sizes have been tested during the study, which were 5, 10, 20 and 50, the batch size value 50 gave the best results. Notably, when training the model EfficientNetB7 with a batch size value 50, the GPU processor couldn't process 50 images simultaneously and forcely finished the training session. As a result, the authors trained the EfficientNetB7 with batch size value 30. Subsequently, the last hyperparameter used in the study was the optimizer, used to reduce the total loss. PyTorch provides many kinds of optimizers, among them are Adagrad, Adam, and RMSProp. In this study, Adam was used as the optimizer since it requires little memory and is efficient in terms of computation [55].

3.5 Evaluating metrics and models' performance

3.5.1 Model's performance

In order to validate the CNN model's efficiency and reliability in detecting brain tumors, many factors have been considered, including time and accuracy. Models were compared based on accuracy and training durations. In the study, the authors employed a computer with a GPU processor.

The authors used output images to provide the results in a visual manner. The images were selected randomly from the test folders and displayed with a corresponding text explaining whether the images were brain tumor affected or not. The images labeled in the text with yes were brain tumor affected, while the ones labeled with no were brain tumor unaffected.

Furthermore, the authors used Loss function that calculates the loss rate which is the difference between the predicted output and the actual output (the loss rate). The loss rate ranges between 0 and 1. If the loss rate is near to 1, then the model has high loss. On the other hand, if the rate is near to 0, then the model has low loss. Loss function is used to observe how the model behave with different values of epochs meaning how much the model is near to the predicted result. In the study, we have binary classification, so we used CrossEntropyLoss function as it is the recommended method for binary classification. The authors plotted the test-loss, train-loss, train-accuracy, and test-accuracy as a function of the number of epochs. These functions were used to determine the lowest test- and train-loss rate compared to the number of epochs used, as well as to determine the highest train- and test-accuracy compared to the number of epochs used.

The authors generated a commonly used matrix to determine the model's performance based on each model's accuracy in classifying brain tumor images namely, confusion matrix. This matrix is explained in detail in section 3.5.2.

3.5.2 Confusion matrix

Confusion matrix is a table that visually represents the actual and predicted values, it was used to calculate and show the results of the experiments. The output "True positive" indicates the number of positive values classified correctly. Likewise, the output "True negative" indicates the number of negative values classified correctly. Whereas the output "False positive" indicates the number of the actual negative values classified as positive. Similarly, the output "False negative" indicates the number of the actual positive values classified as negative [53]. Figure 3.1 illustrates the table for confusion matrix, the values in the table are true positive (TP), false positive (FP), true negative (TN), false negative (FN).

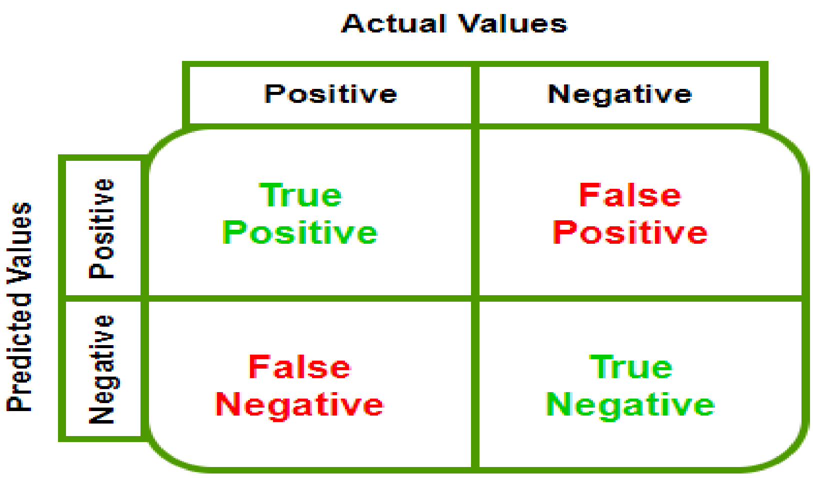

Figure 3.1: Confusion matrix table. Adopted from [54]. Reproduced with permission from GeeksforGeeks.

After getting the output TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) from the table, the authors calculated the accuracy for each model. Accuracy of an model was calculated according to the following form by dividing the correctly classified images (TP+TN) by the total number of images (TP+TN+FP+FN):

$$Accuracy = (\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{FB} + \mathrm{FN} + \mathrm{TN}\ )$$

3.6 The economic impact of CNN models

Economic analysis can be performed to determine the possible benefits on investment of innovative technology in order to make the greatest use of available funds. Brain tumor economic analysis compares the costs and benefits of various therapies. Such analysis that provides a technology's health and economic impact can help with decision-making and encourage further needed investment. In this study, the authors analyzed five different CNN models and compared them in terms of performance in order to determine the economic impact of these five CNN models in the healthcare sector including time and cost saving. In addition, the authors analyzed the social effect of CNN models on the patient's survival and lives saved.

4 Results

This chapter contains an overview of the results of this study training five CNN models. The results contain a confusion matrix table for each model and the graphs of loss- and accuracy rates. Additionally, images generated as output in the experiment are presented in this chapter.

4.1 Confusion Matrix

The figures below illustrate the tables of the confusion matrix for the models AlexNet and ResNet-50. The upper left corner of the table shows the value of the true negative (TN) output, where the actual value is negative and classified as negative. For instance, AlexNet classified 495 images that are unaffected out of 499 total unaffected images. Similarly, the lower right corner of the table shows the value of the true positive (TP), where the actual value is positive and classified as positive. ALexNet classified 468 images that are affected out of 499 total affected images. Likewise, the upper right corner of the table shows the value of the false negative (FP) output, where the actual value is unaffected and classified as affected. For example, ResNet-50 classified 3 images as affected images, where the images were actually unaffected. Finally, the lower left corner of the table shows the value of the false negative (FN), where the actual value is affected and classified as unaffected. For instance, ResNet-50 classified 50 images as unaffected images but actually the images were affected.

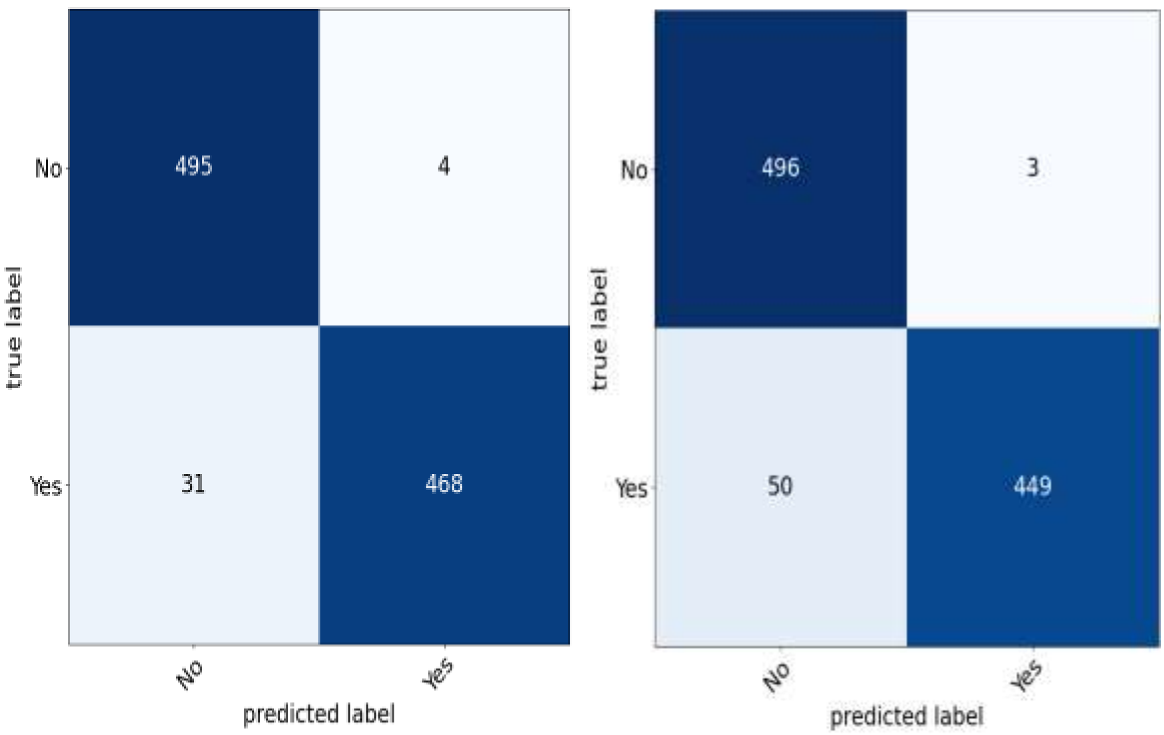

Figure 4.1: Confusion matrix tables for AlexNet (left) and ResNet-50 (right) models.

Figures 4.2, and 4.3 shows the confusion matrix for models VGG-16, EfficientNetB7, and VGG-19.

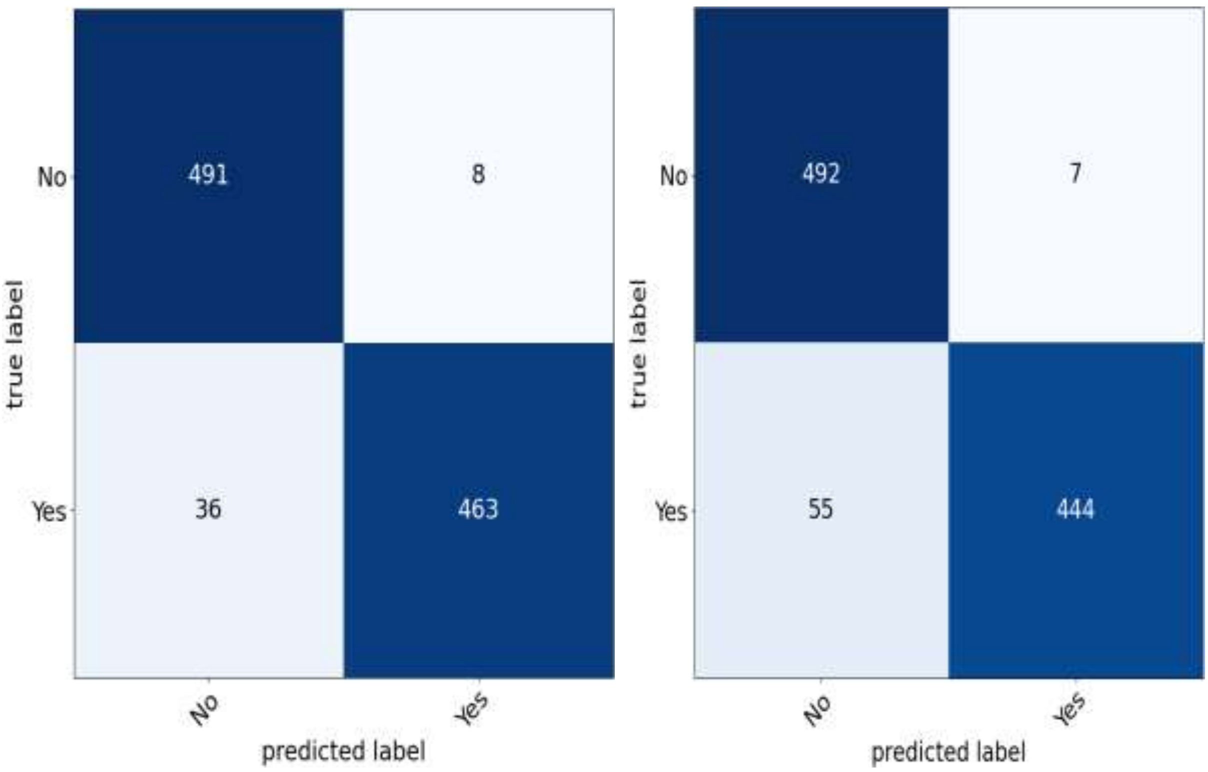

Figure 4.2: Confusion matrices for VGG-16 (left) and EfficientNetB7 (right) models.

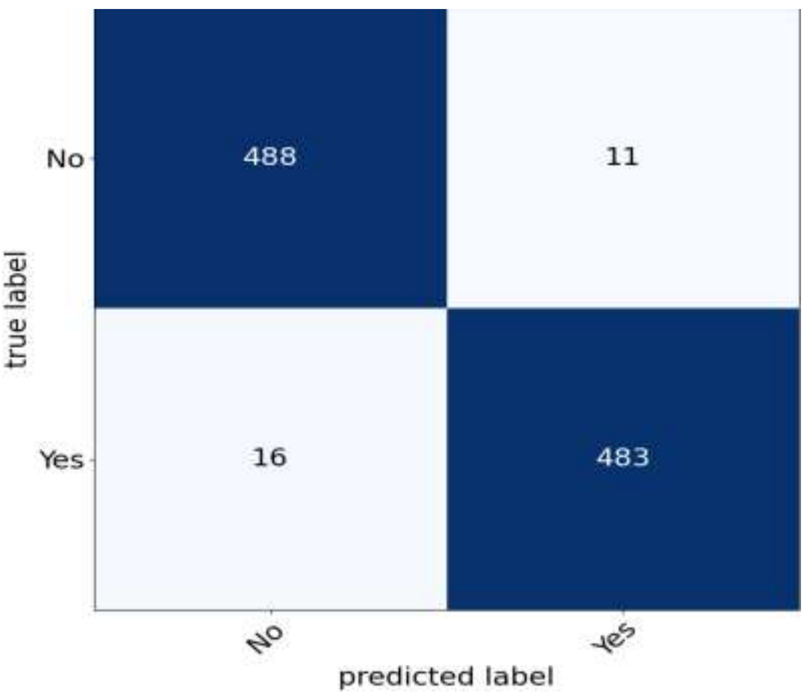

Figure 4.3: Confusion matrices for VGG-19 model

The authors calculated the accuracy for each model based on values (TN), (TP), (FP), and (FN) by using the above mentioned form. For instance, the accuracy of VGG-19 was calculated as follows: VGG-19 accuracy = 488 + 483 / 488 + 483 + 11+ 16 = 0.97. Table 4.1 shows the calculated accuracy of each model.

Table 4.1: Accuracy of each model calculated by using Confusion matrix.

	AlexNet	ResNet-50	VGG-16	EfficientNetB7	VGG-19
Accuracy	0.96	0.94	0.95	0.93	0.97

4.2 Loss and accuracy rates

During the training and testing phase, the authors examined the CNN models' train-loss, test-loss, test-accuracy and train-accuracy rates by plotting their functions based on the number of epochs. These graphs illustrate how the model's performance and accuracy change depending on the number of epochs.

In the following figure 4.4, the left graph shows how the train- and test- loss decreased as the number of epochs increased, where model ResNet-50 achieved the lowest train- and test- loss at the number of epochs 50. While the right graph shows how the train- and test-accuracy increased as the number of epochs increased achieving the highest accuracy at a number of epochs 50.

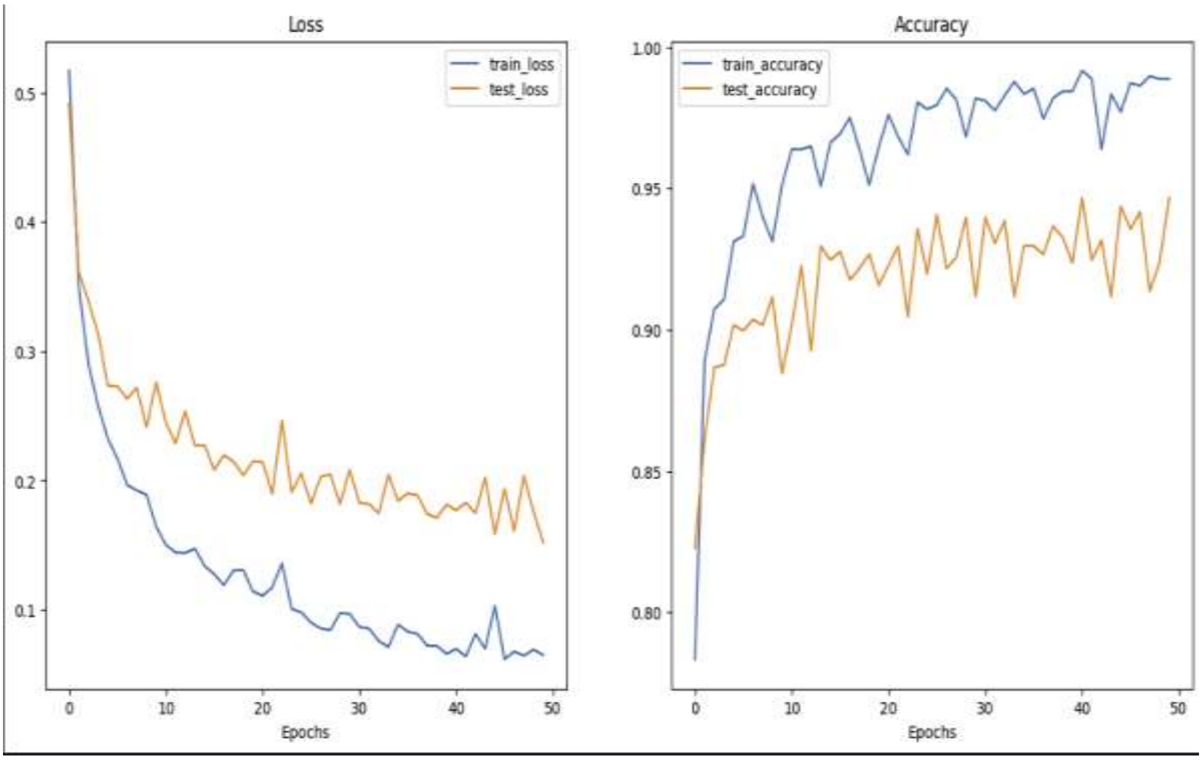

Figure 4.4: The train-loss, test- loss, train-accuracy and test-accuracy for model ResNet-50.

In the figure below 4.5, the left graph illustrates the rates of the test- and train loss for the model AlexNet and the right graph illustrates the rates of the test- and train accuracy. The graphs shows that the model achieved the highest accuracy at a number of epochs 50.

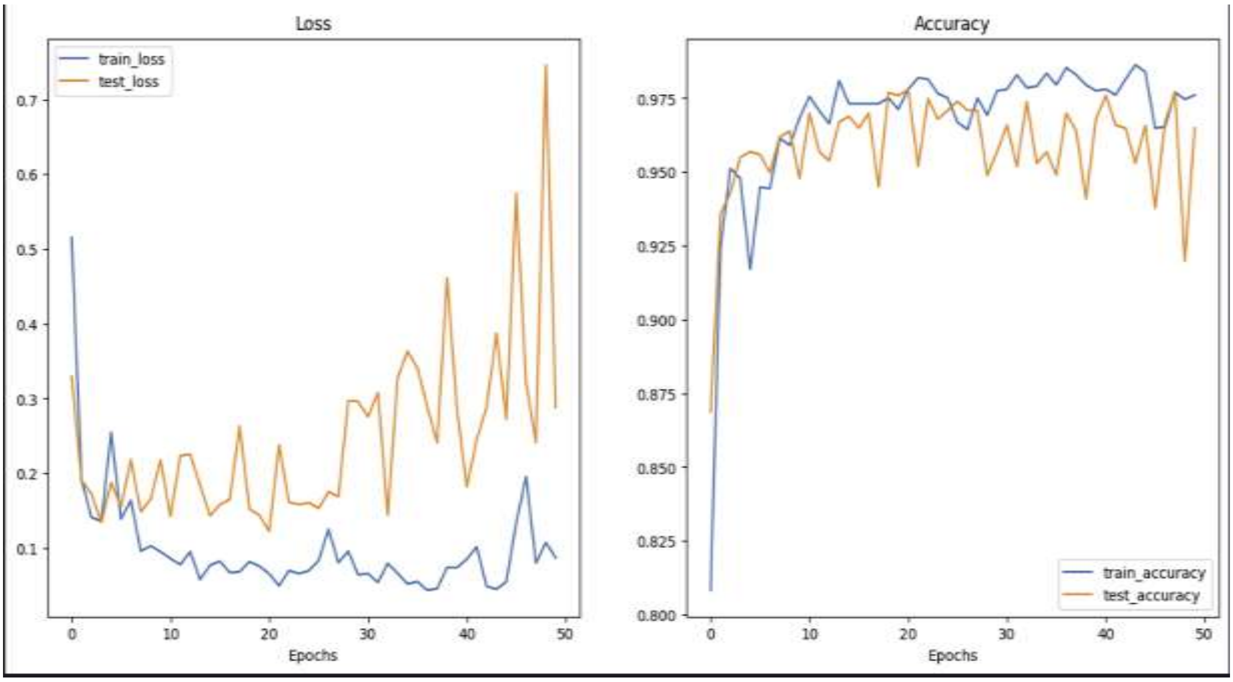

Figure 4.5: The train-loss, test- loss, train-accuracy and test-accuracy for model AlexNet.

figures 4.6, 4.7, and 4.8 shows all rates for the remaining models in the study, namely VGG-16, EfficientNetB7, and VGG-19 respectively.

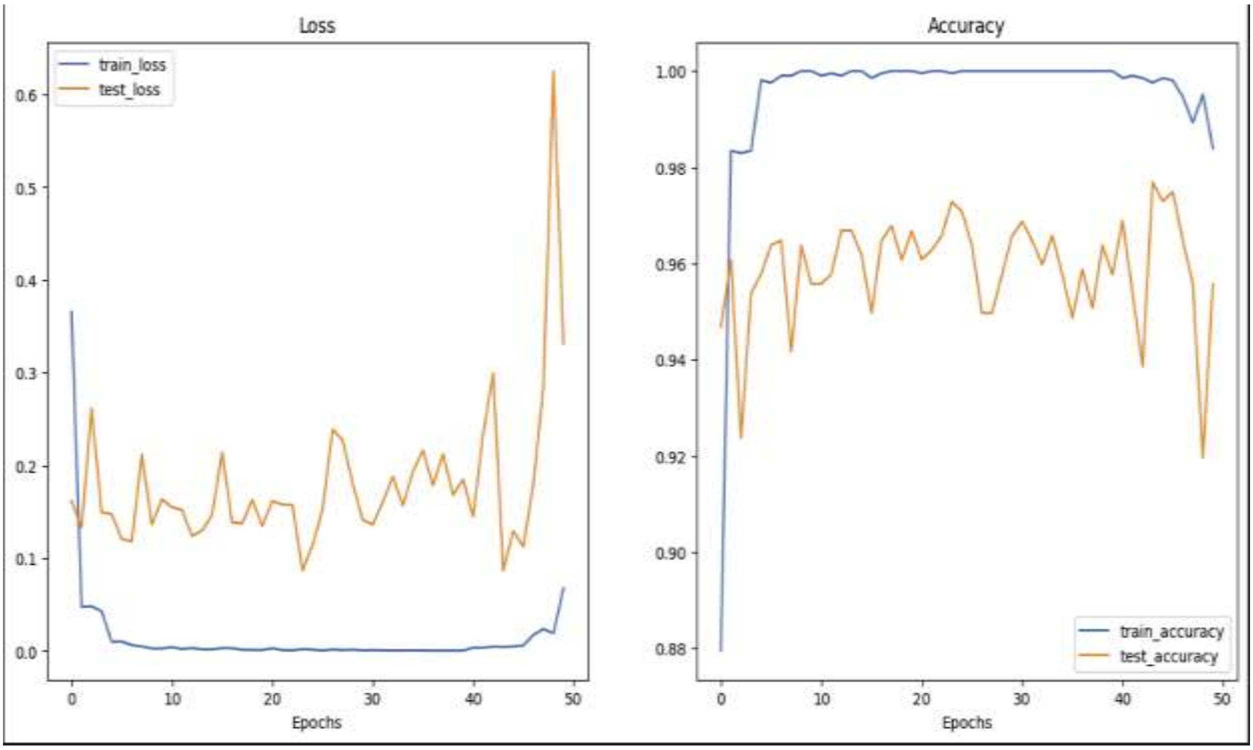

Figure 4.6: The train-loss, test- loss, train-accuracy and test-accuracy for model VGG-16.

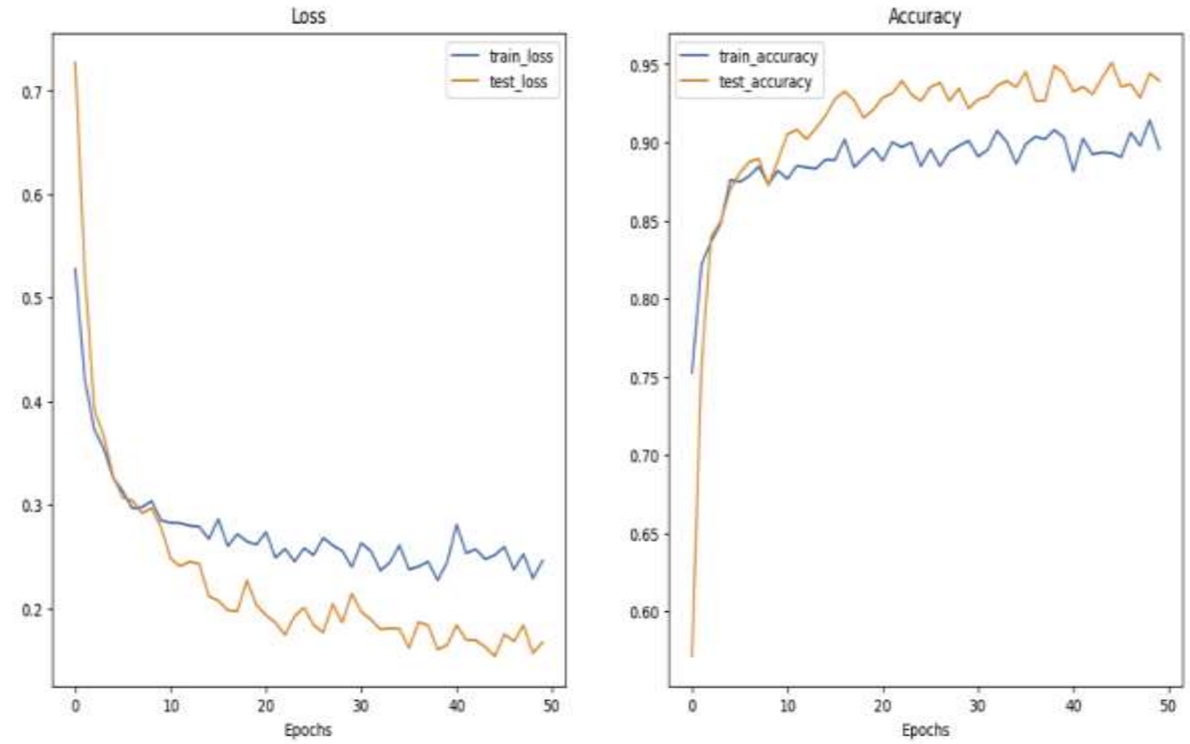

Figure 4.7: The train-loss, test- loss, train-accuracy and test-accuracy for model EfficientNetB7 .

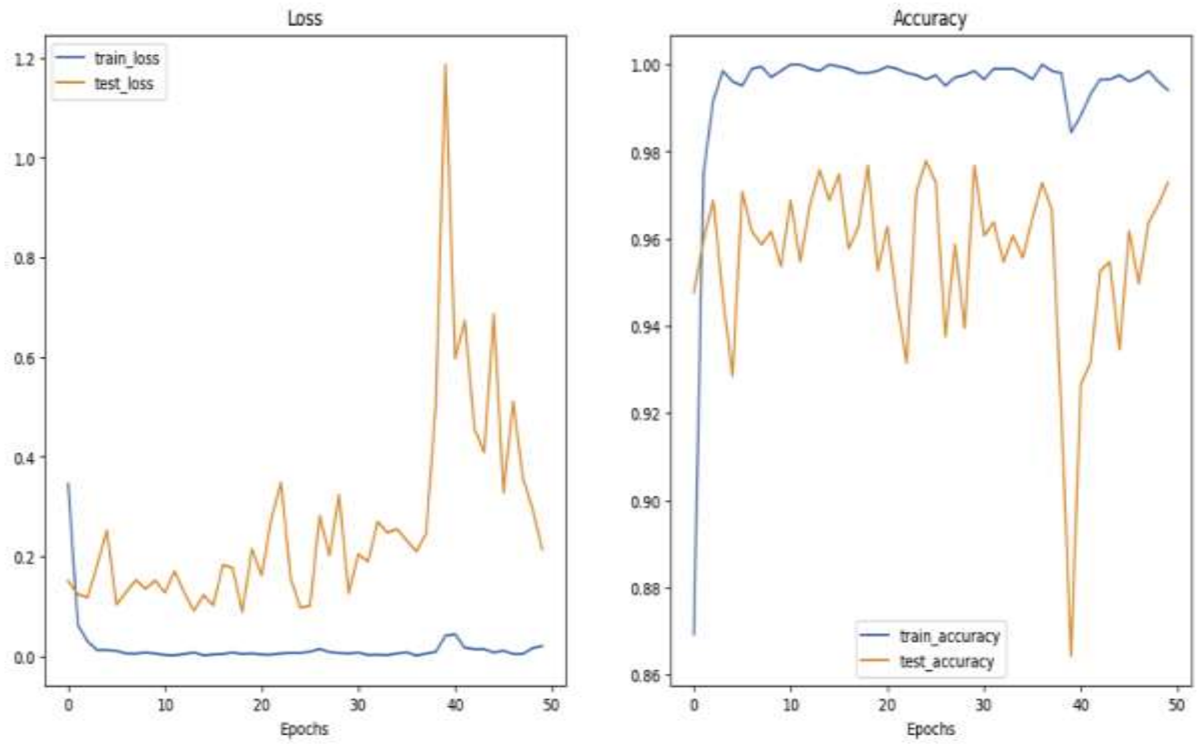

Figure 4.8: The train-loss, test- loss, train-accuracy and test-accuracy for model VGG-19.

4.3 Visualized output

The output in this experiment included images from the most false predictions where the models classified the images incorrectly with a high probability. The figures below show several example images from the output, where each image had a corresponding text. The text explained if the image was actually affected or not, the classification of the image was correct or not, and the probability.

The figure 4.9 shows output images of the model ReaNet-50, for example the first image was actually affected, but the model classified it incorrectly as an unaffected image with a high probability of 0.998. On the other hand, the figure 5.0 shows output images of the model AlexNet, where the first output image was actually affected and the model incorrectly classified it as unaffected and the model was almost sure with its classification with a probability of 1.000.

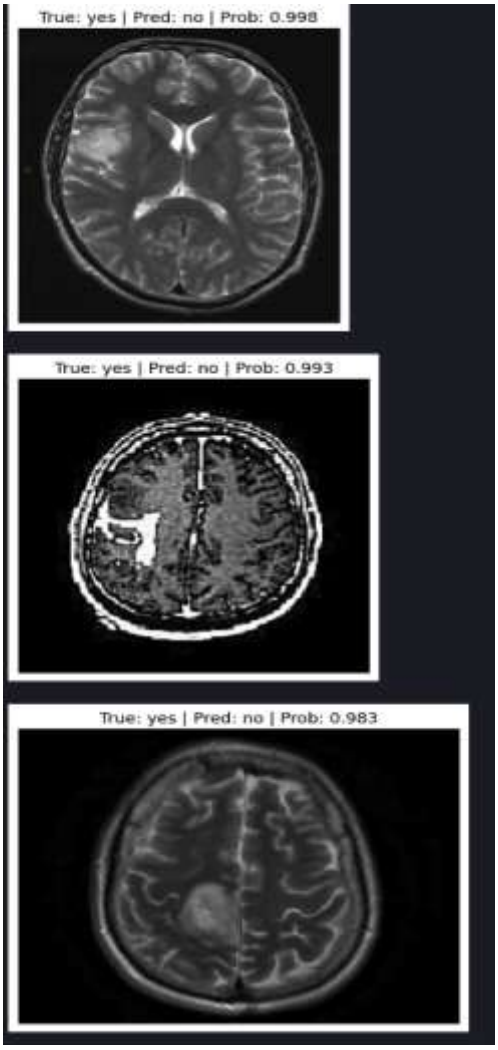

Figure 4.9: Output images for model ResNet-50

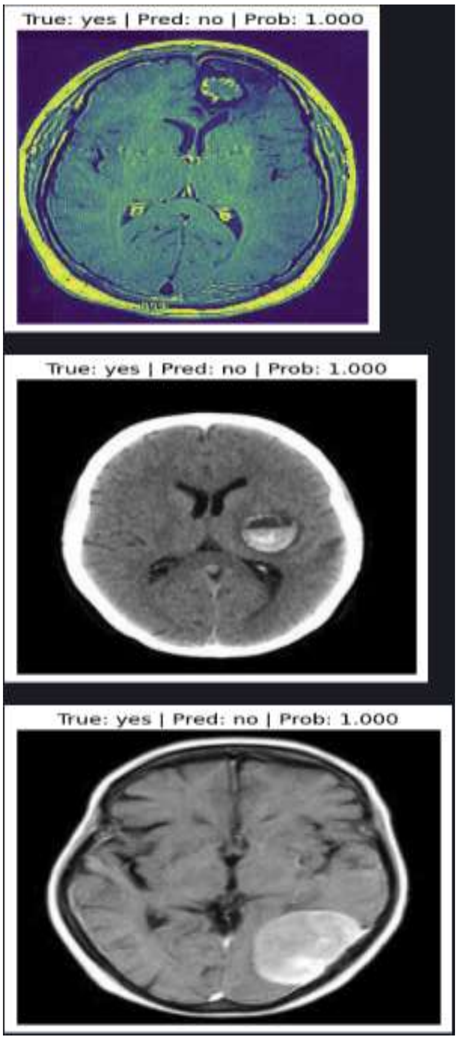

Figure 5.0: Output images for model AlexNet

Similarly, the following figures 5.1, 5.2, and 5.3 show output images of the remaining models, VGG-16, EfficientNetB7, VGG-19.

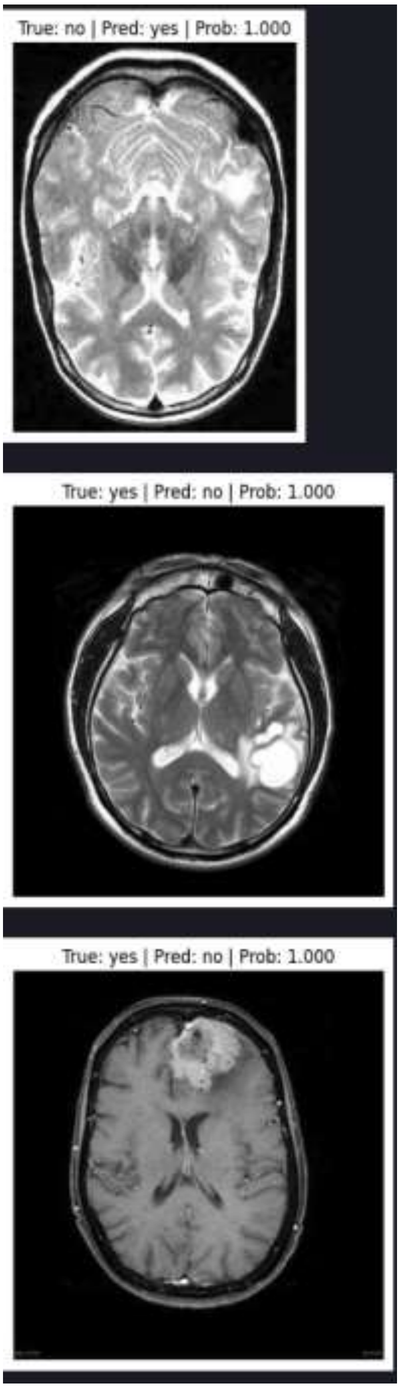

Figure5.1: Output for VGG-16

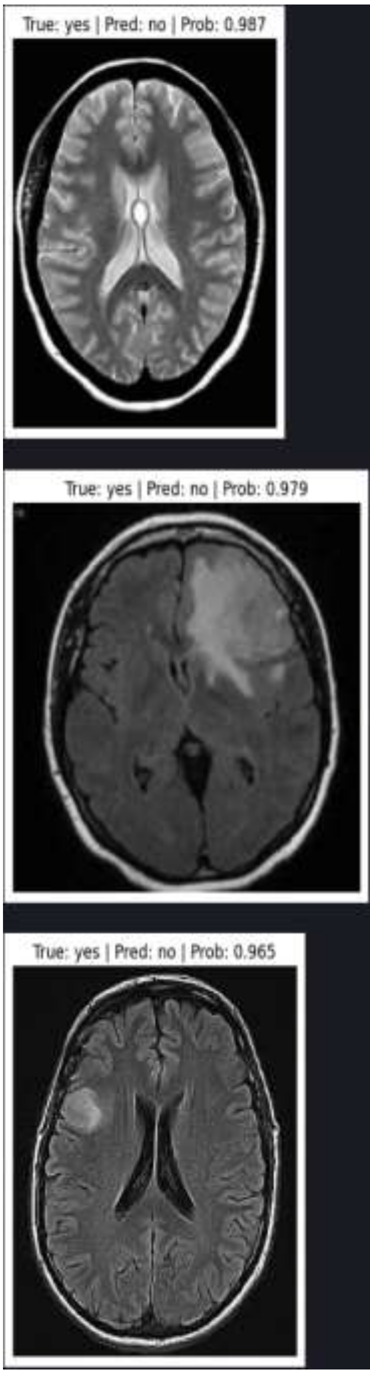

Figure 5.2: EfficientNetB7

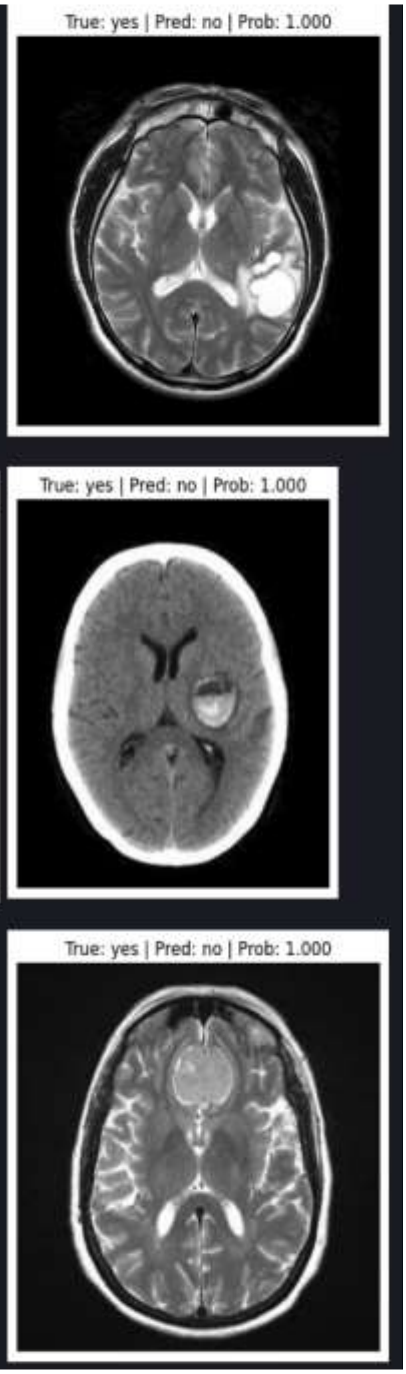

Figure 5.3: VGG-19

4.4 Models' training time

Table 4.2 presents the values of the training time for each model, where the model ResNet-50 took the shortest training time, 50 minutes, while the EfficientNetB7 model took the longest training time, 280 minutes. The training time for each model is also illustrated in the figure 5.8 below as a bar chart.

Table 4.2 Training time for each model in minutes.

Model	ResNet-50	AlexNet	VGG-16	VGG-19	EfficientNetB7
Time	50	90	60	75	280

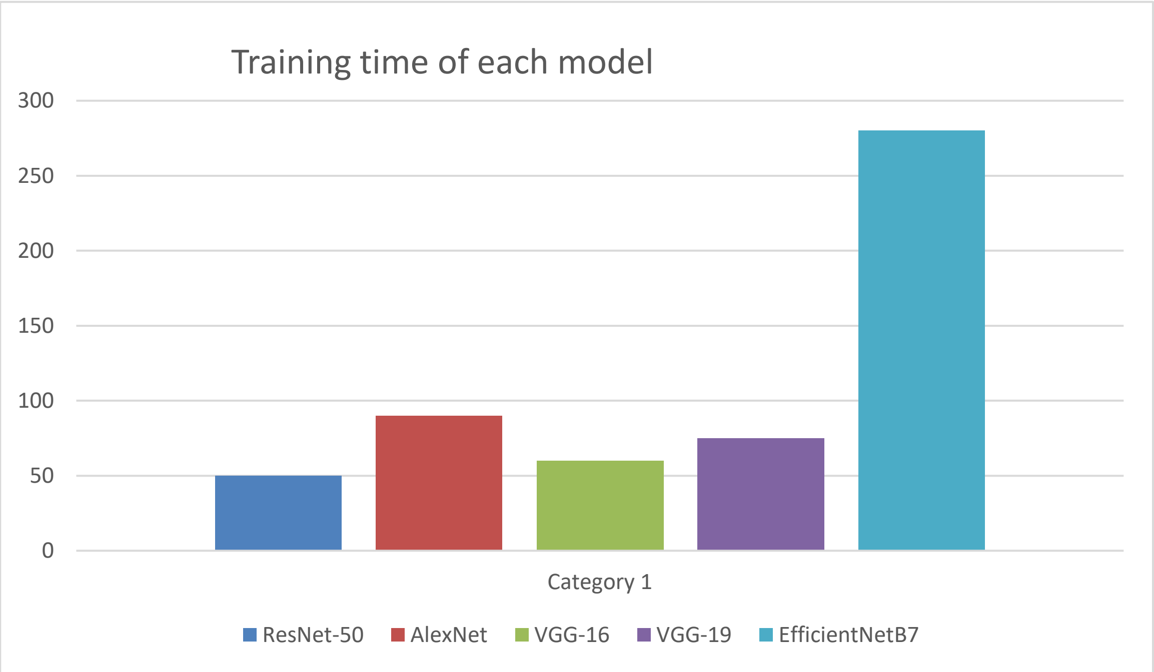

Figure 5.4: Bar chart of the training time for each model

5 Analysis and discussion

This chapter presents a discussion and analysis of this study comparing five CNN models classifying brain tumor images in terms of performance. The authors discuss the experiment results and the used methods' effect on the classification's results of each CNN model. Moreover, the authors discuss in this chapter the economic and social impact of using CNN in the healthcare sector.

5.1 Method analysis

The compared models obtained high results in accuracy, which implies that the used hyperparameters under the experiment had probable values. When training the models with 50 number of epochs, the models had the best obtained accuracy rate and had the lowest loss rate. The training time for all the models were not short and consumed the GPU and the memory using was maximized. The authors stopped training the models with higher number because of the overfitting risk. However, the authors tested to train only one model with 75 number of epochs in order to ensure that 50 epochs are the most suitable value. The tested model was ResNet-50 and resulted in lower accuracy rate than the rate in 50 epochs and had a higher loss rate. Which ensures that 50 epoch is the best epoch number in this study, and it is not needed to train more models with 75 epoch risking overfitting and overload the GPU.

Moreover, data augmentation and transformation helped enriching the dataset and improve the model's ability to learn which in turn helps in getting more precise results in terms of accuracy. As well as using transfer learning assisted the authors by taking advantage of pre-trained models, instead of building a new model from the beginning. The experiment gave high accuracy results for almost all models, which implies that the hyperparameters used in the experiment had appropriate values. Furthermore, using the loss function to calculate the loss rate helped in observing the model's behaviour with different epoch numbers by illustrating how a model's loss is affected by changing the number of epochs. Additionally, the confusion matrix calculated the accuracy for each model and illustrated how accurate each model is by specifying the correctly and incorrectly classified images.

5.2 Result analysis

The experiment conducted in this study comparing five different CNN models in classifying brain tumor images was successful. The results of the experiment shows that all the investigated models achieved high accuracy in the image classification domain. Moreover, all the accuracy values, calculated by using the Confusion matrix, ranged between 93% and 97%. Furthermore, the model VGG-19 achieved the highest performance with 97% accuracy, while the model EfficientNetB7 achieved the lowest performance with 93% accuracy.

In [37], the study compared four models including VGG-16, ResNet-50, Inception-V3, and VGG-19, where the models achieved accuracy 99%, 97.92%, 81.4%, and 99.48 % respectively. Compared to our study, where VGG-16 achieved 95% accuracy, ResNet-50 achieved 94% accuracy, and VGG-19 achieved 97% accuracy. In our study and in [37], the same optimizer was used in conducting the experiment, and data augmentation was relatively similar. The results of both studies were close which supports the reliability of this study.

5.3 Performance and training time

The performance of investigated models in this study were evaluated based on two factors, namely accuracy and time. The authors started the experiment using a CPU processor to train the models.

However, the training time was relatively long, for instance, ResNet-50 took 180 minutes training time. As a result, the authors instead conducted the experiment on a GPU processor to shorten training time and save energy, where the ResNet-50 took 50 minutes.

The performance of a model is a trade off between the accuracy and the training time. The model with the best performance, VGG-19, achieved 97% accuracy with a corresponding training time of 75 minutes, while the model EfficientNetB7 achieved 93% accuracy with a corresponding training time of 280 minutes (the longest training time). On the other hand, ResNet-50 achieved 94% accuracy with a corresponding 50 minutes training time (the shortest training time). The authors find that VGG-19 achieved the best performance, although its training time was not the shortest but it outperformed other models in terms of accuracy with a considerable difference. The high accuracy leads to high reliability in the model's classification ability and the short training time leads to resources and energy saving.

5.4 Loss and accuracy rates

The authors tested several numbers of epochs in order to examine the best corresponding train-loss, test-loss, train-accuracy, and test-accuracy rates for each model. The authors started training all the models with 10 epochs, but the accuracy values were relatively low, for example, ResNet-50 achieved 81% test-accuracy. Then the authors increased the number of epochs and observed that the train-accuracy and test-accuracy increased as well as that the train-loss and test-loss decreased. For example, the test-accuracy value of ResNet-50 was 94% using 50 epochs. However, the tested numbers of epochs did not pass 50 epochs in order to avoid overfitting. For instance, in figure 4.4 it is observed that the test-loss and train-loss rates in ResNet-50 was high at a number of epochs of 10, whereas the rates decreased at the number of epochs 50. On the other hand, the train-accuracy and test-accuracy rates were low at a number of epochs 10, whereas the rates increased at number of epochs 50.

5.5 Validity and reliability

This experiment was performed using five pre-trained models on the ImageNet dataset, all models showed high efficiency in brain tumor image classification. The authors took advantage of these models' features and the knowledge they learned as a start point. Then the authors performed several tests to identify which hyperparameters will be used in the study. For instance, in order to determine the right batch size for the models ResNet-50, VGG-16, AlexNet, and VGG-19, the authors began with 5, 10, 20, and 50. The batch size value 50 gave the best results. On the other hand, when training the model EfficientNetB7 with batch size value 50, the training process was forcibly stopped and a memory error occurred. This was because of the architectural complexity of this model. In order to solve this issue, the authors changed batch size to 30. Whereas, the other hyperparameters including learning rate that has value 0.001, number of epochs that was 50 , and Adam optimizer helped in achieving the best performance of each model, and consequently reliable results.

5.6 The economic, social, ethical, and environmental consequences

The high results the CNN models achieved in this study confirm that CNN is a leading algorithm for image classification. The returns of applying this algorithm in the medical field is discussed below.

A digital image is considered to be data from an image publication perspective. With the use of AI in general and CNN in particular in the medical field, Data protection regulations from an ethical perspective, should control the usage of this data.

Imaging classification might cause ethical issues when the data and images are not handled properly. Ethical Consent Violations occur when images are utilized in research and experiments without the owner's consent, whether they are a person or an organization. On the other hand, if images and data are handled correctly, it can be used in such beneficial algorithms like CNN, leading to improved medical treatments and faster decision-making processes.

As for the economic and social aspects, CNN can provide a variety of benefits to business development by saving cost and time. However, this can be achieved by providing high information accuracy to help medical practitioners in decision making and reducing disease diagnosing time. Accurate diagnosis and decisions implies reducing the human errors and medical mistakes in the healthcare sector. Hence, using CNN helps save financial expenses for both individuals and society by decreasing the number of employees and administrative jobs. Furthermore, by delivering more precise and rapid diagnosis, employing CNN may potentially save more lives or give patients more time to recover.

Image classification in general and CNN in particular have environmental effects in society. The sustainable solutions provided by CNN can support humans and save many sorts of resources. For instance, CNNs have the capacity to work for a long period of time at the same high level.

However, each CNN model requires a strong processor to get the highest performance and accuracy. In this study ,the authors started training CNN models with a CPU processor, where the training process was very slow and consumed a large amount of energy. For example, training the ResNet-50 model consumed 180 minutes taking into consideration that the model has a simple architecture compared to other models such as EfficientNetB7. Then the authors used another computer with a good GPU processor to continue the study. The difference in training time between the CPU and GPU was very clear. For example, the ResNet-50 took 180 minutes with the CPU while it took 50 minutes with the GPU processor. This implies that the GPU has the ability to process huge quantities of data in a relatively brief period of time, which can be translated to significant energy saving. As a result, CNN with a good processor classifies images more accurately than humans, providing improved performance and better outcomes, bringing additional advantages to society such as saving energy and resources.

6 Conclusion

The aim of this study is to examine and compare the performance of five different CNN models, namely VGG-16, VGG-19, ResNet-50, EfficientnetB7, and AlexNet by training and testing the models to classify brain tumor images. The conducted experiment resulted in VGG-19 achieving best performance with 97% accuracy.

Additionally, the study aimed to determine how AI affects resources use and healthcare costs. Employing CNN models as an AI technology in classifying brain cancer and in other medical fields may have benefits in terms of productivity and efficiency in the healthcare sector. It may also assist medical practitioners and professionals in making decisions by providing them with high information accuracy to depend on. Accordingly, increasing the number of lives saved, financial savings, and hours saved up for healthcare practitioners.

6.1 Future work

This study investigates the performance of five different CNN models in classifying brain tumor images. The study was conducted with a relatively limited dataset and specific hyperparameters. Further investigations comparing more CNN models may help in determining the best CNN model in classifying brain tumor images. Additionally, using larger dataset or more hyperparameters may also achieve higher performance. However, further investigations will not be carried out in this paper.

7 References

1. World Health Organization: WHO. (Jul. 12, 2019). Cancer [Internet]. [cited 2023 Mars 29]. Available from: https://www.who.int/health-topics/cancer/
2. R. L. Siegel, K. A. Miller, H. E. Fuchs, A. Jemal. Cancer statistics, 2022. CA: A Cancer Journal for Clinicians [Internet], vol. 72, issue 1, pp. 7–33, (Jan. 2022), doi: 10.3322/caac.21708. [cited 2023 Mars 29]. Available from: https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21708
3. Cancer.Net. (Mar. 09, 2023). Brain Tumor – Statistics [Internet]. [cited 2022 Mars 28]. Available from: https://www.cancer.net/cancer-types/brain-tumor/statistics#/
4. L. Thau, V. Reddy, P Singh. (Oct. 10, 2022). Anatomy, Central Nervous System [Internet]. StatPearls – NCBI Bookshelf.[cited 2023 Mars 28]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK542179/
5. Johns Hopkins Medicine. Brain Tumors and Brain Cancer [Internet]. [cited 2023 Mars 30]. Available from: https://www.hopkinsmedicine.org/health/conditions-and-diseases/brain-tumor/
6. C. Penfold, A. J. Joannides, J. Bell, F. M. Walter. Diagnosing adult primary brain tumours: can we do better?. British Journal of General Practice [Internet], vol. 67, no. 659, pp. 278–279, Jun. 2017, doi: 10.3399/bjgp17x691277. [Cited 2023 Mars 31]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5442949/#b1/
7. A. S. Peddinti, S. Maloji, K. Manepalli. Evolution in diagnosis and detection of brain tumor – review. Journal of Physics [Internet], vol. 2115, no. 1, p. 012039, Nov. 2021, doi: 10.1088/1742-6596/2115/1/012039. [Cited 2023 Mars 31]. Available from: https://iopscience.iop.org/article/10.1088/1742-6596/2115/1/012039/pdf
8. Gomez-Gonzalez E, Gomez E, Marquez-Rivas et al, "Artificial Intelligence in Medicine and Healthcare: a review and classification of current and near-future applications and their ethical and social impact" [Internet]. arXiv:2001.09778, 2020. [cited 2023 Mars 31]. Available from: https://arxiv.org/ftp/arxiv/papers/2001/2001.09778.pdf/
9. A. K. Philip, B. A. Samuel, S. Bhatia, S. a. M. Khalifa, H. R. ElSeedi, "Artificial Intelligence and Precision Medicine: A New Frontier for the Treatment of Brain Tumors [Internet]," Life, vol. 13, no. 1, p. 24, Dec. 22, 2022. doi: 10.3390/life13010024. [cited 2023 Mars 31]. Available from: https://www.mdpi.com/2075-1729/13/1/24
10. S. Saha. (Nov. 16, 2022). A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way [Internet]. [cited 2023 Mars 31]. Available from: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53
11. A. Biswal.(Apr. 01, 2023). Convolutional Neural Network Tutorial. Simplilearn [Internet]. [cited 2023 Mars 31]. Available from: https://www.simplilearn.com/tutorials/deep-learning-tutorial/convolutional-neural-network/

12. R. Manikandan, A. Kumar, D. Gupta. “Hybrid computational intelligence for healthcare and disease diagnosis.” ScienceDirect [Internet], Jan. 2020, doi: 10.1016/b978-0-12-818699-2.00006-8. [cited 2023 Mars 31]. Available from: https://www.sciencedirect.com/science/article/pii/B9780128186992000068/
13. T. Abraham, A. Todd, D. A. Orringer, and R. Levenson. “Applications of artificial intelligence for image enhancement in pathology.” ScienceDirect [Internet], pp. 119–148, Jan. 2021, doi: 10.1016/b978-0-323-67538-3.00007-5. [cited 2023 Mars 31]. Available from: https://www.sciencedirect.com/science/article/pii/B9780323675383000075#bib2/
14. S. Brath. Artificial Intelligence (AI) In Healthcare & Hospitals. ForeSeeMedical [Internet]. [cited 2023 April 02]. Available from : https://www.foreseemed.com/artificial-intelligence-in-healthcare#
15. T. Davenport, R. Kalakota, “The potential for artificial intelligence in healthcare,” Future Healthcare Journal [Internet], vol. 6, no. 2, pp. 94–98, Jun. 2019, doi: 10.7861/futurehosp.6-2-94. [cited 2023 April 02]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181/#CIT0004
16. R. Kumari. (Mar. 05. 2023). Role of Artificial Intelligence in Healthcare. Analytics Steps [Internet]. [cited 2023 April 03]. Available from: https://www.analyticssteps.com/blogs/role-artificial-intelligence-healthcare
17. A. Khan. (Mar. 02. 2023). Artificial Intelligence in Healthcare - AI Applications and Uses. Intellipaat Blog[Internet]. [cited 2023 April 04]. Available from: https://intellipaat.com/blog/artificial-intelligence-in-healthcare
18. S. Yang, F. Zhu, X. Ling, Q. Liu, and P. Zhao. “Intelligent Health Care: Applications of Deep Learning in Computational Medicine”. Frontiers in Genetics [Internet], vol. 12, Apr. 2021, doi: 10.3389/fgene.2021.607471. [cited 2023 April 05]. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2021.607471/full
19. V. Fortunati. (May. 29. 2019). How does deep learning in radiology work?. Quantib [Internet]. [cited 2023. April. 05]. Available from: https://www.quantib.com/blog/how-does-deep-learning-work-in-radiology
20. GeeksforGeeks. (2023). Introduction to Deep Learning. GeeksforGeeks [Internet]. [Cited 2023 May 10]. Available from: https://www.geeksforgeeks.org/introduction-deep-learning/
21. J. Huang, J. Li, Z. Li, Z. Zhu, C. Shen, G. Qi, and G. Yu. “Detection of Diseases Using Machine Learning Image Recognition Technology in Artificial Intelligence,” Computational Intelligence and Neuroscience [Internet], vol. 2022, pp. 1–14, Apr. 2022, doi: 10.1155/2022/5658641. [cited 2023 April 06]. Available from: https://www.hindawi.com/journals/cin/2022/5658641/
22. T. Wood. (Jul. 2020). Convolutional Neural Network. DeepAI [Internet]. [cited 2023 April 07]. Available from: https://deepai.org/machine-learning-glossary-and-terms/convolutional-neural-network
23. R. Yamashita, M. Nishio, R. K. G DO, and K. Togashi. “Convolutional neural networks: an overview and application in radiology”. Insights Into Imaging [Internett], vol. 9, no. 4, pp. 611–629, Jun. 2018, doi: 10.1007/s13244-018-0639-9. [cited 2023 April 07]. Available from: https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9

24. K. Smeda. (Mar. 30. 2022). Understand the architecture of CNN. Towards Data Science [Internet]. [cited 2023 April 08]. Available from: https://towardsdatascience.com/understand-the-architecture-of-cnn-90a25e244c7
25. M. S. Ali. (Jun. 28. 2022). Flattening CNN layers for Neural Network and basic concepts. Medium [Internet]. [cited 2023 April 08]. Available from: https://medium.com/@muhammadshoaibali/flattening-cnn-layers-for-neural-network-694a232eda6a
26. A. P. Bhandari, J. Koppen, and M. Agzarian. "Convolutional neural networks for brain tumour segmentation". Insights Into Imaging [Internet], vol. 11, no. 1, Jun. 2020, doi: 10.1186/s13244-020-00869-4. [cited 2023 April 08]. Available from: https://insightsimaging.springeropen.com/articles/10.1186/s13244-020-00869-4/figures/4
27. Y. Yanagisawa, K. Shido, K. Kojima, and K. Yamasaki. "Convolutional neural network-based skin image segmentation model to improve classification of skin diseases in conventional and non-standardized picture images". Journal of Dermatological Science [Internet], vol. 109, no. 1, pp. 30–36, Jan. 2023, doi: 10.1016/j.jdermsci.2023.01.005. [cited 2023 April 09]. Available from: https://www.sciencedirect.com/science/article/pii/S0923181123000051
28. S. Arooj, S. U. Rehman, A. Imran, A. Almuhaimeed, A. K. Alzahrani, and A. Alzahrani. "A Deep Convolutional Neural Network for the Early Detection of Heart Disease". Biomedi-cines [Internet], vol. 10, no. 11, p. 2796, Nov. 2022, doi: 10.3390/biomedicines10112796. [cited 2023 April 09]. Available from: https://www.mdpi.com/2227-9059/10/11/2796
29. Z. M. Lonseko, P. E. Adjei, W.Du, C. Luo, D. Hu, L. Zhu, T. Gan, and N. Rao. "Gastrointestinal Disease Classification in Endoscopic Images Using Attention-Guided Convolutional Neural Networks". Applied Sciences [Internet], vol. 11, no. 23, p. 11136, Nov. 2021, doi: 10.3390/app112311136. [cited 2023 April 09]. Available from: https://www.mdpi.com/2076-3417/11/23/11136#B1-applsci-11-11136
30. A. Spatharou, S. Hieronimus, J. Jenkins. (Mar. 10. 2020). Transforming healthcare with AI: The impact on the workforce and organizations. McKinsey & Company [Internet]. [cit-ed 2023 April 10]. Available from: https://www.mckinsey.com/industries/healthcare/our-insights/transforming-healthcare-with-ai
31. Medtronic, ResMed. The socio-economic impact of AI in healthcare. MedTech Europe from diagnosis cure [Internet]. Oct. 2020. [cited 2023 2023 April 10]. Available from: https://www.medtecheurope.org/wp-content/uploads/2020/10/mte-ai_impact-in-healthcare_oct2020_report.pdf
32. American Cancer Society. (Jan. 03. 2018). Economic Impact of Cancer [Internet]. [cited 2023 April 11]. Available from: https://www.cancer.org/treatment/finding-and-paying-for-treatment/managing-costs/economic-impact-of-cancer.html
33. M. C. A. Bakke. (Sep. 27. 2022). Using AI during cancer-screening results in more recommended controls. University of Oslo [Internet]. [cited 2023 April 10]. Available from: https://www.med.uio.no/helsam/english/research/news-and-events/news/2022/using-artificial-intelligence-during-cancer-screen.html

34. E. J. Topol. “High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine [Internet], vol. 25, no. 1, pp. 44–56, Jan. 2019, doi: 10.1038/s41591-018-0300-7. [cited 2023 April 11]. Available from: https://www.nature.com/articles/s41591-018-0300-7
35. J. Zhang, X. Shen, T. Zhuo, and H. Zhou. (Dec. 25. 2017). Brain Tumor Segmentation Based on Refined Fully Convolutional Neural Networks with A Hierarchical Dice Loss. arXiv.org [Internet]. [cited 2023 April 11]. Available from: https://arxiv.org/abs/1712.09093
36. J. Seetha, S. S. Raja. “Brain Tumor Classification Using Convolutional Neural Networks”. Biomedical & Pharmacology Journal [Internet], Vol. 11(3), p. 1457-1461, Sep, 2018. [cited 2023 April 12]. Available from: http://proficientsolutions.in/wp-content/uploads/2022/09/BPJ_Vol_11_No_3_p_1457-1461.pdf
37. S. Krishnapriya, and Y. Karuna. (Apr. 20. 2023). Pre-trained deep learning models for brain MRI image classification. Frontiers in Human Neuroscience [Internet]. [cited 2023 May 08]. Available from: https://doi.org/10.3389/fnhum.2023.1150120
38. M. Talo, U. B. Baloglu, Ö. T. Yıldırım, and U. R. Acharya, “Application of deep transfer learning for automated brain abnormality classification using MR images,” Cognitive Sys-tems Research, vol. 54, pp. 176–188, May 2019, doi: 10.1016/j.cogsys.2018.12.007. [cited 2023 April 13]. Available from: https://www.sciencedirect.com/science/article/pii/S1389041718310933
39. A. Rehman, S. Naz, M. I. Razzak, F. Akram, and M. Imran. “A Deep Learning-Based Framework for Automatic Brain Tumors Classification Using Transfer Learning”. Circuits Systems and Signal Processing [Internet], vol. 39, no. 2, pp. 757–775, Jan. 2020, doi: 10.1007/s00034-019-01246-3. [cited 2023 April 15]. Available from: https://dl.acm.org/doi/10.1007/s00034-019-01246-3
40. A. Çinar, and M. Yildirim, “Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture”. Medical Hypotheses [Internet], vol. 139, p. 109684, Jun. 2020, doi: 10.1016/j.mehy.2020.109684. [cited 2023 April 15]. Available from: https://www.sciencedirect.com/science/article/pii/S0306987720301717
41. H. A. Khan, W. Jue, M. Mushtaq, and M. U. Mushtaq. “Brain tumor classification in MRI image using convolutional neural network”. Mathematical Biosciences and Engineering [Internet], vol. 17, no. 5, pp. 6203-6216, Sep. 2020, doi: 10.3934/mbe.2020328. [cited 2023 April 16]. Available from: http://www.aimspress.com/fileOther/PDF/MBE/mbe-17-05-328.pdf
42. A. Saleh, R. Sukaik and S. S. Abu-Naser. "Brain Tumor Classification Using Deep Learn-ing". 2020 International Conference on Assistive and Rehabilitation Technologies (iCare-Tech) [Internet], 2020, pp. 131-136, doi: 10.1109/iCareTech49914.2020.00032. [cited 2023 April 18]. Available from: https://ieeexplore.ieee.org/abstract/document/9328072/authors#authors
43. A. Mittal and D. Kumar. “AiCNNs (Artificially-integrated Convolutional Neural Networks) for Brain Tumor Prediction”. EAI Endorsed Transactions on Pervasive Health and Technology [Internet], vol. 5, no. 17, p. 161976, Feb. 2019, doi: 10.4108/eai.12-2-2019.161976. [cited 2023 Maj 03]. Available from: https://eudl.eu/doi/10.4108/eai.12-2-2019.161976

44. A. Hamada. (Nov. 14. 2021). Br35H :: Brain Tumor Detection 2020. Kaggle [Internet]. [cit-ed 2023 April 10]. Available from: https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection
45. P. Marcelino. (May. 06. 2022). Transfer learning from pre-trained models - Towards Data Science. Medium [Internet]. [cited 2023 May 03]. Available from: https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751
46. G. Boesch. (Mar. 2023). VGG Very Deep Convolutional Networks (VGGNet) – What you need to know. viso.ai [Internet]. [cited 2023 May 04]. Available from: https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks
47. A. Sachan. (Nov. 2019). The Free lunch of 2019 for convolutional neural networks. CV-Tricks.com [Internet]. [cited 2023 May 7]. Available from: https://cv-tricks.com/network-architecture/efficientnets-the-free-lunch-of-2019-for-convolutional-neural-networks/
48. Kaggle. ImageNet Winning CNN Architectures (ILSVRC) [Internet]. [cited 2023 May 04]. Available from: https://www.kaggle.com/getting-started/149448
49. Finetuning Torchvision Models — PyTorch Tutorials 1.2.0 documentation. Pytorch.org [Internet]. [cited 2023 May 04]. Available from: https://pytorch.org/tutorials/beginner/finetuning_torchvision_models_tutorial.html
50. IBM Documentation. (Aug. 12. 2022). Hyperparameter tuning with Deep Learning Impact [Internet]. [cited 2023 May 07]. Available from: https://www.ibm.com/docs/en/wmla/1.2.3?topic=features-hyperparameter-tuning
51. javaTpoint. Overfitting and Underfitting in Machine Learning. Javatpoint [Internet]. [cited 2023 May 10]. Available from: https://www.javatpoint.com/overfitting-and-underfitting-in-machine-learning
52. J. Brownlee (Jan. 25. 2019). Understand the Impact of Learning Rate on Neural Network Performance. MachineLearningMastery [Internet]. [cited 2023 May 10]. Available from: https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/
53. A. Kulkarni, D. Chong, F. A. Batarseh. (2021). Foundations of data imbalance and solutions for a data democracy. In Elsevier eBooks (pp. 83–106) [Internet]. [cited 2023 May 11]. Available from: https://doi.org/10.1016/b978-0-12-818366-3.00005-8
54. GeeksforGeeks. (2023). Visualize Confusion Matrix Using Caret Package in R. GeeksforGeeks [Internet]. [cited 2023 May 11]. Available from: https://www.geeksforgeeks.org/visualize-confusion-matrix-using-caret-package-in-r/
55. D. P. Kingma, J. Ba. (2014, December 22). Adam: A Method for Stochastic Optimization. arXiv.org [Internet]. [cited 2023 May 13]. Available from: https://arxiv.org/abs/1412.6980

TRITA – TRITA-CBH-GRU-2023:105

www.kth.se

TRITA – EECS-EX-2023:357