

**BRAIN TUMOR PROGRESS CLASSIFICATION  
USING DEEP LEARNING**

**A THESIS SUBMITTED TO THE GRADUATE  
SCHOOL OF APPLIED SCIENCES  
OF  
NEAR EAST UNIVERSITY**

**By  
HEDI SYAMAND AZAT**

**In Partial Fulfillment of the Requirements for  
the Degree of Master of Science  
in  
Computer Engineering**

**NICOSIA, 2021**

**HEDI SYAMAND AZAT**

**BRAIN TUMOR PROGRESS CLASSIFICATION USING DEEP LEARNING**

**NEU  
2021**

**BRAIN TUMOR PROGRESS  
CLASSIFICATION USING DEEP LEARNING**

**A THESIS SUBMITTED TO THE  
GRADUATE SCHOOL OF APPLIED SCIENCES  
OF  
NEAR EAST UNIVERSITY**

**By  
HEDI SYAMAND AZAT**

**In Partial Fulfillment of the Requirements for  
the Degree of Master of Science  
in  
Computer Engineering**

**NICOSIA, 2021**

**Hedi Syamand AZAT: BRAIN TUMOR PROGRESS CLASSIFICATION USING DEEP LEARNING**

**Approval of Director of Graduate School of Applied Sciences**



**Prof. Dr. Nadire ÇAVUŞ**

**We certify this thesis is satisfactory for the award of the degree of Masters of Sciences in Computer Engineering**

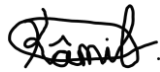
**Examining Committee in Charge:**

Assoc. Prof. Dr. Boran ŞEKEROĞLU



Supervisor, Department of Information Systems Engineering, NEU

Assoc. Prof. Dr. Kamil Dimililer



Committee Chairman, Department of Automotive Engineering, NEU

Assist.Prof. Dr. Elbrus Imanov



Department of Computer Engineering, NEU

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: HEDI, AZAT

Signature: 

Date: 18/01/2021

## **ACKNOWLEDGMENTS**

I would like to express my deep and sincere gratitude to my research supervisor, Assoc. Prof. Dr. Boran ŐEKEROĐLU for his sincere and selfless support, prompt and useful advice during my research. He has given me a lifetime unforgettable memory of his benevolence, patience, intelligence, diligence and erudition. I owe him a lot of my success at this day and I cannot pay for his efforts. I am extremely grateful to my parents for their love, prayers, caring and supporting me spiritually throughout my life.

**To my parents**

## ABSTRACT

Artificial Intelligence tools have become common almost for every kind of application in human lives. The most critical and vital applications are the medical implementations where human health and life is at stake. Brain tumor classification and the progress following is a kind of medical applications. The medical doctors should follow the tumors' progress to observe the effect of the treatment and might direct the therapies about starting, continuing, or changing. In this thesis, the Brain Tumor Progress dataset was considered to classify the progress of brain tumors as initial and progress classes. The convolutional neural networks were implemented and trained using all slices of MRI images. Different convolutional neural network architectures, which were designed for testing the effect of adding/removing layers on classification rates, were considered. The results were validated using a 5-fold cross-validation technique, and four metrics were used as evaluation criteria. The optimal architecture was determined to classify brain tumor progress for a considered particular dataset, and the optimal classification rates were achieved by the median depth of the CNN's implemented in this thesis.

The optimal classification accuracy was achieved as 92.53% and obtained by a CNN with two convolutional (64x32 filters) and two dense layers (128x32 neurons).

**Keywords:** convolutional neural networks, brain tumor, progress follow, MRI images

## OZET

Yapay Zeka araçları, insan hayatındaki hemen hemen her türlü uygulamada yaygınlaşmıştır. En kritik ve hayati uygulamalar, insan sağlığının ve hayatının söz konusu olduğu tıbbi uygulamalardır. Beyin tümörünün sınıflandırılması ve ardından gelen gelişmeler bir tür tıbbi uygulamadır. Tıp doktorları, tedavinin etkisini gözlemlemek için tümörlerin ilerleyişini takip etmelidir ve tedavileri başlatma, devam etme veya değiştirme konusunda yönlendirebilir. Bu tezde, Beyin Tümörü İlerleme veri setinin beyin tümörlerinin ilerlemesini başlangıç ve ilerleme sınıfları olarak sınıflandırdığı düşünülmüştür. Evrişimli sinir ağları, MRI görüntülerinin tüm dilimlerini kullanarak uygulandı ve eğitildi. Katman ekleme / çıkarma işleminin sınıflandırma oranları üzerindeki etkisini test etmek için tasarlanmış farklı evrişimli sinir ağı mimarileri dikkate alınmıştır. Sonuçlar, 5 kat çapraz doğrulama tekniği kullanılarak doğrulanmış ve değerlendirme kriterleri olarak dört ölçüm kullanılmıştır. Optimal mimari, dikkate alınan belirli bir veri kümesi için beyin tümörünün ilerlemesini sınıflandırmak için belirlendi ve optimum sınıflandırma oranları, bu tezde uygulanan CNN'lerin medyan derinliği ile elde edildi.

Optimal sınıflandırma doğruluğu% 92.53 olarak elde edildi ve iki evrişimli (64x32 filtreli) ve iki yoğun katmanlı (128x32 nöron) bir CNN ile elde edildi.

**Anahtar kelimeler:** evrişimli sinir ağları, beyin tümörü, ilerleme takibi, MRI görüntüleri



## TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS</b> .....	<b>i</b>
<b>ABSTRACT</b> .....	<b>iii</b>
<b>OZET</b> .....	<b>iv</b>
<b>TABLE OF CONTENTS</b> .....	<b>v</b>
<b>LIST OF FIGURES</b> .....	<b>vii</b>
<b>LIST OF TABLES</b> .....	<b>ix</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>x</b>

### **CHAPTER 1: INTRODUCTION**

1.1	Background.....	1
1.2	Brain Tumors .....	1
1.3	Deep Learning .....	2
1.4	Problem of the Study .....	3
1.5	Aim of the Study .....	3
1.6	Thesis Outline.....	4

### **CHAPTER 2: LITERATURE REVIEW**

2.1	Overview .....	5
2.2	Deep learning in healthcare .....	5
2.3	Deep Learning in Diagnosis of Brain Tumor .....	8

### **CHAPTER 3: CONVOLUTIONAL NEURAL NETWORK**

3.1	Overview .....	11
3.2	Convolutional Neural Network .....	11
3.3	Convolutional Neural Network Structure.....	11
3.3.1	Convolution Layers .....	12
3.3.2	Pooling Layers.....	15
3.3.3	Fully Connected Layers.....	17
3.4	Convolutional Neural Network Training Process .....	17
3.5	Summary.....	18

## **CHAPTER 4: EXPERIMENTAL DESIGN**

4.1	Overview .....	19
4.2	Dataset and Data Preparation .....	19
4.2.1	Dataset .....	19
4.2.2	Data Preparation .....	20
4.3	Validation of the Experiments .....	21
4.3.1	Hold-Out Method .....	21
4.3.2	Random sub-sampling method .....	22
4.3.3	K-fold cross-validation method .....	22
4.4	Evaluation Metrics for Classification Problems .....	23
4.4.1	Accuracy .....	24
4.4.2	Sensitivity .....	24
4.4.3	Specificity .....	24
4.4.4	Receiver operating characteristic area under curve .....	24
4.5	Design of Experiments .....	25
4.6	Summary.....	29

## **CHAPTER 5: RESULTS AND DISCUSSIONS**

5.1	Overview .....	30
5.2	Results .....	30
5.2.1	Results of architecture “A” .....	30
5.2.2	Results of architecture “B” .....	31
5.2.3	Results of architecture “C” .....	32
5.2.4	Results of architecture “D” .....	33
5.2.5	Results of architecture “E” .....	34
5.2.6	Results of architecture “F” .....	35
5.2.7	Results of architecture “G” .....	36
5.2.8	Comparison of the CNN architectures.....	37
5.3	Discussions .....	40
5.4	Summary.....	42

## **CHAPTER 6: CONCLUSION**

<b>REFERENCES.....</b>	<b>44</b>
------------------------	-----------

<b>APPENDIX 1: ETHICAL APPROVAL DOCUMENT.....</b>	<b>50</b>
---	-----------

<b>APPENDIX 2: SIMILARITY REPORT.....</b>	<b>51</b>
---	-----------

## LIST OF FIGURES

<b>Figure 1. 1:</b> Illustration of deep learning and artificial neural network architectures (June-Goo Lee et al, 2017) .....	2
<b>Figure 3. 1:</b> Structure of convolutional neural network and the training process (Yamashita et al, 2018) .....	12
<b>Figure 3. 2:</b> Training times, comparison between ReLUs and tangent activations functions (Krizhevsky et al., 2012) .....	14
<b>Figure 3. 3:</b> Difference between Average and Max pooling (Waseem Rawat and Zenghui Wang, 2017) .....	16
<b>Figure 3. 4:</b> CNNs training process for image classification (Waseem Rawat and Zenghui Wang, 2017) .....	18
<b>Figure 4. 1:</b> Sample initial class brain MRI exam images.....	20
<b>Figure 4. 2:</b> Sample progress class brain MRI exam images .....	20
<b>Figure 4. 3:</b> Graphical representation of the Hold-Out method for 70% of training and 30% of testing Set.....	21
<b>Figure 4. 4:</b> The process of the Random Sub-sampling method .....	22
<b>Figure 4. 5:</b> The process of the 4-fold cross-validation method.....	23
<b>Figure 5. 1:</b> Results obtained by the architecture “A” .....	30
<b>Figure 5. 2:</b> Confusion matrix for the architecture “A” .....	31
<b>Figure 5. 3:</b> Results obtained by the architecture “B” .....	31
<b>Figure 5. 4:</b> Confusion matrix for the architecture “B” .....	32
<b>Figure 5. 5:</b> Results obtained by the architecture “C” .....	32
<b>Figure 5. 6:</b> Confusion matrix for the architecture “C” .....	33
<b>Figure 5. 7:</b> Results obtained by the architecture “D” .....	33
<b>Figure 5. 8:</b> Confusion matrix for the architecture “D” .....	34
<b>Figure 5. 9:</b> Results obtained by the architecture “E” .....	34
<b>Figure 5. 10:</b> Confusion matrix for the architecture “E” .....	35
<b>Figure 5. 11:</b> Results obtained by the architecture “F” .....	35
<b>Figure 5. 12:</b> Confusion matrix for the architecture “F” .....	36
<b>Figure 5. 13:</b> Results obtained by the architecture “G” .....	36
<b>Figure 5. 14:</b> Confusion matrix for the architecture “G” .....	37

**Figure 5. 15:** Comparison of all architectures for the accuracy, sensitivity, specificity, and ROC AUC scores ..... 39

**Figure 5. 16:** Comparison of all architectures for true positive, true negative, false positive, and false-negative results (Highest bars are optimal results for TP and TN, and lowest bars are optimal results for FP and FN) ..... 41

## LIST OF TABLES

<b>Table 4. 1:</b> Details of CNNs used in experiments (A-B-C) .....	27
<b>Table 4. 2:</b> Details of CNNs used in experiments (D-E-F-G) .....	28
<b>Table 4. 3:</b> CNN architectures and depth levels .....	29
<b>Table 5. 1:</b> Results of all experiments (in %) .....	38
<b>Table 5. 2:</b> Ranking of the CNN architectures for all evaluation metrics .....	39
<b>Table 5. 3:</b> Total TP, TN, FP, and FN results of all experiments .....	40
<b>Table 5. 4:</b> Ranking and the levels of depth of the CNN architectures for the accuracy .	42

## LIST OF ABBREVIATIONS

<b>AI:</b>	Artificial Intelligence
<b>DL:</b>	Deep Learning
<b>ML:</b>	Machine Learning
<b>CNN:</b>	Convolutional Neural Network
<b>RNN:</b>	Recurrent Neural Network
<b>MRI:</b>	Magnetic Resonance Image
<b>CT:</b>	Computed Tomography
<b>DCNN:</b>	Deep Convolutional Neural Network
<b>BTP:</b>	Brain Tumor Progression
<b>CNS:</b>	Central Nervous System
<b>ReLU:</b>	Rectified Linear Unit
<b>TP:</b>	True Positive
<b>TN:</b>	True Negative
<b>FP:</b>	False Positive
<b>FN:</b>	False Negative
<b>ROC AUC:</b>	Receiver Operating Characteristic Area Under Curve
<b>WHO:</b>	World Health Organization

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

One of the most challenging problem in healthcare is the diagnosis of diseases; the medical core used to struggle when it comes to the detection and treatment of viruses, tumors etc. Up to now there is no treatments for more diseases in case of viruses there is only supportive care and for the tumors it depend of the genera of tumors and the most common used method is the surgery this helps to remove the cancer this must be done without any damage. With the development and the recent popularity of intelligent computers also known as artificial intelligence; artificial intelligence techniques have shown good results when it is apply in the diagnosis and detection of diseases. Deep learning is one of the most used method of AI when it comes to the detection of images by using the convolutional neural networks, CNN is mostly used in computer vision for image detection and classification; the recurrent neural network is commonly used for speech recognition.

### 1.2 Brain Tumors

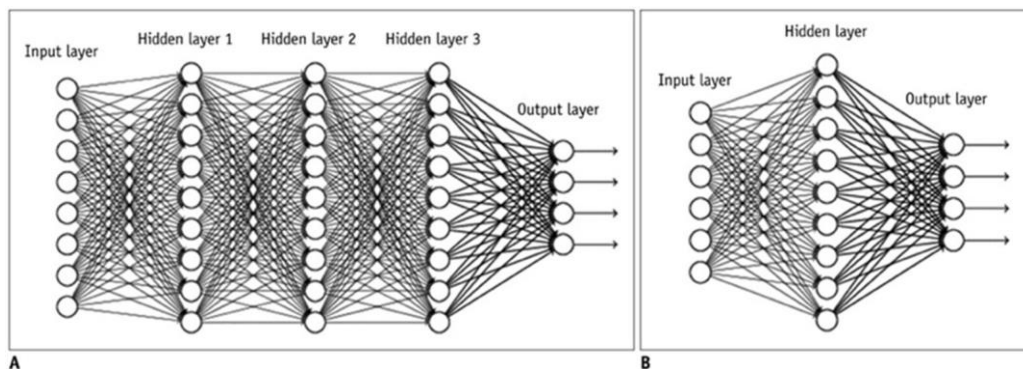
Human most vital organ which operate with billions number of cells is known as the brain; brain tumor can be define as the set, group or a collection of anomalous cells that are stack in the brain. This anomalous stack cells in the brain are one of the main reason that causes brain tumor and they are dangerous to the healthy cells and they can destroy the normal functionality of the healthy cells in the brain. The brain surrounded bone also known as the skull are very fragile if they encounter an inside problem such as growth, this can cause damage to the brain. Brain tumor can be classified into two genera which are cancerous and noncancerous generally known as malignant and benign respectively. Brain tumor can also be classified as low grade for benign (Grade I and Grade II) and high grade for malignant (Grade III and Grade IV). Malignant brain tumors is the fatal and it causes cancer it originated from the rapid growth of boundaries which is not yet defined. Malignant brain tumor can be primary or secondary. Primary malignant are the tumor that begin in the brain itself and the tumor that come from different part of the human body are secondary malignant. Benign brain tumor begin in the brain and they grows slowly only in the brain

not somewhere. Benign is less aggressive compared to malignant (Mohsen et al, 2018, Kalyani et al, 2017).

In 2007 with the update of the Central Nervous System (CNS) by the World Health Organization (WHO), brain tumors can be classified in seven types according to their histology sorts and supposes cellular source. The seven classified major categories of brain tumors by WHO contains the meninges tumor, the neuroepithelial tissue tumor, the paraspinal nerves and the cranial tumor, the hematopoietic neoplasms and the lymphomas tumor, the metastatic tumor, the sellar region tumor and final the germ cell tumor (Strong MJ, Garces J, Vera JC, Mathkour M, Emerson N, et al, 2015).

### 1.3 Deep Learning

Deep learning is the extend version of artificial neural network due to the limitation of artificial neural network and the problem leading to over-fitting and vanishing gradient deep learn was implemented in order to overcome the said limitations. Deep learning is subfield of machine learning which works on two or more hidden layers compared to the traditional neural networks where there is only a single hidden layer.



**Figure 1. 1:** Illustration of deep learning and artificial neural network architectures (June-Goo Lee et al, 2017)

Figure 1.1 above gives an illustration of the difference between deep learning and artificial neural network. The structure A is a deep learning architecture with multiple hidden layers and the structure B is the shallow artificial neural network with a single hidden layer. Deep



learning is commonly used for image and speech recognition tasks (June-Goo Lee et al, 2017). The most popular methods used in deep learning are recurrent neural network (RNN) and convolutional neural network (CNN). RNN is used for speech recognition and the CNN is used for image recognition (June-Goo Lee et al, 2017).

#### **1.4 Problem of the Study**

The CNN is widely applied in medical classification tasks; but, the determination of the optimal architecture is a significant challenge. This directs researchers to implement pre-trained networks and to use transfer learning in their studies to transfer knowledge via pre-defined architecture to their problem domain.

However, the train process of pre-trained networks require huge computational time and this limits the number of experiments to modify architectures and tune the models in order to achieve optimal results.

#### **1.5 Aim of the Study**

One of the fatal and dangerous cancer is the brain tumor therefore detecting it in the early febrile is necessity for the patients and radiologists. Due to the cost and the lack of material when it comes to the diagnosis of the brain tumor we suggest the use deep CNN in order to help the radiologist to diagnose the cancer at low cost and with high accuracy result. The main objective of the study is to implement a deep CNN to classify brain tumor progression with high accuracy to assist medical doctors in following the progress of the tumors.

In addition, several CNN architectures are implemented and comparative evaluation is performed in order to determine the optimal architecture for the considered dataset and to observe the effect of the layer numbers on classification rates of tumor progress.

Due to the massive training time of pre-trained networks, CNN layers are minimized to overcome this problem and to demonstrate the classification ability of the light deep CNN in brain tumor progression.

## **1.6 Thesis Outline**

The rest of the thesis is organized as follows:

Chapter Two presents the literature review on brain tumors, brain tumor progression and deep learning implementations on these.

Chapter Three explains the fundamentals of convolutional neural networks in detail.

Chapter Four presents the design of experiments, considered dataset, evaluation metrics in detail.

Chapter Five presents the experimental results, comparisons and discussions.

Chapter Six concludes the results obtained in this thesis.

## **CHAPTER 2**

### **LITERATURE REVIEW**

#### **2.1 Overview**

Several related research have been conducted in the past concerning this topic. In the same manner, will be reviewing the previous works concerning application of deep learning in healthcare, there is a significant amount of informatics systems that has been developed concerning the usability of deep learning and some domain of artificial intelligence in medical area.

#### **2.2 Deep learning in healthcare**

Due to the financial cost and higher demands in the detection of the Covid-19. Sekeroglu et al. collected chest X-ray images for the diagnosis of Covid-19 using deep learning. The main objectives are to use deep learning due to its ability in the detection of images and to reduce the cost when it comes to the diagnosis of Covid-19. During the experiment Sekeroglu et al, collect three categories of X-ray images; 1583 of healthy patients, 4292 of patients with pneumonia and 225 of patient with Covid-19 confirmed cases. DL and ML classification approaches were implement on the collected images, using the neural network. 38 experiments were concluded; five machine-learning techniques were used in 10 experiments and 14 experiments were concluded using the transfer learning, in order to evaluate the models performance image data and statistical data were separate during the experiments, the eightfold cross-validation was used. As part of the result Sekeroglu et al, find an accuracy mean of 98.52%, a specificity mean of 99.18%, and a sensitivity mean of 93.84% and reached 96.51% under the curve score. Sekeroglu et al, conclude that artificial intelligence techniques can play an important role when it comes to the diagnosis of Covid-19 and in order to improve the results of the conducted study and based on the obtained results more research is needed in the future that will provide more information using CNN architecture.

Chougrad et al 2018, research is based on the challenge and difficulties that radiologists mostly have in the classification of mammography mass lesions that commonly trend to needless breast biopsies. Chougrad et al, implement a computer-aided diagnosis system-

using DCNN, the objective of the developed system is to help radiologists who struggle with classification of mammography mass lesions. Mostly deep learning requires a huge amount of data in order to train the network from scratch, however, transfer learning is the method that deals effectively with a small amount of datasets in the case of medical images. As a result, Chougrad et al. study the use of transfer learning to determine the accurate fine-tuning strategy when it comes to the training of a CNN model. Chougrad et al. find 0.98 AUC and 97.35% accuracy on the DDSM database, 0.97 AUC and 95.50% accuracy in INbreast database and using the BCDR database the following are the results: 96.67% accuracy and 0.96 AUC. For the developed Computer-aided Diagnosis and its efficiency to classify images, the system was tested on an independent database and the accuracy of the system is 98.23% and 0.99 AUC.

In conclusion, due to the accuracy of the system and the results demonstrated, this system can be used in the prediction of mass lesions when they are benign or malignant.

Lee KJ et al. 2020, propose a comparison between the diagnostic performance of a single and multiple trained deep learning algorithms by radiologists in the diagnosis of mastoiditis. Lee KJ et al. used a total of 9,988 mastoid series, the classification was made into two categories: one is normal and the other is abnormal. This is based on the findings of radiographic mastoiditis. From the mastoid series, 792 temporal images were classified as the gold standard test with bone CT, the others were divided into random training and validation with respective numbers (n=8,276) and (n=920) for the development of a deep learning algorithm. The diagnostic performance of the single trained algorithm was compared with the multiple trained algorithm, and the performance of the algorithms was evaluated by two radiologists. Lee KJ et al. conclude that the multiple trained deep learning algorithm provides good results compared to the single trained deep learning algorithm and that the deep learning algorithm's diagnostic performance for the detection of mastoiditis can be compared to that of radiologists and, sometimes, the algorithm provides a better result than radiologists.

Li et al. 2019, propose a fundamental review of deep learning, the general techniques and the importance of deep learning in healthcare. Data were collected from the Institute of Electrical and Electronics Engineering and the PubMed databases publications, these data implement a variation of different deep learning methods. Li et al. classified the collected data into 4

categories which are the electronic health record, biological system, physiological system and medical image. Furthermore Li et al, discuss the challenge of deep learning and the future research that will improve the health management and that will support the new internet technology and physiological signals.

Lee et al 2020, proposal is the transfer learning evaluation with the DCNN based on the layer depths and the degree of the fine-tuning. The aim is to evaluate the performance of DCNN in Chest X-ray based COVID-19 and the efficiency of the transfer learning. The Chest X-ray images used were collect from the available public databases. Images are classified in 3 classes which are normal, Covid-19 and pneumonia, in order to evaluate the depths layer, two methods of CNNs were used the VGG-16 and VGG-19, and these techniques are used as the backbone networks. The result find by Lee et al, shows that the value of the AUC is higher and it is evaluate to 0.950 when it comes to the classification of Covid-19 in fine-tuned with 2/5 block of VGG16 backbone. Lee et al conclude that medical images classification is limited in amount of data and depths layer, this limitations cannot guarantee good results.

Fourcade et al 2019, proposed an investigation on the impact of AI in healthcare, the role of deep learning algorithms in the medical images analysis. The investigation is based on “the role played by deep learning algorithms in image recognition, the improvement of the visual diagnosis in medicine?” Fourcade et al, used the systematic review of the previous related topics published before May 2019, CNNs is the method used for medical images analysis. As of result Fourcade et al, used 352 papers from the PubMed database, from the collected 352 papers, 327 were excluded because they were not related to the topic, the remaining 25 related papers were used to investigate on the role of DL algorithms in the analysis of medical images, these papers were published from 2013 to 2019. Most of the authors were from North America and Asia. Fourcade et al, conclude that CNNs is not the solution in the replacement of medical doctors, however it can be used in the optimization of the routine tasks.

Xin et al 2019, propose and innovative training due to the error of the back propagation algorithm, Xin et al propose a depth neural network in order to find the maximum and

minimum interval separately when it comes to the classification error. Xin et al, used two techniques which will help to obtain the better results, the proposed techniques are the  $M^3CE$  and the cross entropy, these techniques were tested on two well-known deep learning databases, the MNIST and CIFAR-10.  $M^3CE$  have shown better results compared to the cross entropy, the  $M^3CE$  produce better results in both databases. Xin et al, conclude that more research is needed, in terms of comparison between deep convolutional neural network and human brain.

Faust et al 2018, Did a scientific research on deep learning in physiological signal, 53 related papers were collected, the papers collected were based on Electrooculogram, Electrocardiogram, Electromyogram and the Electroencephalogram, these types were used in order to create a structure that provides a content review. During the review Faust et al, find out that the deep learning method provide better result when it is applied to the huge amount and varied datasets compared to the classic analysis and classification method. In conclusion much effort is needed in order to detect diseases with less effort using physiological signals.

### **2.3 Deep Learning in Diagnosis of Brain Tumor**

It is challenging and very difficult to detect brain tumor manually radiologists use to struggle when it comes to the diagnosis of brain tumor using traditional techniques, in order to avoid the said difficulties Khan et al, propose the development of an adopt multimodal automated classification using deep learning method for the classification of brain tumor types. The proposed multimodal brain tumor classification works on 5 steps, which are (i) by the used of the histogram edge-based equalization the linear contrasts stretching is introduced and the transformation of discrete cosine is performed. (ii) the execution of deep learning using the feature extraction, this is done using transfer learning on two CNN pre-trained models VGG16 and VGG19, these two models were used in the extraction process. (iii) consist of the approach of the correntropy-based joint learning, the development and the selection of the finest feature for the extreme learning machine. (iv) the combination in one metric for the partial robust covariance feature square (PLS)-based. (v) final classification was obtained by the combination of matrix in the extreme learning machine. Khan et al, obtained

the following results BraTs2018 produce accuracy result of 92.5%, the BraTs2017 the accuracy result of 96.9% and for the BraTs2015 the accuracy result of 97.8% the above result was validate using BraTs datasets.

Alqudah et al 2019, Due to the higher accuracy result of deep learning numerous informatics system have been developed using this method to solve various and complex problems, mostly the problems that required sensitivity and high accuracy especially in healthcare. Alqudah et al, used CNN due to the fact that this method is the most used in DL for classification. Alqudah et al, used brain tumor dataset, this dataset contains 3064 T1 weight and it includes 3 different classes meningioma, pituitary tumor and the glioma. The propose CNN consist of 18 layers that will enable the classification and the effective evaluation of the brain. Cropped, uncropped and segmented are the 3 methods used on the dataset. The result obtained using the cropped method is efficient and the accuracy grade is 96.9%, 98.4%, 97.4% for the input images of the size 32x32, 64x64, and 128x128 respectively. For the uncropped method, result is efficient and the accuracy grade is 99.2%, 99.0%, 97.4% for the input images of the size 32x32, 64x64, and 128x128 respectively. Finally, the segmented method the result get is efficient and the accuracy grade is 98.4%, 97.6%, 97.5% for the input images of the size 32x32, 64x64, and 128x128 respectively. With the above result Alqudah et al concluded that the system developed could successfully detect brain tumor with the used of the three methods, the developed system can evaluate the tumor in three levels the meningioma, pituitary tumor and glioma using the T1 weight brain MR images. Finally, this system can be improved in the future by adding more brain MR images with different weights.

Sultan et al 2019, proposal is the implement of convolutional neural network based on deep learning model for the classification of distinct types of brain tumor using two public available datasets. DL algorithm which is part of machine learning, it has demonstrate very good result when it is apply in the segmentation problems. It is very crucial and important to categorize brain tumor; the classification will help to make the decision regarding patient treatment can be made by the doctors. Based on the two available public datasets Sultan et al classify the brain into two groups, the first group contains (glioma, meningioma and pituitary tumor) and the second group contain three different grades of the glioma, which are

Grade II, III, and IV. Sultan et al, collected 233 patients and the total number of 3064 images in the first dataset, in the second dataset they collected 73 patients and with total number of 516 images. The performance of the implemented network by Sultan et al, provide the accuracy result of 96.13% for the first collected dataset and 98.7% accuracy result for the second collected dataset.



## **CHAPTER 3**

### **CONVOLUTIONAL NEURAL NETWORK**

#### **3.1 Overview**

This chapter introduces convolutional neural network (CNN) and its components such as layers, pooling operations, activation functions etc. in details

#### **3.2 Convolutional Neural Network**

The combination of newly deep learning methods and artificial neural network also known as convolutional neural network is the most used deep learning technique. It has shown considerable results when it is used in computer vision to perform different tasks and it has provided various interest in several domains that include radiology. (Yamashita et al, 2018; Waseem Rawat and Zenghui Wang, 2017).

Convolutional neural network is a best model that can fulfill tasks efficiently and it plays an important role in the field of image recognition and since it provides less training process, due to the simple architecture that it provides and its robustness adaptivity these characteristics make convolutional neural network less complicated. Since it can reduce the node weight, the convolutional neural network attracts most researchers in the field. There is a tremendous number of convolutional neural networks structures such as the VGGNet, GoogleNet, LeNet, AlexNet and the ZFNet. CNN can be used in classification of images, recognition and detection of targets, semantic segmentation and so forth. CNN is considered as an updated version of old neural networks; it uses feedforward, the neurons are not completely connected in the neural network (Hang Zhou and Qichen Sun, 2020).

#### **3.3 Convolutional Neural Network Structure**

Convolutional neural network also known as the feedforward networks where the flow of information moves in one direction that is from the inputs to the outputs convolutional neural networks are biological inspiration as the artificial neural network. The architecture of convolutional neural network may have several variations or modules, but in general they consist of convolution layers, pooling layers and the fully connected layer these are the

blocks that composed the convolutional neural network, backpropagation algorithm is the most used algorithm in convolutional neural network (Waseem Rawat and Zenghui Wang, 2017; Yamashita et al, 2018).

Figure 3.1 below gives a picture of an overview structure of CNN under training; CNN is a set of different blocks of layers, which are the convolution, the pooling and the fully connected layers. Through one or more fully connected layers the performance of the kernels and weights are calculated using the loss function. The cost function is also known as the loss function, it have the ability to measure the compatibility amid the network output prediction via the forward propagation, the most and commonly used loss function is cross entropy it used for multiclass classification. Forward propagation is the step where the inputs are transformed into outputs. It is to be notify that the convolution and pooling layers used are 2D-CNN however the same operation can be used for 3D-CNN.



**Figure 3. 1:** Structure of convolutional neural network and the training process (Yamashita et al, 2018).

### 3.3.1 Convolution Layers

The most important component of the convolutional neural network structure is the convolution layer, it is used to achieve feature extraction which lies on combination of linear and nonlinear procedures in other words, activation function and operations, neurons that

are in the convolution layer are arranged in form of feature maps and every neuron in the map has a receptive area and these are connected to the previous nearby neurons using the weights group that can be trained and sometimes this can be called the filter bank (Yamashita et al, 2018; Waseem Rawat and Zenghui Wang, 2017 ).

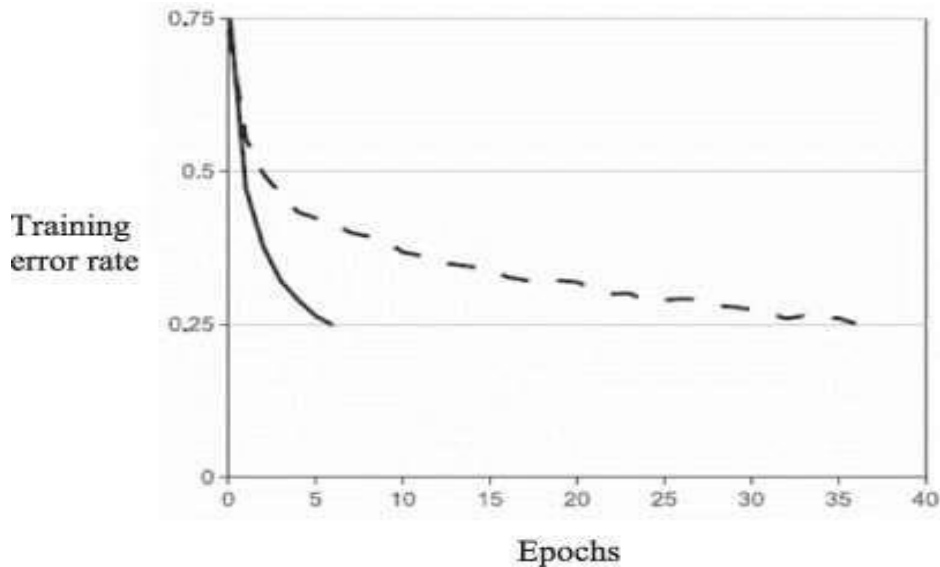
The convolution layer twists the learning weights and inputs so that they can process new feature maps and the result of the twisted learning weights and inputs are sent via the nonlinear activation function. Every neuron contained in the feature map is compelled to be equal; but the convolution layer has different feature maps and these feature maps contain different weights. This allows the extraction of feature maps at each location (Yamashita, et al, 2018; Waseem Rawat and Zenghui Wang, 2017; Hang Zhou and Qichen Sun, 2020).

$$Y_k = f(W_k * x) \tag{3.1}$$

Where  $x$  is the image input; the multiplication sign is used in the context that refers to the 2D convolutional operator, it calculates the result of the inner filter model. This process is done at every location of the input image;  $W_k$  denotes the  $k$ th feature map for the related convolutional filter and the  $f(\cdot)$  is the representation of the activation function of a nonlinear function. This function permits the extraction of nonlinear features; hyperbolic and sigmoid tangent functions were traditionally used, the newest rectified linear units (ReLU) have become more popular, due to their success and popularity more research in the area has been opened and it is mostly focused on the implementation and application of new DCNN activation functions that can increase the performance of numerous sorts of DCNN. Activation function choice has an impact and effect on the training time of the network and it plays an important role in the influence and the performance of huge DCNN datasets. The introduction of ReLUs by (Nair and Hinton, 2010). Krizhevsky et al, made the ReLU function more popular in 2012 and when it has to be used in DCNNs. Even the trio of Glorot, Boerdes and Bengio state that compared to the other functions and when it comes to the timing the ReLUs provide a faster training process when it is applied to the complete supervised networks, this can be made without the need of the unsupervised pretraining (Glorot et al, 2011).

Figure 3.2 below shows the training times comparison between ReLUs with the solid line in the figure and the hyperbolic tangent with a dashed line in the figure on the four-layers

DCNN, the comparison were made by Krizhevsky in 2009 by using the CIFAR-10 dataset. According to the results get ReLUs were trained 6 times much faster compared to an equivalent network using a tangent hyperbolic activations.



**Figure 3. 2:** Training times, comparison between ReLUs and tangent activations functions (Krizhevsky et al., 2012).

In the traditional activations functions like the hyperbolic and the sigmoid tangent are given using the following formula

$$f(x) = 1/(1 + e^{-x}) \quad (3.2)$$

Where  $f(x) = \tanh(x)$ ; the output neuron denoted by  $f$ ;  $x$  an input function; this representation is used as a reminder to activations functions. But for the ReLUs the linear function form as been simplified, in the ReLUs reduce the negative part to zero and only the positive part of the activation function is retained, the maximum integrate operator endorse fast computation. Numerous state-of-art image classification systems used the ReLUs. Below is the simplified ReLUs activation function formula

$$f(x) = \max(x, 0) \quad (3.3)$$

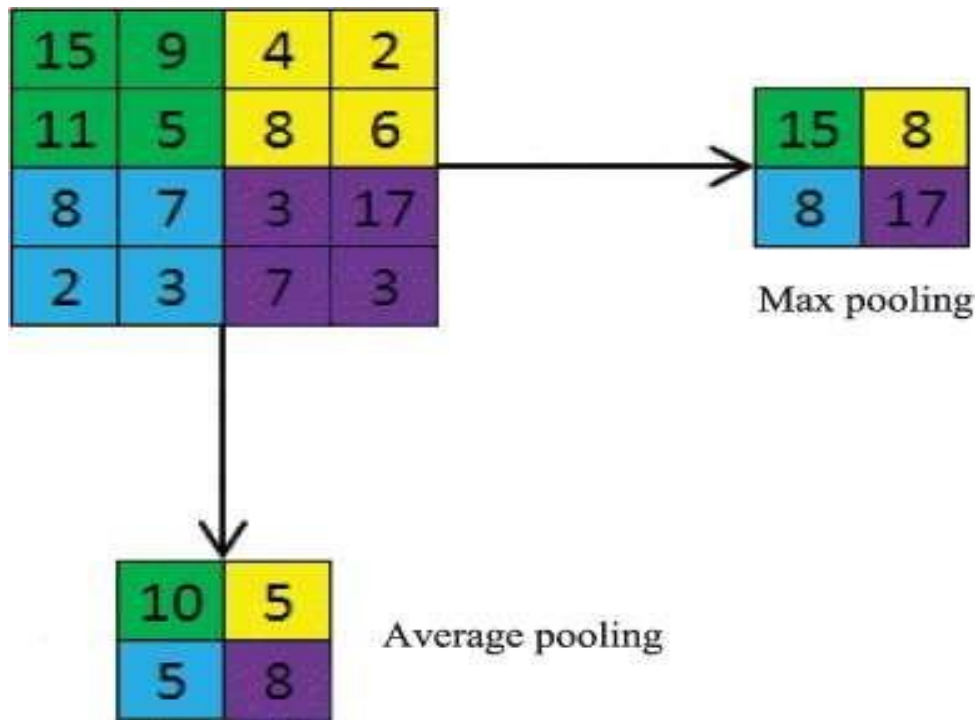
### 3.3.2 Pooling Layers

It has the ability to decrease the dimension of data and it must maintain the invariance of the network structure, furthermore the pooling layers must reduce the dimension of feature maps so that it can be able to provide the invariance translation to the distortion and small shifts, it must also provide the decrease number of learnable parameters subsequently. Max pooling is the most used and most popular form of pooling's, in feature map it has the aim to extract patches, and then maximize the value of the outputs in each patch and discard all the remaining values. Average pooling is another form of pooling but it worth nothing it only used once during the training process it applied before the fully connected layers. It has two main advantages one is it reduce the amount of learnable parameters and the other one is it allowed the convolutional neural network to accept inputs of different size (Yamashita, et al, 2018; Waseem Rawat and Zenghui Wang, 2017; Hang Zhou and Qichen Sun, 2020).

$$Y_{kij} = \max_{(p,q) \in \mathcal{A}_{ij}} x_{kpq}, \quad (3.4)$$

Where  $Y_{kij}$  denotes the pooling operator output connected with the  $k$ th feature map, the elements where the location are  $(p,q)$  are denoted by  $x_{kpq}$  contained pooling area  $\mathcal{A}_{ij}$  this represents a receptive area near  $(i,j)$  position.

Concerning the deference between average pooling and max pooling. Figure 3.3 below gives a clear explanation where a  $4 \times 4$  input image. In case  $2 \times 2$  filter that applies max pooling outputs provides the maximum value of every  $2 \times 2$  area. Where the average pooling outputs provides the average value of each area. In other words in the max pooling the maximum values of every piece of the feature map is calculate and when it comes to the average pooling, it provides the average value of every piece on the feature map is calculate. (Yamashita et al, 2018; Waseem Rawat and Zenghui Wang, 2017; Hang Zhou and Qichen Sun, 2020).



**Figure 3. 3:** Difference between Average and Max pooling (Waseem Rawat and Zenghui Wang, 2017)

Backpropagation was applied in deep convolutional neural network architecture for the first time in 2007 using max pooling. In the year 2010 Scherer, Müller, and Behnke proved that max pooling operator was empirical and far superior compared to other subsampling operation when it comes to capture invariance in the images. By using max pooling they proved and accomplished best results that was published on the normalized-uniform NORB dataset and this overcame by half percent the previous best result obtained by Nair and Hinton in 2009. In 2009 Jarret et al, shows rectification layer demand was promoted by the max pooling. Max pooling layer is not normally part of the deep convolutional neural network structure, and they conclude that the average pooling does not provide the same benefits as the max pooling since the average pooling suffers from the effect among neighboring filter output cancellation. (Yamashita, et al, 2018; Waseem Rawat and Zenghui Wang, 2017; Hang Zhou and Qichen Sun, 2020).

### 3.3.3 Fully Connected Layers

During the training process, more convolution and pooling layers are mostly loaded on top of each other in order to pull out additional feature abstract representations that are moving in the network. It follow convolution and pooling layers and it has provide the interpretation of the feature representations also execute the high-level reasoning function. When it comes to classification softmax is the standard and common operator that is use on the top of the deep convolutional neural network. Softmax is the most used activation function in the DCCNs fully connected layers due it ability of interpretation and it simplicity and probabilistic.

When the  $i^{th}$  input feature  $x_i$  that has the label corresponding to  $y_i$  softmax loss can be written as follow

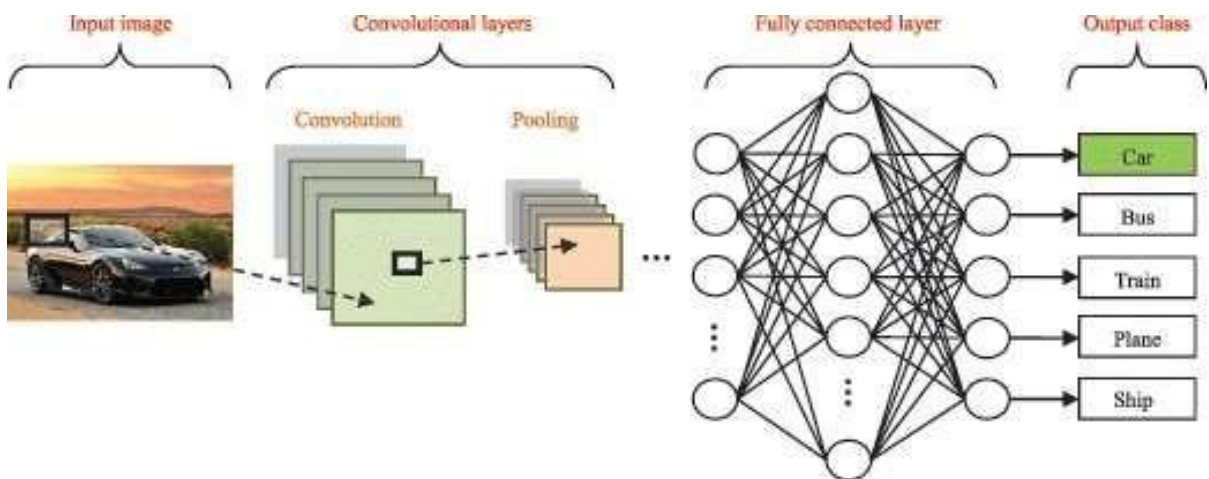
$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i - \log \left( \frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \right), \quad (3.5)$$

Where the number of element is denoted by  $j$  ( $j \in [1, K]$  where  $K$ , denotes the number of classes);  $f$  is the class score of the vector and it is represented by  $f_j$  and  $N$  represent the number of training data. In 2013 Tang found out that replace softmax with support vector machine conduct to the best result and improve the accuracy result of the classification however it is very challenging to use SVM in the fully connected layers due to their compute-to-data ratio instead average pooling can be used as an alternative for the simple classifier (Yamashita et al, 2018).

### 3.4 Convolutional Neural Network Training Process

CNN is an update version of old neural network in order to adjust their free parameters they used learning algorithms and that is the weights and biases; so that it can provide the output network desired. The most used learning algorithm is the backpropagation, in order to minimize errors and performance effect backpropagation is used for the computation of the gradient this can referred to the performance lost or cost function that determine how the parameters of the network can be adjust. Overfitting is the common and most experienced

problems when it comes to the training of Deep convolutional network in particular and Convolutional neural networks, the overfitting problem is the poor performance on a held-out test set when the network trained a small amount of data or huge amount of data, this affect the ability of the model (Yamashita, et al, 2018; Waseem Rawat and Zenghui Wang, 2017; Hang Zhou and Qichen Sun, 2020).



**Figure 3. 4:** CNNs training process for image classification (Waseem Rawat and Zenghui Wang, 2017)

Figure 3.4 above gives an illustration of how the convolutional network architecture classified images. First an image input is directly input into the network, the image must go through several process of convolutional and pooling layers after this process the representation form of convolutional and pooling layers operations provide one or more fully connected layers and after providing the fully connected layers it should go to the final fully connected layer output the class label of the input image as we can see the above figure the class label of the input image is car. (Waseem Rawat and Zenghui Wang, 2017)

### 3.5 Summary

In this chapter, basic and structure of the CNNs, layer and pooling operations and activation functions were explained in detail using recent development and example.



## **CHAPTER 4**

### **EXPERIMENTAL DESIGN**

#### **4.1 Overview**

This chapter presents the details about the design of experiments, considered dataset, evaluation criteria, and other information used in the thesis.

#### **4.2 Dataset and Data Preparation**

##### **4.2.1 Dataset**

In this thesis, the Brain Tumor Progression (BTP) dataset is considered to classify the progresses as initial and progress phases of brain tumors. The BTP dataset consisted of 20 subjects with primary newly diagnosed glioblastoma. The dataset used were retrieved from <https://wiki.cancerimagingarchive.net/display/Public/Brain-Tumor-Progression>. Two MRI exams are included for each patient: within 90 days following chemo-radiation therapy completion and at progression. The images of the first MRI exam is used for the initial phase after treatment, and the second MRI exam is used for the progress phase.

All images were obtained in DICOM format. Image sets consisted of T1w (pre and post-contrast agent), FLAIR, T2w, ADC, normalized cerebral blood flow, normalized relative cerebral blood volume, standardized relative cerebral blood volume, and binary tumor masks (generated using T1w images). The perfusion images of the dataset were generated from dynamic susceptibility contrast (GRE-EPI DSC) imaging following a preload of contrast agents. All of the series are co-registered with the T1+C images.

Two MRI exams for twenty patients with 383 series were performed, and a total of 8798 images were obtained.

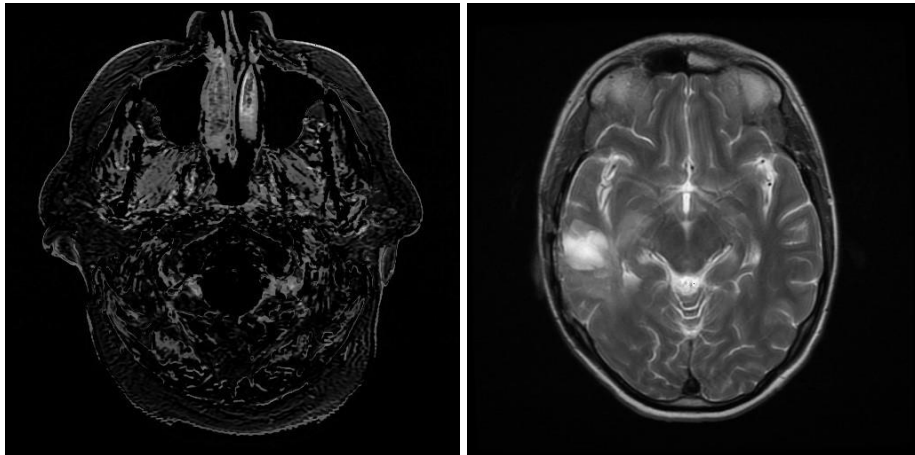
Sample images of initial and progress classes are shown in Figure 4.1 and Figure 4.2, respectively.

#### 4.2.2 Data Preparation

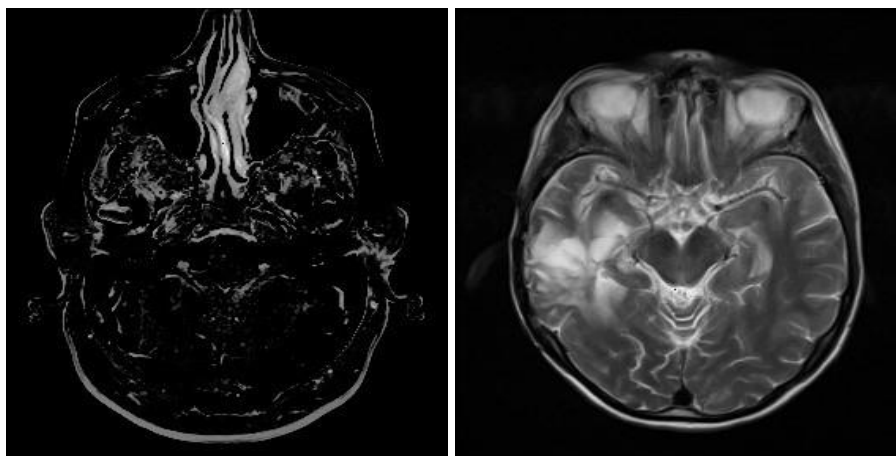
In the first step, Dicom images were converted to PNG format, which causes the minimal loss of the details within an image. Image resizing was not applied to the images not to change the intensity values of the original images, and the resizing was only applied in the input layer of CNN when the images were fed to the network.

After obtaining the PNG images, it was observed that both initial and progress classes had black slices, which cause false recognition of images for both classes. Therefore, the slices that consisted of only black pixels were removed from the dataset.

Finally, 3479 images were obtained for the initial class and 3436 images for the progress class. The number of the obtained images demonstrate the balanced nature of the dataset used in this thesis.



**Figure 4. 1:** Sample initial class brain MRI exam images



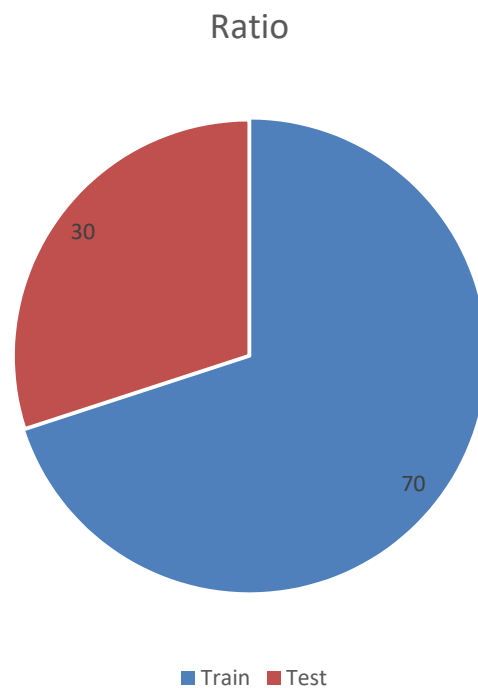
**Figure 4. 2:** Sample progress class brain MRI exam images

### 4.3 Validation of the Experiments

There are several validation techniques used in the model evaluation and hyperparameter determination. These techniques can be listed as Hold-out Method, Random Sub-sampling method, and k-Fold Cross-Validation methods.

#### 4.3.1 Hold-Out Method

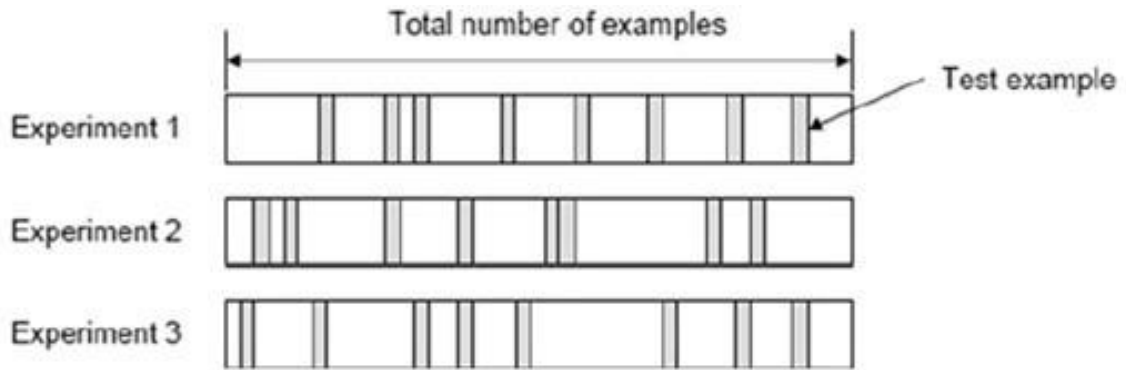
The Hold-Out method is a basic validation technique where the dataset is divided into two sets, a training set and a test set. The training set is used to teach the data to the model, and the test set, which is untrained data, is used to test the ability of the model for the considered task. The disadvantage of the Hold-Out method is not considering all the data for training and testing. Figure 4.3 shows the graphical representation of the Hold-Out method.



**Figure 4. 3:** Graphical representation of the Hold-Out method for 70% of training and 30% of testing Set.

### 4.3.2 Random sub-sampling method

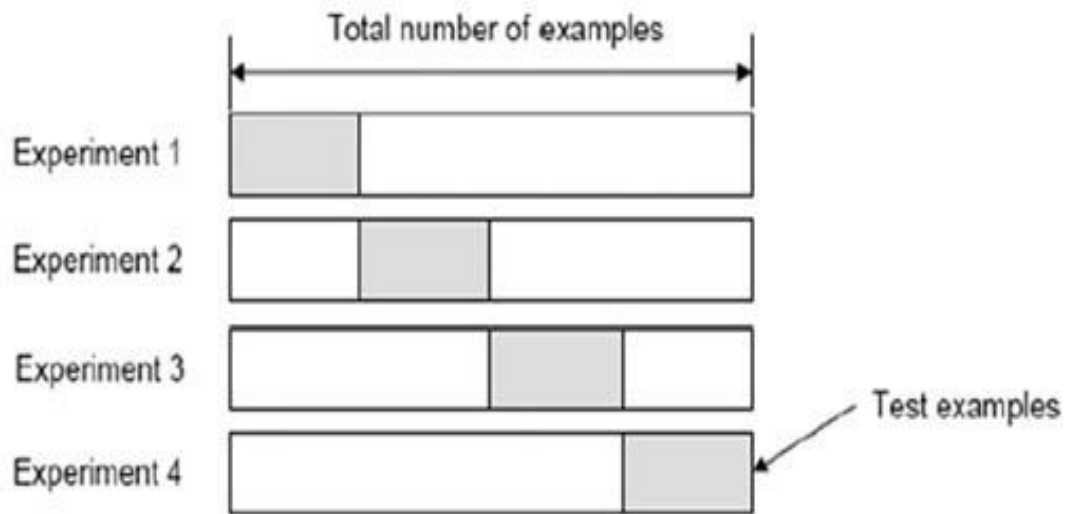
The Random sub-sampling method is similar to the Hold-out Method, but it is repeated several times to consider more data for both training and testing phases. This increases the robustness of the evaluation of the considered models. Figure 4.4 demonstrates the process of random sub-sampling.



**Figure 4. 4:** The process of the Random Sub-sampling method.

### 4.3.3 K-fold cross-validation method

In k-fold cross-validation, the dataset is divided into k equal-sized subsets, one for testing and the rest for training. The process is repeated k-times until each subset is used for testing exactly once. The total success of the model is obtained by averaging the rates for all the runs. The main advantage of this method is considering all the data for testing and training phases and providing a more robust evaluation of the models on the data. For this reason, the k-fold cross-validation is considered in this thesis. Figure 4.5 shows the example process of 4-fold cross-validation.



**Figure 4. 5:** The process of the 4-fold cross-validation method.

#### 4.4 Evaluation Metrics for Classification Problems

The evaluation metrics for the classification problems are based on predicted and observed outputs of the dataset. Predicted outputs are the results obtained by the model, and the observed outputs are real labels of the data.

The correctness of the results is indicated by "True" and "False," if the predicted result is the same as in observed outputs, it is called "True," and if the predicted result is not the same as in observed outputs, it is called "False."

A "True Positive" (TP) result is used to indicate the positively predicted result and is also observed as a "Positive." Similarly, a "True Negative" (TN) result indicates the negatively predicted result and is also observed as a "Negative."

A "False Positive" (FP) result indicates the positively predicted result but is observed as a "Negative," and a "False Negative" (FN) result indicates the negatively predicted result but is observed as a "Positive."

The abovementioned results form the basis for the evaluation metrics of two-label or multi-label classification problems. Because of the two-class nature of the considered dataset of this thesis, only the evaluation metrics that can be used for all kinds of classification problems will be presented in this chapter.

#### 4.4.1 Accuracy

The accuracy is used to determine the general ability of the model in assigning the test samples into proper classes. However, it should also be noticed that it has a limitation if the considered dataset is imbalanced, which means not equally distributes samples of classes. The accuracy considers all the correctly classified samples ( $TP + TN$ ) and finds the general ratio of these samples to the whole dataset. Equation 4.1 shows the formulae of the accuracy.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4.1)$$

where  $TP$  and  $TN$  indicates True Positive and True Negative, respectively, and  $FP$  and  $FN$  demonstrates the False Positive and False Negative, respectively.

#### 4.4.2 Sensitivity

The Sensitivity is a metric to measure how the model is capable of predicting observed positives correctly. It uses True Positives and False Negatives to consider all positively observed data. Equation 4.2 shows the formulae of the Sensitivity.

$$Sensitivity = \frac{TP}{TP+FN} \quad (4.2)$$

#### 4.4.3 Specificity

The specificity is a metric to measure how the model is capable of predicting observed negatives correctly. It uses True Negatives and False Positives to consider all negatively observed data. Equation 4.3 shows the formulae of the specificity.

$$Specificity = \frac{TN}{TN+FP} \quad (4.3)$$

#### 4.4.4 Receiver operating characteristic area under curve

The receiver operating characteristic (ROC) area under the curve (AUC) score is the

common metric for the models that use imbalanced data. It is based on sensitivity and

specificity rates, and a ROC curve plots TP rate vs. FP rate. The area under the curve measures the area under this curve, and it is scale-invariant because of considering sensitivity and specificity rates. Therefore, the number of instances or images becomes insignificant and stable evaluation for the general detection ability of the model can be performed. Equation 4.4 shows the formula of ROC AUC score.

$$ROC\ AUC\ Score = \frac{1}{2 * Sensitivity} + \frac{1}{2 * Specificity} \quad (4.4)$$

In this thesis, the accuracy is used as a main evaluation metric because of the balanced data; however, the ROC AUC score is also considered to support the obtained results. In addition, the sensitivity and specificity rates are considered to evaluate the implemented models' ability to predict TP and TN instances.

#### 4.5 Design of Experiments

Seven Convolutional Neural Networks with different architectures are implemented in the experiments to detect the tumor types mentioned above. Five-fold cross-validation is applied in all experiments, and the average accuracy, average specificity, average Sensitivity, and average ROC AUC Scores are considered in the evaluation.

The selection of CNN's was based on the systematic increment and decrement of convolutional and dense layers to achieve the optimal classification rate. The architecture "A" consisted of two convolutional layers with 64 and 32 filters, respectively. The dense (or fully connected) layers include two layers with 128 and 32 neurons, respectively.

The architecture "B" had the same convolutional layer number and properties as in "A"; however, the number of dense layers increased to three with 128, 64, and 32 neurons, respectively.

The architecture "C" consisted of the same dense layer and neuron numbers as in "B", but the number of convolutional layers were increased to three with 64, 32, and 16 filters, respectively. Table 4.1 present the details of the CNNs (A-B-C) in detail.

The architecture "D" consisted of two convolutional and dense layers with 64 and 32 filters and 64 and 32 neurons, respectively. In the architecture "E", the number of layers did not increase; however, the number of neurons in dense layers were increased to 128 and 64. In



the architecture "F", the number of dense layers and their neurons left the same as in "E", but the number of filters in convolutional layers was decreased to 32 and 16, respectively. In the final architecture considered in this thesis, which is "G", had the same convolutional layer number and properties as in "F", but the number of neurons in dense layers were considered as 128 and 16, respectively. Table 4.2 present the details of the CNNs (D-E-F-G) in detail.

The  $3 \times 3$  filter size was used in all architectures, and a 0.2 dropout was used for each layer to avoid overfitting. The maximum pooling was applied as a pooling operation, and  $2 \times 2$  pooling was considered for each layer except the last convolutional layer of each architecture. The pooling was applied as  $1 \times 1$  in the last convolutional layer of each architecture not to minimize the extracted features.

**Table 4. 1:** Details of CNNs used in experiments (A-B-C).

<b>Name</b>	<b>Conv. Layer No.</b>	<b>Filters and Size</b>	<b>Pooling and Size</b>	<b>Dropout</b>	<b>Dense Layer No.</b>	<b>Neurons</b>
A	Conv. L.1	64 (3x3)	Max-pooling (2x2)	0.2	Dense L.1	128
	Conv. L.2	32 (3x3)	Max-pooling (1x1)	0.2	Dense L.2	32
B	Conv. L.1	64 (3x3)	Max-pooling (2x2)	0.2	Dense L.1	128
	Conv. L.2	32 (3x3)	Max-pooling (1x1)	0.2	Dense L.2	64
	-	-	-	-	Dense L.3	32
C	Conv. L.1	64 (3x3)	Max-pooling (2x2)	0.2	Dense L.1	128
	Conv. L.2	32 (3x3)	Max-pooling (2x2)	0.2	Dense L.2	64
	Conv. L.3	16 (3x3)	Max-pooling (1x1)	0.2	Dense L.3	32

---

**Table 4. 2:** Details of CNNs used in experiments (D-E-F-G).

<b>Name</b>	<b>Conv. Layer No.</b>	<b>Filters and Size</b>	<b>Pooling and Size</b>	<b>Dropout</b>	<b>Dense Layer No.</b>	<b>Neurons</b>
D	Conv. L.1	64 (3x3)	Max-pooling (2x2)	0.2	Dense L.1	64
	Conv. L.2	32 (3x3)	Max-pooling (1x1)	0.2	Dense L.2	32
E	Conv. L.1	64 (3x3)	Max-pooling (2x2)	0.2	Dense L.1	128
	Conv. L.2	32 (3x3)	Max-pooling (1x1)	0.2	Dense L.2	64
F	Conv. L.1	32 (3x3)	Max-pooling (2x2)	0.2	Dense L.1	128
	Conv. L.2	16 (3x3)	Max-pooling (1x1)	0.2	Dense L.2	64
G	Conv. L.1	32 (3x3)	Max-pooling (2x2)	0.2	Dense L.1	128

Conv. L.2	16 (3x3)	Max- pooling (1x1)	0.2	Dense L.2	16
-----------	----------	--------------------------	-----	-----------	----

---

The implemented architectures are also categorized based on their depth levels. The categorization is performed by considering the number of convolutional layers, the number of filters in the convolutional layers, the number of dense layers, and the neurons in dense layers, respectively. Table 4.3 shows the architectures and their determined depth levels. All architectures were implemented using Python 3.0 programming language and Keras Module.

**Table 4.3:** CNN architectures and depth levels

Accuracy	Level of Depth
A	4
B	6
C	7
D	3
E	5
F	2
G	1

#### 4.6 Summary

In this chapter, the design of experiments, considered evaluation metrics, and the implemented CNN architectures were presented in detail.

## CHAPTER 5

### RESULTS AND DISCUSSIONS

#### 5.1 Overview

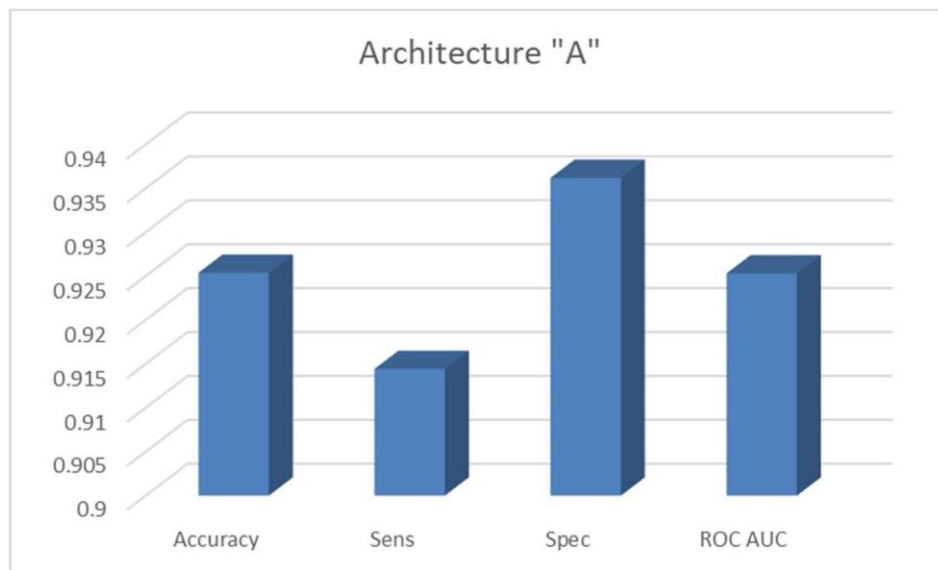
This chapter presents the results obtained in the experiments in detail and discussions on these results.

#### 5.2 Results

Seven different experiments were performed to achieve the optimal classification rates on the considered dataset and determine the best-fitted architecture.

##### 5.2.1 Results of architecture “A”

A CNN with two convolutional (64x32) and two dense layers (128x32) was trained using five-fold cross-validation. It achieved 92.53% accuracy, 91.44% sensitivity, 93.61% specificity, and 92.53% of ROC AUC score. Figure 5.1 shows the graphical representation of the obtained results, and Figure 5.2 shows the confusion matrix that demonstrates the TP, TN, FP, and FN results obtained by the architecture “A”.



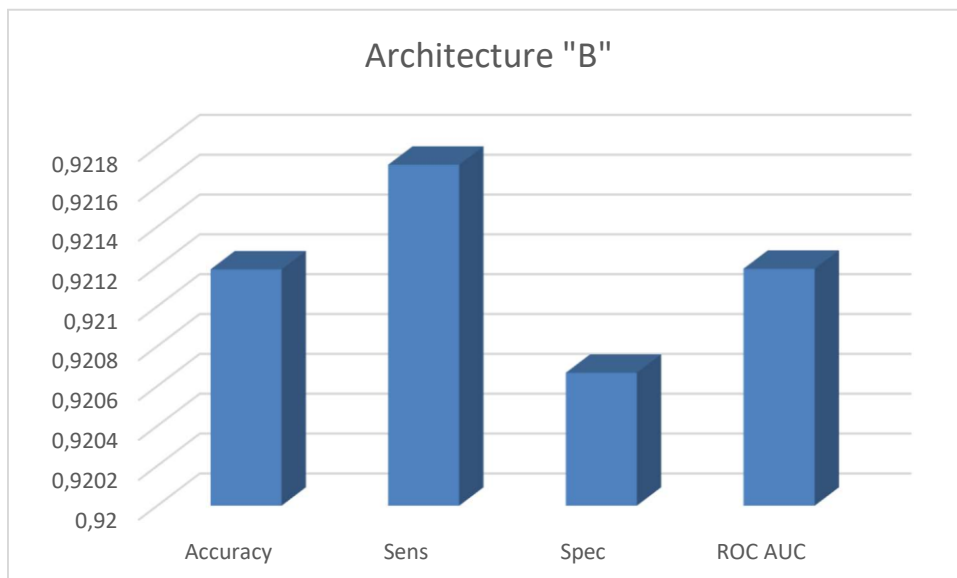
**Figure 5. 1:** Results obtained by the architecture “A”.

	Actually Positive	Actually Negative
Predicted Positive	3142	222
Predicted Negative	294	3257

**Figure 5. 2:** Confusion matrix for the architecture “A”.

### 5.2.2 Results of architecture “B”

A CNN with two convolutional (64x32) and three dense layers (128x64x32) was trained using five-fold cross-validation. It achieved 92.11% accuracy, 92.17% sensitivity, 92.06% specificity, and 92.11% of ROC AUC score. Figure 5.3 shows the graphical representation of the obtained results, and Figure 5.4 shows the confusion matrix of the architecture “B”.



**Figure 5. 3:** Results obtained by the architecture “B”.

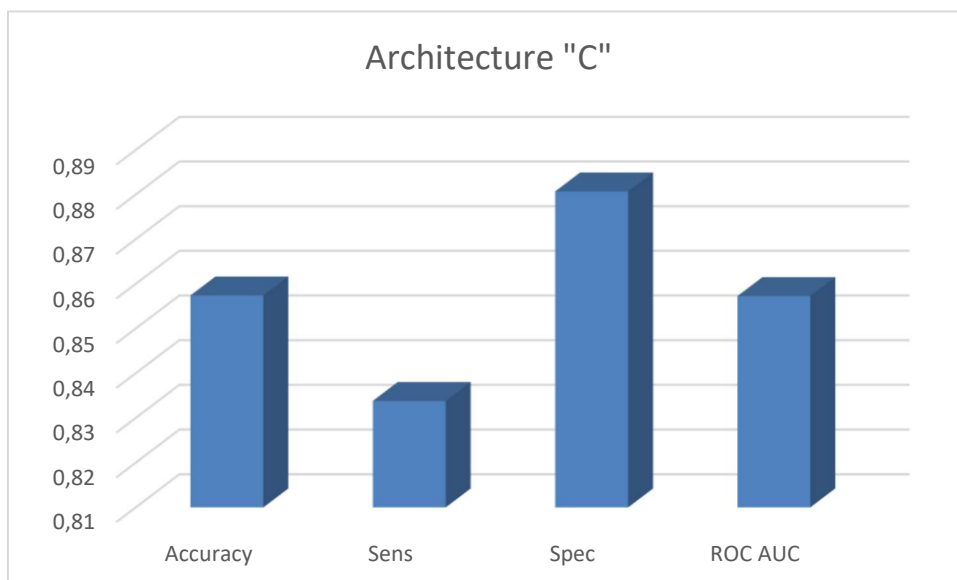


	Actually Positive	Actually Negative
Predicted Positive	3167	276
Predicted Negative	269	3203

**Figure 5. 4:** Confusion matrix for the architecture “B”.

### 5.2.3 Results of architecture “C”

A CNN with three convolutional (64x32x16) and three dense layers (128x64x32) was trained using five-fold cross-validation. It achieved 85.74% accuracy, 83.38% sensitivity, 88.07% specificity, and 85.72% of ROC AUC score. Figure 5.5 shows the graphical representation of the obtained results, and Figure 5.6 shows the confusion matrix of the architecture “C”.



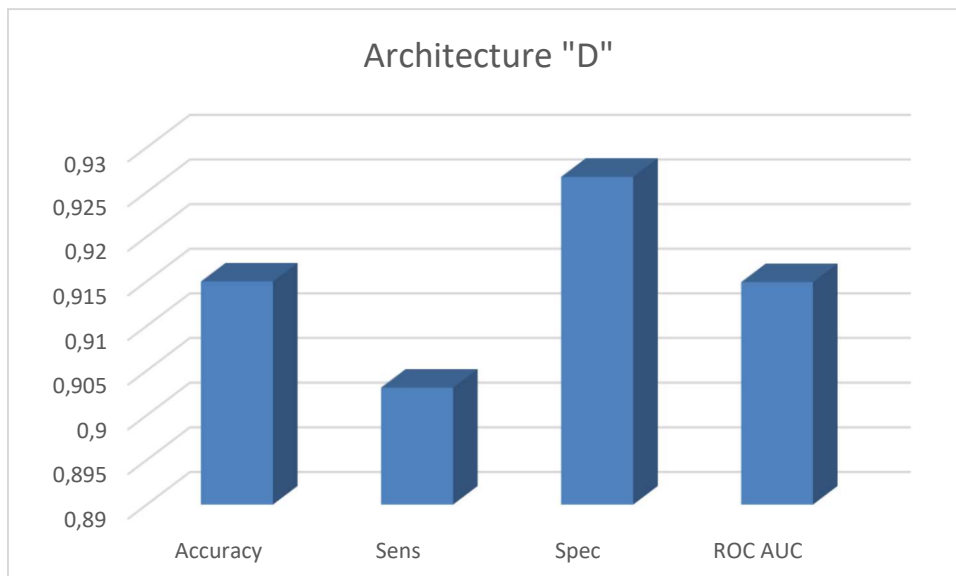
**Figure 5. 5:** Results obtained by the architecture “C”.

	Actually Positive	Actually Negative
Predicted Positive	2865	415
Predicted Negative	571	3064

**Figure 5. 6:** Confusion matrix for the architecture “C”.

### 5.2.4 Results of architecture “D”

A CNN with two convolutional (64x32) and two dense layers (64x32) was trained using five-fold cross-validation. It achieved 91.49% accuracy, 90.30% sensitivity, 92.67% specificity, and 91.48% of ROC AUC score. Figure 5.7 shows the graphical representation of the obtained results, and Figure 5.8 shows the confusion matrix of the architecture “D”.



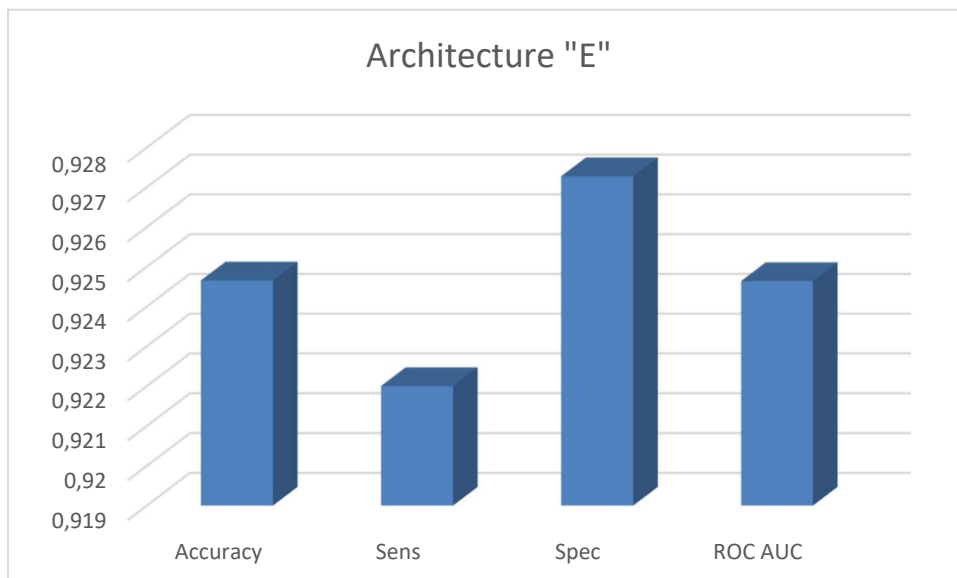
**Figure 5. 7:** Results obtained by the architecture “D”.

	Actually Positive	Actually Negative
Predicted Positive	3103	255
Predicted Negative	333	3224

**Figure 5. 8:** Confusion matrix for the architecture “D”.

### 5.2.5 Results of architecture “E”

A CNN with two convolutional (64x32) and two dense layers (128x64) was trained using five-fold cross-validation. It achieved 92.46% accuracy, 92.20% sensitivity, 92.72% specificity, and 92.46% of ROC AUC score. Figure 5.9 shows the graphical representation of the obtained results, and Figure 5.10 shows the confusion matrix of the architecture “E”.



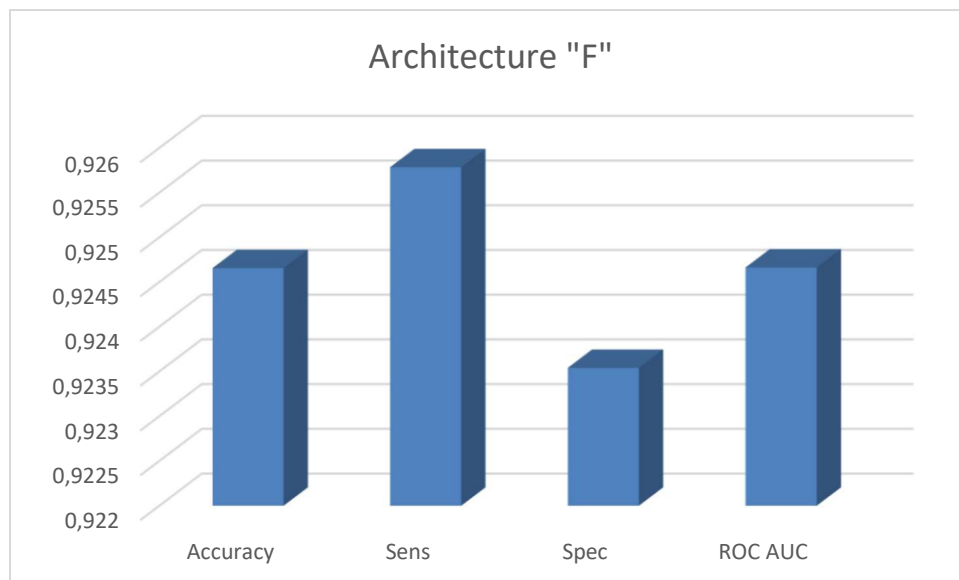
**Figure 5. 9:** Results obtained by the architecture “E”.

	Actually Positive	Actually Negative
Predicted Positive	3168	253
Predicted Negative	268	3226

**Figure 5. 10:** Confusion matrix for the architecture “E”.

### 5.2.6 Results of architecture “F”

A CNN with two convolutional (32x16) and two dense layers (128x64) was trained using five-fold cross-validation. It achieved 92.46% accuracy, 92.57% sensitivity, 92.35% specificity, and 92.46% of ROC AUC score. Figure 5.11 shows the graphical representation of the obtained results, and Figure 5.12 shows the confusion matrix of the architecture “F”.



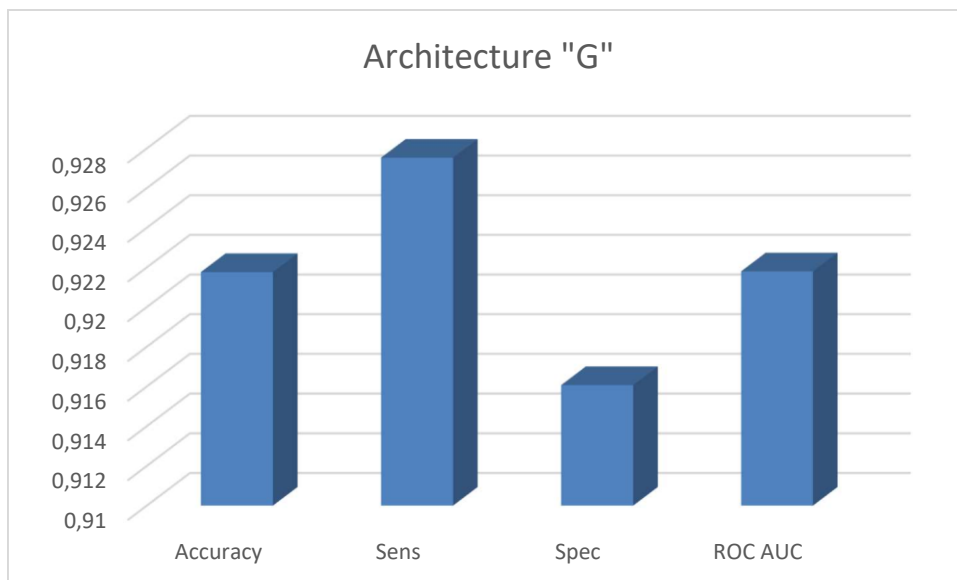
**Figure 5. 11:** Results obtained by the architecture “F”.

	Actually Positive	Actually Negative
Predicted Positive	3181	266
Predicted Negative	255	3213

**Figure 5. 12:** Confusion matrix for the architecture “F”.

### 5.2.7 Results of architecture “G”

A CNN with two convolutional (32x16) and two dense layers (128x16) was trained using five-fold cross-validation. It achieved 92.17% accuracy, 92.75% sensitivity, 91.60% specificity, and 92.17% of ROC AUC score. Figure 5.13 shows the graphical representation of the obtained results, and Figure 5.14 shows the confusion matrix of the architecture “G”.



**Figure 5. 13:** Results obtained by the architecture “G”.

	Actually Positive	Actually Negative
Predicted Positive	3187	292
Predicted Negative	249	3187

**Figure 5. 14:** Confusion matrix for the architecture “G”.

### 5.2.8 Comparison of the CNN architectures

As mentioned above, seven CNN architectures were implemented to determine the optimal architecture for tumor detection.

Architecture A, which consisted of two convolutional and dense layers, achieved the highest specificity rate (93.61%) and was followed by architecture E (92.72%). The only difference between architecture A and E was the number of neurons in the last dense layer, where A and E had 32 and 64 neurons, respectively.

The lowest specificity rate was obtained by the architecture C (88.07%), which was the deepest CNN considered in this thesis (three convolutional and three dense layers).

The highest sensitivity rates were achieved by the lightest CNN architectures considered in this thesis. The architecture G, which is the lightest architecture, achieved 92.75% sensitivity and was followed by F (92.57%). The architecture G consisted of two convolutional layers (32 and 16 filters, respectively) and two dense layers (128 and 16 neurons, respectively). The difference between the architectures F and G was only the last dense layer, where F had 64 neurons, and G had 16 neurons.

The lowest sensitivity rate (83.38%) was obtained by the deepest CNN architecture C, similar to obtained specificity rates.

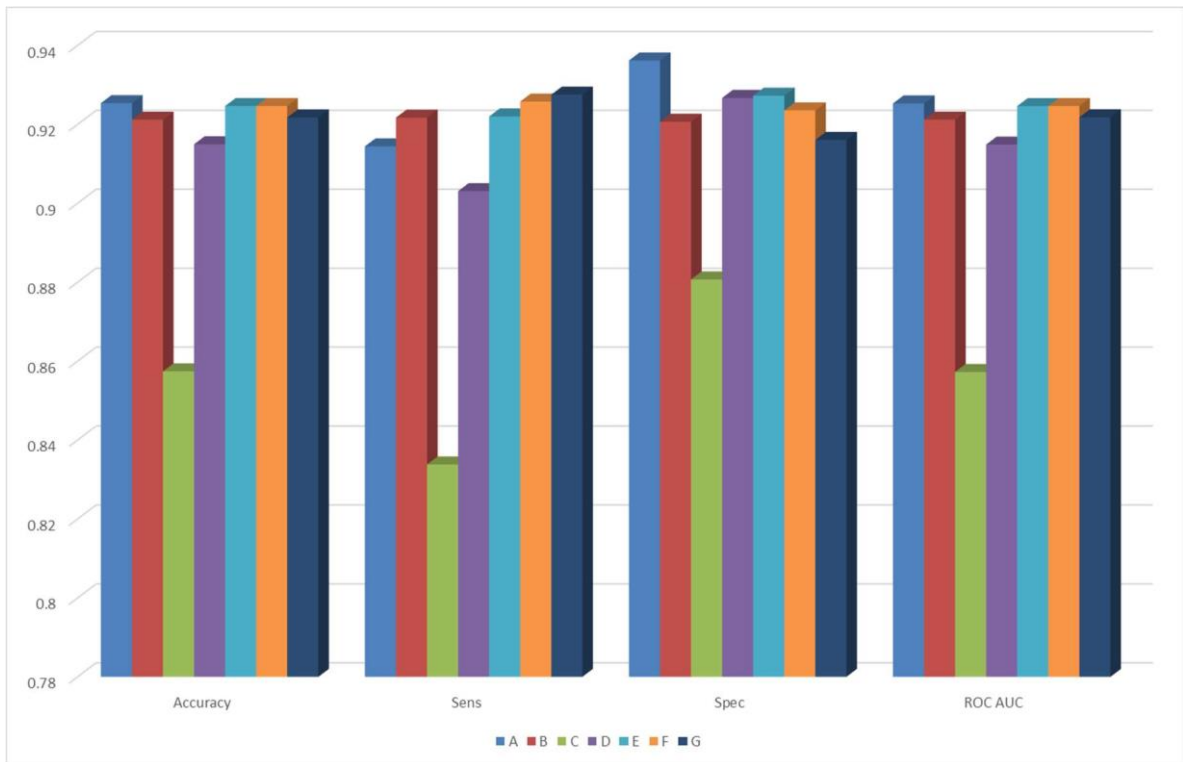
When the accuracy is considered, the highest rate was achieved by the architecture A (92.53%) and followed by the E (92.46%) and F (92.46%). The lowest accuracy was

obtained by the architecture C. The ROC AUC scores were at the same level and percentages with the accuracy rates because of the balanced dataset.

Table 5.1 presents the details of all results obtained in this thesis. Figure 5.15 visualizes the results obtained for each CNN architecture. Table 5.2 shows the sorted list of the CNN architectures according to ranks of obtained rates.

**Table 5. 1:** Results of all experiments (in %)

Architecture	Accuracy	Sensitivity	Specificity	ROC AUC Score
A	<b>92.53</b>	91.44	<b>93.61</b>	<b>92.53</b>
B	92.11	92.17	92.06	92.11
C	85.74	83.38	88.07	85.72
D	91.49	90.30	92.67	91.48
E	92.46	92.20	92.72	92.46
F	92.46	92.57	92.35	92.46
G	92.17	<b>92.75</b>	91.60	92.17



**Figure 5. 15:** Comparison of all architectures for the accuracy, sensitivity, specificity, and ROC AUC scores.

**Table 5. 2:** Ranking of the CNN architectures for all evaluation metrics

Accuracy	Sensitivity	Specificity	ROC AUC Score
A	G	A	A
E	F	E	E
F	E	D	F
G	B	F	G
B	A	B	B
D	D	G	D
C	C	C	C



### 5.3 Discussions

The obtained results showed that the different depth levels produced different results for predicting true positive and true negative samples. Table 5.3 presents the total true positive, true negative, false positive, and false-negative results obtained by the considered architectures, and Figure 5.16 visualizes these values.

**Table 5. 3:** Total TP, TN, FP, and FN results of all experiments

Architecture	TP	TN	FP	FN
A	3142	<b>3257</b>	<b>222</b>	294
B	3167	3203	276	269
C	2865	3064	415	571
D	3103	3224	255	333
E	3168	3226	253	268
F	3181	3213	266	255
G	<b>3187</b>	3187	292	<b>249</b>

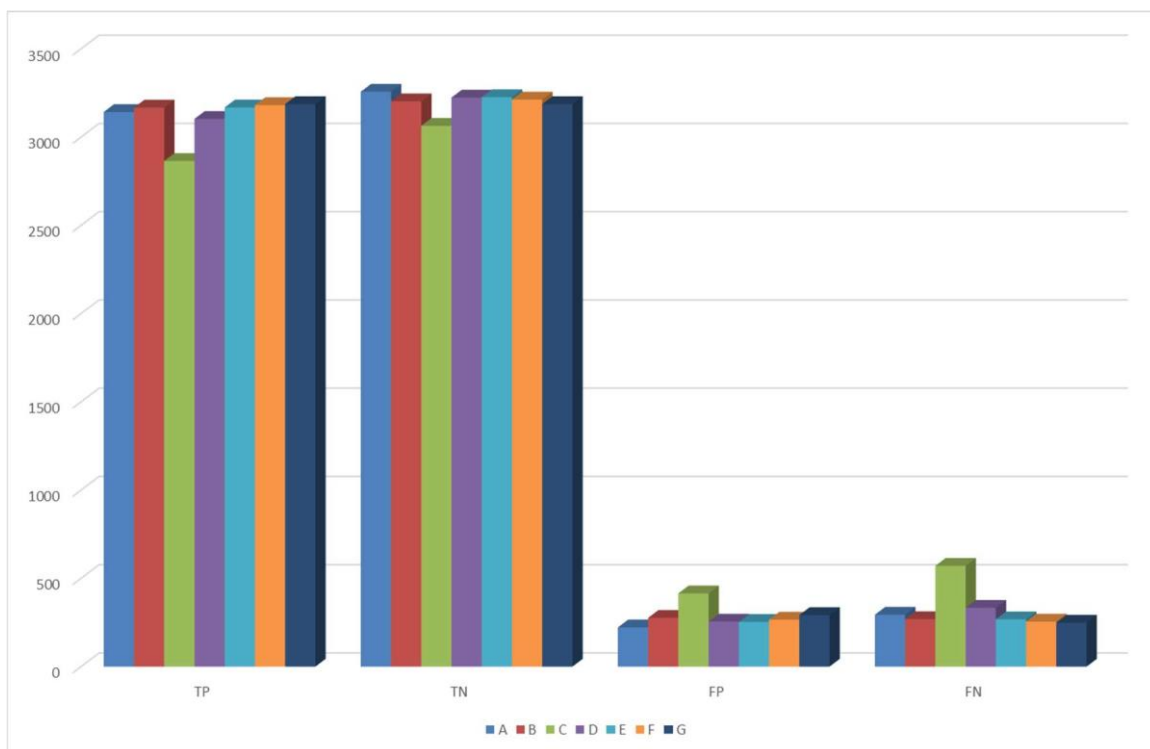
Lighter CNN architectures achieved better rates than the deeper ones in terms of sensitivity rates, which shows the better performance in predicting true positives. The highest sensitivity (TP = 3187) was achieved by the architecture G, which consisted of the lowest number of filters in convolutional layers. It also outperformed the architecture F (TP = 3181), which only has more neurons in its' last dense layer. However, they could not achieve the highest specificity rates compared to other architectures.

The increment of filter numbers in convolution layers improved the specificity rates, which indicates the better ability of the model for predicting true negatives. The optimal specificity rates were obtained by A (TN = 3257) and E (TN = 3226).

When the optimal models of specificity and sensitivity rates were compared, the architecture G outperformed A by 1.32% in terms of sensitivity. However, Architecture A outperformed G by 2.01% in terms of specificity rates.

When the accuracy and ROC AUC scores, which are the primary evaluation metrics used in this thesis, were considered, it was seen that the architecture A, which had the median level of depth, achieved the highest rates and followed by E. The deepest architecture C produced the lowest rates for all metrics.

The rank of the architectures in terms of accuracy and the level of depths are shown in Table 5.4.



**Figure 5. 16:** Comparison of all architectures for true positive, true negative, false positive, and false-negative results (Highest bars are optimal results for TP and TN, and lowest bars are optimal results for FP and FN).

**Table 5. 4:** Ranking and the levels of depth of the CNN architectures for the accuracy

Accuracy	Level of Depth
A	4
E	5
F	2
G	1
B	6
D	3
C	7

The obtained results showed that the overfitting occurred in the deepest model C. The increment of the number of convolution and dense layers is not suitable for the classification of images considered in this dataset. Similarly, the increment of only the dense layer number reduced the success level of the model.

In addition, the obtained results of the architectures G and F indicated that the decrement of the filter numbers in the convolutional layer might increase the sensitivity level of the model but decreases the specificity level.

When all the obtained results and the architectures considered, it could be suggested to be considered a CNN architecture with two convolution layers (64 and 32 filters, respectively) and two dense layers (128 and 32 neurons, respectively) as an initial point of experimentation for the binary brain tumor classification in MRI.

#### **5.4 Summary**

This chapter presented the details of experimental results, a comparison of the considered models, and the discussions on the results.

## **CHAPTER 6**

### **CONCLUSION**

Brain tumors require following their progress after the treatment process. The progress of the tumors has vital importance for the patients to observe the influence of the treatment. Besides, the progress of the tumors directs the doctors about starting, continuing, or changing the applied therapy.

In this thesis, the progress of brain tumors were classified into two groups as initial and progress using convolutional neural networks. The Brain Tumor Progress dataset was considered, and all slices of MRI images were fed to the network. Seven different convolutional neural network architectures, which were considered for testing the lighter and deeper architectures, were implemented and compared in order to achieve optimal classification rates. The results were evaluated using four evaluation metrics.

The obtained results showed that the use of incremented depth of the convolutional neural networks caused overfitting and could not produce reasonable rates. Similarly, the use of overly simplified architecture caused the network to learn one class better than another class. In this thesis, the proper architectures were determined to classify brain tumor progress for a particular dataset. The optimal classification rates were achieved by the median depth of the CNN's considered in this thesis.

The optimal accuracy rate (92.53%), which was the primary indicator considered in this thesis, was obtained by a CNN with two convolutional (64x32 filters) and two dense layers (128x32 neurons).

The results suggested that the implementation of an appropriate CNN architecture can achieve reasonable rates and help the medical doctors follow-up the progression of the brain tumors after the therapy.

The future work will include the use of transfer learning by implementing pre-trained networks and compare the optimal results obtained in this thesis. In addition, the level of progression will be taken into consideration using deep learning approaches.

## REFERENCES

- Alqudah, Ali & Alquran, Hiam & Abuqasmieh, Isam & Alqudah, Amin & Al-Sharu, Wafaa. (2020). Brain Tumor Classification Using Deep Learning Technique - A Comparison between Cropped, Uncropped, and Segmented Lesion Images with Different Sizes. *International Journal of Advanced Trends in Computer Science and Engineering*, 8, 3684 - 3691.
- A Fourcade, R.H. Khonsari. ( 2019). Deep learning in medical image analysis: A third eye for doctors, *Journal of Stomatology, Oral and Maxillofacial Surgery*, 120(4), 279-288.
- Chougrad H, Zouaki H, Alheyane O. (2018). Deep Convolutional Neural Networks for breast cancer screening. *Comput Methods Programs Biomed*, [doi: 10.1016/j.cmpb.2018.01.011](https://doi.org/10.1016/j.cmpb.2018.01.011)
- C. Szegedy, W. Liu, Y. Jia and P. Sermanet. (2015). Going deeper with convolutions *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-9). Boston, MA.
- Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. (2013). The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, *Journal of Digital Imaging*, 26(6), 1045-1057.
- Dr. P. Kalyani, Dr. D. Murali, T. Munireddy (2018). Identifying the Brain Tumors and Classified Using a New Approach with The Support of Random Forest Decision Tree. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 3(6), 270-273.
- Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics* (pp. 315-323). [www.jmlr.org/proceedings/papers/v15/glorot11a/glorot11a.pdf](http://www.jmlr.org/proceedings/papers/v15/glorot11a/glorot11a.pdf) [Google Scholar](#)

- H. H. Sultan, N. M. Salem and W. Al-Atabany. (2019) . Multi-Classification of Brain Tumor Images Using Deep Neural Network, *in IEEE Access*, [doi: 10.1109/ACCESS.2019.2919122](https://doi.org/10.1109/ACCESS.2019.2919122)
- Hang Zhou and Qichen Sun. (2020). Research on Principle and Application of Convolutional Neural Networks. *IOP Conf. Ser.: Earth Environ. Sci.* 440 042055
- June-Goo Lee, Sanghoon Jun, Young-Won Cho, Hyunna Lee, Guk Bae Kim, Joon Beom Seo, Namkug Kim (2017). Deep Learning in Medical Imaging. *Korean Journal of Radiology* 18(4), 570-584.
- Jarrett, K., Kavukcuoglu, K., & LeCun, Y. (2009). What is the best multi-stage architecture for object recognition? In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2146–2153). Red Hook, NY: Curran. [Crossref](#), [Google Scholar](#)
- Khan, M.A.; Ashraf, I.; Alhaisoni, M.; Damaševičius, R.; Scherer, R.; Rehman, A.; Bukhari, S.A.C. (2020). Multimodal Brain Tumor Classification Using Deep Learning and Robust Feature Selection: A Machine Learning Application for Radiologists. *Diagnostics*, 10(8), 565.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, (pp. 1097–1105). Red Hook, NY: Curran. [Google Scholar](#)
- Krizhevsky, A. (2009). Learning multiple layers of features from tiny images. *Master's thesis*, University of Toronto, Canada. [Google Scholar](#)
- Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*, <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neuralnetworks.pdf>.

- Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, Scheithauer BW, Kleihues P. (2007). The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol*, 114(2), 97-109. Epub 2007 Jul 6. Erratum in: *Acta Neuropathol*. 114(5), 547. [doi: 10.1007/s00401-007-0243-4](https://doi.org/10.1007/s00401-007-0243-4).
- Lee, K.-S.; Kim, J.Y.; Jeon, E.-T.; Choi, W.S.; Kim, N.H.; Lee, K.Y. (2020). Evaluation of Scalability and Degree of Fine-Tuning of Deep Convolutional Neural Networks for COVID-19 Screening on Chest X-ray Images Using Explainable Deep-Learning Algorithm. *J. Pers. Med.* 10(4), 213.
- Lee KJ, Ryoo I, Choi D, Sunwoo L, You SH, Jung HN. (2020). Performance of deep learning to detect mastoiditis using multiple conventional radiographs of mastoid. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0241796>
- LeCun Y., Haffner P., Bottou L., Bengio Y. (1999) Object Recognition with Gradient-Based Learning. In: Shape, Contour and Grouping in Computer Vision. *Lecture Notes in Computer Science, 1681*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/3-540-46805-6\\_19](https://doi.org/10.1007/3-540-46805-6_19)
- LeCun, Y., Huang, F. J., & Bottou, L. (2004). Learning methods for generic object recognition with invariance to pose and lighting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 97–104). Red Hook, NY: Curran. [Crossref](#), [Google Scholar](#)
- Mohsen, Heba & El-Dahshan, El-Sayed & El-Horbarty, El-Sayed & M.Salem, Abdel-Badeeh. (2017). Classification using Deep Learning Neural Networks for Brain Tumors. *Future Computing and Informatics Journal*, 3(1), <https://doi.org/10.1016/j.fcij.2017.12.001>.
- Muhammad Sajjad, Salman Khan, Khan Muhammad, Wanqing Wu, Amin Ullah, Sung Wook Baik. (2019). Multi-grade brain tumor classification using deep CNN with extensive data augmentation, *Journal of Computational Science*, 30, 174-182, <https://doi.org/10.1016/j.jocs.2018.12.003>.

- M.D. Zeiler and R. Fergus. (2014). Visualization and understanding convolutional networks, *European Conference on Computer Vision*, ( pp. 818-833). Zurich
- Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning* (pp. 807–814). N.p.: International Machine Learning Society. [Google Scholar](#)
- Oliver Faust, Yuki Hagiwara, Tan Jen Hong, Oh Shu Lih, U Rajendra Acharya. (2018). Deep learning for healthcare applications based on physiological signals: A review, *Computer Methods and Programs in Biomedicine*, 161, 1-13.
- Sekeroglu B, Ozsahin I. (2020). Detection of COVID-19 from Chest X-Ray Images Using Convolutional Neural Networks. *SLAS TECHNOLOGY: Translating Life Sciences Innovation*, 25(6), 553-565. [doi:10.1177/2472630320958376](https://doi.org/10.1177/2472630320958376)
- Strong MJ, Garces J, Vera JC, Mathkour M, Emerson N, et al. (2015) Brain Tumors: Epidemiology and Current Trends in Treatment. *Brain Tumors Neurooncol*, 1(1), 1-21.
- Simonyan, K.; Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv*. arXiv:14091556.
- Schmainda KM, Prah M (2018). Data from Brain-Tumor-Progression. *The Cancer Imaging Archive*. <http://doi.org/10.7937/K9/TCIA.2018.15quzvnb>
- Tobore I, Li J, Yuhang L, Al-Handarish Y, Kandwal A, Nie Z, Wang L (2019). Deep Learning Intervention for Health Care Challenges: Some Biomedical Domain Considerations. *JMIR Mhealth Uhealth*, 7(8), [DOI: 10.2196/11966](https://doi.org/10.2196/11966)
- Waseem Rawat and Zenghui Wang Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*, 29(9), 2352-2449. [https://doi.org/10.1162/neco\\_a\\_00990](https://doi.org/10.1162/neco_a_00990)



- Xiaofeng Han and Yan Li. (2015). The Application of Convolution Neural Networks in Handwritten Numeral Recognition. *International Journal of Database Theory and Application*, 8(3), 367-376. <http://dx.doi.org/10.14257/ijdta.2015.8.3.32>
- Xin, M., Wang, Y. (2019). Research on image classification model based on deep convolution neural network. *J Image Video Proc.* 2019 (40), <https://doi.org/10.1186/s13640-019-0417-8>.
- Yamashita, R., Nishio, M., Do, R.K.G. et al. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging*, 9(4), 611–629. <https://doi.org/10.1007/s13244-018-0639-9>
- Y. Lecun, L.Bottou, Bengio and P. Haffner. (1998) Gradient-based learning applied to document recognition, *Proc. IEEE*, 86(11), 2278-2324.
- <https://wiki.cancerimagingarchive.net/display/Public/Brain-Tumor-Progression>

## **APPENDICES**

## APPENDIX 1: ETHICAL APPROVAL DOCUMENT



### ETHICAL APPROVAL DOCUMENT

Date: 23/12/2020

To the Graduate School of Applied Sciences

For the thesis project entitled as “Brain Tumor Progress Classification Using Deep Learning” the researchers declare that they did not collect any data from human/animal or any other subjects. Therefore, this project does not need to go through the ethics committee evaluation.

Title: Assoc. Prof. Dr.

Name Surname: Boran Şekeroğlu

Signature:



Role in the Research Project: Supervisor

## APPENDIX 2 SIMILARITY REPORT

turnitin.com/t\_inbox.asp?aid=101142726&lang=en\_us&session-id=87c3414883cc47669e5d53e4235a2e3e

Uygulamalar Gmail YouTube Haritalar Installing a Python... evaluation - Micro... Applied Sciences |... Course Offer for 20... ASYU 2020 | Akilli S... ASYU 2020 | Akilli S... Indicators — Europ...

Assignments Students Grade Book Libraries Calendar Discussion Preferences

NOW VIEWING: HOME > HEDI SYAMAND AZAT AZAT > HEDI SYAMAND AZAT AZAT

About this page  
This is your assignment inbox. To view a paper, select the paper's title. To view a Similarity Report, select the paper's Similarity Report icon in the similarity column. A ghosted icon indicates that the Similarity Report has not yet been generated.

Hedi Syamand Azat Azat  
INBOX | NOW VIEWING: NEW PAPERS ▾

Submit File Online Grading Report | Edit assignment settings | Email non-submitters

<input type="checkbox"/>	AUTHOR	TITLE	SIMILARITY	GRADE	RESPONSE	FILE	PAPER ID	DATE
<input type="checkbox"/>	Hedi Syamand Azat Az...	Abstract	0%	--	--		1480848916	23-Dec-2020
<input type="checkbox"/>	Hedi Syamand Azat Az...	Chapter 6 - Conclusion	0%	--	--		1480848904	23-Dec-2020
<input type="checkbox"/>	Hedi Syamand Azat Az...	Chapter 2	4%	--	--		1480848135	23-Dec-2020
<input type="checkbox"/>	Hedi Syamand Azat Az...	Chapter 5 - Results and Discussions	4%	--	--		1480848340	23-Dec-2020
<input type="checkbox"/>	Hedi Syamand Azat Az...	Chapter 1 - Introduction	6%	--	--		1480848075	23-Dec-2020
<input type="checkbox"/>	Hedi Syamand Azat Az...	All Thesis	8%	--	--		1480848032	23-Dec-2020
<input type="checkbox"/>	Hedi Syamand Azat Az...	Chapter 3	8%	--	--		1480848213	23-Dec-2020
<input type="checkbox"/>	Hedi Syamand Azat Az...	Chapter 4	14%	--	--		1480848266	23-Dec-2020

Assoc.Prof.Dr. Boran Şekeroğlu

