



Faculty of Engineering

**DETECTION AND CLASSIFICATION OF BRAIN CANCER
USING DEEP LEARNING**

Asvin Kumar A/L Moghan

Bachelor of Engineering

Electrical and Electronics Engineering with Honours

2023

UNIVERSITI MALAYSIA SARAWAK

Grade: _____

Please tick (✓)

Final Year Project Report

Masters

PhD

DECLARATION OF ORIGINAL WORK

This declaration is made on the 25th day of July 2023.

Student's Declaration:

I, ASVIN KUMAR A/L MOGHAN, MATRIC NO.: 69148, FACULTY OF ENGINEERING hereby declare that the work entitled BRAIN CANCER DETECTION AND CLASSIFICATION USING DEEP LEARNING is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

25 JULY 2023

Date submitted



ASVIN KUMAR A/L MOGHAN (69148)

Supervisor's Declaration:

I, ASSOCIATE PROFESSOR IR DR DAVID BONG BOON LIANG hereby certifies that the work entitled BRAIN CANCER DETECTION AND CLASSIFICATION USING DEEP LEARNING was prepared by the above named student, and was submitted to the "FACULTY" as a * partial/full fulfillment for the conferment of BACHELOR OF ENGINEERING (HONS) IN ELECTRICAL AND ELECTRONICS ENGINEERING, and the aforementioned work, to the best of my knowledge, is the said student's work.

Received for examination by: _____



AP IR DR DAVID BONG BOON LIANG

Date: 25 JULY 2023


I declare that Project/Thesis is classified as (Please tick (√)):


- CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*
 RESTRICTED (Contains restricted information as specified by the organisation where research was done)*
 OPEN ACCESS

Validation of Project/Thesis

I therefore duly affirm with free consent and willingly declare that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abiding interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies for the purpose of academic and research only and not for other purpose.
- The Centre for Academic Information Services has the lawful right to digitalise the content for the Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic exchange between Higher Learning Institute.
- No dispute or any claim shall arise from the student itself neither third party on this Project/Thesis once it becomes the sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student except with UNIMAS permission.

Student signature 
(25 JULY 2023)

Supervisor signature: 
(25 JULY 2023)

Current Address: FACULTY OF ENGINEERING, UNIVERSITI MALAYSIA SARAWAK, JALAN DATUK MOHAMMAD MUSA, 94300 KOTA SAMARAHAN, SARAWAK

Notes: * If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the period and reasons of confidentiality and restriction.

[The instrument is duly prepared by The Centre for Academic Information Services]

DETECTION AND CLASSIFICATION OF BRAIN CANCER
USING DEEP LEARNING

ASVIN KUMAR A/L MOGHAN

A dissertation submitted in partial fulfilment
of the requirement for the degree of
Bachelor of Engineering
Electrical and Electronics Engineering with Honours

Faculty of Engineering
Universiti Malaysia Sarawak

2023

ACKNOWLEDGEMENT

First and foremost, I would like to thank our Almighty God for His grace, wisdom and protection throughout the research and writing process. Without His guidance, this project would not have been feasible.

I would also like to extend my deepest gratitude to Associate Professor Ir. Dr. David Bong Boon Liang, my Final Year Project supervisor, for his unwavering guidance and support, as well as for generously sharing his skills from the beginning to the completion of this academic journey.

Additionally, I must thank Universiti Malaysia Sarawak for providing access to research facilities and resources. Especially, I am very grateful to the Department of Electrical and Electronics Engineering at Universiti Malaysia Sarawak for their support and resources throughout this academic endeavour.

Last but not least, I would like to express my heartfelt thanks to my family for their love, support and encouragement over the course of conducting research and writing the paper.

ABSTRACT

Brain cancer is a serious medical condition that requires an accurate and timely diagnosis for effective treatment planning. In recent years, deep learning techniques have shown great potential in the field of medical image analysis. In this study, a brain cancer detection and classification system based on deep learning algorithms is proposed. The system utilises a convolutional neural network (CNN) architecture trained on a large dataset of brain MRI images. The images were preprocessed to enhance relevant features and remove noise. The CNN architecture chosen was GoogleNet. To validate the robustness of the system, a 5-fold cross-validation approach was employed, ensuring reliable and consistent results. The proposed system has the potential to assist medical professionals in the early detection and classification of brain tumours, aiding in accurate diagnosis and treatment decision-making. By automating the classification process, it reduces the burden of manual analysis, potentially saving time and improving the overall efficiency of the diagnostic process. The proposed model achieved an accuracy of $97.5522 \pm 0.2739\%$, a precision of 0.9498 ± 0.0054 , a recall of 0.9494 ± 0.0057 , a specificity 0.9839 ± 0.0018 and an F1 Score of 0.9493 ± 0.0057 across the 5-fold cross-validation iterations, demonstrating its effectiveness in accurately classifying brain MRI.

ABSTRAK

Kanser otak merupakan keadaan perubatan yang serius yang memerlukan diagnosis yang tepat, dan tepat pada waktunya untuk perancangan rawatan yang berkesan. Dalam beberapa tahun terakhir, teknik pembelajaran mendalam telah menunjukkan potensi besar dalam bidang analisis imej perubatan. Dalam kajian ini, satu sistem pengesanan dan pengelasan kanser otak berdasarkan algoritma pembelajaran mendalam dicadangkan. Sistem ini menggunakan jaringan saraf terkonvolusi (CNN) berdasarkan senibina GoogleNet yang dilatih menggunakan dataset yang besar dari imej MRI otak. Imej-imej tersebut telah melalui pra-pemprosesan untuk meningkatkan ciri-ciri yang relevan dan mengurangkan kebingaran data. Pendekatan 5-lipatan validasi persilangan diterapkan untuk mengesahkan kebolehpercayaan sistem dengan hasil yang boleh diandalkan dan konsisten. Sistem yang dicadangkan mempunyai potensi untuk membantu profesional perubatan dalam pengesanan awal dan pengelasan tumor otak, membantu dalam diagnosis yang tepat dan membuat keputusan rawatan. Dengan mengautomatikkan proses pengelasan, ia mengurangkan beban analisis manual, berpotensi menjimatkan masa dan meningkatkan kecekapan keseluruhan proses diagnostik. Model yang dicadangkan mencapai ketepatan sebanyak $97.5522 \pm 0.2739\%$, kefahaman sebanyak 0.9498 ± 0.0054 , pemulihan sebanyak 0.9494 ± 0.0057 , khususiti sebanyak 0.9839 ± 0.0018 dan Skor F1 sebanyak 0.9493 ± 0.0057 , melintasi iterasi persilangan, sekali gus membuktikan menunjukkan keberkesanannya dalam mengelaskan imej MRI otak secara tepat.

TABLE OF CONTENTS

| | |
|--|------------|
| ACKNOWLEDGEMENT | i |
| ABSTRACT | ii |
| ABSTRAK | iii |
| TABLE OF CONTENTS | iv |
| LIST OF TABLES | vii |
| LIST OF FIGURES | x |
| LIST OF FORMULAS | xii |
| LIST OF ABBREVIATIONS | xiv |
| Chapter 1 INTRODUCTION | 1 |
| 1.1 Background | 1 |
| 1.1.1 Brain Cancer | 1 |
| 1.1.2 Deep Learning | 3 |
| 1.2 Problem Statement | 4 |
| 1.3 Objectives | 5 |
| 1.4 Project Contributions | 6 |
| 1.5 Project Scope | 6 |
| 1.6 Chapter Outline | 7 |
| Chapter 2 LITERATURE REVIEW | 8 |
| 2.1 Overview | 8 |
| 2.2 Related Studies | 8 |
| 2.2.1 Brain Cancer | 8 |
| 2.2.2 DL for Cancer Detection and Classification | 10 |
| 2.2.3 Imaging Modality | 11 |
| 2.2.4 Dataset | 12 |

| | | | |
|------------------|-------|---|-----------|
| | 2.2.5 | Pre-processing | 15 |
| | 2.2.6 | Modelling Approach | 19 |
| | 2.3 | Research gap | 30 |
| | 2.4 | Summary | 30 |
| Chapter 3 | | METHODOLOGY | 32 |
| | 3.1 | Overview | 32 |
| | 3.2 | Preparation | 33 |
| | 3.2.1 | Hardware and Software | 33 |
| | 3.2.2 | Dataset | 33 |
| | 3.2.3 | Image Pre-processing | 34 |
| | 3.3 | Proposed Detection and Classification Model | 36 |
| | 3.3.1 | GoogleNet | 36 |
| | 3.3.2 | Transfer Learning | 37 |
| | 3.3.3 | Modified GoogleNet | 38 |
| | 3.3.4 | Fine-tuning | 39 |
| | 3.3.5 | Experimental Settings | 40 |
| | 3.3.6 | Cross-validation | 42 |
| | 3.4 | Testing | 43 |
| | 3.4.1 | Confusion Matrix | 43 |
| | 3.4.2 | Class-specific Metrics | 45 |
| | 3.4.3 | Overall Performance Metrics | 47 |
| | 3.5 | Summary | 48 |
| Chapter 4 | | RESULTS AND DISCUSSION | 50 |
| | 4.1 | Overview | 50 |
| | 4.2 | Experimental Results | 50 |
| | 4.2.1 | First Cross-validation Iteration | 50 |

| | | |
|------------------|--|-----------|
| 4.2.2 | Second Cross-validation Iteration | 55 |
| 4.2.3 | Third Cross-validation Iteration | 59 |
| 4.2.4 | Fourth Cross-validation Iteration | 63 |
| 4.2.5 | Fifth Cross-validation Iteration | 68 |
| 4.2.6 | Average | 72 |
| 4.3 | Analysis and Discussion | 77 |
| 4.3.1 | Modifications to Key Settings | 77 |
| 4.3.2 | GoogleNet Architecture | 78 |
| 4.3.3 | Interpretation of Classification Metrics | 80 |
| 4.3.4 | Misclassifications as Non-meningioma | 80 |
| 4.4 | Summary | 81 |
| Chapter 5 | CONCLUSIONS | 82 |
| 5.1 | General Conclusions | 82 |
| 5.2 | Limitations | 83 |
| 5.3 | Future Scopes | 84 |
| | REFERENCES | 86 |
| | Appendix A | 91 |
| | Appendix B | 93 |

LIST OF TABLES

| Table | Page |
|---|-------------|
| Table 2.1: Baseline Characteristics and Standard Mortality Rates for Brain Cancer Patients Who Died of Suicide [7] | 10 |
| Table 2.2: Brain Tumour Dataset | 14 |
| Table 2.3: Pre-processing Methods | 18 |
| Table 2.4: Papers which used CNN..... | 22 |
| Table 2.5: Papers which used Non-CNN..... | 28 |
| Table 3.1: Brain MRI Classes and Their Respective Number of Images..... | 33 |
| Table 3.2: Training Settings | 40 |
| Table 3.3: Weightage of Each Class in the Testing Set of Every Cross-validation Iteration..... | 47 |
| Table 4.1: Details of the Training and Validation Process for the First Cross-validation Iteration..... | 51 |
| Table 4.2: Confusion Matrix (where G, M, P and N refer to Glioma, Meningioma, Pituitary Tumour and Normal, respectively) for the First Cross-validation Iteration..... | 52 |
| Table 4.3: No. of Samples and Counts of TP, TN, FP and FN for the First Cross- validation Iteration..... | 53 |
| Table 4.4: Class-specific Metrics for the First Cross-validation Iteration | 53 |
| Table 4.5: Overall Performance Metrics for the First Cross-validation Iteration | 54 |
| Table 4.6: Details of the Training and Validation Process for the Second Cross- validation Iteration..... | 55 |
| Table 4.7: Confusion Matrix (where G, M, P and N refer to Glioma, Meningioma, Pituitary Tumour and Normal, respectively) for the Second Cross-validation Iteration..... | 56 |
| Table 4.8: No. of Samples and Counts of TP, TN, FP and FN for the Second Cross- validation Iteration..... | 57 |
| Table 4.9: Class-specific Metrics for the Second Cross-validation Iteration | 57 |
| Table 4.10: Overall Performance Metrics for the Second Cross-validation Iteration | 58 |

| | |
|--|----|
| Table 4.11: The Details of the Training and Validation Process for the Third Cross-validation Iteration..... | 59 |
| Table 4.12: The Confusion Matrix (where G, M, P and N refer to Glioma, Meningioma, Pituitary Tumour and Normal, respectively) for the Third Cross-validation Iteration..... | 60 |
| Table 4.13: No. of Samples Counts of TP, TN, FP and FN for the Third Cross-validation Iteration..... | 61 |
| Table 4.14: Class-specific Metrics for the Third Cross-validation Iteration..... | 62 |
| Table 4.15: Overall Performance Metrics for the Third Cross-validation Iteration..... | 62 |
| Table 4.16: The Details of the Training and Validation Process for the Fourth Cross-validation Iteration..... | 64 |
| Table 4.17: The Confusion Matrix (where G, M, P and N refer to Glioma, Meningioma, Pituitary Tumour and Normal, respectively) for the Fourth Cross-validation Iteration..... | 65 |
| Table 4.18: No. of Samples and Counts of TP, TN, FP and FN for the Fourth Cross-validation Iteration..... | 66 |
| Table 4.19: Class-specific Metrics for the Fourth Cross-validation Iteration..... | 66 |
| Table 4.20: Overall Performance Metrics for the Fourth Cross-validation Iteration..... | 67 |
| Table 4.21: The Details of the Training and Validation Process for the 5 th Cross-validation Iteration..... | 68 |
| Table 4.22: The Confusion Matrix (where G, M, P and N refer to Glioma, Meningioma, Pituitary Tumour and Normal, respectively) for the Fifth Cross-validation Iteration..... | 69 |
| Table 4.23: No. of Samples and Counts of TP, TN, FP and FN for the Fifth Cross-validation Iteration..... | 70 |
| Table 4.24: Class-specific Metrics in the Fifth Cross-validation Iteration..... | 70 |
| Table 4.25: Overall Performance Metrics for the Fifth Cross-validation Iteration..... | 71 |
| Table 4.26: Training and Validation Results..... | 72 |
| Table 4.27: Class-specific Metrics Average..... | 73 |
| Table 4.28: Averages of Overall Performance Metrics across the 5-fold Cross-validation Iterations..... | 75 |
| Table 4.29: Misclassifications for Each Iteration..... | 76 |
| Table 4.30: Training Settings..... | 79 |

| | |
|---|----|
| Table 4.31: The Differences in Performance between a Non-fine-tuned GoogleNet Architecture and a Fine-tuned GoogleNet Architecture..... | 79 |
|---|----|

LIST OF FIGURES

| Figure | Page |
|--|-------------|
| Figure 1.1: Human Brain [1] | 1 |
| Figure 1.2: Euler’s Diagram on AI, ML, NN and DL [3] | 3 |
| Figure 1.3: DL can be used to detect the presence of brain cancer in this MRI [4] | 4 |
| Figure 2.1: The Steps Taken by Arunachalam and Royappan [23] for Image Enhancement..... | 16 |
| Figure 2.2: Data Augmentation that was Performed by Francisco et al [26] | 16 |
| Figure 2.3: The Methods Used by Kang et al [27] to Crop the MRIs | 17 |
| Figure 2.4: The Procedure Used by Saxena et al [29] to Crop the MRIs | 17 |
| Figure 3.1: The Flowchart of the Proposed Approach | 33 |
| Figure 3.2: Examples of the Brain MRIs in the Dataset [4]. a) Axial View, b) Coronal View and c) Sagittal View | 34 |
| Figure 3.3: Image Pre-processing Steps | 35 |
| Figure 3.4: Architecture of GoogleNet..... | 36 |
| Figure 3.5: Transfer Learning..... | 38 |
| Figure 3.6: Modified Architecture of GoogleNet..... | 38 |
| Figure 3.7: Fine-tuned GoogleNet Architecture..... | 39 |
| Figure 3.8: 5-fold Cross-validation | 42 |
| Figure 3.9: Binary Confusion Matrix | 43 |
| Figure 3.10: One-vs-all Confusion Matrix (where G, M, P and N refer to Glioma, Meningioma, Pituitary Tumour and Normal, respectively) for the First Cross-validation Iteration)..... | 44 |
| Figure 4.1: First Fold as Testing Set | 51 |
| Figure 4.2: Training and Validation Accuracy for the First Cross-validation Iteration . | 52 |
| Figure 4.3: Training and Validation Loss for the First Cross-validation Iteration..... | 52 |
| Figure 4.4: Second Fold as Testing Set | 55 |
| Figure 4.5: Training and Validation Accuracy for the Second Cross-validation Iteration | 56 |
| Figure 4.6: Training and Validation Loss for the Second Cross-validation Iteration | 56 |
| Figure 4.7: Third Fold as Testing Set..... | 59 |

| | |
|--|----|
| Figure 4.8: Training and Validation Accuracy for the Third Cross-validation Iteration | 60 |
| Figure 4.9: Training and Validation Loss for the Third Cross-validation Iteration | 60 |
| Figure 4.10: Fourth Fold as Testing Set | 63 |
| Figure 4.11: Training and Validation Accuracy for the Fourth Cross-validation Iteration | 64 |
| Figure 4.12: Training and Validation Loss for the Fourth Cross-validation Iteration ... | 65 |
| Figure 4.13: Fifth Fold as Testing Set | 68 |
| Figure 4.14: Training and Validation Accuracy for the Fifth Cross-validation Iteration | 69 |
| Figure 4.15: Training and Validation Loss for the Fifth Cross-validation Iteration | 69 |
| Figure 4.16: The Folds Combination used for the Comparison | 78 |

LIST OF FORMULAS

| Equation | Formulas | |
|----------|---------------------------------------|---|
| 3.1 | Softmax Activation Function | $\frac{e^{z_i}}{\sum e^z}$ |
| 3.2 | Accuracy for Glioma (G) | $\frac{TP_G + TN_G}{TP_G + TN_G + FP_G + FN_G}$ |
| 3.3 | Accuracy for Meningioma (M) | $\frac{TP_M + TN_M}{TP_M + TN_M + FP_M + FN_M}$ |
| 3.4 | Accuracy for Pituitary Tumour (P) | $\frac{TP_P + TN_P}{TP_P + TN_P + FP_P + FN_P}$ |
| 3.5 | Accuracy for Normal (N) | $\frac{TP_N + TN_N}{TP_N + TN_N + FP_N + FN_N}$ |
| 3.6 | Precision for Glioma (G) | $\frac{TP_G}{TP_G + FP_G}$ |
| 3.7 | Precision for Meningioma (M) | $\frac{TP_M}{TP_M + FP_M}$ |
| 3.8 | Precision for Pituitary Tumour (P) | $\frac{TP_P}{TP_P + FP_P}$ |
| 3.9 | Precision for Normal (N) | $\frac{TP_N}{TP_N + FP_N}$ |
| 3.10 | Recall for Glioma (G) | $\frac{TP_G}{TP_G + FN_G}$ |
| 3.11 | Recall for Meningioma (M) | $\frac{TP_M}{TP_M + FN_M}$ |
| 3.12 | Recall for Pituitary Tumour (P) | $\frac{TP_P}{TP_P + FN_P}$ |
| 3.13 | Recall for Normal (N) | $\frac{TP_N}{TP_N + FN_N}$ |
| 3.14 | Specificity for Glioma (G) | $\frac{TN_G}{TN_G + FP_G}$ |

| | | |
|------|---|---|
| 3.15 | Specificity for Meningioma (M) | $\frac{TN_M}{TN_M + FP_M}$ |
| 3.16 | Specificity for Pituitary Tumour (P) | $\frac{TN_P}{TN_P + FP_P}$ |
| 3.17 | Specificity for Normal (N) | $\frac{TN_N}{TN_N + FP_N}$ |
| 3.18 | F1 Score for Glioma (G) | $\frac{2(Precision_G * Recall_G)}{Precision_G + Recall_G}$ |
| 3.19 | F1 Score for Meningioma (M) | $\frac{2(Precision_M * Recall_M)}{Precision_M + Recall_M}$ |
| 3.20 | F1 Score for Pituitary Tumour (P) | $\frac{2(Precision_P * Recall_P)}{Precision_P + Recall_P}$ |
| 3.21 | F1 Score for Normal (N) | $\frac{2(Precision_N * Recall_N)}{Precision_N + Recall_N}$ |
| 3.22 | Weighted Accuracy (Acc) | $\frac{Acc_G N_G + Acc_M N_M + Acc_P N_P + Acc_N N_N}{W_G + W_M + W_P + W_N}$ |
| 3.23 | Weighted Precision (Prc) | $\frac{Prc_G N_G + Prc_M N_M + Prc_P N_P + Prc_N N_N}{W_G + W_M + W_P + W_N}$ |
| 3.24 | Weighted Recall (Rcl) | $\frac{Rcl_G N_G + Rcl_M N_M + Rcl_P N_P + Rcl_N N_N}{W_G + W_M + W_P + W_N}$ |
| 3.25 | Weighted Specificity (Spc) | $\frac{Spc_G N_G + Spc_M N_M + Spc_P N_P + Spc_N N_N}{W_G + W_M + W_P + W_N}$ |
| 3.26 | Weighted F1 Score (F1 S) | $\frac{F1S_G N_G + F1S_M N_M + F1S_P N_P + F1S_N N_N}{W_G + W_M + W_P + W_N}$ |
| 4.1 | Recall | $\frac{TP}{TP + FN}$ |
| 4.2 | F1 Score | $\frac{2(Precision * Recall)}{Precision + Recall}$ |

LIST OF ABBREVIATIONS

| No. | Abbreviations | Full Sentence |
|-----|---------------|---|
| 1 | AI | Artificial Intelligence |
| 2 | ML | Machine Learning |
| 3 | NN | Neural Network |
| 4 | DL | Deep Learning |
| 5 | MRI | Magnetic Resonance Imaging |
| 6 | CT | Computed Tomography |
| 7 | CNN | Convolutional Neural Network |
| 8 | SVM | Support Vector Machine |
| 9 | 3D | Three Dimensional |
| 10 | FLAIR | Fluid Attenuated Inversion Recovery |
| 11 | CE | Contrast-enhanced |
| 12 | SNR | Signal-to-noise Ratio |
| 13 | SIST | Shift-invariant Shearlet Transform |
| 14 | RGB | Red-green-blue |
| 15 | FC | Fully Connected |
| 16 | RBF | Radial Basis Function |
| 17 | NB | Naïve Bayes |
| 18 | K-NN | K-nearest Neighbour |
| 19 | RF | Random Forest |
| 20 | DAE | Deep Autoencoder |
| 21 | BFC | Bayesian Fuzzy Clustering |
| 22 | JOA | Jaya Optimization Algorithm |
| 23 | DNN | Deep Neural Network |
| 24 | ANN | Artificial Neural Network |
| 25 | ReLU | Rectified Linear Unit |
| 26 | GLCM | Gray Level Co-occurrence Matrix |
| 27 | OFPA | Oppositional Flower Pollination Algorithm |
| 28 | PFCM | Possibilistic Fuzzy C-means |
| 29 | DWT | Discrete Wavelet Transformation |

| | | |
|----|---------|--|
| 30 | KMFCM | K-means with Fuzzy C-means |
| 31 | PCA | Principal Component Analysis |
| 32 | ACLS | Active Contour by Level Set |
| 33 | BWT | Berkeley Wavelet Transformation |
| 34 | ERT | Extremely Randomised Trees |
| 35 | EM | Expectation-maximisation |
| 36 | FFT | Fast Fourier Transform |
| 37 | MRMR | Minimal-redundancy-maximal-relevance |
| 38 | ELM | Extreme Learning Machine |
| 39 | LRF | Local Receptive Fields |
| 40 | ELM-LRF | Extreme Learning Machine with Local Receptive Fields |
| 41 | FFBPNN | Feed Forward Back Propagation Neural Network |
| 42 | WHO | World Health Organisation |
| 43 | RAM | Random-access Memory |
| 44 | Adam | Adaptive Moment Estimation |
| 45 | TP | True Positive |
| 46 | FP | False Positive |
| 47 | TN | True Negative |
| 48 | FN | False Negative |

CHAPTER 1

INTRODUCTION

1.1 Background

1.1.1 Brain Cancer

The brain, which is displayed in Figure 1.1, is the control centre of the human body and its most complex part. It is responsible for all the body's functions. This organ is the centre of intelligence and controls behaviour, movement, and the interpretation of the senses. It is made up of billions of cells called neurons and glial cells, which work together to transmit and process information.



Figure 1.1: Human Brain [1]

One of the most serious and deadly diseases that can affect the brain is brain cancer. Brain cancer arises when abnormal cells in the brain divide and grow out of control. These cells can combine to generate benign or malignant tumours. Brain tumours that are

malignant are more aggressive and can spread to other regions of the body, whereas benign tumours are often less dangerous and do not spread.

According to the Cancer Council of New South Wales [2], brain tumours can be classified into four grades based on their level of aggressiveness. These grades are used to help patients with brain tumours forecast their prognosis and guide treatment options. Brain cancer is classified into four categories based on their malignancy: Grade I, Grade II, Grade III, and Grade IV. Grade I tumours are the least aggressive and are characterised by slow growth and a low risk of recurrence. They can be treated surgically or with a mix of surgery, radiation treatment, and chemotherapy. Grade II tumours are more aggressive than grade I tumours but still have a relatively low risk of recurrence. They may be treated with surgery, radiation therapy, or chemotherapy, depending on the specific characteristics of the tumour. Grade III tumours are more aggressive than grade II tumours and have a higher risk of recurrence. They may be treated with surgery, radiation therapy, and chemotherapy, but the prognosis for these tumours is generally worse than for lower-grade tumours. Grade IV tumours are the most aggressive and have a high risk of recurrence. They may be treated with surgery, radiation therapy, and chemotherapy, but the prognosis for these tumours is generally poor.

Some common types of brain tumours include gliomas, meningiomas, pituitary tumours, craniopharyngiomas, primary brain lymphomas, and metastatic brain tumours. Gliomas are tumours that originate in the glial cells, which support and protect the nerve cells of the brain. Gliomas can be either benign or malignant and include astrocytomas, oligodendrogliomas, and ependymomas. Meningiomas are tumours that develop in the meninges. Meningiomas are usually benign but can sometimes be aggressive. Pituitary tumours are cancers that originate from the pituitary gland, a tiny endocrine gland situated at the base of the brain. Pituitary tumours can be either benign or malignant. Craniopharyngiomas are tumours that arise from cells in the pituitary gland and are usually benign. Primary brain lymphoma is a type of cancer that originates in the brain and is usually associated with non-Hodgkin's lymphoma. Tumours that have progressed to the brain from another area of the body, such as the lungs or breast, are known as metastatic brain tumours.

For this research and writing process, the brain cancer detection and classification will cover three types of brain cancer, which are gliomas, meningiomas, and pituitary tumours. Symptoms of brain cancer can differ based on the tumour's location and size.

Common symptoms include headaches, seizures, changes in mood or behaviour, weakness or numbness in the limbs, and difficulty with balance and coordination. A physical assessment, imaging tests including MRI or CT scans, and a biopsy to study a sample of tumour tissue are all used to detect brain cancer. The specific treatment and prognosis for a brain tumour will depend on the type and stage of the tumour, as well as the overall health of the patient.

1.1.2 Deep Learning

DL is a sort of ML that analyses and recognises patterns in data using ANNs with several layers. The difference between AI, ML, NN, and DL is shown in Euler's diagram in Figure 1.2.

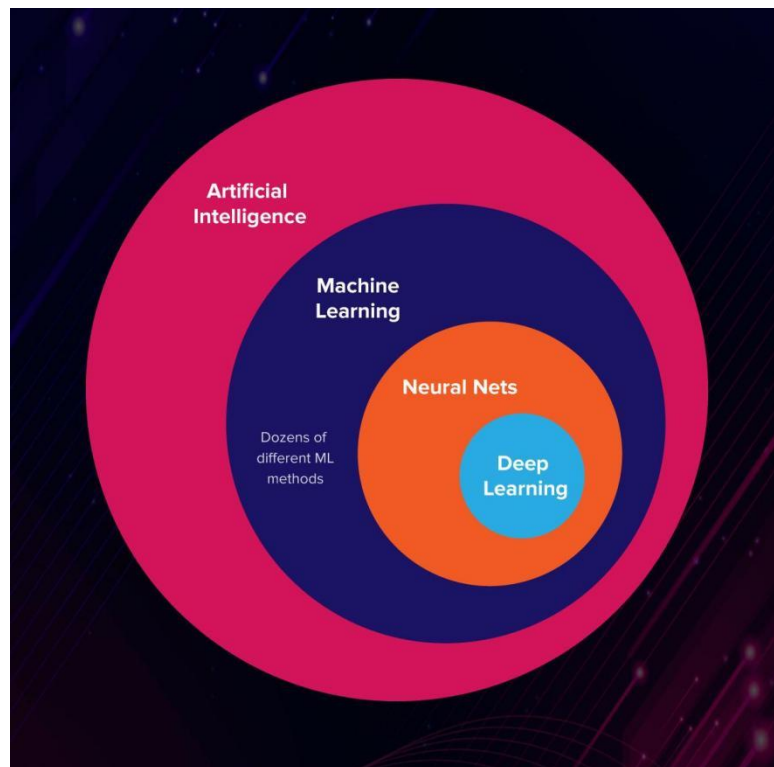


Figure 1.2: Euler's Diagram on AI, ML, NN and DL [3]

The key characteristic of DL is the utilisation of DNNs, which are composed of multiple layers of interconnected neurons. These networks are designed to mimic the structure and functioning of the human brain, allowing them to learn and model highly complex relationships within the data. Each layer of neurons in a DNN performs specific computations and progressively learns more abstract and high-level representations as information flows through the network.

One of the major breakthroughs in deep learning has been in the field of computer vision, where CNNs have demonstrated exceptional performance in tasks such as image classification, object detection, and segmentation. CNNs can automatically learn spatial hierarchies of features from images, enabling them to identify and distinguish objects with remarkable accuracy. This has opened new possibilities in areas such as autonomous driving, facial recognition, and medical imaging.

In recent years, DL techniques have emerged as powerful tools for medical image analysis, offering the potential to revolutionise brain cancer detection and classification. By leveraging large amounts of data, as can be seen in Figure 1.3, and complex NN architectures, DL models have demonstrated remarkable capabilities in extracting meaningful features from medical images and making accurate predictions.

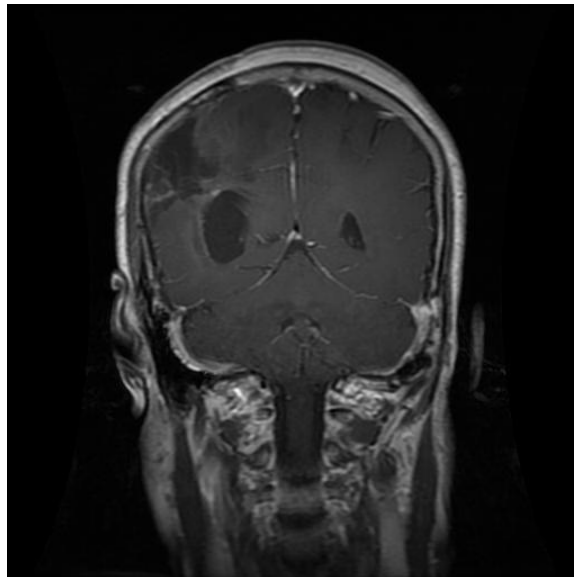


Figure 1.3: DL can be used to detect the presence of brain cancer in this MRI [4]

The implementation of DL in the identification and classification of brain cancer holds great promise for improving the accuracy and efficiency of these processes and could greatly enhance patient outcomes. However, further study is required to properly comprehend the capabilities and limitations of this approach and to develop robust and reliable methods for implementing it in clinical practise.

1.2 Problem Statement

Detecting and classifying brain cancer manually is a time-consuming process that requires extensive analysis and interpretation of medical imaging data. Radiologists and

medical professionals typically examine brain MRI's to identify the presence of brain cancer in them and classify them into specific types. However, this manual approach can be slow and labour-intensive, leading to delays in diagnosis and treatment planning. The interpretation of brain MRI images involves visually analysing intricate details, patterns, and abnormalities in the scans. Radiologists need to meticulously examine multiple slices of the brain to accurately identify and characterise tumors. This process is not only time-consuming but also susceptible to human error and subjectivity, which can impact the accuracy and consistency of diagnoses. Addressing the time-consuming nature of manual brain cancer detection and classification is crucial to improve patient outcomes and optimize healthcare resources. By leveraging automated techniques, such as DL algorithms, we can develop a system that assists radiologists in the efficient and accurate detection and classification of brain cancer, decreasing the time required for diagnosis and enabling timely treatment interventions.

Other than that, detecting and classifying brain cancer often requires the processing of vast amounts of medical imaging data. This task can present challenges due to the substantial computational resources required. The size and complexity of the datasets, along with the computational demands of sophisticated analysis algorithms, can strain the capabilities of available resources. Limited computational power, memory, or storage capacity may hinder the timely and efficient analysis of brain cancer imaging data. These constraints can lead to longer processing times, delays in diagnosis, and potential limitations in the accuracy and effectiveness of the detection and classification processes. Overcoming these computational resource constraints is crucial for developing efficient and scalable brain cancer detection and classification systems that can handle the increasing volume and complexity of medical imaging data.

1.3 Objectives

There is a need for further research to address these challenges and develop reliable and effective DL approaches for the detection and classification of brain cancer. This will require a multidisciplinary approach that brings together experts in DL, medical imaging, and brain cancer research, as well as clinicians who can provide valuable insights into the needs and constraints of real-world clinical practise. The following are the primary goals of a study on the detection and classification of brain tumours with DL:

1. To develop a reliable and robust deep learning-based approach for detecting and classifying brain cancers.
2. To investigate the robustness and generalisability of the proposed model.
3. To evaluate and compare the performance of the proposed model to state-of-the-art methods.

1.4 Project Contributions

In this project, a comprehensive review of existing research and literature related to brain cancer detection and classification using deep learning methods was conducted. Other than that, relevant techniques, algorithms, and datasets for the project were identified. Thirdly, the model's performance was analysed, limitations were identified, and strategies to improve the model's accuracy and generalisation capabilities were also proposed. Next, the project's methodology, including data pre-processing steps, model architectures, training configurations, and evaluation results was documented. Finally, a clear and concise report summarising the project's findings, insights, and conclusions was written.

1.5 Project Scope

The primary focus of the project was to address the challenge of brain cancer detection and classification using deep learning techniques. The project aimed to investigate the feasibility and performance of a proposed approach involving the fine-tuning of a pre-trained CNN model. The fine-tuning process involved optimising the model's parameters and architecture to enhance its performance specifically for brain cancer detection and classification tasks.

The research centred specifically on the analysis and evaluation of this modified model using a dataset comprising brain cancer and normal brain MRI scans. It is important to note that the scope of this project was limited to the utilisation of a publicly available online dataset at Kaggle, which was created by Masoud Nickparvar [4], rather than collecting data firsthand from hospitals or clinics.

To assess the effectiveness of the modified CNN model, various metrics for evaluation were employed, including accuracy, precision, recall, specificity, and F1-

score. The objective was to examine the model's performance and determine its capability for accurately detecting and classifying brain cancer cases.

The project's scope was limited to the analysis, fine-tuning, and evaluation stages of the modified pre-trained model using transfer learning techniques. Consequently, the deployment and implementation of the selected model in a real-world setting were not within the project's scope. Moreover, due to computational constraints, the project focused on utilising a single pre-trained model, namely GoogleNet. The implementation of the proposed DL model for brain cancer detection and classification had been carried out using the MATLAB R2022a software.

1.6 Chapter Outline

The organisation of this Final Year Project report is divided into several chapters. In the second chapter, related works from the literature were reviewed. The third chapter describes the workflow and steps of the proposed method. The fourth chapter presents the results that were obtained. Finally, the fifth chapter includes the conclusion and discusses limitations and future directions for this work.

Chapter 2

LITERATURE REVIEW

2.1 Overview

This chapter covers relevant studies on the use of DL for brain cancer detection and classification for the past six years (2016–2022). It begins with a brief analysis of past research papers on the roles of DL in cancer detection and classification. It then explores the imaging modality that may be put to use for detecting and classifying a vast range of medical conditions, including brain cancer. The chapter also covers the publicly available datasets, including their problems. The chapter concludes with a discussion on the research gap and future measures that may be taken to improve the quality of the existing brain cancer detection and classification DL models.

The results of this review provide strong evidence for the potential of DL in improving brain cancer detection and classification. However, the quality of the studies and the small sample sizes of some studies suggest that further research is needed to confirm these findings.

2.2 Related Studies

2.2.1 Brain Cancer

According to Rebecca et al. [5], cancer deaths caused by tumours in the brain and other parts of the nervous system are the highest among men under 40 and women under 20. Brain cancer may be incurable and/or result in death. This means that the prognosis for a person with brain cancer can greatly differ based on the specific type of cancer and how advanced it is at the time of detection. Some types of brain tumours, such as benign tumours, may be completely curable with surgery alone. Other types, such as malignant tumours, may be more difficult to treat and may not respond to standard treatments like surgery, radiation, and chemotherapy. In these cases, the cancer may be considered incurable, and the focus of treatment may shift to managing symptoms and providing

palliative care. Additionally, brain cancer can be fatal, depending on the location, size, and progression of the tumour.

In the paper by Kalan Farmanfarma et al. [6], the key factors that increase the risk of developing brain cancer are a family history of the disease, hormonal imbalances, a weakened immune system, a poor diet, alcohol consumption, smoking, exposure to aspartame, radiation, and prolonged use of cell phones.

Brain cancer can have a significant impact on a person's emotional and psychological well-being, and there is evidence to suggest that brain cancer patients may be at an increased risk of suicide. Saad et al. [7] investigated the association between brain cancer and suicide risk. They analysed data from 87,785 brain cancer patients between the years 2000 and 2016. Out of these patients, 29 (0.03%) committed suicide during the first year of their diagnosis. The study found that the possibility of suicide was significantly higher in the first year after a brain cancer diagnosis, with an ratio of the observed value to the expected value of 3.05 (95% CI, 2.04–4.37). The study investigated the traits and typical mortality rates of brain cancer patients who committed suicide, as can be seen in Table 2.1. Factors that may contribute to this increased risk include depression and hopelessness. It is important for brain cancer patients to be screened for depression and other emotional and psychological issues and to have access to supportive care and counselling.

Table 2.1: Baseline Characteristics and Standard Mortality Rates for Brain Cancer Patients Who Died of Suicide [7]

| Characteristic | Within the first year | | | | After more than a year | | | |
|----------------------------------|-------------------------|-------------------|---------------|-------------------------------|-------------------------|-------------------|---------------|--------------------|
| | Total patients, No. (%) | Observed, No. (%) | Expected, No. | O/E (95% CI) | Total patients, No. (%) | Observed, No. (%) | Expected, No. | O/E (95% CI) |
| Total | 87 785 (100) | 29 (100) | 9.52 | 3.05 (2.04-4.37) ^a | 48 185 (100) | 35 (100) | 29.79 | 1.17 (0.82-1.63) |
| Sex | | | | | | | | |
| Men | 48 908 (55.7) | 27 (93.1) | 7.98 | 3.38 (2.23-4.92) ^a | 27 035 (56.1) | 29 (82.9) | 24.08 | 1.20 (0.81-1.73) |
| Women | 38 877 (44.3) | 2 (6.9) | 1.54 | 1.30 (0.16-4.68) | 21 150 (43.9) | 6 (14.1) | 5.71 | 1.05 (0.39-2.29) |
| Age at diagnosis, y | | | | | | | | |
| <18 | 11 657 (13.3) | 0 | 0.18 | 0 (0-19.98) | 9420 (19.5) | 3 (8.6) | 3.7 | 0.81 (0.17-2.37) |
| 18-44 | 17 921 (20.4) | 5 (17.2) | 2.57 | 1.94 (0.63-4.53) | 14 402 (29.9) | 17 (48.6) | 13.72 | 1.24 (0.72-1.98) |
| 45-64 | 30 059 (34.2) | 12 (41.4) | 4.38 | 2.74 (1.42-4.79) ^a | 17 140 (35.6) | 10 (28.6) | 9.87 | 1.01 (0.49-1.86) |
| >64 | 28 148 (32.1) | 12 (41.4) | 2.38 | 5.04 (2.60-8.80) ^a | 7223 (15.0) | 5 (14.2) | 2.49 | 2.01 (0.65-4.69) |
| Race | | | | | | | | |
| White | 75 616 (86.1) | 26 (89.8) | 8.94 | 2.91 (1.90-4.26) ^a | 40 967 (85.0) | 30 (85.5) | 27.72 | 1.08 (0.73-1.55) |
| Black | 6511 (7.4) | 1 (3.4) | 0.26 | 3.89 (0.10-21.66) | 3801 (7.9) | 1 (2.9) | 0.97 | 1.03 (0.03-5.71) |
| Asian or Pacific Islander | 5258 (6.0) | 2 (6.8) | 0.3 | 6.71 (0.81-24.25) | 3175 (6.6) | 3 (8.7) | 1.01 | 2.97 (0.61-8.68) |
| American Indian or Alaska Native | 400 (0.5) | 0 | 0.02 | 0 (0-161.85) | 242 (0.5) | 1 (2.9) | 0.09 | 11.32 (0.29-63.06) |
| Marital status | | | | | | | | |
| Married | 44 040 (50.2) | 17 (58.8) | 6 | 2.83 (1.65-4.54) ^a | 23 283 (48.3) | 12 (34.3) | 15.54 | 0.77 (0.40-1.35) |
| Single | 26 487 (30.2) | 4 (13.7) | 2.06 | 1.94 (0.53-4.96) | 18 493 (38.4) | 20 (57.2) | 10.76 | 1.86 (1.14-2.87) |
| Separated | 710 (0.8) | 0 | 0.08 | 0 (0-45.89) | 368 (0.8) | 1 (2.9) | 0.25 | 4.08 (0.10-22.73) |
| Divorced | 5966 (6.8) | 2 (6.8) | 0.64 | 3.13 (0.38-11.31) | 2783 (5.8) | 1 (2.9) | 1.67 | 0.60 (0.02-3.34) |
| Widowed | 7278 (8.3) | 1 (3.4) | 0.34 | 2.91 (0.07-16.24) | 1583 (3.3) | 1 (2.9) | 0.42 | 2.37 (0.06-13.20) |
| Unknown | 3185 (3.6) | 5 (17.3) | NA | NA | 1609 (3.3) | 0 | NA | NA |
| Brain tumor histology | | | | | | | | |
| Glioblastoma | 39 248 (44.7) | 18 (62.1) | 4.5 | 4 (2.37-6.32) ^a | 14 609 (30.3) | 7 (20) | 3.8 | 1.84 (0.74-3.79) |
| Other brain histologies | 48 537 (55.3) | 11 (37.9) | 5.02 | 2.19 (1.09-3.92) ^a | 33 576 (69.9) | 28 (80) | 25.99 | 1.08 (0.72-1.56) |

Abbreviation: NA, not applicable; O/E, observed to expected ratio.

^a $P < .05$.

2.2.2 DL for Cancer Detection and Classification

DL techniques have been applied to the area of cancer detection and classification with positive outcomes. One common approach is to use deep CNNs to analyse medical images such as CT scans or MRIs and predict the likelihood that a tumour is present. These models can be trained on huge datasets of labelled medical images to learn features that are indicative of cancer.

Makaju et al [8] developed a system for the detection of cancerous nodules in lung CT scans by applying a watershed segmentation for detection and an SVM for classification as benign or malignant. The proposed model had a 92% accuracy rate in detecting cancer, which is very competitive. In the study by Tasnim et al [9], they employed CNN to analyse colon cell imaging data in their investigation. Specifically, they used two CNN models that employed max-pooling and average pooling layers and a MobileNetV2 model to classify colon cell images. To determine the best model, the authors trained and examined the models at various epochs (iterations over the training dataset) and varied the learning rate (a hyperparameter that controls the size of the model

weights that update after each epoch). They found that the CNN coupled with max-pooling layers had an accuracy of 97.49%, the CNN with average pooling layers had an accuracy of 95.48%, and the MobileNetV2 model had the highest accuracy of 99.67% and a data loss rate of 1.24. This indicates that the MobileNetV2 model had better accuracy compared to the other two models.

2.2.3 Imaging Modality

2.2.3.1 Computed Tomography Imaging

CT scans are a type of medical imaging test that uses X-rays to generate detailed photos of the body's interior. CT scans are often used to diagnose diseases and injuries and can be used to visualise a wide range of tissues in the body, including bones, organs, and blood vessels. During a CT scan, the patient is placed on a table, which glides into a big, donut-shaped piece of equipment known as a CT scanner. The scanner contains an X-ray source and a series of detectors that measure the X-rays as they pass through the body. The CT scanner takes a series of detailed pictures of the body from different angles, which are then combined by a computer to generate a detailed three-dimensional internal body image. CT scans can be performed quickly and are widely available, providing detailed images that can be used to diagnose a wide range of medical conditions. However, they do expose the patient to a small amount of ionising radiation, so they should be used judiciously. In a study by Smith-Bindman et al [10], the chance of acquiring cancer as a result of a CT scan varies based on various factors, including the type of CT examination, the patient's age, and gender. For example, it is predicted that one in every 270 women who have had a CT coronary angiography (a type of CT examination that visualises the coronary arteries) at age 40 will be diagnosed with cancer from that CT scan, while the risk is lower for men (1 in 600).

2.2.3.2 Magnetic Resonance Imaging

MRI is a type of medical imaging technique that utilises a strong magnetic field and radio waves to generate detailed images of the internal parts of the body. MRI is often used to visualise the body's soft tissues, such as the brain, muscles, and organs, and it is particularly useful for imaging the brain and spine. The three MRI sequences most commonly used for analysing the brain are T1-weighted, T2-weighted, and FLAIR [11]. During an MRI scan, the patient lies down on a table that is inserted into a large, cylinder-shaped machine called an MRI scanner. The scanner has a powerful magnetic field that

aligns protons inside the body's cells. Then, radio waves are used to temporarily disturb the protons' alignment, and when they return to their previous alignment, they release a weak signal that the MRI scanner detects. This signal is used to generate detailed photographs of the body's inside. MRI is a non-invasive procedure that does not expose the patient to ionising radiation. However, it can be time-consuming and may not be suitable for patients who are claustrophobic or who have certain types of metal in their bodies. Despite these limitations, MRI is an important tool for diagnosing and monitoring a wide range of medical conditions.

2.2.3.3 X-ray Imaging

An X-ray is a medical imaging method that creates images of the inside of the body by utilising a small amount of ionising radiation. X-rays are commonly used to visualise bones, but they can also be used to visualise other tissues in the body, such as the lungs and organs. During an X-ray exam, the patient is positioned on a table, and an X-ray machine is used to produce a beam of ionising radiation that passes through the body. Radiation is taken in at varying rates by different tissues, with denser tissues like bones receiving more radiation than softer tissues such as muscles and organs. This results in an image that shows the relative densities of different tissues within the body. X-rays are non-invasive and relatively quick, and they are widely available. However, they do expose the patient to a small amount of ionising radiation, so they should be used judiciously. According to Rajan and Sundar [12], X-ray is effective at visualising dense objects such as bones, but is not very effective at diagnosing bone degeneration, bone fractures, or the specific location of a tumour.

2.2.4 Dataset

Dataset is of vital importance in the creation of a DL model for brain cancer detection and classification. A dataset is a set of data utilized to train and evaluate the efficacy of a DL model. In the context of brain cancer detection and classification, the dataset will typically consist of images of brain tissue that have been annotated with information about the presence and type of cancer. The success of a DL model for brain cancer detection and classification heavily depends on the quality and diversity of the dataset used. A high-quality dataset will include a diverse range of images that accurately represent the range of brain tumours that are likely to be encountered in practice. It will also include accurate annotations of the presence and type of cancer in each image. The size of the dataset is also important, as a larger dataset will typically lead to a more robust and accurate model.

However, it is also important to ensure that the dataset is balanced, meaning that it includes a similar number of examples of each class of cancer. This is particularly important in the context of brain cancer detection and classification, as some types of brain tumours may be rare and therefore underrepresented in the dataset. Overall, the dataset is a critical component of a DL model for brain cancer detection and classification, and the quality and diversity of the dataset will significantly impact the model's performance.

All of the datasets found utilised MRIs rather than CT and X-rays. CT scans expose patients to higher radiation levels than MRIs or X-rays. MRI is widely regarded as the safest choice since it creates images using a strong magnetic field and radio waves rather than ionising radiation. The publicly available brain tumour dataset on Figshare by Cheng [13] includes 3064 T1-weighted CE images from 233 patients, with 708 Meningioma MRI, 1426 Glioma MRI, and 930 Pituitary Tumour MRI. These images are presented in axial, coronal, and sagittal views. The images are available in MATLAB data format (.mat file). Meanwhile, the dataset by Sartaj et al [14] contains brain MRIs that were divided into Training and Testing folders, each containing four subfolders. These subfolders contain brain MRIs of three tumour classes: Glioma, Meningioma, and Pituitary Tumour, as well as non-tumour samples. Their dataset has 3264 images stored in .jpg format. In Ahmed's dataset [15], titled Br35H :: Brain Tumour Detection 2020, he divided the MRIs into three folders: "yes" contains tumorous brain MRIs, "no" contains non-tumorous brain MRIs, and "pred" is reserved for testing purposes. This dataset includes 3060 images stored as .jpg files. Masoud [4] took the initiative to combine these brain tumour datasets but noticed that the Glioma class images in Sartaj's dataset had been misclassified and removed them. He then used the Glioma class images only from Cheng's dataset [13]. The No Tumour class images were taken from Ahmed's dataset [15]. The resulting combined dataset includes 7022 human brain MRIs in .jpg format, which can be divided into four classes: Glioma, Meningioma, Pituitary Tumour, and Normal.

A summary of the brain tumour dataset can be found in the form of a matrix in Table 2.2.

Table 2.2: Brain Tumour Dataset

| Titles of Dataset | Brain Tumour Dataset | Brain Tumour Classification (MRI) | Br35H :: Brain Tumour Detection 2020 | Brain Tumour MRI Dataset |
|---------------------------------|--|---|--|---|
| Authors | Jun Cheng [13] | Sartaj Bhuvaji [14] | Ahmed Hamada [15] | Masoud Nickparvar [4] |
| Source | Nanfang Hospital, China General Hospital and Tianjing Medical University, China | Not stated | Not stated | This dataset is a combination of Cheng [13], Sartaj [14] and Ahmed's [15] datasets. |
| No. of Patients Involved | 233 | Not stated | Not stated | Not stated |
| Imaging Modality | MRI | MRI | MRI | MRI |
| No. of Images | 3064 | | | 7022 |
| Format of Images | .mat | .jpg | .jpg | .jpg |
| Classes | <ol style="list-style-type: none"> 1. Glioma 2. Meningioma 3. Pituitary | <ol style="list-style-type: none"> 1. Glioma 2. Meningioma 3. Pituitary 4. Normal | <ol style="list-style-type: none"> 1. Tumorous 2. Normal | <ol style="list-style-type: none"> 1. Glioma 2. Meningioma 3. Pituitary 4. Normal |

2.2.5 Pre-processing

Pre-processing is essential in ML because raw data is often incomplete, irregular, noisy, or not in a format that can be easily used for analysis. Pre-processing is a step that is applied to prepare the data for analysis and modelling.

Some of the authors [16] and [17] did not apply any pre-processing techniques. Although both papers reported high accuracies, pre-processing is still required in DL. DL models are often able to handle raw and unstructured data better than traditional ML models; however, image pre-processing can lead to improvements in the model's performance and robustness.

In the papers [18] [19] [20] [21] [22] [23], the authors used several techniques for noise removal from the brain MRI. In image processing, noise can appear as random speckles or grains that can degrade the quality of the image. The choice of technique will depend on the type of noise and the desired level of noise reduction. Alfonse and Salem [18] applied enhancement and cropping for pre-processing. They used a median filter and high-pass filter enhancement, which removed noise and cleaned up the background of the image. Reema et al [19] explained in their paper that the idea behind using the median filter is to preserve all the edge information after noise removal. Ali and Davut [20] carried out several pre-processing steps such as noise filtering, cropping, scaling, and histogram equalisation. They used specific techniques for noise filtering, such as nonlocal means and local smoothing methods. Preethi and Aishwarya [21] employed the Gaussian filter to eliminate Gaussian noise from brain MRIs and obtained a noise-free image. Bahadure et al [22] applied adaptive contrast enhancement based on a modified sigmoid function to enhance the SNR and raw MRI clarity in their work. Arunachalam and Royappan [23] used the SIST to fuse two brain images before applying the inverse SIST transform to the merged image to obtain an improved quality of MRI. The flowchart of their method can be seen in Figure 2.1.

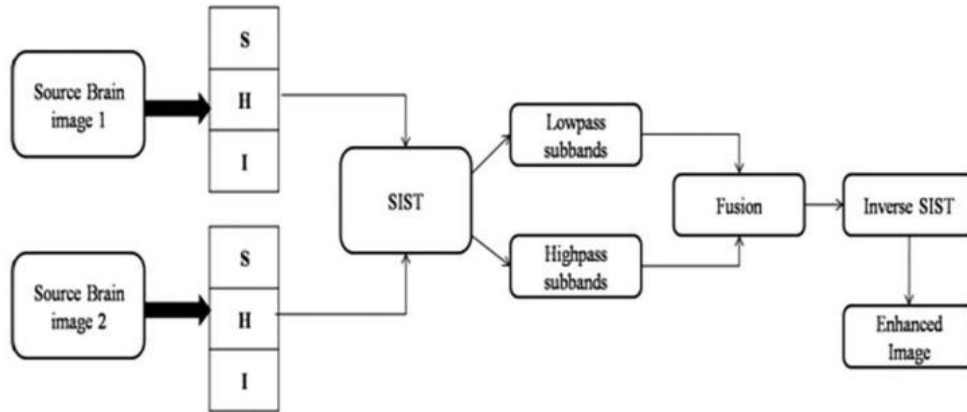


Figure 2.1: The Steps Taken by Arunachalam and Royappan [23] for Image Enhancement

Deepak and Ameer [24] and Zar et al [25] used the same pre-processing techniques. They first utilised the min-max normalisation technique to scale the intensity of the images between 0 and 1. They also resized the brain MRIs to $224 \times 224 \times 3$ to make them compatible with the pre-trained models used, which were originally designed for RGB colour images. Since the brain MRIs were in grayscale, they had to replicate the grayscale values three times to create three channels. Francisco et al [26] performed data augmentation by applying an elastic transformation to double the size of their dataset. In this case, data augmentation is a method used to artificially enlarge the brain MRI dataset size by applying multiple types of random transformations to the existing images, which include adding noise or stretching the image. An example of this transformation can be seen in Figures 2.2. After data augmentation, they extracted 65×65 -pixel training examples from each picture in the training set. Then, they scaled the pixels using pixel standardisation (zero mean and unit variance) throughout their training set.

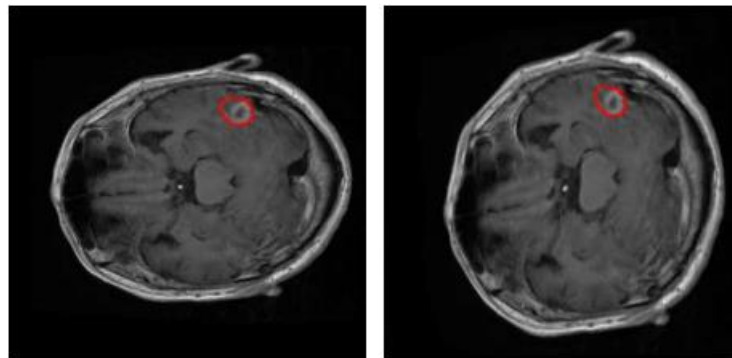


Figure 2.2: Data Augmentation that was Performed by Francisco et al [26]

Kang et al [27] first used extreme point calculations to crop their images. The methods are unique and can be seen in Figures 2.3. They also utilised image augmentation (rotation and horizontal flipping) to create an artificial dataset. In their paper, the input was randomly rotated by 90 degrees zero or more times before being horizontally flipped. Finally, the MRIs were resized using bicubic interpolation, to either 224x224 or 299x299 pixels depending on the pre-trained CNN models being used.

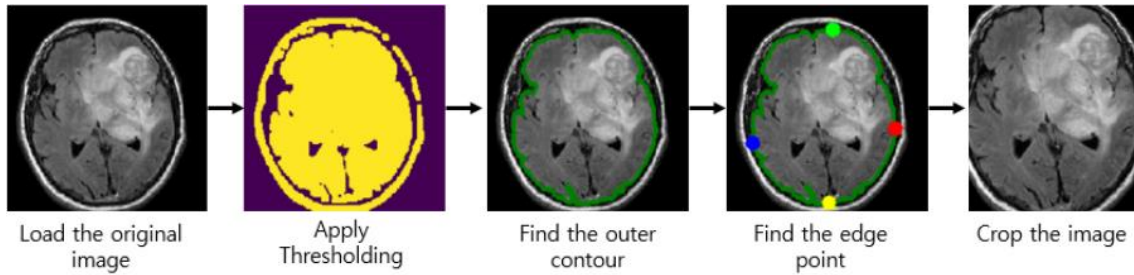


Figure 2.3: The Methods Used by Kang et al [27] to Crop the MRIs

Khawaldeh et al [28] pre-processed the MRIs by normalising them using the mean and standard deviation and then resizing them to 160 x 160 pixels. Sarah et al [X] converted the images from .mat into .jpg, then resized and cropped them. They also used various data augmentation methods, for example, horizontal flipping, vertical flipping, rotation, and zooming. They attempted to use random cropping as a pre-processing technique, but it did not significantly improve the accuracy of their model. Saxena et al [29] employed a technique called crop normalisation, which involves identifying the furthest points of a contour in the north, south, east, and west directions to crop the brain MRIs before resizing them and performing data augmentation. The crop normalisation method used is shown in Figure 2.4.

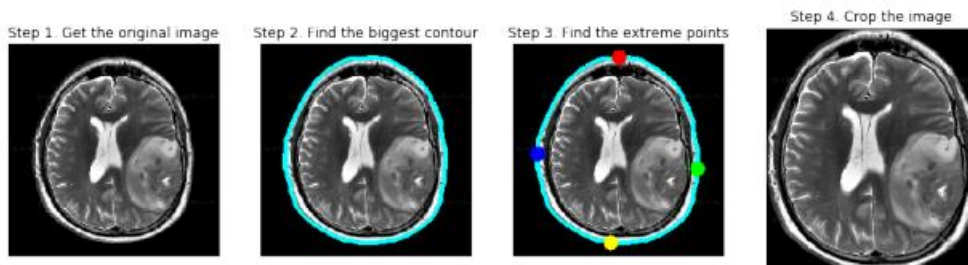


Figure 2.4: The Procedure Used by Saxena et al [29] to Crop the MRIs

Various methods of pre-processing were used in these studies. It is understood that having a small dataset can be a challenge when training an ML model, as there may not be enough data to accurately capture the underlying patterns in them. A small dataset can lead to a model with high variance, meaning that it has a high level of accuracy on the training data but performs poorly on new data. This is a phenomenon known as overfitting. To prevent overfitting, techniques such as data augmentation, which artificially increases the number of images in an image dataset by applying various types of random transformations to the existing data, can be employed. In summary, pre-processing is crucial for ML as it improves the accuracy and performance of the model. Similarly, it is also important for DL for data preparation and model robustness. A matrix representation of the pre-processing methods used in these papers can be found in Table 2.3 as a summary.

Table 2.3: Pre-processing Methods

| No | Paper | Dataset | Pre-processing |
|----|--|-----------|--|
| 1 | Abiwinanda et al, 2019 [16] | 3064 MRIs | None |
| 2 | Ahmet and Muhammed, 2020 [17] | 253 MRIs | None |
| 3 | Alfonse and Salem, 2016 [18] | 100 MRIs | Enhancement and cropping |
| 4 | Reema et al, 2017 [19] | 60 MRIs | Noise removal |
| 5 | Ali and Davut, 2018 [20] | 16 MRIs | Noise filtering, cropping, scaling and histogram equalization |
| 6 | Preethi and Aishwarya, 2017 [21] | 20 MRIs | Noise removal |
| 7 | Bahadure et al, 2017 [22] | 135 MRIs | Adaptive contrast enhancement based on modified sigmoid function |
| 8 | Arunachalam and Royappan, 2017 [23] | 100 MRIs | Enhancement |

| | | | |
|----|------------------------------------|--|-------------------------------------|
| 9 | Deepak and Ameer, 2019 [24] | 3064 MRIs | Normalising and resizing |
| 10 | Zar et al, 2019 [25] | 3064 MRIs | Normalising and resizing |
| 11 | Francisco et al, 2021 [26] | 3064 MRIs | Augmentation and scaling |
| 12 | Kang et al, 2021 [27] | 6317 MRIs (three different datasets were used in this study) | Cropping, augmentation and resizing |
| 13 | Khawaldeh et al, 2018 [28] | 4069 MRIs | Resizing and normalising |
| 14 | Sarah et al, 2019 [30] | 3064 MRIs | Cropping |
| 15 | Saxena et al, 2019 [29] | 253 MRIs | Cropping and resizing |

2.2.6 Modelling Approach

2.2.6.1 Convolutional Neural Network

Several researchers have employed CNNs for classifying brain tumours from MRIs and have tested their methods on brain tumour classification datasets. CNNs are ANNs that are extensively used for image and video recognition. They are able to analyse and interpret visual data by extracting features and patterns from the input images or videos and using them to classify or predict outcomes.

Saxena et al [29] classified brain MRIs into normal and abnormal. They used Inception V3, ResNet-50, and VGG-16 pre-trained CNN models in conjunction with transfer learning methods for brain tumour data classification. The ResNet-50 model, with an accuracy rate of 95%, came out on top of the other pre-trained models. In their paper, Kang et al [27] conducted experiments using three public brain MRI datasets for the task of brain tumour classification. They employed 13 pre-trained deep CNNs (ResNet-50, ResNet-101, DenseNet-121, DenseNet-169, VGG-16, VGG-19, AlexNet, Inception V3, ResNet-50, ResNet-101, ShuffleNet, MobileNet, and MnasNet) as feature extractors. They used nine different ML classifiers (FC layer, Gaussian NB, AdaBoost, k-NN, RF, SVM with three different kernels (linear, sigmoid, and, RBF), and ELM) and the outcomes showed that SVM with the RBF kernel was superior to the other ML classifiers.

Ahmet and Muhammad [17] used a range of CNN models, including GoogleNet, Inception V3, DenseNet-201, AlexNet, and ResNet-50, for brain MRI classification and achieved impressive accuracy results. They proposed a modified pre-trained ResNet-50 CNN by eliminating its final five layers and adding eight new layers, resulting in an accuracy of 97.2%, the highest of all the pre-trained models they evaluated. In studies by Khawaldeh et al [28], the authors suggested using a CNN model, specifically a modified version of the AlexNet model, to identify if brain MRIs were normal or abnormal and if Glioma tumours were high or low grade. They reported that this approach had an accuracy rate of 91%.

Raja et al [31] developed a hybrid DAE approach for classifying brain MRIs using BFC segmentation. BFC was used to identify positive regions in the MRIs, which were then analysed using information-theoretic measures, wavelet packet Tsallis entropy, and scattering transform to extract features related to the existence of oedema and the absence of tumours. These features were used to classify the images using a hybrid DAE with JOA and softmax regression. The proposed method achieved the highest accuracy of 98.5% compared to DNN, KNN, ANN, and Multi-SVM classifiers. Deepak and Ameer [24] focused on a three-class classification problem to distinguish between three of the most prevalent kinds of brain tumours: Glioma, Meningioma, and Pituitary Tumour. Deepak and Ameer's algorithm uses deep transfer learning and employs a modified and fine-tuned version of GoogleNet to learn the features of brain MRIs with tumours and has a mean classification accuracy of 98%.

In their paper, Francisco et al [26] introduced a multi-pathway CNN architecture for tumour segmentation. In their method, the classification of tumours was directly determined by the resulting segmentation. They employed an MRI dataset comprising 3064 images sourced from 233 patients. 80% of the photos are used for training, while the remaining 20% are used to evaluate performance. The NN was tested using a 5-fold cross-validation technique, which resulted in an excellent tumour classification accuracy of 97.3%. Abiwinanda et al [16] presented an automatic brain tumour segmentation and classification method using CNN that can identify Glioma, Meningioma, and Pituitary Tumour, without the need for pre-processing steps. They discovered the most optimal CNN architecture, which comprises of two convolutional layers, activation (ReLU), and max-pooling, after which comes a hidden layer of 64 neurons, which achieved a training accuracy and validation accuracy of 98.51% and 84.19% at best, respectively.

In the paper by Hemanth et al [32], they examined the use of a modified deep CNN for classifying brain tumour images in MRIs, using abnormal brain tumour images from four different classes, and obtained promising results with the proposed approach. This approach, in which the weights are not adjusted in the FC layer, reduced the computational complexity, thus making it more suitable for practical applications. Paul et al [33] examined the use of two types of NN to classify brain images with three types of tumours: Glioma, Meningioma, and Pituitary Tumour. They used both FC and CNN in their approach.

While CNNs have been successful in many image and video recognition tasks, the semantic gap between human perception and computer understanding of visual data is still a challenge. CNNs are limited in their ability to understand the context and meaning of visual content in the same way that humans do, and they may struggle to recognise or classify images or videos that contain unusual or unexpected elements. However, CNNs can still be an effective tool for overcoming the semantic gap in many applications, particularly when they are trained on large, diverse datasets and when they are used in conjunction with other techniques, such as transfer learning or data augmentation. By using these techniques, it is possible to improve the ability of CNNs to recognise and classify a wide range of images and videos, even those that contain unusual or unexpected elements. Table 2.4 presents a summary of the literature review in the form of a matrix.

Table 2.4: Papers which used CNN

| No. | Paper | Modelling Approach | Objective | Feature Extraction Method | Accuracy |
|------------|-------------------------------|---------------------------------|---|---|-----------------|
| 1 | Saxena et al, 2019 [29] | CNN with Transfer Learning | Detection of brain tumours | CNN | 95% |
| 2 | Kang et al, 2021 [27] | Deep CNN with Transfer Learning | Classification of brain tumours | CNN | 97.85% |
| 3 | Ahmet and Muhammed, 2020 [17] | CNN | Detection and classification of brain tumours | CNN | 97.01% |
| 4 | Khawaldeh et al, 2017 [28] | CNN | Detection of brain tumours | Modified AlexNet | 91.16% |
| 5 | Raja et al, 2020 [31] | DAE, JOA and Softmax Regression | Classification of brain tumours | Information-theoretic Measures, Scattering Transform and Wavelet Packet Tsallis Entropy | 98.5% |
| 6 | Deepak and Ameer, 2019 [24] | Deep Transfer Learning | Classification of brain tumours into Glioma, Meningioma, and Pituitary Tumour | Modified GoogleNet | 98% |
| 7 | Francisco et al, 2021 [26] | Multi-pathway CNN | Classification of brain tumours into Glioma, Meningioma, and Pituitary Tumour | CNN | 97.3% |

| | | | | | |
|-----------|-----------------------------|-------------|---|-----|--------|
| 8 | Abiwinanda et al, 2019 [16] | CNN | Classification of brain tumours into Glioma, Meningioma, and Pituitary Tumour | CNN | 98.51% |
| 9 | Hemanth et al, 2019 [32] | CNN | Classification of brain tumours into Normal and Abnormal | CNN | 94.5% |
| 10 | Paul et al, 2017 [33] | CNN with FC | Classification of brain tumours into Normal and Abnormal | CNN | 91.43% |

2.2.6.2 Non-Convolutional Neural Network

An ANN is a particular kind of ML algorithm that draws inspiration from the design and operation of the human brain. It is made up of many linked processing nodes, often known as "neurons," that collaborate to analyse incoming data and produce output. ANNs can be trained to recognise patterns and relationships in data. There are many different types of ANNs, including feedforward networks, CNNs, RNNs, and more. ANNs are trained using large amounts of labelled data and a process called backpropagation, which adjusts the network's internal weights and biases to reduce error and improve performance. Arunachalam and Royappan [23] developed a CAD, fully automatic brain tumour detection and segmentation method with four modules: enhancement, transformation, feature extraction, and classification. They used an FFBPNN network for the purpose of classifying the brain MRIs as either normal or abnormal. They used the SIST to improve the quality of the brain image. They then extracted texture features using the GLCM, Gabor, and DWT. The extracted features were trained and classified as either normal or glioblastoma through the application of FFBPNN. Arunachalam and Royappan's proposed method has an accuracy of 99.8%.

DNNs are ANNs with multiple layers and are modelled after the human brain's information processing method. They are called "deep" because they have many layers of interconnected nodes, which allows them to learn and represent very complex patterns and relationships in data. DNNs are particularly good at handling large and complex datasets. DNNs are trained using large amounts of labelled data and a process called backpropagation, which modifies the network's internal weights and biases to decrease error and increase performance. Preethi and Aishwarya [21] presented a brain tumour classification method that combines wavelet texture features and DNNs. The system consists of feature extraction, feature selection, tumour classification, and segmentation. First and foremost, the noise was removed from the image, and by using wavelet texture features (GLCM and wavelet-based GLCM), they extracted the features from the images. The OFPA was used to select relevant features, and DNN was used for image classification. The PFCM clustering algorithm was then used to extract the tumour area from the abnormal MR image. They achieved a maximum accuracy of 92% with the proposed DNN and 99.3% accuracy with the proposed PFCM algorithm.

A PNN is a type of ANN that is used for classification tasks. It is based on the idea of using probability distributions to model the connection between input features and output

classes. The basic structure of a PNN consists of an input layer, a hidden layer, and an output layer. Shree and Kumar [34] used a PNN classifier for classifying tumours into normal brain tissue (Unaffected) and abnormal brain tissue (Tumorous) from brain MRIs. They used DWT to decompose the images and extract textural features using GLCM. They found that their classifier achieved nearly 100% accuracy on the training set and 95% accuracy on the testing set. Ural [35] developed a method for automatic brain tumour detection and analysis using advanced image processing techniques and a PNN approach. His proposed integrated image processing algorithm, which is KMFCM, resulted in less computational time. His approach obtained a performance rate of approximately 85–90% when tested in MATLAB.

In the context of classification, an SVM is a supervised learning algorithm that takes a set of labelled training data as input and tries to find the hyperplane that best separates the different classes. The hyperplane is defined as the line that is equidistant from the nearest points of each class, known as the support vectors. Once the hyperplane is found, new data can be classified by seeing on which side of the hyperplane it falls. Reema et al [19] proposed a method for segmenting and classifying brain tumours from MRIs to aid in treatment planning and surgery. The technique involves pre-processing the MRIs using Ostu's thresholding, detecting the tumours using K-Means clustering, extracting features using DWT, Gabor wavelet, and GLCM, and reducing the feature set using PCA. The tumours are then classified as benign or malignant using SVM with linear, polynomial, and RBF kernels. The results show that the linear kernel produces the most accurate results.

Rajaguru et al [36] presented a model that uses an SVM classifier to detect cancer in earlier stages by analysing MRIs. To differentiate normal from abnormal MRIs, the authors used a technique called GLCM to extract features for classification. It was determined that the quadratic and RBF ($\sigma = 0.5$) kernel methods in SVM produced the highest classification accuracy when compared to other kernel methods used in SVM. According to their paper, evaluating this approach on larger datasets is expected to not only test its generalisation ability but also further improve its performance through more extensive training, as classification systems typically perform better with more data. Rajan and Sundar [12] developed a hybrid system for detecting and segmenting brain cancer that combines SVM with KMFCM. The system used ACLS to segment tumours and an adaptive GLCM method to extract features.

Bahadure et al [22] proposed a brain tumour detection model that uses a biologically inspired combination of BWT and an SVM classifier to improve detection accuracy. The technique was found to have 96.51% accuracy, 94.2% specificity, and 97.72% sensitivity in experiments using brain MRIs, showing it to be effective at distinguishing normal and abnormal tissues. Soltaninejad et al [37] presented a fully automated method for detecting and segmenting abnormal tissue related to brain tumours (including the tumour core and oedema) from FLAIR MRIs. The proposed method utilised the superpixel technique and classifies each superpixel using a range of novel image features, which included intensity-based, Gabor textons, fractal analysis, and curvature measures. The method compares the performance of an ERT classifier and an SVM for classifying each superpixel as tumorous or non-tumorous.

In their paper, Alfonse and Salem [18] developed a model for automatically classifying MRIs of the brain as benign or malignant using an SVM classifier. The model includes steps for dataset acquisition, pre-processing, segmentation, feature extraction, feature selection, and classification. After pre-processing, they used the EM algorithm and adaptive thresholding to perform image segmentation. They extracted features using FFT and used the MRMR criterion to select the most valuable features. When these selected features were used as the inputs of the SVM classifier, the model achieved an accuracy of 98.9%.

SVMs are particularly useful in cases where the data has a high number of dimensions, as they can still perform well even when the number of dimensions is much greater than the number of samples. They are also robust to noise and can handle non-linear decision boundaries. One of the key advantages of SVMs is that they provide a unique solution, thanks to the use of the hinge loss function and the optimisation of the margin. This makes them less prone to overfitting compared to other algorithms, such as NNs. In summary, SVM is a powerful ML algorithm that can handle high-dimensional data and is particularly useful for classification tasks.

ELM-LRF is a type of ML algorithm that combines the fast training and generalisation capabilities of ELM with the ability to learn local features from the data using LRF. ELM is a type of feedforward NN that can be trained very quickly and has good generalisation performance, but it is not as good at learning local features as some other types of NNs. By combining ELM with LRF, the resulting ELM-LRF algorithm can learn local features from the data while still maintaining the fast training and good generalisation of ELM.

Ali and Davut [20] presented a model for identifying and classifying brain tumours from cranial MRIs based on ELM-LRF. During pre-processing, nonlocal means and local smoothing methods were utilised to remove potential noise. The classification stage involved using an ELM-LRF to classify brain tumours as benign or malignant. Finally, the cancers were detected using watershed segmentation, and the suggested ELM-LRF structure obtained a classification accuracy of 97.18%, compared to AlexNet's accuracy of 96.91%.

The findings of the literature review on non-CNN models are summarised in a matrix, which can be found in Table 2.5.

Table 2.5: Papers which used Non-CNN

| No. | Paper | Modelling Approach | Objective | Feature Extraction Method | Accuracy |
|-----|-------------------------------------|--------------------|--|------------------------------|---|
| 1 | Arunachalam and Royappan, 2017 [23] | FFBPNN | Classification of brain tumours into normal and abnormal | DWT, Gabor Wavelet, and GLCM | 99.8% |
| 2 | Preethi and Aishwarya, 2019 [21] | DNN | Classification of brain tumours into normal and abnormal | GLCM with Wavelet GLCM | 99.3% |
| 3 | Shree and Kumar, 2018 [34] | PNN | Classification of brain tumours into normal and abnormal | GLCM | 95% |
| 4 | B. Ural, 2018 [35] | PNN | Detection of brain tumours | KMFCM | 90% |
| 5 | Reema et al, 2017 [19] | SVM | Segmentation and classification of brain tumours | DWT, Gabor Wavelet, and GLCM | Not stated (although classifiers with linear kernel were shown to have the highest accuracies for majority of data) |
| 6 | Rajaguru et al, 2016 [36] | SVM | Detection of brain tumours | DWT, Gabor Wavelet, and GLCM | 100% (for quadratic and RBF ($\sigma = 0.5$) kernel methods in SVM) |

| | | | | | |
|-----------|-------------------------------|---------------|---|--|--------|
| 7 | Rajan and Sundar, 2019 [12] | SVM and KMFCM | Detection and segmentation of brain tumours | Adaptive GLCM | 98% |
| 8 | Bhadure et al, 2017 [22] | SVM with BWT | Detection of brain tumours | GLCM | 96.51% |
| 9 | Soltaninejad et al, 2016 [37] | SVM | Detection and segmentation of brain tumours | Calculation of Gabor texton feature, fractal analysis, curvature and statistical intensity features from superpixels | 98.22% |
| 10 | Alfonse and Salem, 2016 [18] | SVM | Classification of brain tumours | FFT | 98.9% |
| 11 | Ali and Davut, 2018 [20] | ELM-LRF | Detection and classification of brain tumours | ELM-LRF | 97.18% |

2.3 Research gap

Despite the potential for DL to facilitate brain cancer detection and classification, there is still a research gap in the field due to the lack of focus on subtyping Gliomas, Meningiomas, and Pituitary Tumours. Most of the papers reviewed in this area have focused on classifying these tumours into broad categories but have not addressed the issue of subtyping them into more specific subtypes based on their location, histology, and behaviour. This is a significant gap in the literature, as accurate subtyping of brain tumours is vital for planning treatments and determining prognosis.

In addition, none of the papers reviewed have discussed the issue of classifying the severity of brain tumours according to the WHO guidelines, which categorise tumours into four grades based on their degree of aggressiveness. This is another important gap in the literature, as accurate grading of brain tumours is crucial for predicting prognosis and planning treatment.

To address these gaps in the literature, further research is needed to develop DL models that can classify brain cancer into a larger number of subtypes and that can accurately grade tumours according to the WHO guidelines. This will require the development of new, more comprehensive image datasets that include a wider variety of brain tumours and are annotated with detailed information about the location, histology, and behaviour of each tumour. With these resources, it will be possible to build more accurate and comprehensive models for classifying and grading brain cancer, which will ultimately lead to better outcomes for patients.

2.4 Summary

This chapter presents a comprehensive review of the relevant literature on brain cancer detection and classification using DL. It covers background information on brain cancer, including its types, causes, and consequences. It also discusses the current methods of detection and their limitations. The chapter delves into the application of DL in medical imaging, specifically for cancer detection and classification. It reviews the literature on different imaging modalities such as CT, X-ray, and MRI and their suitability for brain cancer detection. It also covers the various datasets used in the literature for brain cancer detection and classification, pre-processing techniques that have been used

to improve the performance of the models, and the modelling approaches used in the literature (CNN vs. Non-CNN). Additionally, the chapter identifies the research gaps and the areas that need further exploration in the field of brain cancer detection and classification using DL. By using the knowledge gained from previous research and incorporating new techniques, better results in the detection and classification of brain cancer could be achieved.

Chapter 3

METHODOLOGY

3.1 Overview

This chapter presents the methodology of this Final Year Project, in which a DL model for brain cancer detection and classification was proposed. A robust DL model for accurate detection and classification of brain cancer is crucial for successful treatment and survival rates. This research utilised a methodology that involves several stages to build a DL model for brain cancer detection and classification. The process began with obtaining a brain MR image dataset. Then, the images underwent pre-processing steps to improve the quality of the dataset and prepare it for the next stages. The dataset was split into five subsets of equal size to prepare for 5-fold cross-validation. Next, GoogleNet, a pre-trained CNN, was used as a base model. This model underwent fine-tuning to suit the specific characteristics of the brain MRIs. After the brain MRIs were fed into the network, a feature extraction stage was then performed using the fine-tuned GoogleNet model. The features extracted were then used for the brain MRI classification stage, where the model classified the brain MRI into four classes: Glioma, Meningioma, Pituitary Tumour, and Normal. Finally, the model's predictions and evaluations were carried out. The proposed methods are displayed in Figure 3.1.

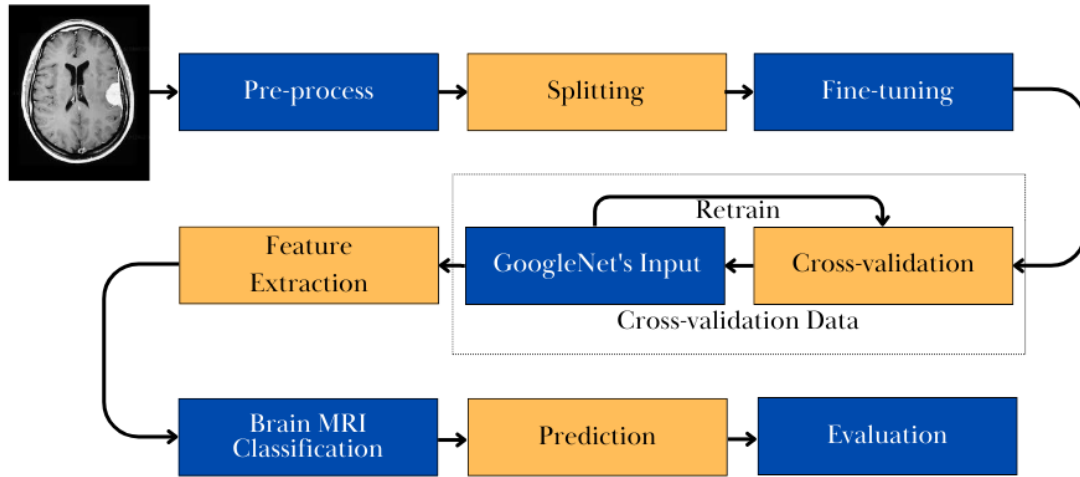


Figure 3.1: The Flowchart of the Proposed Approach

3.2 Preparation

3.2.1 Hardware and Software

The proposed brain cancer detection and classification model will be implemented on a computer with 8 GB of RAM and a 7th Gen Intel Core i5-7300HQ processor using the MATLAB R2022a software.

3.2.2 Dataset

The dataset used in this research was sourced from Kaggle and has a Kaggle usability score of 8.75. Masoud Nickparvar [4] created this dataset, which had 7022 brain MRIs that were divided into 4 categories: Glioma, Meningioma, Pituitary Tumour, and Normal. The specifications of the dataset can be seen in Table 3.1.

Table 3.1: Brain MRI Classes and Their Respective Number of Images

| Classes | No. of Images | Sizes |
|------------------|---------------|---------|
| Glioma | 1621 MRI's | Varying |
| Meningioma | 1645 MRI's | |
| Pituitary Tumour | 1757 MRI's | |
| Normal | 2000 MRI's | |

The brain MRIs were of JPEG format and included coronal, sagittal, and axial views, which can be found in Figure 3.2. The dataset was already split into a training set and a testing set.

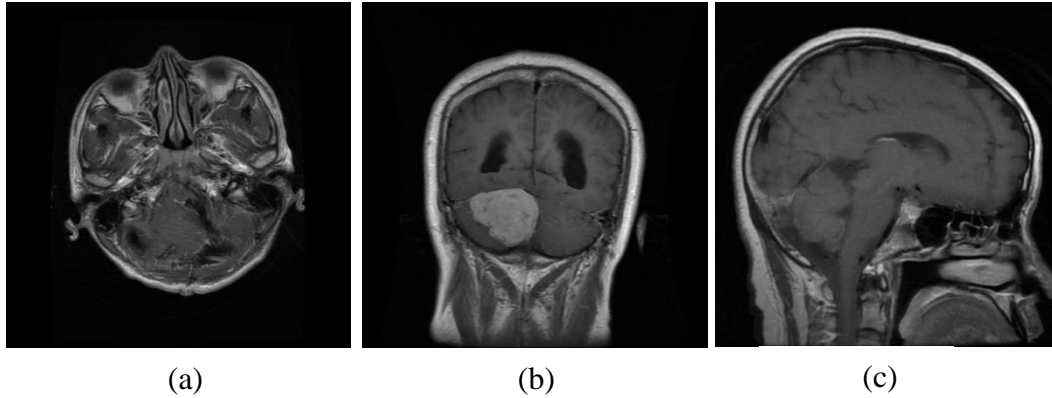


Figure 3.2: Examples of the Brain MRIs in the Dataset [4]. a) Axial View, b) Coronal View and c) Sagittal View

3.2.3 Image Pre-processing

Raw data is often incomplete, noisy, or inconsistent, and pre-processing can help to clean and prepare the data for analysis. It involves a series of steps that are designed to clean and transform the data into a format that is suitable for the algorithm. Pre-processing is an important step in the ML pipeline, as it can have a massive impact on the performance and accuracy of the model. In this paper, the brain MRIs from the dataset underwent pre-processing steps such as cropping, resizing, and normalisation, as presented in Figure 3.3. The images were pre-processed using the Image Batch Processor application in MATLAB R2022a.

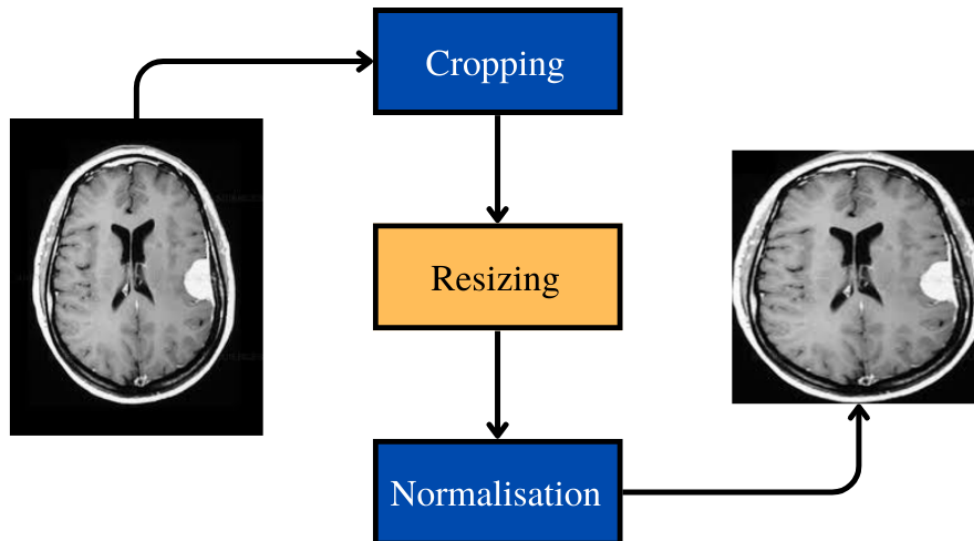


Figure 3.3: Image Pre-processing Steps

The black region around the brain MRIs may not be relevant to the analysis and can introduce noise into the data. Cropping the images to remove this region can help focus the analysis on the relevant parts and remove unnecessary background information. First, the images were checked to determine if they were in grayscale or RGB. If the image was in RGB, it was then converted to grayscale. Then, a median filter was applied to remove noise. Otsu's thresholding was then utilised to obtain a binary image. Dilation and erosion operations were applied to the binary image using a disc-shaped structuring element. The bounding box of the non-zero pixels in the thresholded image was determined by finding the minimum and maximum row and column indices. From these indices, the width and height of the bounding box were calculated. The top-left corner coordinates of the bounding box were determined, and finally, the original image was cropped using the calculated coordinates and dimensions of the bounding box.

In order to use GoogleNet as the pre-trained CNN model for this research, the resulting cropped images were resized to a uniform size (224 x 224). Varying sizes of MRIs can cause issues during model development. This step was required as it allowed the model to process the data more efficiently, improving the training stage. It is also important to note that GoogleNet is a model that was trained on RGB images, which have three colour channels, and the images in the dataset were in grayscale, which have only one channel. In order to make it compatible with the model's architecture, the single channel in the image needed to be replicated into three channels. This process is also known as "colour channel replication" or "fake RGB."

Once the images were cropped and resized, normalisation in intensity values was carried out to scale the brain MRIs between 0 and 1 using the min-max normalisation method by subtracting the minimum value of each pixel of the image and then dividing the result by the range of the image, which is the difference between the maximum and minimum values.

3.3 Proposed Detection and Classification Model

3.3.1 GoogleNet

A CNN is a DL algorithm designed to automatically extract and learn hierarchical patterns and features from input data, primarily used for tasks such as image recognition and classification. It is outlined to process data through multiple layers of arrays, called convolutional layers, by performing convolutions with filters over the input data, passing it through nonlinear activation functions, and pooling the output to obtain a compact representation of the input data. The compact representation is then passed through one or more FC layers, which perform classification on the input data.

GoogleNet is a CNN architecture that was developed by Google and introduced in a 2014 paper titled “Going deeper with convolutions” by Christian Szegedy et al [29]. GoogleNet was the first CNN to achieve outstanding results on the ImageNet dataset, which is a massive dataset of images containing 10,000 classes used to train and evaluate image recognition algorithms. The architecture of GoogleNet can be found in Figure 3.4.

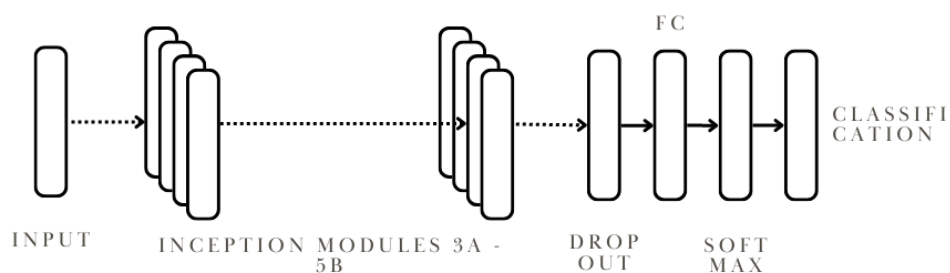


Figure 3.4: Architecture of GoogleNet

GoogleNet, also known as Inception v1, is a deep CNN architecture designed to achieve high accuracy while minimising computational complexity. It introduced the concept of the Inception module, which employs parallel convolutional filters of different sizes to capture diverse scale-invariant features. The GoogleNet architecture consists of multiple stacked Inception modules, with each of them coupled with a max-pooling layer

for spatial down-sampling at the end. It also includes auxiliary classifiers at intermediate layers to encourage the flow of gradients during training and mitigate the vanishing gradient problem. Additionally, GoogleNet incorporates dropout layers, which randomly remove a portion of the neurons during training, to prevent overfitting and improve the generalisation capabilities of the network. The final layer uses global average pooling to reduce spatial dimensions and a fully connected layer for classification. This unique architecture enables GoogleNet to achieve state-of-the-art performance on various image recognition tasks with significantly fewer parameters compared to traditional deep networks.

3.3.2 Transfer Learning

In the context of building a DL model for brain cancer detection and classification, transfer learning can be a powerful technique to improve the performance of the model. Transfer learning is an ML technique where a model's knowledge acquired from solving one task is utilised to enhance the performance of a related task, instead of training the model from scratch.

By using a pre-trained model, such as GoogleNet, as the starting point for this model, the features learned by the model on a large dataset, such as ImageNet, were leveraged to improve the performance of the model on a smaller brain MRI dataset. This approach accelerated the training process and improved the proposed model's performance. This process is presented in Figure 3.5.

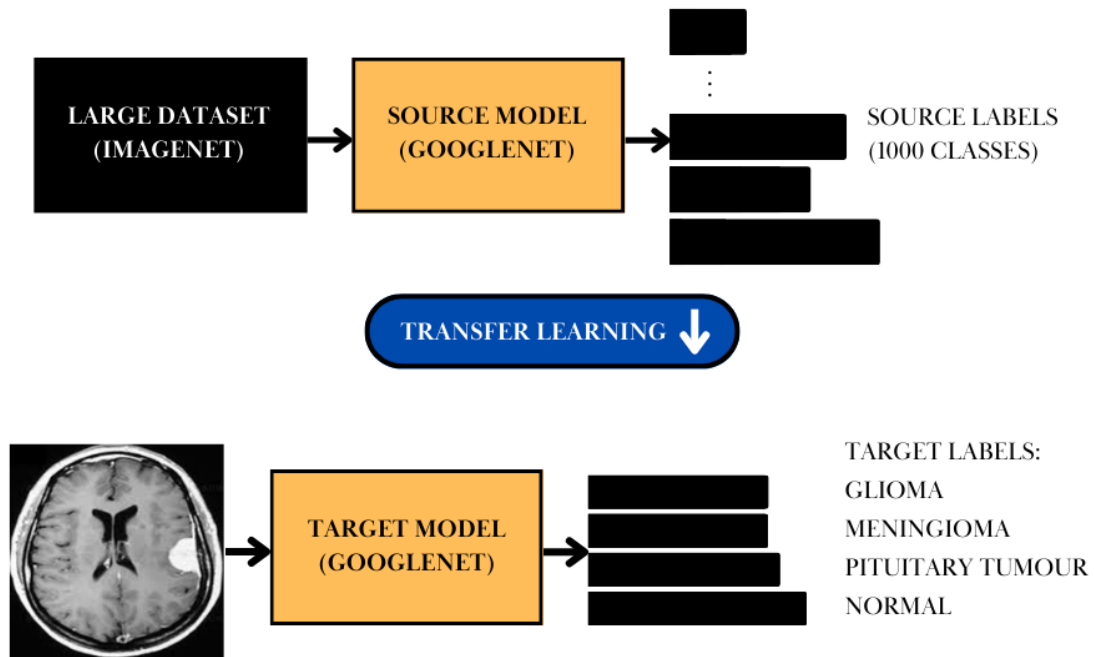


Figure 3.5: Transfer Learning

3.3.3 Modified GoogleNet

GoogleNet is a CNN architecture that was trained on the ImageNet dataset, which has a different number of classes and different image resolutions than the brain MR image dataset. To achieve the goal of this project, the architecture of GoogleNet was modified to suit the dataset, as can be seen in Figure 3.6.

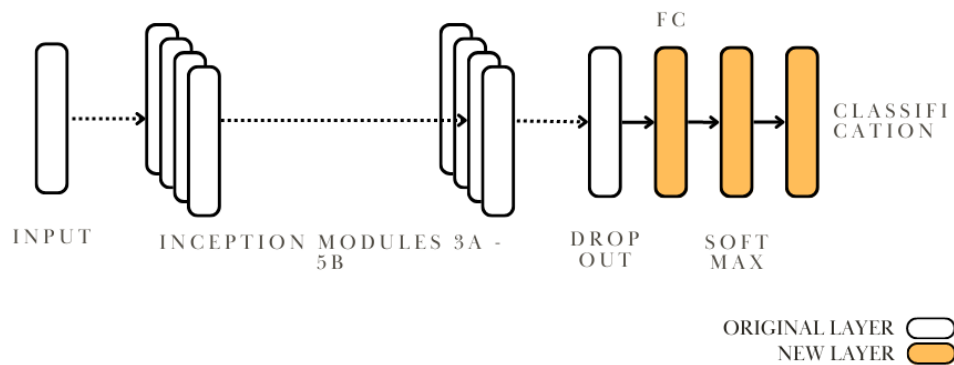


Figure 3.6: Modified Architecture of GoogleNet

To adapt GoogleNet for this task, the FC layer of GoogleNet that was used for classifying the ImageNet dataset was removed and replaced with a new FC layer that has an output size of 4, one for each of the 4 classes: Glioma, Meningioma, Pituitary Tumour, and Normal. The softmax activation function and classification layer were also removed and replaced with new ones. The softmax activation function mapped the output of the

neuron to a probability distribution over multiple classes. The output of the softmax function may be seen as the likelihood of the input image belonging to each of the classes. It can be defined as:

$$\text{softmax}(z) = \frac{e^{z_i}}{\sum e^z} \dots \dots \dots (3.1)$$

where z is a vector of the outputs of the neurons in the final layer, z_i is the output of the i -th neuron, and e is the exponential function.

3.3.4 Fine-tuning

In this project, the earlier layers of GoogleNet, the Inception modules, and the dropout layer, as displayed in Figure 3.7, were fine-tuned before the training phase. Modifications were made to the weight and bias learning rate factors within them to fine-tune the learning process and enhance the model's performance.

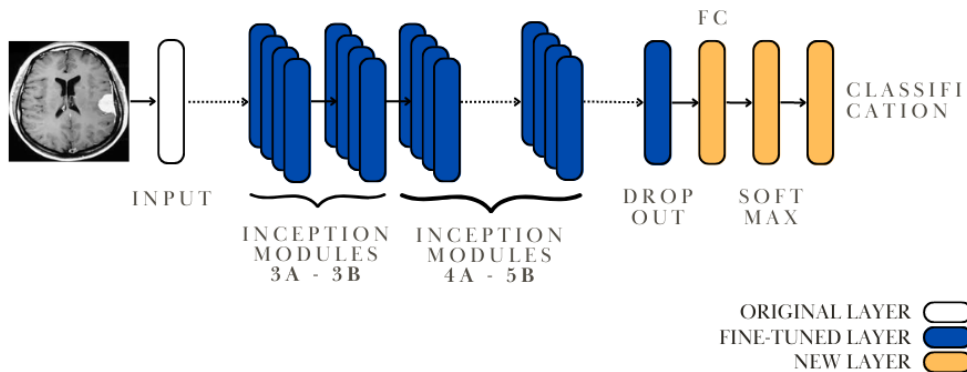


Figure 3.7: Fine-tuned GoogleNet Architecture

For the layers from the first convolutional layer until the Inception 3b module of GoogleNet, the weight and bias learning rate factors were modified from their default values of 1 to 0.001. This refers to the early stages of the network, where low-level features were extracted from the input data. This reduction in the learning rate factors allowed for a more gradual update of the weights and biases in these layers, enabling better convergence and improved generalisation.

Similarly, for the layers spanning from the Inception 4a module until the 5b module, the weight and bias learning rate factors were adjusted from the default value of 1 to 0.01. This modification aimed to fine-tune the learning process within these specific layers, ensuring that the model could effectively capture and leverage the intricate patterns and

features present in the data. These layers represent the deeper and more complex layers of the network, where higher-level features and abstractions were learned.

Additionally, the dropout value was reduced to 0.3 from the default 0.4 after several experiments. Setting a lower value for the dropout rate may reduce the regularisation effect, making the model more prone to overfitting. On the other hand, setting a higher value may excessively regularise the model, leading to underfitting and reduced performance. However, the dropout value of 0.3 was the optimal dropout rate that maximised the model's performance.

By customising the learning rate factors for different layers of the GoogLeNet architecture, the model's ability to learn and adapt to the specific characteristics of the dataset was enhanced. These modifications allowed for more targeted and efficient learning, ultimately leading to improved performance and generalisation capabilities.

3.3.5 Experimental Settings

The experimental settings for this project, which are displayed in Table 3.2, were carefully chosen to train and evaluate the DL model for brain cancer detection and classification. These experimental settings were chosen based on prior knowledge and empirical studies to strike a balance between model complexity, training efficiency, and generalisation performance.

Table 3.2: Training Settings

| Settings | Values |
|---------------------------|--------|
| Optimizer | Adam |
| Initial Learning Rate | 0.001 |
| Learning Rate Drop Period | 10 |
| Learning Rate Drop Factor | 0.1 |
| Maximum Epochs | 20 |
| Mini-batch Size | 32 |
| Validation Frequency | 20 |
| Validation Patience | 10 |

The Adam optimizer, known for its effectiveness in training deep neural networks, was selected as the optimisation algorithm. It utilises adaptive learning rates and momentum to efficiently update the model's parameters during training.

The initial learning rate was set to 0.001, determining the step size at the beginning of the training process. A higher initial learning rate allows for faster convergence, while a lower value ensures more stable learning. By tuning this parameter, the model struck a balance between learning speed and accuracy.

The learning rate drop factor of 0.1 and the learning rate drop period of 10 were used to schedule learning rate reductions. This technique helped the model fine-tune its parameters by decreasing the learning rate at predefined intervals. It allowed the model to make smaller updates to the weights, ensuring better convergence and preventing overshooting of the optimal solution. Hence, after 10 epochs, the learning rate was automatically reduced to 0.0001.

The maximum number of epochs was set to 20, representing the total number of complete passes through the entire training dataset. This ensures enough training iterations to capture the underlying patterns in the data without overfitting. Each epoch consists of multiple mini-batches, and a mini-batch size of 32 was chosen to balance computational efficiency and model performance.

The validation frequency of 20 indicates that the model's performance on the validation set is evaluated every 20 mini-batches during training. This allows for regular monitoring of the model's progress and early detection of any signs of overfitting or underfitting.

The validation patience parameter was set to 10. Validation patience is a technique used during training to trigger early stopping based on validation losses. It determines the number of consecutive epochs with no improvement in the validation loss before training is halted. In this project, when the validation loss did not improve for 10 consecutive epochs in each cross-validation iteration, the training process was automatically stopped early due to this setting to prevent overfitting.

To prevent overfitting, L2 regularisation with a coefficient of 0.001 was applied. This regularisation technique adds a penalty term to the loss function, encouraging the model to learn simpler and more generalised representations.

3.3.6 Cross-validation

Cross-validation is a technique used to evaluate the performance of an ML model by dividing the data into a training set and a test set. The main goal of cross-validation is to provide an estimate of the model's performance on unseen data and to help prevent overfitting by assessing how well the model generalises to new data. In this Final Year Project, a cross-validation technique named k-fold cross-validation was used, in which 'k' can be defined as the number of folds (in this paper, $k = 5$). The brain MR image dataset was split into 5 subsets of equal size, and the model was trained on 'k - 1' (in this paper, $k - 1 = 4$) subsets while being tested on the remaining subset. This process was repeated five times, with a different subset being used as the test set in each iteration. The model's performance was then averaged across all five cross-validation iterations. This process can be seen in Figure 3.8.

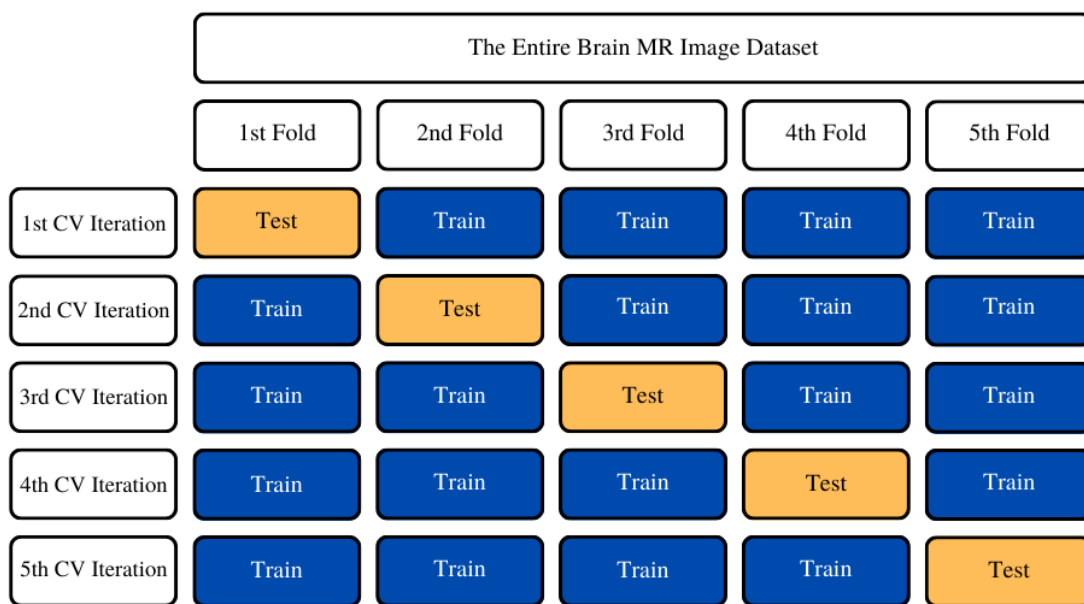


Figure 3.8: 5-fold Cross-validation

3.4 Testing

3.4.1 Confusion Matrix

A confusion matrix is a tabular representation that could be used to assess the performance of a classification model, primarily in ML and statistics, where the predicted class labels are compared with the actual class labels. Figure 3.9 presents the confusion matrix. In the context of brain cancer detection and classification, the confusion matrix can be used to evaluate the performance of the model in classifying brain MRIs into 4 classes: Glioma, Meningioma, Pituitary Tumour, and Normal.

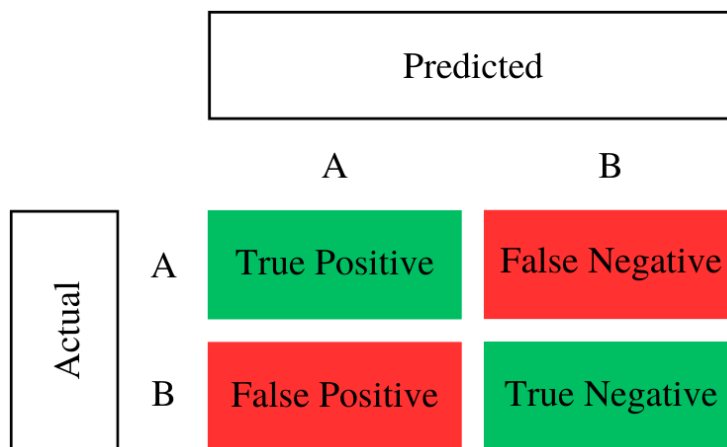


Figure 3.9: Binary Confusion Matrix

True Positive (TP) is the count of instances that the model correctly identifies as positive. False Positive (FP) is the count of instances that the model incorrectly identifies as positive. True Negative (TN) represents the count of instances that the model correctly identifies as negative. False Negative (FN) is the count of instances that the model incorrectly identifies as negative.

In the case of a multi-class classification model, a binary confusion matrix could not be used. However, a one-vs.-all confusion matrix, which is displayed in Figure 3.9, could be used as a substitute.

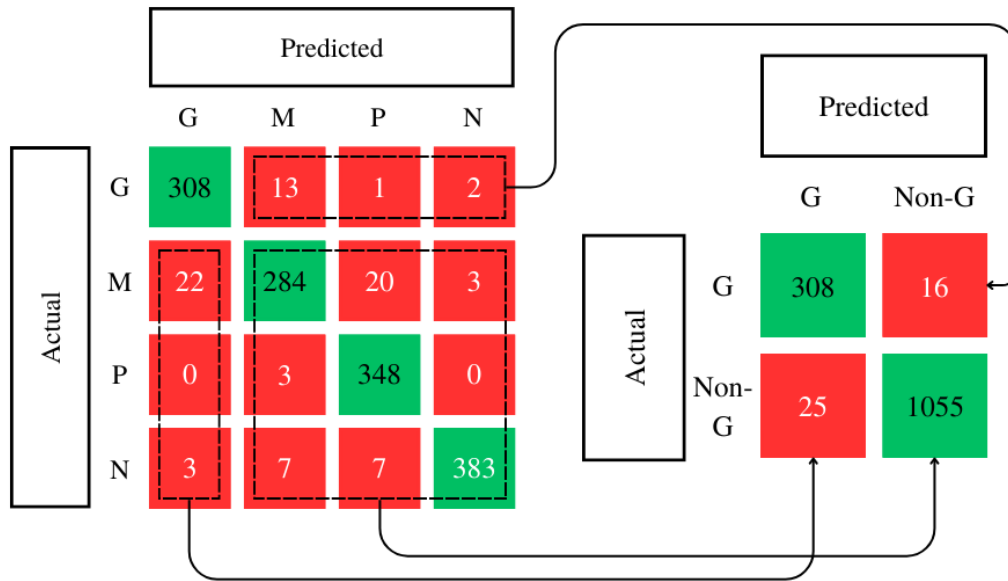


Figure 3.10: One-vs-all Confusion Matrix (where G, M, P and N refer to Glioma, Meningioma, Pituitary Tumour and Normal, respectively) for the First Cross-validation Iteration)

In a multi-class classification problem, this approach is often used to extend binary classifiers to handle multiple classes. In this approach, each class (e.g., Glioma (G)) is treated as a positive class, and the remaining classes (e.g., Meningioma (M), Pituitary Tumour (P), and Normal (N)) are treated as a single negative class. The rows represent the actual classes, while the columns represent the predicted classes.

This type of confusion matrix provides insights into the classification accuracy, precision, recall, specificity, and F1 score for each class separately, which can be useful for assessing the model's performance in a multi-class setting. Since each class is treated as a positive class in a separate evaluation, metrics such as accuracy, precision, recall, specificity, and F1 score can be calculated individually for each class.

The overall metrics for the model can be obtained by considering the weight of each class. To obtain the overall metrics for the model, the metrics for each class can be weighted based on the proportion of samples in the dataset belonging to that class. This weightage considers the imbalance in class distribution and provides a more representative measure of the model's performance.

3.4.2 Class-specific Metrics

There are several class-specific metrics that could be obtained from the one-vs.-all confusion matrix. They provide a detailed evaluation of the performance of a multi-class classification model for each individual class. Some commonly used class-specific metrics include accuracy, precision, recall (sensitivity), specificity, and F1 score.

The accuracy of a class represents the proportion of correctly classified instances for that specific class. It is calculated as the ratio of true positives (TP) and the sum of true positives and false negatives (TP + FN). Accuracy measures the overall correctness of the predictions for a given class.

$$\text{Accuracy for Glioma (G)} = \frac{TP_G + TN_G}{TP_G + TN_G + FP_G + FN_G} \dots \dots \dots (3.2)$$

$$\text{Accuracy for Meningioma (M)} = \frac{TP_M + TN_M}{TP_M + TN_M + FP_M + FN_M} \dots \dots \dots (3.3)$$

$$\text{Accuracy for Pituitary Tumour (P)} = \frac{TP_P + TN_P}{TP_P + TN_P + FP_P + FN_P} \dots \dots \dots (3.4)$$

$$\text{Accuracy for Normal (N)} = \frac{TP_N + TN_N}{TP_N + TN_N + FP_N + FN_N} \dots \dots \dots (3.5)$$

Precision for a class quantifies the proportion of correctly classified instances out of all instances predicted to belong to that specific class. It is calculated as the ratio of true positives (TP) and the sum of true positives and false positives (TP + FP). Precision reflects the classifier's ability to avoid false positives.

$$\text{Precision for Glioma (G)} = \frac{TP_G}{TP_G + FP_G} \dots \dots \dots (3.6)$$

$$\text{Precision for Meningioma (M)} = \frac{TP_M}{TP_M + FP_M} \dots \dots \dots (3.7)$$

$$\text{Precision for Pituitary Tumour (P)} = \frac{TP_P}{TP_P + FP_P} \dots \dots \dots (3.8)$$

$$\text{Precision for Normal (N)} = \frac{TP_N}{TP_N + FP_N} \dots \dots \dots (3.9)$$

Recall, also known as sensitivity, measures the proportion of correctly classified instances of a class out of all instances that exactly belong to that class. It is calculated as

the ratio of true positives (TP) and the sum of true positives and false negatives (TP + FN). Recall indicates the classifier's ability to identify positive instances.

$$\text{Recall for Glioma (G)} = \frac{TP_G}{TP_G + FN_G} \dots \dots \dots (3.10)$$

$$\text{Recall for Meningioma (M)} = \frac{TP_M}{TP_M + FN_M} \dots \dots \dots (3.11)$$

$$\text{Recall for Pituitary Tumour (P)} = \frac{TP_P}{TP_P + FN_P} \dots \dots \dots (3.12)$$

$$\text{Recall for Normal (N)} = \frac{TP_N}{TP_N + FN_N} \dots \dots \dots (3.13)$$

Specificity refers to the proportion of correctly classified instances of non-target classes out of all instances that do not belong to the specific class. It is calculated as the ratio of true negatives (TN) and the sum of true negatives and false positives (TN + FP). Specificity reflects the classifier's ability to avoid false positives for non-target classes.

$$\text{Specificity for Glioma (G)} = \frac{TN_G}{TN_G + FP_G} \dots \dots \dots (3.14)$$

$$\text{Specificity for Meningioma (M)} = \frac{TN_M}{TN_M + FP_M} \dots \dots \dots (3.15)$$

$$\text{Specificity for Pituitary Tumour (P)} = \frac{TN_P}{TN_P + FP_P} \dots \dots \dots (3.16)$$

$$\text{Specificity for Normal (N)} = \frac{TN_N}{TN_N + FP_N} \dots \dots \dots (3.17)$$

The F1 score is the harmonic mean of precision and recall, providing a balanced measure of a classifier's performance. It combines both precision and recall into a single metric. The F1 score is calculated as:

$$\text{F1 Score for Glioma (G)} = \frac{2(\text{Precision}_G * \text{Recall}_G)}{\text{Precision}_G + \text{Recall}_G} \dots \dots \dots (3.18)$$

$$\text{F1 Score for Meningioma (M)} = \frac{2(\text{Precision}_M * \text{Recall}_M)}{\text{Precision}_M + \text{Recall}_M} \dots \dots \dots (3.19)$$

$$\text{F1 Score for Pituitary Tumour (P)} = \frac{2(\text{Precision}_P * \text{Recall}_P)}{\text{Precision}_P + \text{Recall}_P} \dots \dots \dots (3.20)$$

$$F1 \text{ Score for Normal } (N) = \frac{2(Precision_N * Recall_N)}{Precision_N + Recall_N} \dots \dots \dots (3.21)$$

These class-specific metrics provide valuable insights into the performance of a multi-class classification model for each individual class, allowing for a more detailed evaluation and understanding of the model's strengths and weaknesses in classifying different categories.

3.4.3 Overall Performance Metrics

In a multi-class classification problem, it is crucial to consider the potential class imbalances when evaluating the performance of a model. Class imbalances occur when the number of instances in different classes is not evenly distributed. Simply dividing the metrics by the number of classes may not provide an accurate representation of the model's overall performance. The metrics may be heavily biased towards the larger classes.

The number of samples for every class in this project, which is displayed in Table 3.3 is not balanced. Fortunately, weighted metrics consider the varying sizes of each class by assigning appropriate weights based on the proportion of instances in each class. By incorporating these weights, the evaluation metrics provide a more accurate and meaningful assessment of the model's performance across all classes. When classes are imbalanced, the majority class may dominate the evaluation metrics, overshadowing the performance of the minority class. Weighted metrics mitigate this issue by giving more weight to the minority classes, ensuring that the evaluation is balanced and reflective of the model's effectiveness in handling class imbalances.

Table 3.3: Weightage of Each Class in the Testing Set of Every Cross-validation Iteration

| Weightage of Each Class (W) | | | | | |
|-----------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Classes | 1st CV Iteration | 2st CV Iteration | 3st CV Iteration | 4st CV Iteration | 5st CV Iteration |
| Glioma | 324 | 324 | 324 | 325 | 324 |
| Meningioma | 329 | 329 | 329 | 329 | 329 |

| | | | | | |
|------------------|------|------|------|------|------|
| Pituitary Tumour | 351 | 351 | 352 | 352 | 351 |
| Normal | 400 | 400 | 400 | 400 | 400 |
| Total | 1404 | 1404 | 1405 | 1406 | 1404 |
| Weightage | | | | | |

The formulas of the weighted metrics are as follow:

Weighted Accuracy (Acc)

$$= \frac{Acc_G N_G + Acc_M N_M + Acc_P N_P + Acc_N N_N}{W_G + W_M + W_P + W_N} \dots \dots \dots (3.22)$$

Weighted Precision (Prc)

$$= \frac{Prc_G N_G + Prc_M N_M + Prc_P N_P + Prc_N N_N}{W_G + W_M + W_P + W_N} \dots \dots \dots (3.23)$$

$$Weighted Recall (Rcl) = \frac{Rcl_G N_G + Rcl_M N_M + Rcl_P N_P + Rcl_N N_N}{W_G + W_M + W_P + W_N} \dots \dots \dots (3.24)$$

Weighted Specificity (Spc)

$$= \frac{Spc_G N_G + Spc_M N_M + Spc_P N_P + Spc_N N_N}{W_G + W_M + W_P + W_N} \dots \dots \dots (3.25)$$

Weighted F1 Score (F1S)

$$= \frac{F1S_G N_G + F1S_M N_M + F1S_P N_P + F1S_N N_N}{W_G + W_M + W_P + W_N} \dots \dots \dots (3.26)$$

By aggregating the weighted metrics from each class, the overall metrics for the model can be computed. These metrics provide an overall assessment of the model's performance in terms of its ability to classify samples across all classes, considering both the individual class performance and their relative weightage based on the class distribution.

3.5 Summary

This chapter presents the methodology used in the development of a DL model for brain cancer detection and classification. The hardware and software used, as well as details of the brain MRI dataset chosen, are discussed. The steps taken during image pre-

processing, which were vital to improving the model's performance and enhancing the accuracy of predictions, are discussed in this chapter. The proposed approach used to detect and classify brain tumours according to their types, which includes transfer learning and modifications to GoogleNet architecture, is described in detail. The chapter also addresses the training stage of the DL model in this research, specifically the measures taken to prevent overfitting by making changes to the optimizer and value of parameters. Additionally, the chapter discusses the classification metrics that were used to determine the efficacy of the proposed model.

Chapter 4

RESULTS AND DISCUSSION

4.1 Overview

The chapter presents a comprehensive analysis of the performance of the classification model for brain cancer detection. This chapter highlights various metrics, including training and validation accuracies, training and validation losses, and classification metrics like accuracy, precision, recall, specificity, and F1 score. Additionally, it provides insights into the confusion matrix and engages in a detailed discussion of the obtained results.

4.2 Experimental Results

To assess the performance of the DL model for brain cancer detection and classification, a 5-fold cross-validation approach was utilised. This technique involved splitting the dataset into five folds of equal size and performing training and testing on five different combinations of these folds. By conducting cross-validation, more robust and reliable results could be obtained. At the conclusion of the cross-validation process, the averages for all the evaluation metrics of each iteration, including accuracy, precision, recall, specificity, and F1 score, were calculated. These averaged metrics provide a comprehensive overview of the model's performance across different folds and allow for a more accurate assessment of its effectiveness in detecting and classifying brain cancer cases.

4.2.1 First Cross-validation Iteration

In the first cross-validation iteration as shown in Figure 4.1, the first fold was used as the testing set, while the other folds served as the training set. The model was trained on the training set and evaluated on the first fold to assess its performance.

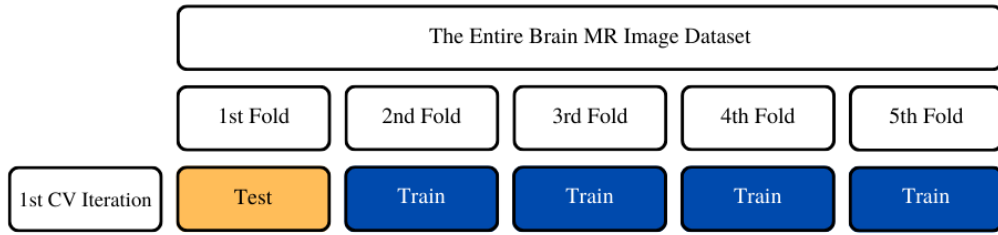


Figure 4.1: First Fold as Testing Set

4.2.1.1 Training and Validation

Table 4.1: Details of the Training and Validation Process for the First Cross-validation Iteration

| Training | Value |
|----------------------|-----------------------------|
| Completed Epochs | 14 out of 20 Epochs |
| Completed Iterations | 1740 out of 2620 Iterations |
| Training Time | 56 Minutes and 1 Second |
| Training Accuracy | 93.7500% |
| Validation Accuracy | 94.5869% |
| Training Loss | 0.2520 |
| Validation Loss | 0.1469 |

The training process took 56 minutes and 1 second to complete, involving 14 epochs out of 20 total epochs. Throughout the training, the model underwent 1740 iterations out of 2620 iterations. The details are displayed in Table 4.1.

Figures 4.1 and 4.2 provide visual representations of the training and validation accuracies as well as the training and validation losses. The final training accuracy was 93.7500%, with a corresponding training loss of 0.2520. The validation accuracy reached

94.5869%, while the validation loss achieved 0.1469, indicating the model's performance on unseen data.

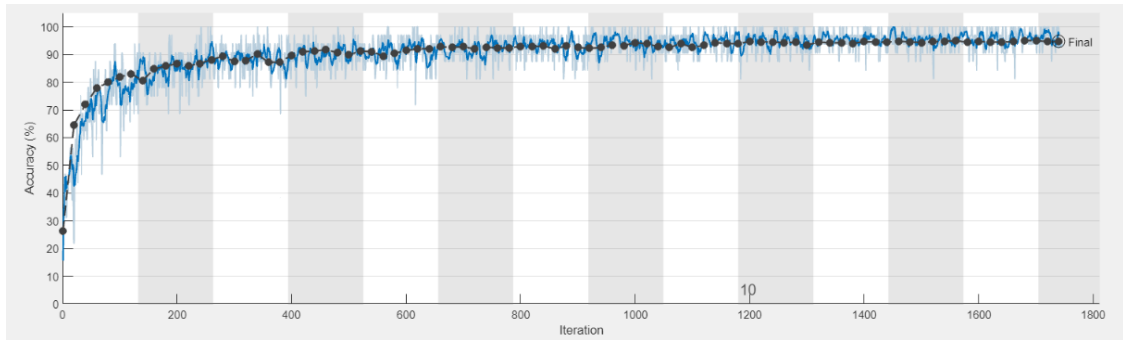


Figure 4.2: Training and Validation Accuracy for the First Cross-validation Iteration

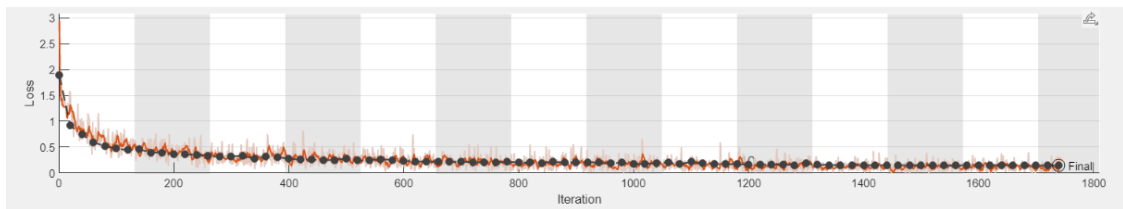


Figure 4.3: Training and Validation Loss for the First Cross-validation Iteration

4.2.1.2 Confusion Matrix

The confusion matrix for the first cross-validation iteration are displayed in Table 4.2.

Table 4.2: Confusion Matrix (where G, M, P and N refer to Glioma, Meningioma, Pituitary Tumour and Normal, respectively) for the First Cross-validation Iteration

| | | Predicted | | | |
|--------|---|-----------|-----|-----|-----|
| | | G | M | P | N |
| Actual | G | 308 | 13 | 1 | 2 |
| | M | 22 | 284 | 20 | 3 |
| | P | 0 | 3 | 348 | 0 |
| | N | 3 | 7 | 7 | 383 |

The model correctly predicted 308 out of 324 samples from the Glioma class. There were 13 misclassifications as Meningioma, 1 misclassification as Pituitary Tumour, and 2 misclassifications as Normal.

Out of 329 samples from the Meningioma class, the model correctly predicted 284. There were 22 misclassifications as Glioma, 20 misclassifications as Pituitary Tumour, and 3 misclassifications as Normal.

348 out of 351 samples from the Pituitary Tumour class were correctly classified. There were 3 misclassifications as Meningioma but none as Glioma or Normal.

The model correctly predicted 383 out of 400 samples from the Normal class. There were 3 misclassifications as Glioma, 7 misclassifications as Meningioma, and 7 misclassifications as Pituitary Tumour. A total of 81 images were misclassified in the first cross-validation iteration.

Table 4.3: No. of Samples and Counts of TP, TN, FP and FN for the First Cross-validation Iteration

| | Samples | TP | TN | FP | FN |
|---|---------|-----|------|----|----|
| G | 324 | 308 | 1055 | 25 | 16 |
| M | 329 | 284 | 1052 | 23 | 45 |
| P | 351 | 348 | 1025 | 28 | 3 |
| N | 400 | 383 | 999 | 5 | 17 |

4.2.1.3 Model Performance

The class-specific metrics in the first cross-validation iteration are displayed in Table 4.4.

Table 4.4: Class-specific Metrics for the First Cross-validation Iteration

| | Accuracy | Precision | Recall | Specificity | F1 Score |
|---|----------|-----------|--------|-------------|----------|
| G | 97.0798% | 0.9249 | 0.9506 | 0.9769 | 0.9376 |
| M | 95.1567% | 0.9251 | 0.8632 | 0.9786 | 0.8931 |
| P | 97.7920% | 0.9255 | 0.9915 | 0.9734 | 0.9574 |
| N | 98.4330% | 0.9871 | 0.9575 | 0.9950 | 0.9721 |

The overall performance metrics of the first cross-validation iteration for the Brain Cancer Detection and Classification Model are displayed in Table 4.5 and were calculated using the weighted metrics formulas (Formulas 3.22–3.26).

Table 4.5: Overall Performance Metrics for the First Cross-validation Iteration

| Metric | Value |
|-------------|----------|
| Accuracy | 97.1927% |
| Precision | 0.9428 |
| Recall | 0.9423 |
| Specificity | 0.9816 |
| F1 Score | 0.9419 |

The model's accuracy of approximately 97.1927% indicates its effectiveness in correctly classifying brain MRI images into their respective categories. This demonstrates its strong ability to differentiate between the four classes: Glioma, Meningioma, Pituitary Tumour, and Normal brain scans.

With a precision value of 0.9428, when the model predicts a specific class, it is accurate around 94.28% of the time. This suggests a low false positives rate in the classification of that particular class, which is crucial for providing reliable information and minimising misclassifications.

The recall metric, measured at 0.9423, indicates that the model correctly identifies approximately 94.23% of the actual positive instances for a specific class. This high recall or sensitivity ensures that a significant majority of positive cases for that specific class are correctly identified.

The specificity value of 0.9816 implies that the model accurately identifies around 98.16% of the instances that do not belong to the specific class. This high specificity minimises false positives for the given class and reduces the chances of misclassification.

The F1 score, computed at 0.9419, represents the harmonic mean of precision and recall for the specific class. It provides a comprehensive measure of the model's performance in accurately classifying instances belonging to that specific class, taking into account both precision and recall.

4.2.2 Second Cross-validation Iteration

In the second cross-validation iteration as shown in Figure 4.4, the second fold was used as the testing set, while the other folds served as the training set. The model was trained on the training set and evaluated on the second fold to assess its performance.

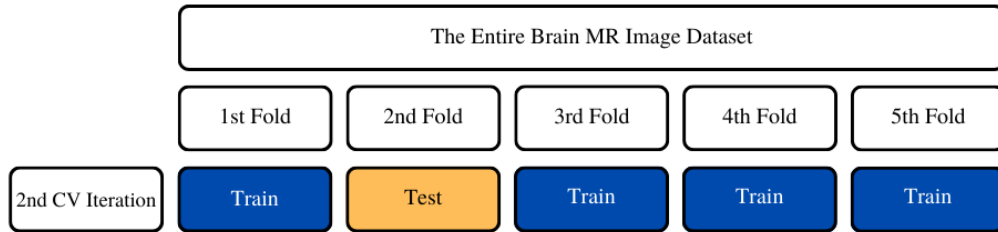


Figure 4.4: Second Fold as Testing Set

4.2.2.1 Training and Validation

Table 4.6: Details of the Training and Validation Process for the Second Cross-validation Iteration

| Training | Value |
|----------------------|-----------------------------|
| Completed Epochs | 16 out of 20 Epochs |
| Completed Iterations | 2000 out of 2620 Iterations |
| Training Time | 63 Minutes and 7 Seconds |
| Training Accuracy | 96.8750% |
| Validation Accuracy | 95.5160% |
| Training Loss | 0.1252 |
| Validation Loss | 0.1380 |

The training process lasted for 63 minutes and 7 seconds, encompassing 16 epochs out of the total 20 epochs. It underwent 2000 iterations out of 2620 iterations, showcasing the model's progressive refinement. Figure 4.5 illustrates the training and validation accuracies, while Figure 4.6 visualises the training and validation losses. The training process concluded with a final training accuracy of 96.8750% and a training loss of 0.1252. The validation accuracy at the end of training was 95.5160%, accompanied by a validation loss of 0.1380, demonstrating the model's performance on unseen data.

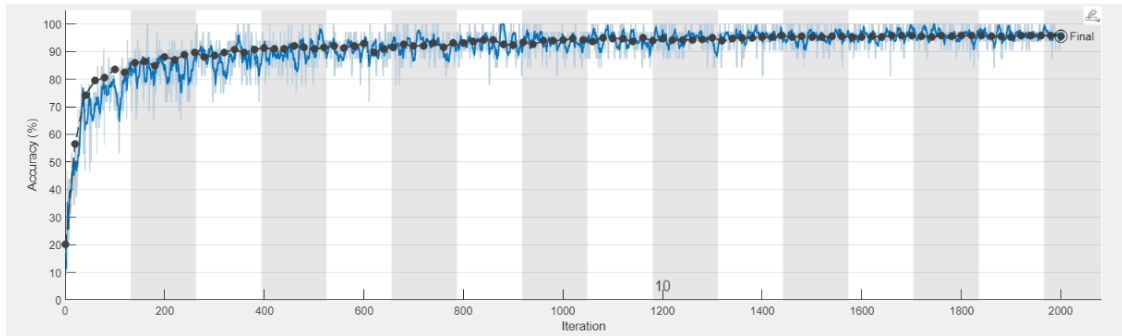


Figure 4.5: Training and Validation Accuracy for the Second Cross-validation Iteration

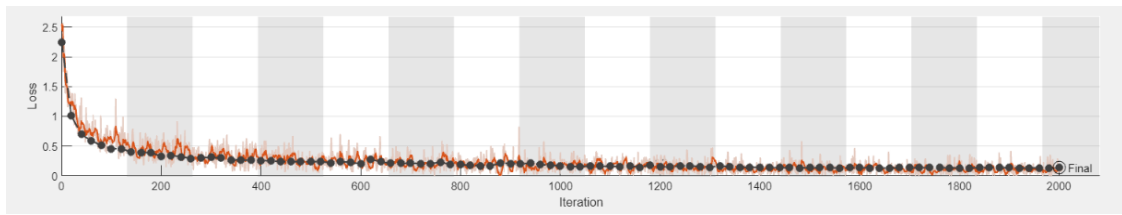


Figure 4.6: Training and Validation Loss for the Second Cross-validation Iteration

4.2.2.2 Confusion Matrix

The confusion matrix for the second cross-validation iteration are displayed in Table 4.7.

Table 4.7: Confusion Matrix (where G, M, P and N refer to Glioma, Meningioma, Pituitary Tumour and Normal, respectively) for the Second Cross-validation Iteration

| | | Predicted | | | |
|--------|---|-----------|-----|-----|-----|
| | | G | M | P | N |
| Actual | G | 296 | 17 | 11 | 0 |
| | M | 15 | 299 | 15 | 0 |
| | P | 1 | 8 | 342 | 0 |
| | N | 2 | 6 | 3 | 389 |

The model correctly predicted 296 out of 324 samples from the Glioma class. There were 17 misclassifications as Meningioma, 11 misclassifications as Pituitary Tumour, and no misclassifications as Normal.

Out of 329 samples from the Meningioma class, the model correctly predicted 299. There were 15 misclassifications as Glioma, 15 misclassifications as Pituitary Tumour, and no misclassification as Normal.

342 out of 351 samples from the Pituitary Tumour class were correctly classified. There were 1 misclassification as Glioma, 8 misclassifications as Meningioma and no misclassification as Normal.

The model correctly predicted 389 out of 400 samples from the Normal class. There were 2 misclassifications as Glioma, 6 misclassifications as Meningioma, and 3 misclassifications as Pituitary Tumour. 78 images were misclassified in the second cross-validation iteration.

Table 4.8: No. of Samples and Counts of TP, TN, FP and FN for the Second Cross-validation Iteration

| | Samples | TP | TN | FP | FN |
|---|---------|-----|------|----|----|
| G | 324 | 296 | 1062 | 18 | 28 |
| M | 329 | 299 | 1044 | 31 | 30 |
| P | 351 | 342 | 1024 | 29 | 9 |
| N | 400 | 389 | 1004 | 0 | 11 |

4.2.2.3 Model Performance

The class-specific metrics in the second cross-validation iteration are displayed in Table 4.9.

Table 4.9: Class-specific Metrics for the Second Cross-validation Iteration

| | Accuracy | Precision | Recall | Specificity | F1 Score |
|---|----------|-----------|--------|-------------|----------|
| G | 96.7236% | 0.9427 | 0.9136 | 0.9833 | 0.9279 |
| M | 95.6553% | 0.9061 | 0.9088 | 0.9712 | 0.9074 |
| P | 97.2934% | 0.9218 | 0.9744 | 0.9725 | 0.9474 |
| N | 99.2165% | 1.0000 | 0.9725 | 1.0000 | 0.9861 |

The overall performance metrics of the second cross-validation iteration for the Brain Cancer Detection and Classification Model are displayed in Table 4.10 and were calculated using the weighted metrics formulas (Formulas 3.22–3.26).

Table 4.10: Overall Performance Metrics for the Second Cross-validation Iteration

| Metric | Value |
|-------------|----------|
| Accuracy | 97.3260% |
| Precision | 0.9452 |
| Recall | 0.9444 |
| Specificity | 0.9825 |
| F1 Score | 0.9445 |

The model's accuracy of approximately 97.3260% indicates its effectiveness in accurately classifying brain MRI images into their respective categories. This demonstrates its strong ability to differentiate between the four classes: Glioma, Meningioma, Pituitary Tumour, and Normal brain scans.

With a precision value of 0.9452, when the model predicts a specific class, it is accurate around 94.52% of the time. This suggests a low false positives rate in the classification of that particular class, which is important for providing reliable information and minimising misclassifications.

The recall metric, measured at 0.9444, indicates that the model correctly identifies approximately 94.44% of the actual positive instances for a specific class. This high recall or sensitivity ensures that a significant majority of positive cases for that specific class are correctly identified.

The specificity value of 0.9825 implies that the model accurately identifies around 98.25% of the instances that do not belong to the specific class. This high specificity minimises false positives for the given class and reduces the chances of misclassification.

The F1 score, computed at 0.9445, represents the harmonic mean of precision and recall for the specific class. It provides a comprehensive measure of the model's performance in accurately classifying instances belonging to that specific class, taking into account both precision and recall.

4.2.3 Third Cross-validation Iteration

In the third cross-validation iteration as shown in Figure 4.7, the third fold was used as the testing set, while the other folds served as the training set. The model was trained on the training set and evaluated on the third fold to assess its performance.

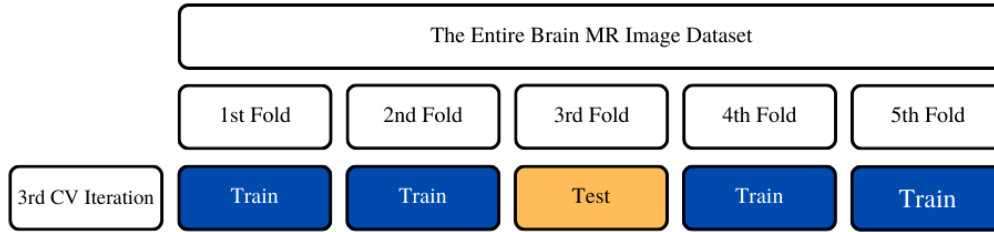


Figure 4.7: Third Fold as Testing Set

4.2.3.1 Training and Validation

Table 4.11: The Details of the Training and Validation Process for the Third Cross-validation Iteration

| Training | Value |
|----------------------|-----------------------------|
| Completed Epochs | 13 out of 20 Epochs |
| Completed Iterations | 1700 out of 2620 Iterations |
| Training Time | 53 Minutes and 30 Seconds |
| Training Accuracy | 100.0000% |
| Validation Accuracy | 95.0925% |
| Training Loss | 0.0395 |
| Validation Loss | 0.1596 |

Throughout the training process, which spanned 53 minutes and 30 seconds, 13 epochs were completed out of the total 20 epochs. The model underwent 1700 iterations. Figures 4.8 and 4.9 showcase the training and validation accuracies, as well as the training and validation losses, giving a visual overview of the model's performance over time. After completion, the training phase yielded a final training accuracy of 100.0000% and a training loss of 0.0395. The validation accuracy reached 95.0925%, with a

corresponding validation loss of 0.1596, indicating the model's ability to generalise to new data.

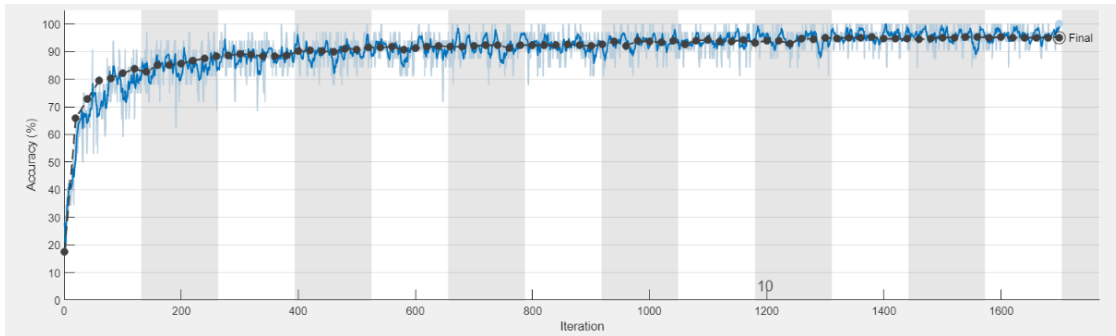


Figure 4.8: Training and Validation Accuracy for the Third Cross-validation Iteration

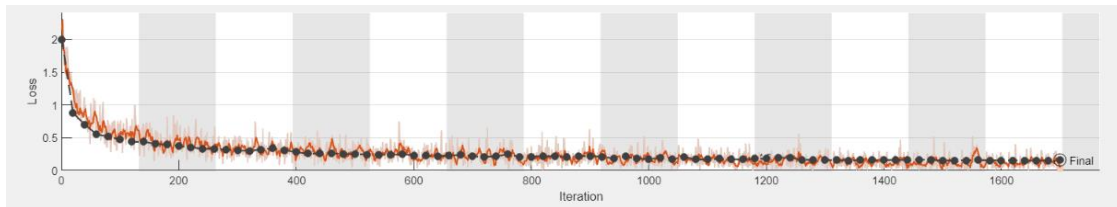


Figure 4.9: Training and Validation Loss for the Third Cross-validation Iteration

4.2.3.2 Confusion Matrix

The confusion matrix for the third cross-validation iteration are displayed in Table 4.12.

Table 4.12: The Confusion Matrix (where G, M, P and N refer to Glioma, Meningioma, Pituitary Tumour and Normal, respectively) for the Third Cross-validation Iteration

| | | Predicted | | | |
|--------|---|-----------|-----|-----|-----|
| | | G | M | P | N |
| Actual | G | 310 | 13 | 1 | 0 |
| | M | 15 | 290 | 17 | 7 |
| | P | 0 | 7 | 345 | 0 |
| | N | 2 | 2 | 3 | 393 |

The model correctly predicted 310 out of 324 samples from the Glioma class. There were 13 misclassifications as Meningioma, 1 misclassification as Pituitary Tumour, and no misclassification as Normal.

Out of 329 samples from the Meningioma class, the model correctly predicted 290. There were 15 misclassifications as Glioma, 17 misclassifications as Pituitary Tumour, and 7 misclassifications as Normal.

345 out of 352 samples from the Pituitary Tumour class were correctly classified. There were 7 misclassifications as Meningioma but none as Glioma or Normal.

The model correctly predicted 393 out of 400 samples from the Normal class. There were 2 misclassifications as Glioma, 2 misclassifications as Meningioma, and 3 misclassifications as Pituitary Tumour. In the third cross-validation iteration, the number of misclassified images was 67.

Table 4.13: No. of Samples Counts of TP, TN, FP and FN for the Third Cross-validation Iteration

| | Samples | TP | TN | FP | FN |
|---|---------|-----|------|----|----|
| G | 324 | 310 | 1064 | 17 | 14 |
| M | 329 | 290 | 1054 | 22 | 39 |
| P | 352 | 345 | 1032 | 21 | 7 |
| N | 400 | 393 | 998 | 7 | 7 |

4.2.3.3 Model Performance

The class-specific metrics in the third cross-validation iteration are displayed in Table 4.14.

Table 4.14: Class-specific Metrics for the Third Cross-validation Iteration

| | Accuracy | Precision | Recall | Specificity | F1 Score |
|---|----------|-----------|--------|-------------|----------|
| G | 97.7936% | 0.9480 | 0.9568 | 0.9843 | 0.9524 |
| M | 95.6584% | 0.9295 | 0.8815 | 0.9796 | 0.9048 |
| P | 98.0071% | 0.9426 | 0.9801 | 0.9801 | 0.9610 |
| N | 99.0036% | 0.9825 | 0.9825 | 0.9930 | 0.9825 |

The overall performance metrics of the third cross-validation iteration for the Brain Cancer Detection and Classification Model are displayed in Table 4.15 and were calculated using the weighted metrics formulas (Formulas 3.22–3.26).

Table 4.15: Overall Performance Metrics for the Third Cross-validation Iteration

| Metric | Value |
|-------------|----------|
| Accuracy | 97.6916% |
| Precision | 0.9521 |
| Recall | 0.9523 |
| Specificity | 0.9846 |
| F1 Score | 0.9520 |

The model's accuracy of approximately 97.6916% indicates its effectiveness in accurately classifying brain MRI images into their respective categories. This high accuracy demonstrates its strong ability to distinguish between the four classes: Glioma, Meningioma, Pituitary Tumour, and Normal brain scans.

The precision value of 0.9521 suggests that when the model predicts a specific class, it is accurate around 95.21% of the time. This implies a low rate of false positives in the

classification of that particular class, which is crucial for providing reliable information and minimising misclassifications.

The recall metric, also known as sensitivity, of 0.9523, indicates that the model correctly identifies approximately 95.23% of the actual positive instances for a specific class. This high sensitivity ensures that a significant majority of positive cases for that specific class are correctly identified.

The specificity value of 0.9846 suggests that the model accurately identifies around 98.46% of the instances that do not belong to the specific class. This high specificity minimises false positives for the given class and reduces the chances of misclassification.

The F1 score of 0.9520 represents the harmonic mean of precision and recall for the specific class. It provides a comprehensive measure of the model's performance in accurately classifying instances belonging to that specific class, taking into account both precision and recall.

4.2.4 Fourth Cross-validation Iteration

In the fourth cross-validation iteration as shown in Figure 4.10, the fourth fold was used as the testing set, while the other folds served as the training set. The model was trained on the training set and evaluated on the fourth fold to assess its performance.

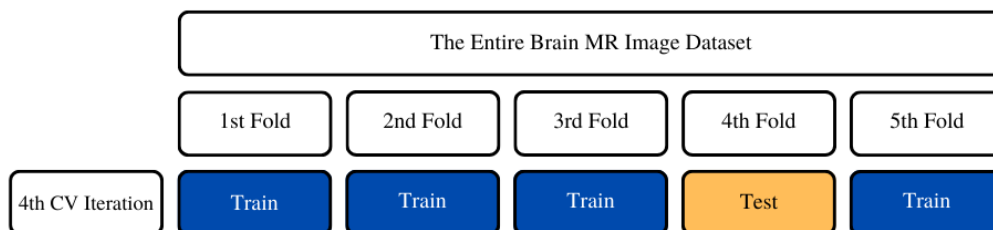


Figure 4.10: Fourth Fold as Testing Set

4.2.4.1 Training and Validation

Table 4.16: The Details of the Training and Validation Process for the Fourth Cross-validation Iteration

| Training | Value |
|----------------------|-----------------------------|
| Completed Epochs | 14 out of 20 Epochs |
| Completed Iterations | 1760 out of 2620 Iterations |
| Training Time | 55 Minutes and 25 Seconds |
| Training Accuracy | 90.6250% |
| Validation Accuracy | 94.7293% |
| Training Loss | 0.1720 |
| Validation Loss | 0.1410 |

The training duration amounted to 55 minutes and 25 seconds, during which 14 epochs were completed out of the overall 20 epochs. Through 1760 iterations, the model steadily progressed. The training and validation accuracies can be observed in Figure 4.11, whereas Figure 4.12 presents the training and validation losses. At the conclusion of training, the model achieved a final training accuracy of 90.6250% and a training loss of 0.1720. The validation accuracy stood at 94.7293%, accompanied by a validation loss of 0.1410, underscoring the model's effectiveness in handling unseen data.

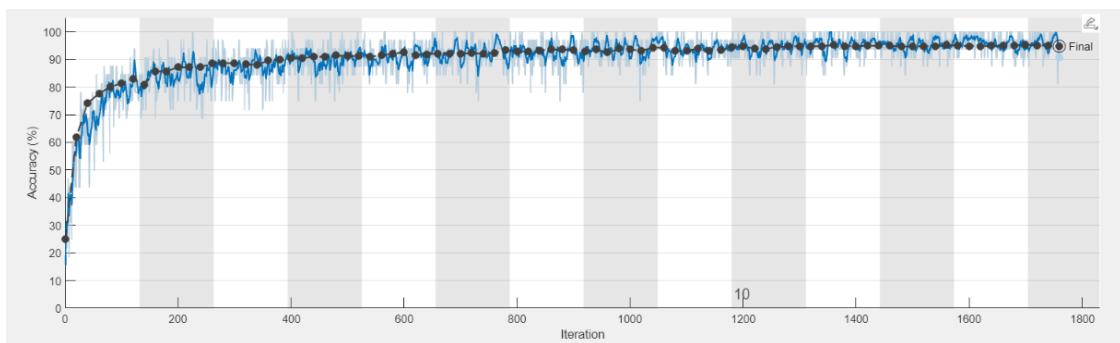


Figure 4.11: Training and Validation Accuracy for the Fourth Cross-validation Iteration

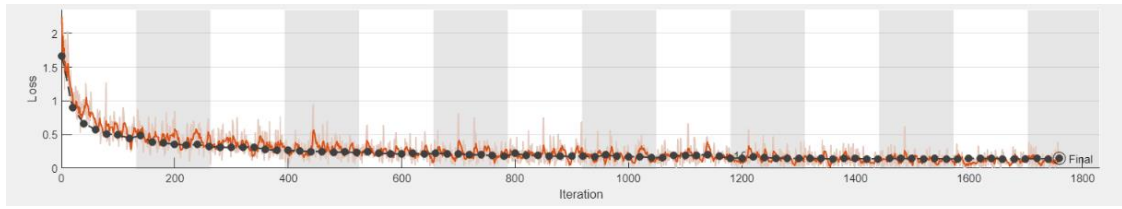


Figure 4.12: Training and Validation Loss for the Fourth Cross-validation Iteration

4.2.4.2 Confusion Matrix

The confusion matrix for the fourth cross-validation iteration are displayed in Table 4.17.

Table 4.17: The Confusion Matrix (where G, M, P and N refer to Glioma, Meningioma, Pituitary Tumour and Normal, respectively) for the Fourth Cross-validation Iteration

| | | Predicted | | | |
|--------|---|-----------|-----|-----|-----|
| | | G | M | P | N |
| Actual | G | 306 | 18 | 1 | 0 |
| | M | 9 | 298 | 16 | 6 |
| | P | 1 | 5 | 345 | 1 |
| | N | 3 | 2 | 3 | 392 |

The model correctly predicted 306 out of 325 samples from the Glioma class. There were 18 misclassifications as Meningioma, 1 misclassification as Pituitary Tumour, and no misclassification as Normal.

Out of 329 samples from the Meningioma class, the model correctly predicted 298. There were 9 misclassifications as Glioma, 16 misclassifications as Pituitary Tumour, and 6 misclassifications as Normal.

345 out of 352 samples from the Pituitary Tumour class were correctly classified. There were 5 misclassifications as Meningioma and 1 misclassification each as Glioma and Normal.

The model correctly predicted 392 out of 400 samples from the Normal class. There were 3 misclassifications as Glioma, 2 misclassifications as Meningioma, and 3

misclassifications as Pituitary Tumour. In the fourth cross-validation iteration, a total of 65 images were misclassified.

Table 4.18: No. of Samples and Counts of TP, TN, FP and FN for the Fourth Cross-validation Iteration

| | Samples | TP | TN | FP | FN |
|---|---------|-----|------|----|----|
| G | 325 | 306 | 1068 | 13 | 19 |
| M | 329 | 298 | 1052 | 25 | 31 |
| P | 352 | 345 | 1034 | 20 | 7 |
| N | 400 | 392 | 999 | 7 | 8 |

4.2.4.3 Model Performance

The class-specific metrics in the fourth cross-validation iteration are displayed in Table 4.19.

Table 4.19: Class-specific Metrics for the Fourth Cross-validation Iteration

| | Accuracy | Precision | Recall | Specificity | F1 Score |
|---|----------|-----------|--------|-------------|----------|
| G | 97.7240% | 0.9592 | 0.9415 | 0.9880 | 0.9503 |
| M | 96.0171% | 0.9226 | 0.9058 | 0.9768 | 0.9141 |
| P | 98.0797% | 0.9452 | 0.9801 | 0.9810 | 0.9623 |
| N | 98.9331% | 0.9825 | 0.9800 | 0.9930 | 0.9812 |

The overall performance metrics of the fourth cross-validation iteration for the Brain Cancer Detection and Classification Model are displayed in Table 4.20 and were calculated using the weighted metrics formulas (Formulas 3.22–3.26).

Table 4.20: Overall Performance Metrics for the Fourth Cross-validation Iteration

| Metric | Value |
|-------------|----------|
| Accuracy | 97.7576% |
| Precision | 0.9538 |
| Recall | 0.9538 |
| Specificity | 0.9851 |
| F1 Score | 0.9536 |

The model achieves an accuracy of approximately 97.7576%, indicating its strong performance in accurately classifying brain MRI images into their respective categories. This high accuracy demonstrates its ability to effectively distinguish between the four classes: Glioma, Meningioma, Pituitary Tumour, and Normal brain scans.

With a precision value of 0.9538, the model accurately predicts a specific class around 95.38% of the time. This suggests a low rate of false positives in the classification of that particular class, contributing to reliable results and minimising misclassifications.

The recall metric, also known as sensitivity, is measured at 0.9538, indicating that the model correctly identifies approximately 95.38% of the actual positive instances for a specific class. This high recall ensures that a significant majority of positive cases for that specific class are correctly identified, providing valuable information for medical professionals.

The specificity value of 0.9851 suggests that the model accurately identifies around 98.51% of the instances that do not belong to the specific class. This high specificity minimises false positives for the given class, reducing the chances of misclassification and providing reliable results.

The F1 score, calculated at 0.9536, represents the harmonic mean of precision and recall for the specific class. It provides a comprehensive measure of the model's performance in accurately classifying instances belonging to that specific class, considering both precision and recall simultaneously.

4.2.5 Fifth Cross-validation Iteration

In the fifth cross-validation iteration as shown in Figure 4.13, the fifth fold was used as the testing set, while the other folds served as the training set. The model was trained on the training set and evaluated on the fifth fold to assess its performance.

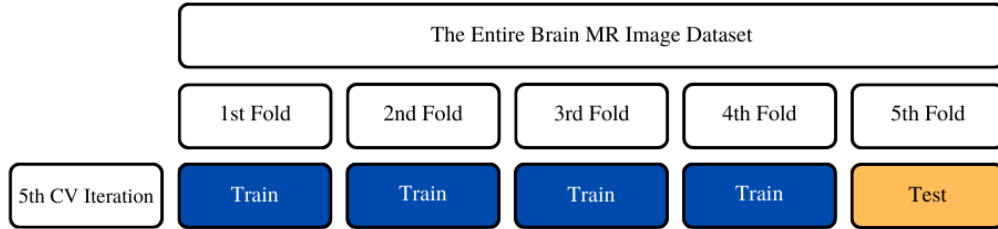


Figure 4.13: Fifth Fold as Testing Set

4.2.5.1 Training and Validation

Table 4.21: The Details of the Training and Validation Process for the 5th Cross-validation Iteration

| Training | Value |
|----------------------|-----------------------------|
| Completed Epochs | 15 out of 20 Epochs |
| Completed Iterations | 1900 out of 2620 Iterations |
| Training Time | 60 Minutes and 40 Seconds |
| Training Accuracy | 90.6250% |
| Validation Accuracy | 94.7293% |
| Training Loss | 0.1648 |
| Validation Loss | 0.1465 |

Over the course of 60 minutes and 40 seconds, the training process unfolded, encompassing 15 epochs out of the total 20 epochs. The model underwent 1900 iterations out of 2620. Figure 4.14 provides a visualisation of the training and validation accuracies, while Figure 4.15 displays the training and validation losses. The training process resulted in a final training accuracy of 90.6250% and a training loss of 0.1648. The validation accuracy achieved 94.7293%, along with a validation loss of 0.1465, showcasing the model's performance on previously unseen data.

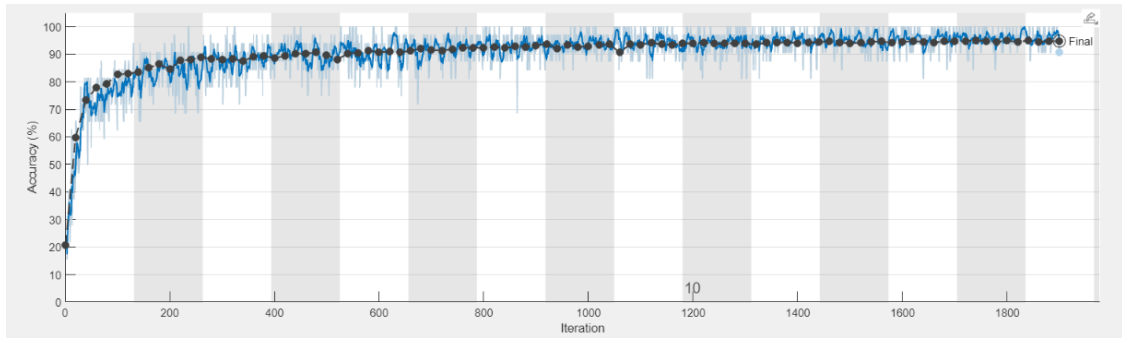


Figure 4.14: Training and Validation Accuracy for the Fifth Cross-validation Iteration

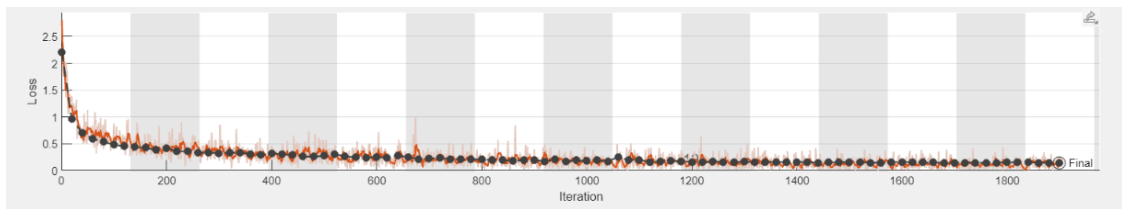


Figure 4.15: Training and Validation Loss for the Fifth Cross-validation Iteration

4.2.5.2 Confusion Matrix

The confusion matrix for the fifth cross-validation iteration are displayed in Table 4.22.

Table 4.22: The Confusion Matrix (where G, M, P and N refer to Glioma, Meningioma, Pituitary Tumour and Normal, respectively) for the Fifth Cross-validation Iteration

| | | Predicted | | | |
|--------|---|-----------|-----|-----|-----|
| | | G | M | P | N |
| Actual | G | 313 | 10 | 0 | 1 |
| | M | 16 | 299 | 0 | 14 |
| | P | 6 | 4 | 340 | 1 |
| | N | 2 | 7 | 3 | 388 |

Out of a total of 324 samples from the Glioma class, the model correctly predicted 313 of them. There were 10 misclassifications as Meningioma and 1 misclassification as Pituitary Tumour, but no misclassification as Normal.

For the 329 samples from the Meningioma class, the model correctly predicted 299 of them. There were 16 misclassifications as Glioma, 14 misclassifications as Pituitary Tumour, but no misclassification as Normal.

340 out of 351 samples from the Pituitary Tumour class were correctly classified. There were 6 misclassifications as Glioma, 4 misclassifications as Meningioma and 1 misclassification as Normal.

Out of the 400 samples from the Normal class, the model correctly predicted 388 of them. There were 2 misclassifications where samples from the Normal class were wrongly predicted as Glioma, 7 misclassifications as Meningioma, and 3 misclassifications as Pituitary Tumour. The fifth iteration resulted in 64 misclassified images.

Table 4.23: No. of Samples and Counts of TP, TN, FP and FN for the Fifth Cross-validation Iteration

| | Samples | TP | TN | FP | FN |
|---|---------|-----|------|----|----|
| G | 324 | 313 | 1056 | 24 | 11 |
| M | 329 | 299 | 1054 | 21 | 30 |
| P | 351 | 340 | 1035 | 18 | 11 |
| N | 400 | 388 | 1003 | 1 | 12 |

4.2.5.3 Model Performance

The class-specific metrics in the fifth cross-validation iteration are displayed in Table 4.24.

Table 4.24: Class-specific Metrics in the Fifth Cross-validation Iteration

| | Accuracy | Precision | Recall | Specificity | F1 Score |
|---|----------|-----------|--------|-------------|----------|
| G | 97.5071% | 0.9288 | 0.9660 | 0.9778 | 0.9470 |
| M | 96.3675% | 0.9344 | 0.9088 | 0.9805 | 0.9214 |
| P | 97.9345% | 0.9497 | 0.9687 | 0.9829 | 0.9591 |
| N | 99.0741% | 0.9974 | 0.9700 | 0.9990 | 0.9835 |

The overall performance metrics of the fifth cross-validation iteration for the Brain Cancer Detection and Classification Model are displayed in Table 4.25 and were calculated using the weighted metrics formulas (Formulas 3.22–3.26).

Table 4.25: Overall Performance Metrics for the Fifth Cross-validation Iteration

| Metric | Value |
|-------------|----------|
| Accuracy | 97.7933% |
| Precision | 0.9549 |
| Recall | 0.9544 |
| Specificity | 0.9857 |
| F1 Score | 0.9544 |

The model achieves an impressive accuracy of approximately 97.7933%, indicating its strong performance in accurately classifying brain MRI images across the four classes: Glioma, Meningioma, Pituitary Tumour, and Normal brain scans. This high accuracy demonstrates its effectiveness in distinguishing between these classes and providing reliable results.

With a precision value of 0.9549, the model accurately predicts a specific class around 95.49% of the time. This indicates a low rate of false positives in the classification of that particular class, which is essential for minimising misclassifications and providing reliable information for medical professionals.

The recall metric, measured at 0.9544, indicates that the model correctly identifies approximately 95.44% of the actual positive instances for a specific class. This high recall ensures that a significant majority of positive cases for that specific class are correctly identified, contributing to effective diagnosis and treatment planning.

The specificity value of 0.9857 suggests that the model accurately identifies around 98.57% of the instances that do not belong to the specific class. This high specificity reduces the chances of false positives for the given class and helps minimise misclassifications.

The F1 score, computed at 0.9544, represents the harmonic mean of precision and recall for the specific class. It provides a comprehensive measure of the model's performance in accurately classifying instances belonging to that specific class, considering both precision and recall simultaneously.

4.2.6 Average

4.2.6.1 Training and Validation Average

The averages of the training accuracy, training loss, validation accuracy, and validation loss for all five cross-validation iterations are displayed in Table 4.26.

Table 4.26: Training and Validation Results

| | Training | | Validation | |
|---------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| | Accuracy | Loss | Accuracy | Loss |
| 1 st CV Iteration | 93.7500% | 0.2520 | 94.5869% | 0.1469 |
| 2 nd CV Iteration | 96.8750% | 0.1252 | 95.5160% | 0.1380 |
| 3 rd CV Iteration | 100.0000% | 0.0395 | 95.0925% | 0.1596 |
| 4 th CV Iteration | 90.6250% | 0.1720 | 94.7293% | 0.1410 |
| 5 th CV Iteration | 90.6250% | 0.1648 | 94.7293% | 0.1465 |
| Mean | 94.3750% | 0.1507 | 94.9308% | 0.1464 |
| Variance | 1.6602 $\times 10^{-3}$ | 5.9825 $\times 10^{-3}$ | 1.4202 $\times 10^{-5}$ | 6.8555 $\times 10^{-5}$ |
| Standard Deviation | 0.0407 | 0.0773 | 0.0038 | 0.0083 |

The final training accuracy, training loss, validation accuracy, and validation loss averages were $94.3750 \pm 0.0407\%$, 0.1507 ± 0.0773 , $94.9308 \pm 0.0038\%$, and

0.1464±0.0083, respectively. These averages were obtained after completing a 5-fold cross-validation in which the models were trained and scored on different subsets of the dataset.

4.2.6.2 Model Performance Average

The class-specific metrics averages are displayed in Table 4.27.

Table 4.27: Class-specific Metrics Average

| | Classes | Accuracy | Precision | Recall | Specificity | F1 Score |
|---------------------------------|---------|----------|-----------|--------|-------------|----------|
| 1 st CV Iteration | G | 97.0798% | 0.9249 | 0.9506 | 0.9769 | 0.9376 |
| | M | 95.1567% | 0.9251 | 0.8632 | 0.9786 | 0.8931 |
| | P | 97.7920% | 0.9255 | 0.9915 | 0.9734 | 0.9574 |
| | N | 98.4330% | 0.9871 | 0.9575 | 0.9950 | 0.9721 |
| 2 nd CV Iteration | G | 96.7236% | 0.9427 | 0.9136 | 0.9833 | 0.9279 |
| | M | 95.6553% | 0.9061 | 0.9088 | 0.9712 | 0.9074 |
| | P | 97.2934% | 0.9218 | 0.9744 | 0.9725 | 0.9474 |
| | N | 99.2165% | 1.0000 | 0.9725 | 1.0000 | 0.9861 |
| 3 rd CV Iteration | G | 97.7936% | 0.9480 | 0.9568 | 0.9843 | 0.9524 |
| | M | 95.6584% | 0.9295 | 0.8815 | 0.9796 | 0.9048 |
| | P | 98.0071% | 0.9426 | 0.9801 | 0.9801 | 0.9610 |
| | N | 99.0036% | 0.9825 | 0.9825 | 0.9930 | 0.9825 |
| 4 th CV Iteration | G | 97.7240% | 0.9592 | 0.9415 | 0.9880 | 0.9503 |
| | M | 96.0171% | 0.9226 | 0.9058 | 0.9768 | 0.9141 |
| | P | 98.0797% | 0.9452 | 0.9801 | 0.9810 | 0.9623 |
| | N | 98.9331% | 0.9825 | 0.9800 | 0.9930 | 0.9812 |
| 5 th CV Iteration | G | 97.5071% | 0.9288 | 0.9660 | 0.9778 | 0.9470 |
| | M | 96.3675% | 0.9344 | 0.9088 | 0.9805 | 0.9214 |
| | P | 97.9345% | 0.9497 | 0.9687 | 0.9829 | 0.9591 |

| | | | | | | |
|-----------------------|---|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| | N | 99.0741% | 0.9974 | 0.9700 | 0.9990 | 0.9835 |
| Average | G | 97.3656% | 0.9407 | 0.9457 | 0.9821 | 0.9430 |
| | M | 95.7710% | 0.9235 | 0.8936 | 0.9773 | 0.9082 |
| | P | 97.8213% | 0.9370 | 0.9790 | 0.9780 | 0.9574 |
| | N | 98.9321% | 0.9899 | 0.9725 | 0.9960 | 0.9811 |
| Variance | G | 2.0638×10^{-1} | 1.9770×10^{-4} | 4.0184×10^{-4} | 2.1653×10^{-5} | 1.0370×10^{-4} |
| | M | 2.0495×10^{-1} | 1.1523×10^{-4} | 4.2037×10^{-4} | 1.3668×10^{-5} | 1.1231×10^{-4} |
| | P | 9.8411×10^{-2} | 1.5579×10^{-4} | 7.1478×10^{-5} | 2.2207×10^{-5} | 3.4963×10^{-5} |
| | N | 8.8815×10^{-2} | 6.8905×10^{-5} | 9.6875×10^{-5} | 1.1000×10^{-5} | 2.8432×10^{-5} |
| Standard Deviation | G | 0.4543 | 0.0141 | 0.0200 | 0.0047 | 0.0102 |
| | M | 0.4527 | 0.0107 | 0.0205 | 0.0037 | 0.0106 |
| | P | 0.3137 | 0.0125 | 0.0085 | 0.0047 | 0.0059 |
| | N | 0.2980 | 0.0083 | 0.0098 | 0.0033 | 0.0053 |

The averages of the overall performance metrics across the 5-fold cross-validation iterations of the Brain Cancer Detection and Classification Model can be seen in Table 4.28.

Table 4.28: Averages of Overall Performance Metrics across the 5-fold Cross-validation Iterations

| | Accuracy | Precision | Recall | Specificity | F1 Score |
|---------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| 1 st CV Iteration | 97.1927% | 0.9428 | 0.9423 | 0.9816 | 0.9419 |
| 2 nd CV Iteration | 97.3260% | 0.9452 | 0.9444 | 0.9825 | 0.9445 |
| 3 rd CV Iteration | 97.6916% | 0.9521 | 0.9523 | 0.9846 | 0.9520 |
| 4 th CV Iteration | 97.7576% | 0.9538 | 0.9538 | 0.9851 | 0.9536 |
| 5 th CV Iteration | 97.7933% | 0.9549 | 0.9544 | 0.9857 | 0.9544 |
| Average | 97.5522% | 0.9498 | 0.9494 | 0.9839 | 0.9493 |
| Variance | 7.5039×10^{-2} | 2.9363×10^{-5} | 3.2043×10^{-5} | 3.1050×10^{-6} | 3.2397×10^{-5} |
| Standard Deviation | 0.2739 | 0.0054 | 0.0057 | 0.0018 | 0.0057 |

The accuracy consistently remained high throughout all cross-validation iterations, ranging from 97.1927% to 97.7933%. This indicates the model's robustness in accurately classifying brain MRI images across different folds. The accuracy average of the brain cancer detection and classification model is $97.5522 \pm 0.0027\%$.

The precision values showed an increasing trend from the first to the fifth cross-validation iteration. Starting at 0.9428, it reached a peak value of 0.9549 in the final iteration. This suggests that the model's ability to accurately predict a specific class improved with each iteration. The precision average of the brain cancer detection and classification model is 0.9498 ± 0.0054 .

Similar to precision, the recall values also displayed an increasing pattern across the cross-validation iterations. Starting at 0.9423 in the first iteration, it gradually rose to 0.9544 in the fifth iteration. This indicates that the model's sensitivity to identifying positive instances for each class improved over time. The recall average of the brain cancer detection and classification model is 0.9839 ± 0.0018 .

The specificity values remained consistently high across all cross-validation iterations, ranging from 0.9816 to 0.9857. This indicates the model's ability to accurately identify negative instances (non-class) and avoid false positives. The specificity average of the brain cancer detection and classification model is 0.9498 ± 0.0054 .

The F1 score, which is a measure of the model's overall performance, showed a similar increasing trend like precision and recall. It started at 0.9419 and reached 0.9544 in the final iteration, reflecting the model's improved balance between precision and recall over the course of the iterations. The F1 score average of the brain cancer detection and classification model is 0.9493 ± 0.0057 .

By analysing these patterns, the model's consistent performance and improvement in accurately classifying brain MRI images across the 5-fold cross-validation process can be observed.

Table 4.29 shows the misclassifications of brain MRI across five iterations. Notably, the Meningioma class demonstrated a significantly higher rate of misclassifications compared to the other three classes: Glioma, Pituitary Tumour and Normal.

Table 4.29: Misclassifications for Each Iteration

| | Classes | No. of Misclassifications | Total Misclassifications |
|------------------------------|---------|------------------------------|-----------------------------|
| 1 st CV Iteration | G | 16 | 81 out of 1404 images |
| | M | 45 | |
| | P | 3 | |
| | N | 17 | |
| 2 nd CV Iteration | G | 28 | |
| | M | 30 | |

| | | | |
|------------------------------|---|------|------------------|
| | P | 9 | 78 out of 1404 |
| | N | 11 | images |
| <hr/> | | | |
| 3 rd CV Iteration | G | 14 | 67 out of 1405 |
| | M | 39 | images |
| | P | 7 | |
| | N | 7 | |
| <hr/> | | | |
| 4 th CV Iteration | G | 19 | 65 out of 1406 |
| | M | 31 | images |
| | P | 7 | |
| | N | 8 | |
| <hr/> | | | |
| 5 th CV Iteration | G | 11 | 64 out of 1404 |
| | M | 30 | images |
| | P | 11 | |
| | N | 12 | |
| <hr/> | | | |
| Average | G | 17.6 | 71 out of 1404.6 |
| | M | 35.0 | images |
| | P | 7.4 | |
| | N | 11.0 | |
| <hr/> | | | |

4.3 Analysis and Discussion

4.3.1 Modifications to Key Settings

To enhance the generalisation of the DL model, several modifications were implemented. Firstly, a validation patience of 10 was employed, allowing for early stopping based on the model's performance on the validation set. This ensured that the model was not trained for an excessive number of epochs, mitigating the risk of overfitting and facilitating better generalisation. Additionally, the initial learning rate was set to 0.001, striking a balance between learning speed and stability. A learning rate drop

factor of 0.1 and a learning rate drop period of 10 were also utilised, periodically reducing the learning rate during training. This strategy facilitated the fine-tuning of model weights, leading to convergence on a more optimal solution and improved generalisation. Furthermore, L2 regularisation with a strength of 0.001 was applied to the model. By penalising large weight values, this regularisation technique encouraged the model to avoid extreme weight configurations that could contribute to overfitting. Incorporating this regularisation term into the loss function fostered the learning of more robust and generalizable features.

4.3.2 GoogleNet Architecture

To optimise the learning process in GoogleNet, the weight and bias learning rate factors were adjusted for different layers of the network. Specifically, for the initial layers up to the inception of the 3b module, the learning rate factors were reduced from 1 to 0.001. This change facilitated a gradual update of weights and biases in these early layers, promoting better convergence and enhancing generalisation. Similarly, for the deeper layers spanning from the inception of the 4a module to the 5b module, the learning rate factors were adjusted from 1 to 0.01. This modification aimed to fine-tune the learning process within these complex layers, enabling the model to effectively capture intricate patterns and higher-level features. By customising the learning rate factors across different layers, the model demonstrated improved performance and generalisation capabilities, as observed in Table 4.30.

In order to compare the performance between the non-fine-tuned GoogleNet and fine-tuned GoogleNet architectures, the folds combination from the first cross-validation iteration, as shown in Figure 4.16, was utilised.

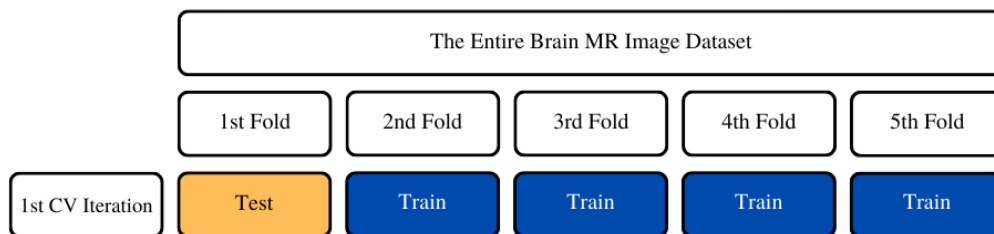


Figure 4.16: The Folds Combination used for the Comparison

To ensure a fair comparison between the non-fine-tuned GoogleNet and fine-tuned GoogleNet architectures, both models were trained using the same set of key parameters, as detailed in Table 4.30. This approach aimed to eliminate any potential biases that could

arise from differences in parameter settings. By maintaining consistency in the selection and configuration of these key parameters, a reliable assessment of the impact of fine-tuning on the model's performance could be achieved.

Table 4.30: Training Settings

| Settings | Values |
|---------------------------|--------|
| Optimizer | Adam |
| Initial Learning Rate | 0.001 |
| Learning Rate Drop Period | 10 |
| Learning Rate Drop Factor | 0.1 |
| Maximum Epochs | 20 |
| Mini-batch Size | 32 |
| Validation Frequency | 20 |
| Validation Patience | 10 |
| L2 Regularisation | 0.001 |

After the training processes for both the non-fine-tuned GoogleNet and fine-tuned GoogleNet models were completed, they were evaluated using the first fold of the dataset. The performance metrics of both are shown in Table 3.1.

Table 4.31: The Differences in Performance between a Non-fine-tuned GoogleNet Architecture and a Fine-tuned GoogleNet Architecture

| Metrics | | Non-fine-tuned GoogleNet Architecture | Fine-tuned GoogleNet Architecture |
|------------|----------|---|---|
| Training | Accuracy | 84.3750% | 93.7500% |
| | Loss | 0.5109 | 0.2520 |
| Validation | Accuracy | 86.8234% | 94.5869% |

| | | | |
|---------|-------------|----------|----------|
| | Loss | 0.4044 | 0.1469 |
| Testing | Accuracy | 93.7322% | 97.3260% |
| | Precision | 0.9436 | 0.9452 |
| | Recall | 0.9290 | 0.9444 |
| | Specificity | 0.9396 | 0.9825 |
| | F1 Score | 0.9362 | 0.9445 |

4.3.3 Interpretation of Classification Metrics

The Brain Cancer Detection and Classification Model consistently achieved high classification accuracies ranging from 97.1927% to 97.7933% across the 5 iterations, indicating its effectiveness in correctly classifying brain MRIs into the 4 classes. The incremental improvement from the 1st to the 5th iteration highlights the robustness of k-fold cross-validation in evaluating model performance. The precision values exhibited minor variations from 0.9428 to 0.9549 but maintained a reasonably consistent trend, demonstrating the model's stable ability to accurately identify positive cases and minimise false positives. Similarly, the recall values displayed consistency, ranging from 0.9423 to 0.9544, reflecting the model's ability to consistently capture a high proportion of actual positive cases and avoid false negatives. The specificity values, ranging from 0.9816 to 0.9857, consistently indicated the model's ability to correctly identify negative cases, ensuring reliable predictions of healthy individuals. The F1 scores, ranging from 0.9419 to 0.9544, showed a consistent pattern, reflecting a balanced trade-off between precision and recall and showcasing the model's reliable performance in accurately classifying brain cancer cases.

4.3.4 Misclassifications as Non-meningioma

In all five cross-validation iterations, it was noticed that the recalls and F1 scores of Meningioma were lower than the recalls and F1 scores of the remaining classes. This is due to Meningioma having significantly higher rates of false negatives (FN), which means a high number of brain MRIs from the Meningioma class were incorrectly classified as non-Meningioma (Glioma, Pituitary Tumour, and Normal). The formulas for recall and F1 score are as follows:

$$Recall = \frac{TP}{TP + FN} \dots \dots \dots (4.1)$$

$$F1\ Score = \frac{2(Precision * Recall)}{Precision + Recall} \dots \dots \dots (4.2)$$

As shown by the formulas, the high false negative rates directly impacted the recalls and F1 scores. The features used by the proposed DL model to distinguish Meningioma cases from other classes might not be as informative as those for the other three classes. It is possible that the extracted features by the model did not capture the specific patterns of meningioma cases effectively, leading to a high rate of false negatives and comparatively lower recall value.

4.4 Summary

This chapter focuses on unveiling the results obtained from the DL model's application to the selected brain MRI dataset. The chapter commences by presenting the achieved results in a systematic and organised manner, employing various graphical representations and tables to provide a clear representation of the model's performance. The obtained metrics, such as testing accuracies, precision, recall, specificity, and F1 scores, were analysed and discussed in detail. Additionally, the chapter delves into the consistencies observed across the iterations, showcasing the stability and reliability of the model's predictions. The chapter also highlights the significance of the confusion matrix in assessing the model's performance. Moreover, the chapter provides a comprehensive analysis of the modifications made to the GoogleNet architecture, key parameters, and techniques during the training stage.

Chapter 5

CONCLUSIONS

5.1 General Conclusions

This Final Year Project presents a robust and accurate system based on DL for brain cancer detection and classification. In the proposed framework, a fine-tuned GoogleNet, a pre-trained DL model, was employed to classify brain MR images into 4 classes: Glioma, Meningioma, Pituitary Tumour, and Normal. The Brain Cancer Detection and Classification Model obtained an accuracy of $97.5522 \pm 0.2739\%$, a precision of 0.9498 ± 0.0054 , a recall of 0.9494 ± 0.0057 , a specificity 0.9839 ± 0.0018 and an F1 Score of 0.9493 ± 0.0057 across the 5-fold cross-validation iterations, demonstrating its effectiveness in accurately classifying brain MRI. The consistencies observed in the accuracies, precision, recall, specificity, and F1 score across multiple iterations in 5-fold cross-validation highlight the stability and reliability of the model. These consistencies indicate that the model consistently performs well, providing reliable and accurate predictions.

All three objectives were successfully accomplished. Objective 1, which aimed to develop a reliable and robust DL-based approach for brain cancer detection and classification, was achieved through the fine-tuning of the proposed DL model and experimental settings. By carefully adjusting parameters such as learning rate, L2 regularisation, and optimisation algorithms like Adam, the model was optimised to achieve higher accuracy and better convergence. This process involved iteratively refining the model's architecture and training hyperparameters to achieve optimal results.

Objective 2, which aimed to investigate the robustness and generalisability of the proposed model, was also met. It is essential to assess its performance across different datasets and ensure its effectiveness in real-world scenarios. By using a 5-fold cross-validation technique, the model's performance was tested on multiple subsets of the data, allowing for a more comprehensive assessment of its generalisability. This analysis

helped identify potential limitations, improve model stability, and provide insights into the model's ability to accurately detect and classify brain cancers in diverse cases.

Objective 3, which aimed to evaluate and compare the performance of the proposed model to state-of-the-art methods, was achieved through conducting a thorough performance evaluation and comparing the proposed model to state-of-the-art methods. It allows for an objective assessment of its effectiveness. By benchmarking against existing approaches, the strengths and weaknesses of the developed deep learning model can be identified. Performance metrics such as accuracy, precision, recall, specificity, and F1 score were computed to quantitatively measure the model's performance. This evaluation provided insights into the model's competitiveness, potential advantages, and areas for improvement in comparison to other advanced techniques for brain cancer detection and classification.

Building an accurate DL model requires a continuous cycle of experimentation, as there is no one-size-fits-all approach that guarantees optimal performance across all scenarios. Due to the complexity and the diverse nature of data and architectures, different architectures and techniques may be required to achieve the best results. Each dataset has its own unique characteristics, such as size, distribution, noise levels, and class imbalance. These variations can significantly impact the choice of model architecture, hyperparameters, and regularisation techniques. For instance, a dataset with a high class imbalance might benefit from techniques like oversampling or class weighting to address the imbalance issue, while a dataset with noisy data might require more robust regularisation methods.

Another aspect to consider is the availability of computational resources. DL models often require substantial computational power, memory, and time for training and inference. Not all individuals have access to high-performance hardware, which can limit the choice of architectures and optimisation techniques that can be employed. This resource constraint further highlights the need for customised approaches that strike a balance between model complexity and computational feasibility.

5.2 Limitations

The main challenge faced in building a DL model for brain cancer detection and classification is the imbalanced nature of the dataset used in this project. With imbalanced

data, the model can become biased towards the majority class (the Normal class), as it has more samples for learning. This bias can result in minority classes being overlooked or misclassified, leading to lower accuracy and recall for these classes. The limited number of samples available for the minority classes can hinder the model's ability to learn their distinguishing features, resulting in higher misclassification rates. Therefore, traditional evaluation metrics may not provide an accurate assessment of the model's performance. The class weighting technique, which was used in this project, may be necessary to mitigate the impact of class imbalance. It is important to consider these limitations when interpreting the results and drawing conclusions based on the model's performance.

The computational requirements involved in the training and classification stages was also an obstacle. DL models demand significant computational resources, such as powerful GPUs or specialised hardware accelerators, to handle the large datasets and complex computations required. The computational cost also presents scalability issues. As models become larger and more intricate, the computational demands grow exponentially, making it challenging to train and implement extensive models on standard hardware setups. This limitation hampers the widespread adoption of DL techniques for brain cancer detection and classification, especially in places with limited resources. To overcome these limitations, advancements in both hardware technology and algorithmic optimisation are needed. Developing more efficient model architectures, designing algorithms that reduce computational complexity, and exploring techniques like model compression and quantization can help mitigate the computational burden. Additionally, leveraging cloud computing and distributed computing frameworks can provide scalable resources, making deep learning more accessible and practical for brain tumour detection and classification tasks.

5.3 Future Scopes

Ensemble learning is an ML technique that involves combining multiple models to make more accurate predictions than individual models. Each model in the ensemble is trained separately, and its predictions are aggregated to produce the final prediction. Ensemble methods leverage the diversity of multiple models to overcome limitations and biases, harnessing the concept of "wisdom of the crowds." It encompasses various forms such as bagging, boosting, and stacking. In the context of brain tumour detection and

classification, incorporating ensemble learning can be a promising area for future research. By combining predictions from different DL models trained with diverse architectures or initializations, the overall accuracy and generalisation of the model can be further improved. Utilising ensemble learning techniques opens up new possibilities for projects exploring the benefits of model combination in brain tumour detection and classification. This approach can enhance prediction robustness, handle uncertain cases more effectively, and elevate overall system accuracy.

Online learning is a machine learning paradigm where the model continuously learns and updates its knowledge as new data becomes available in real-time. It is particularly valuable in the context of brain tumour detection and classification, enabling real-time analysis of new patient data. Real-time detection and classification of brain tumours are crucial in clinical settings, where timely diagnoses are essential for prompt treatment decisions. Future research can focus on developing algorithms and architectures that can handle the computational challenges of online learning in real-time scenarios. However, it is important to consider the ethical implications and potential risks associated with real-time analysis of patient data. Safeguarding data privacy and security, as well as establishing robust protocols for validation and verification of online learning models, are crucial considerations for future research in this area.

REFERENCES

- [1] “The Brain of a Spiteful Person | HuffPost Latest News.” https://www.huffpost.com/entry/the-brain-of-a-spiteful-person_b_3292298 (accessed Jan. 18, 2023).
- [2] “Staging and Prognosis for Brain Cancer | Cancer Council NSW.” <https://www.cancercouncil.com.au/brain-cancer/diagnosis/staging-prognosis/> (accessed Jan. 18, 2023).
- [3] “AI vs. ML vs. DL: What’s the Difference.” <https://serokell.io/blog/ai-ml-dl-difference> (accessed Jan. 18, 2023).
- [4] “Brain Tumor MRI Dataset | Kaggle.” <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset?select=Training> (accessed Jan. 13, 2023).
- [5] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2019,” *CA Cancer J Clin*, vol. 69, no. 1, pp. 7–34, Jan. 2019, doi: 10.3322/caac.21551.
- [6] H. Salehiniya and K. K. Farmanfarma, “BRAIN CANCER IN THE WORLD: AN EPIDEMIOLOGICAL REVIEW,” 2020. [Online]. Available: <https://www.researchgate.net/publication/340088881>
- [7] A. M. Saad, A. M. Elmatboly, M. M. Gad, M. J. Al-Husseini, K. A. Jazieh, M. A. Alzuabi, and A. Samir Alfaar, “Association of Brain Cancer with Risk of Suicide,” *JAMA Netw Open*, vol. 3, no. 5, p. e203862, May 2020, doi: 10.1001/jamanetworkopen.2020.3862.
- [8] S. Makaju, P. W. C. Prasad, A. Alsadoon, A. K. Singh, and A. Elchouemi, “Lung Cancer Detection using CT Scan Images,” in *Procedia Computer Science*, Elsevier B.V., 2018, pp. 107–114. doi: 10.1016/j.procs.2017.12.016.
- [9] Z. Tasnim, S. Chakraborty, F. M. J. M. Shamrat, A. N. Chowdhury, H. A. Nuha, A. Karim, S. B. Zahir, and M. M. Billah, “Deep Learning Predictive Model for Colon Cancer Patient using CNN-based Classification,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 8, pp. 687–696, 2021, doi: 10.14569/IJACSA.2021.0120880.
- [10] R. Smith-Bindman, J. Lipson, R. Marcus, K.P Kim, Mahesh, M., R. Gould, A. Berrington, and D. L. Miglioretti, PhD, “Radiation Dose Associated with Common

Computed Tomography Examinations and the Associated Lifetime Attributable Risk of Cancer.”

- [11] Dong, Q., R. C. Welsh, T. L. Chenevert, R. C. Carlos, MS, P. Maly-Sundgren, D. M. Gomez-Hassan, and Mukherji, S. K., MD, “Clinical Applications of Diffusion Tensor Imaging,” *Journal of Magnetic Resonance Imaging*, vol. 19, no. 1. pp. 6–18, Jan. 2004. doi: 10.1002/jmri.10424.
- [12] P. G. Rajan and C. Sundar, “Brain Tumor Detection and Segmentation by Intensity Adjustment,” *J Med Syst*, vol. 43, no. 8, Aug. 2019, doi: 10.1007/s10916-019-1368-4.
- [13] “brain tumor dataset.” https://figshare.com/articles/dataset/brain_tumor_dataset/1512427?file=7953679 (accessed Jan. 13, 2023).
- [14] “Brain Tumor Classification (MRI) | Kaggle.” <https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri> (accessed Jan. 13, 2023).
- [15] “Br35H :: Brain Tumor Detection 2020 | Kaggle.” <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection> (accessed Jan. 13, 2023).
- [16] Abiwinanda, N., M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, “Brain tumor classification using convolutional neural network,” in *IFMBE Proceedings*, Springer Verlag, 2019, pp. 183–189. doi: 10.1007/978-981-10-9035-6_33.
- [17] A. Çinar and M. Yildirim, “Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture,” *Med Hypotheses*, vol. 139, Jun. 2020, doi: 10.1016/j.mehy.2020.109684.
- [18] M. Alfonse and A.-B. M. Salem, “An Automatic Classification of Brain Tumors through MRI Using Support Vector Machine,” 2016.
- [19] Reema, M.A., Babu, A.P., and Thara, N. K., “Brain Tumor Segmentation And Classification using DWT, Gabour Wavelet And GLCM,” *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT)*. IEEE, 2017.

- [20] A. Ari and D. Hanbay, “Deep learning-based brain tumor classification and detection system,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 26, no. 5, pp. 2275–2286, 2018, doi: 10.3906/elk-1801-8.
- [21] Preethi, S. and Aishwarya, P., “Combining wavelet texture features and deep neural network for tumor detection and segmentation over MRI,” *Journal of Intelligent Systems*, vol. 28, no. 4, pp. 571–588, 2021, doi: 10.1515/jisys-2017-0090.
- [22] N. B. Bahadure, A. K. Ray, and H. P. Thethi, “Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM,” *Int J Biomed Imaging*, vol. 2017, 2017, doi: 10.1155/2017/9749108.
- [23] Arunachalam, M. and Savarimuthu, S. R., “An efficient and automatic glioblastoma brain tumor detection using shift-invariant shearlet transform and neural networks,” *Int J Imaging Syst Technol*, vol. 27, no. 3, pp. 216–226, Sep. 2017, doi: 10.1002/ima.22227.
- [24] Deepak, S. and P. M. Ameer, “Brain tumor classification using deep CNN features via transfer learning,” *Comput Biol Med*, vol. 111, Aug. 2019, doi: 10.1016/j.combiomed.2019.103345.
- [25] Swati, Z. N. K., Q. H. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. F. Lu, “Brain tumor classification for MR images using transfer learning and fine-tuning,” *Computerized Medical Imaging and Graphics*, vol. 75, pp. 34–46, Jul. 2019, doi: 10.1016/j.compmedimag.2019.05.001.
- [26] F. J. Díaz-Pernas, M. Martínez-Zarzuela, D. González-Ortega, and M. Antón-Rodríguez, “A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network,” *Healthcare (Switzerland)*, vol. 9, no. 2, Feb. 2021, doi: 10.3390/healthcare9020153.
- [27] J. Kang, Z. Ullah, and J. Gwak, “Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers,” *Sensors*, vol. 21, no. 6, pp. 1–21, Mar. 2021, doi: 10.3390/s21062222.
- [28] S. Khawaldeh, U. Pervaiz, A. Rafiq, and R. S. Alkhawaldeh, “Noninvasive grading of glioma tumor using magnetic resonance imaging with convolutional neural

- networks,” *Applied Sciences (Switzerland)*, vol. 8, no. 1, Dec. 2017, doi: 10.3390/app8010027.
- [29] P. Saxena, A. Maheshwari, S. Tayal, and S. Maheshwari, “Predictive modeling of brain tumor: A Deep learning approach,” Nov. 2019, [Online]. Available: <http://arxiv.org/abs/1911.02265>
- [30] S. A. Abdelaziz Ismael, A. Mohammed, and H. Hefny, “An enhanced deep learning approach for brain cancer MRI images classification using residual networks,” *Artif Intell Med*, vol. 102, Jan. 2020, doi: 10.1016/j.artmed.2019.101779.
- [31] Siva Raja, P. M. and Rani, A. V., “Brain tumor classification using a hybrid deep autoencoder with Bayesian fuzzy clustering-based segmentation approach,” *Biocybern Biomed Eng*, vol. 40, no. 1, pp. 440–453, Jan. 2020, doi: 10.1016/j.bbe.2020.01.006.
- [32] Hemanth, D. J., Anitha, J., Naaji, A., O. Geman, D. E. Popescu, and L. Hoang Son, “A Modified Deep Convolutional Neural Network for Abnormal Brain Image Classification,” *IEEE Access*, vol. 7, pp. 4275–4283, 2019, doi: 10.1109/ACCESS.2018.2885639.
- [33] J. S. Paul, A. J. Plassard, B. A. Landman, and D. Fabbri, “Deep learning for brain tumor classification,” in *Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging*, SPIE, Mar. 2017, p. 1013710. doi: 10.1117/12.2254195.
- [34] N. Varuna Shree and T. N. R. Kumar, “Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network,” *Brain Inform*, vol. 5, no. 1, pp. 23–30, Mar. 2018, doi: 10.1007/s40708-017-0075-5.
- [35] B. Ural, “A Computer-Based Brain Tumor Detection Approach with Advanced Image Processing and Probabilistic Neural Network Methods,” *J Med Biol Eng*, vol. 38, no. 6, pp. 867–879, Dec. 2018, doi: 10.1007/s40846-017-0353-y.

- [36] Rajaguru, H., Ganesan, K., and Bojan, V. K., “Earlier detection of cancer regions from MR image features and SVM classifiers,” *Int J Imaging Syst Technol*, vol. 26, no. 3, pp. 196–208, Sep. 2016, doi: 10.1002/ima.22177.
- [37] M. Soltaninejad, Y. Guang, T. Lambrou, N. Allinson, T. L. Jones, T. R. Barrick, F. A. Howe, and X. J. Ye, “Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI,” *Int J Comput Assist Radiol Surg*, vol. 12, no. 2, pp. 183–203, Feb. 2017, doi: 10.1007/s11548-016-1483-3.

APPENDIX A

EXAMPLES OF BRAIN MRI IN THE DATASET

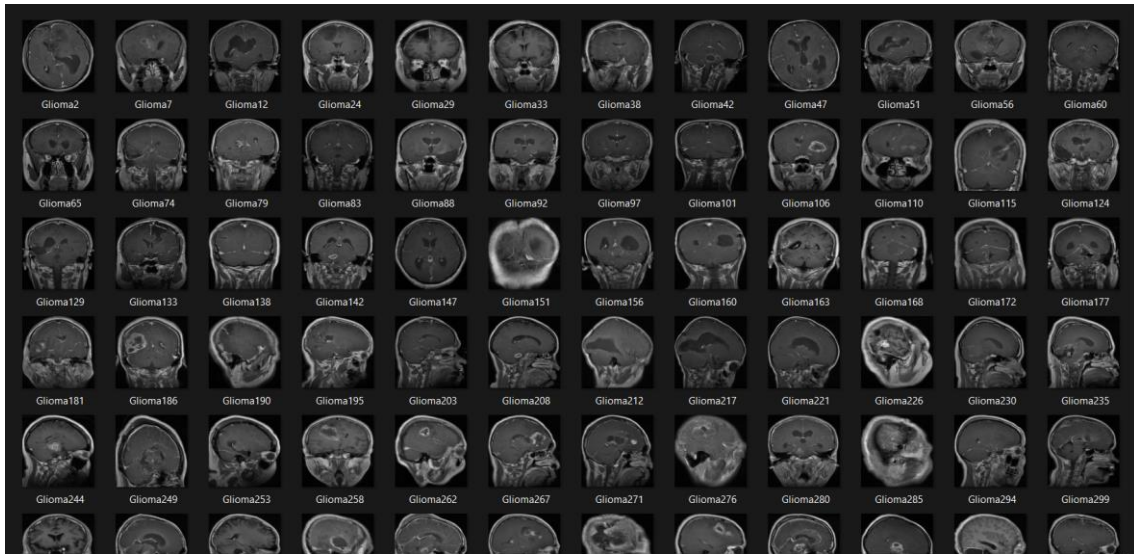


Figure A.1: Glioma Brain MRI

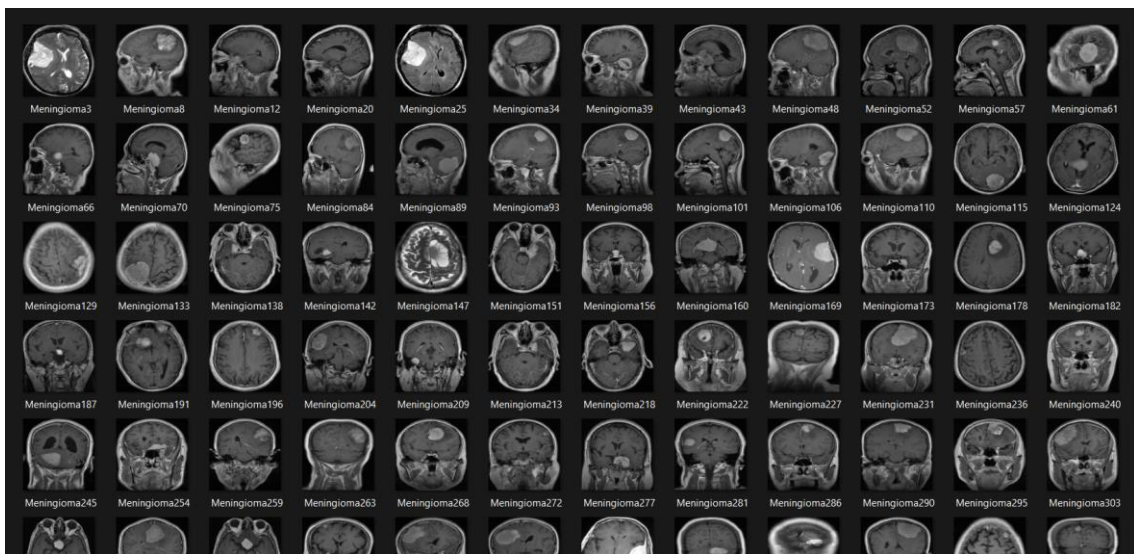


Figure A.2: Meningioma Brain MRI

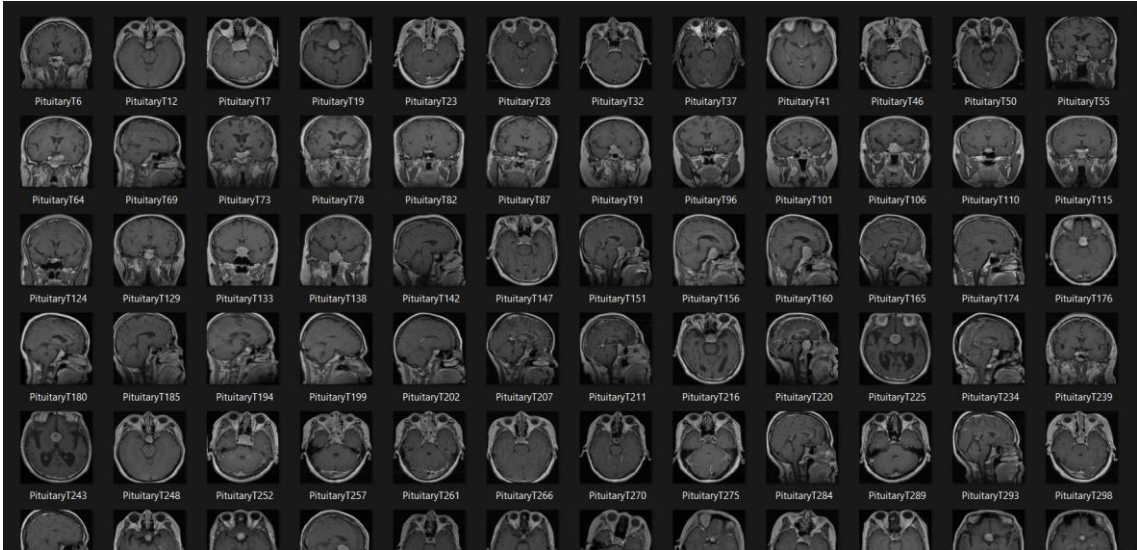


Figure A.3: Pituitary Tumour Brain MRI



Figure A.4: Normal Brain MRI

APPENDIX B

CROPPING VIA MATLAB IMAGE BATCH PROCESSOR

The code snippet used to crop the black region around the brain MRI.

```
function results = myimfcn_cropping(im)
%Image Processing Function
%
% IM      - Input image.
% RESULTS - A scalar structure with the processing results.
%
%-----
% Auto-generated by imageBatchProcessor App.
%
% When used by the App, this function will be called for every input image
% file automatically. IM contains the input image as a matrix. RESULTS is a
% scalar structure containing the results of this processing function.
%
%-----

% Replace the sample below with your code-----

    % Check if the image is grayscale or RGB, and convert it to grayscale if
    necessary
    if numel(size(im)) == 3

        % Convert the image to grayscale
        grayImg = rgb2gray(im);
    else
        grayImg = im;
    end

    % Apply a median filter to remove noise
    filteredImg = medfilt2(grayImg, [7, 7]);

    % Apply Otsu's thresholding to the filtered image
    binaryImg = imbinarize(filteredImg, 'global');

    % Apply dilation and erosion to the binary image
    se = strel('disk', 3);
    binaryImg = imdilate(binaryImg, se);
    binaryImg = imerode(binaryImg, se);

    % Get the bounding box of the non-zero pixels in the thresholded image
    [r, c] = find(binaryImg);
    x1 = min(c);
    y1 = min(r);
    x2 = max(c);
    y2 = max(r);
    boxWidth = x2 - x1 + 1;
    boxHeight = y2 - y1 + 1;

    % Calculate the coordinates of the top-left corner of the bounding box
```

```

boxTop = y1;
boxLeft = x1;

% Crop the image using the bounding box
results = imcrop(im, [boxLeft, boxTop, boxWidth, boxHeight]);

```

%-----

RESIZING VIA MATLAB IMAGE BATCH PROCESSOR

The code snippet used to resize the brain MRIs to a uniform size (224 x 224 x 3).

```

function results = myimfcn_resizing(im)
%Image Processing Function
%
% IM      - Input image.
% RESULTS - A scalar structure with the processing results.
%
%-----
% Auto-generated by imageBatchProcessor App.
%
% When used by the App, this function will be called for every input image
% file automatically. IM contains the input image as a matrix. RESULTS is a
% scalar structure containing the results of this processing function.
%
%-----

% Replace the sample below with your code-----

% Resize the image to a size that is compatible with the desired output size
resizedImg = imresize(im, [224, 224]);

% Check the dimensions of the loaded image
img_size = size(resizedImg);
if numel(img_size) == 2
    % The image is in grayscale
    results = cat(3, resizedImg, resizedImg, resizedImg);
else
    % The image is in RGB
    results = resizedImg;
end

%-----

```

NORMALISATION VIA MATLAB IMAGE BATCH PROCESSOR

The code snippet used to normalise the brain MRIs.

```

function results = myimfcn_normalisation(im)
%Image Processing Function
%
% IM      - Input image.
% RESULTS - A scalar structure with the processing results.
%

```

```

%-----
% Auto-generated by imageBatchProcessor App.
%
% When used by the App, this function will be called for every input image
% file automatically. IM contains the input image as a matrix. RESULTS is a
% scalar structure containing the results of this processing function.
%
%-----

```

```

% Replace the sample below with your code-----

```

```

% Convert to double precision
im = im2double(im);

% Compute the minimum and maximum pixel intensities
min_intensity = min(im(:));
max_intensity = max(im(:));

% Normalize the pixel intensities using min-max normalization
results = (im - min_intensity) / (max_intensity - min_intensity);

```

```

%-----

```

SPLITTING THE DATASET FOR 5-FOLD CROSS-VALIDATION

The code snippet used to split the dataset into five to prepare for 5-fold cross-validation.

```

%-----

```

```

% Specify the path to the folder containing the images
folder_path = 'insert folder directory';

% Specify the path to the folder where you want to save the divided images
save_folder_path = 'insert folder directory';

% Get a list of all the image files in the folder
image_files = dir(fullfile(folder_path, '*.png'));

% Create 5 subfolders inside the save folder
subfolder_names = {'xyz 1', 'xyz 2', 'xyz 3', 'xyz 4', 'xyz 5'};
for i = 1:5
    subfolder_path = fullfile(save_folder_path, subfolder_names{i});
    if ~exist(subfolder_path, 'dir')
        mkdir(subfolder_path)
    end
end

% Loop through each image file and save it to a separate subfolder
for i = 1:length(image_files)
    % Read the image file
    img = imread(fullfile(folder_path, image_files(i).name));

    % Calculate the subfolder index to save the image
    subfolder_index = mod(i, 5) + 1;

    % Save the image to the corresponding subfolder

```

```

    subfolder_path = fullfile(save_folder_path,
subfolder_names{subfolder_index});
    imwrite(img, fullfile(subfolder_path, image_files(i).name));
end
%-----

```

CONFUSION MATRIX

The code snippet used to generate a confusion matrix (rows: actual labels, columns: predicted labels).

```

%-----
% Load the trained network
net = trainedNetwork;

% Specify the directory containing the testing images
testDir = 'insert folder directory';

% Create an imageDatastore for the testing images
testData = imageDatastore(testDir, 'IncludeSubfolders', true, 'LabelSource',
'foldernames');

% Calculate and display the confusion matrix
confusionMat = confusionmat(testData.Labels, testPreds);
disp('Confusion Matrix:');
disp(confusionMat);
%-----

```