



**Department of Computer Science**

**MSc Data Science and Analytics**

**Academic Year 2023-2024**

*Enhancing Brain Cancer Prognosis: Predictive Modelling for Progression and Survival Rates*

*Adwait Yogesh Kumbhare (2357954)*

A report submitted in partial fulfilment of the requirement for the degree of Master of Science

Brunel University

Department of Computer Science

Uxbridge, Middlesex UB8 3PH

United Kingdom

Tel: +44 (0) 1895 203397

Fax: +44 (0) 1895 251686

## Abstract

The purpose of this study is to identify the methods of using image data in conjunction with quantitative features of patient data, for prognosis of brain cancer progression and survival rates. Based on structural and quantitative tumour characteristics as well as the second set of images-only data, the work employs machine learning algorithms for the prognosis of tumour classifications and overall survival. Two primary models were evaluated: two conventional classifiers, to be precise, a Random Forest classifier and a simple neural network. Moreover, when implementing the Random Forest model, the highest accuracy was considered with a value of 98.41% and proved that it can be used as a good discriminator between tumour and non-tumour classes. This high accuracy is an indication that the model is built to deal with the imbalanced features and make actual, dependable predictions hence the chance that it be used clinically. However, since the performance is less accurate, such objectives uncover the possibility of developing deep learning models and provide guidance for researching this field further. The study also highlights the considerations of data cleaning for missing values, scaling, and feature construction for complete categorical features to obtain accurate output. About main patterns and feature interconnections, it is necessary to note that they were established during EDA and the exploration of feature space with the help of visualization. The use of the CNN model in this research provides an accuracy of 100% which is the highest among all. The use of quantitative and image-based data is a major advancement in the study of brain cancer, it gives a new direction to the classification and growth of the tumour. Further studies could focus on broadening the dataset, fine-tuning the models' complexity, adding more data sources, as well as testing the models in live patient populations. Responsibility and moral issues, including data protection, are also mentioned as important factors for applying predictive modelling in healthcare. The current study is amongst the few initial investigations towards integrating machine learning in oncology with an aim of delivering precise and applicable prognoses for handling brain cancer in the future.

## Acknowledgement

I would like to show my acknowledgement and appreciation to the individuals who have assisted me in the course of this study. I am most grateful to my Prof. Kezhi Wang for the guidance and encouragement they provided me in this course of the study. I also want to thank the staff and faculty of the Brunel University London for their support that ensures the development of such research. I am thankful to my friends and colleagues for discussing and motivating me. Finally, my special thanks to my family for supporting me in my career and encouraging me to study this research.

## Table of Contents

Abstract.....	2
Acknowledgement.....	3
Chapter 1: Introduction.....	6
1.1 Research Background.....	6
1.2 Research Problem.....	7
1.3 Aim and Objectives .....	7
1.3.1 Aim:.....	7
1.3.2 Objectives: .....	7
1.4 Research Significance.....	8
1.5 Structure of Report.....	8
Chapter 2: Literature Review.....	11
2.1 Introduction .....	11
2.2 Technological advancement in brain cancer prediction .....	11
2.3 Predictive Modelling for Brain Cancer Prognosis.....	14
2.4 Challenges Available in Brain Cancer Prediction .....	16
2.5 Summary of Literature.....	18
Chapter 3: Research Methodology.....	19
3.1 Introduction .....	19
3.2 Research Method Used.....	19
3.3 Data Collection [ <a href="https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor/data">https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor/data</a> ]...	20
3.4 Data Analysis .....	20
3.5 Methods and Tools Used.....	21
3.6 Ethical Considerations.....	21
3.7 Limitations.....	22
Chapter 4: Implementation and Results .....	23
4.1 Introduction .....	23
4.2 Data Preparation and Cleaning .....	23
4.3 Exploratory Data Analysis .....	24

4.4 Feature and Target Definition .....	26
4.5 Model Training and Evaluation .....	27
Chapter 5: Discussion and Analysis .....	32
Chapter 6: Conclusion and Recommendations .....	39
6.1 Conclusion .....	39
6.2 Recommendations .....	40
References .....	42

## List of Figures

Figure 1 Libraries Imported .....	23
Figure 2 Dataset Loading .....	24
Figure 3 Descriptive Statistics .....	26
Figure 4 Normalizing Features .....	27
Figure 5 Simple Neural Network Building .....	28
Figure 6 Model Evaluation .....	28
Figure 7 Training Random Forest Model .....	29
Figure 8 Building CNN Model .....	30
Figure 9 Model Evaluation .....	30
Figure 10 Histogram of class distribution .....	33
Figure 11 Accuracy of SNN model .....	34

# Chapter 1: Introduction

## 1.1 Research Background

Brain tumours are a heterogeneous and complex group of diseases. There are many types of brain tumours available such as pituitary, meningiomas, and gliomas. Gliomas are the general types of brain tumours. Gliomas are the general kinds of brain tumours responsible for almost 80% of the brain tumours. Gliomas are still one of the most intricate and lethal types of cancer because of the organ it affects and its functions". Brain cancer patients' outcomes also differ based on the type of tumour, its location, genetic characteristics, and the patient's age or gender. Over the years, different patients' prognoses and survival rates have remained elusive because the progression of brain cancer is challenging to predict and the tumours are diverse in characteristics (Owonikoko et al., 2014). The newer technologies of medical imaging, genomics, and data analysis have conversely unveiled other ways of improving the prognosis by using analytical models.

The regular ways of evaluating the prognosis of brain cancer include clinical evaluation, histological examination and imaging MRI or CT scan. These two approaches although useful are crude in giving an overall plan of the tumour conduct and patient status. The biological nature of tumours and the influence of genetic and epigenetic factors further disregard the homogeneous nature for the purpose of prognosis and therefore researchers are unable to approximate the different progress and survival tendencies (Kaifi, 2023). Therefore, there is an increasing demand for the development of a more complex and advanced model that is able to utilise multiple data inputs to enhance the accuracy of the prediction information provided. Improvement of late in the field of analytics adopted advanced tools such as ML and AI in developing the potential for anticipating the severity of brain cancer and survival. These techniques can provide the opportunity to analyze a large amount of data of different origins, such as genomic, imaging, and clinical data. Applying such algorithms that are able to detect relationships within these data sets, one could create models to describe tumour characteristics, responses to treatment, or prognosis of the patient's condition. For instance, by analyzing various forms of imaging data, it is possible to accurately categorize different types of tumours or, based on the patient's genomic markers, to forecast how a particular person is likely to react to a specific treatment.

### 1.2 Research Problem

Even with advanced medical technology and treatment modalities, brain cancer is considered one of the most challenging cancers to diagnose and treat to this date due to its nature and variability. Currently, most approaches that are used to estimate the progression and survival of diseases are clinical investigations supported by imaging analysis and historical examinations that can offer a discontinuous picture of the tumour behaviour and patients' prognosis (Yousefi et al., 2017). Current methods of prognostication do not fully account for the variation in tumour biology and response to factors and can be further from or closer to reality in terms of how the disease progresses. The challenge is again made significantly more complex by the nature of the data sources that need to be integrated: genomic profiles, medical images, and clinical records, to name just a few. Many current models used to predict phenomena cannot synchronously combine multi-omics data and the latest machine learning algorithms, and has shortcomings in precision and applicability. Therefore, it is crucial to initiate work and build better models to include these disparate sets of data into proactive biomarkers, as well as to improve predictive and prognostic capabilities and yield patient-specific knowledge regarding tumour dynamics and fates. This serves as a rationale for solving this particular problem of putting predictive modelling into practice in the context of brain cancer management for better treatment plans and improved patient survival (Sun et al., 2019). The application of AI in the analysis of the said integrated datasets also enables discovery of new biomarkers and therapeutic targets, which can be useful in precise and effective treatment approaches.

### 1.3 Aim and Objectives

#### 1.3.1 Aim:

The aim of this research study is to identify the complex factors affecting patient outcomes by combining various datasets, such as genomic data and clinical records, from valid sources like cancer registries and electronic health records (EHRs).

#### 1.3.2 Objectives:

- To collect and preprocess a comprehensive dataset of genomic, imaging, and clinical information from reputable sources such as the Cancer Imaging Archive (TCIA) and The Cancer Genome Atlas (TCGA), while assuring data quality and completeness.
- Advanced feature engineering and selection will be used to discover the most significant features for predicting brain cancer outcomes, including Principal Component Analysis (PCA) and imaging biomarker analysis.

- To implement and train a variety of machine learning models, such as Convolutional Neural Networks (CNNs), Random Forests, Gradient Boosting Machines (GBMs), and Long Short-Term Memory (LSTM) networks, using hyperparameter improvement and cross-validation.
- To assess the performance of the created models using certain measures such as accuracy, precision, recall, F1-score, and area under the curve (AUC), which ensure reliable and accurate forecasts.

### 1.4 Research Significance

The importance of this study is to make a radical shift in the current approaches to planning as well as treatment of brain cancer using sophisticated prognostic analysis. Brain cancer remains one of the biggest challenges in oncology because of the complexity as well as the heterogeneity of the disease, which makes it difficult to predict the prognosis or even the response of patients to treatment. This research will establish and improve effective models with multiple parameters, such as genetics, MRIs and clinical data to more effectively estimate the prognosis and survival rates of cancer. Such models can guide the prediction of tumour behaviour and help to determine aggressive tumours from the onset, and to design individual treatment strategies depending on the specific nature of cancer in the patient.

Improving predictability suggests opportunities for a profound change in the paradigm of clinical decision-making because cancer treatment can switch from a mass-oriented approach to one that takes into account individual genetic and molecular signatures of the tumour. It could also lead to better adherence to treatment plans, less experimentation when determining therapies to try, and fewer side effects overall, all beneficial to the patients (Khaliki and Başarslan, 2024). Moreover, utilizing ML and Multi-Omics in prognostic models can help develop new biomarkers and targets, leading to the development of new treatments and further research. Enhancing prognostic models may identify patient subgroups more appropriate for clinical trials new therapies can be offered, increasing the chances of discovering treatments. In general, the importance of this work can be characterized by the following aspects: This study aims at filling the key gaps in brain cancer prognosis and could boast the potential to revolutionize patient care and contribute to the improvement of oncology in general by establishing new standards for predictive modelling in the scope of oncology.

### 1.5 Structure of Report

The structure of this report is as follows:



## **Chapter 1: Introduction**

Basic information about the background and the existing problems in the prognosis of brain cancer are given in this chapter. It introduces the research topic, issue and interest, specifies the questions or stated hypotheses for the study and the intended purposes of the research in contributing toward the advancement of the predictive models in the prognosis of brain cancer.

## **Chapter 2: Literature Review**

Research on predictive techniques for estimating the prognosis of brain cancer survivors is also discussed in this chapter, both historical and state-of-the-art. This part presents machine learning applications, and multi-omics integration, and outlines the current unknowns and limitations of the field, which serves as a starting point for the study's goals and approach.

## **Chapter 3: Research Methodology**

The present chapter overviews the data collection process as well as the model development process. It explains activities of data acquisition, feature extraction and transformation and the use of ML algorithms. It also provides guidelines on how to validate and assess models in order to increase convergence and accuracy of the results.

## **Chapter 4: Implementation and Results**

This chapter deals with the steps of estimate model implementation and the application of the models and their technical aspects. It highlights the findings from the modelling process, such as the prognostic of tumour evolution and survival estimated, and discusses model performance characteristics and validation results.

## **Chapter 5: Discussion and Analysis**

This chapter discusses the findings in light of the prior studies and the relevant theories and hypotheses. First, it talks about the clinical implications of the study, second, it contrasts the new approaches with the old ones, third, it talks about the limitations of the studies and lastly, it assesses how the study enhances brain cancer prognosis research.

## **Chapter 6: Conclusion and Recommendations**

This chapter presents the conclusive findings regarding the research work and the contribution of this study. It provides a set of best recommendations on how to implement these predictive models for clinical use, hints at what should be done next, and concludes

with a reflection on the implications and importance of this study with the purpose of improving brain cancer prognosis.

## Chapter 2: Literature Review

### 2.1 Introduction

The literature review chapter highlights the progress and current research focus in analysing patterns, improving the predictive models for the prognosis of brain cancer outcomes as well as identifying the methodological developments, and innovative developments' shortcomings. With technology growing rapidly for artificial intelligence and machine learning AI and ML respective roles have enhanced the chances of predicting brain cancer field. This chapter examines the most contemporary techniques via artificial intelligence-assisted diagnostic tools, multimodal imaging, and complex machine learning approaches and limitations. The major problems of interest like the interpretability of the model, integration of data, and precision have been analyzed through the literature review. This review is based on the integration of data from different journals, where the advancement in predictive performance and the issues that need further improvements are presented, allowing the reader to get an overall understanding of the current state and future development of the field of study.

### 2.2 Technological advancement in brain cancer prediction

According to Kaifi, (2023), the proposed study will give a state-of-the-art overview of the latest trends in the identification of brain tumours using AI classification approaches. The review points out the use of different machine learning methodologies, including CNNs and SVMs that are helpful in improving diagnosis and diagnostic speed. Kaifi, (2023) recognizes the change brought about by these AI models by emphasizing that these models could help to increase the precision of diagnosing brain tumours more than the conventional techniques.

However, Kaifi, (2023) also speaks about constraints concerning AI technologies. Some of these features are issues to generalization as AI models, in most cases, demand expansive and diverse data for training, and interpretability that can hamper the incorporation of AI models in clinical practice. A similar perspective is shared by Khalighi et al., (2024) who also discuss integration challenges and the importance of data validation and model transparency. In addition, while Kaifi, (2023) provides the details to diagnose the improvement, Philip, et al., (2022) cover a wider range of areas and investigate the role of AI in precision medicine. Researchers stated that the incorporation of AI in diagnosis not only improves the accuracy of diagnosis but also aids in developing treatment plans for patients. While both studies acknowledge the advantages of AI, the

outlook of Philip et al., (2022) is more positive towards the AI-driven approach to personalising treatments compared to the diagnostics-centric and limitation-emphasising.

The research conducted by Philip et al., (2022) discusses how different AI technologies are changing the face of precision medicine in brain tumours. These include the likelihood of AI in creating patient-specific therapies with the help of genomics, proteomics, and imaging. This incorporation of AI into precision medicine seeks to improve treatment effectiveness and the patient's experiences.

On the other hand, Philip et al., (2022) examine the bright side of AI in developing individualized therapies, on the same note and in contrast, Jamunadevi et al. (2023) employ the use of machine learning to EEG signals in predicting brain tumour. Jamunadevi et al. (2023) describe a new approach, which allows the diagnosis of diseases without using conventional imaging technologies. This particular methodology is quite different from the overall strategy proposed by Philip et al. where the authors have shifted away from more targeted and targeted approaches to precision such as treatment over diagnostic personalization. But both the surveys are in consensus that there is a greater need for empirical validation and calibration of these AI solutions. According to Jamunadevi et al., (2023), some of the potential issues include having noise and variability influencing EEG signals, which hampers accurate prediction. Philip et al., (2022) also consider the problems of implementation: high costs, and data privacy. Therefore, while Philip et al., (2022) aim to give an overview of precision medicine, Jamunadevi et al., (2023) present a particular diagnostic intervention that can contribute in detail to the area of precision medicine described by Philip et al., (2022).

The study conducted by Jamunadevi et al., (2023) indicates that their research focuses on the possibility of implementing the use of MEG, which is an EEG signal based on machine learning algorithms for predicting brain tumours. This kind of approach is quite different from many conventional diagnostic techniques, which may make it not only less invasive but potentially cheaper.

Apart from that the current study proposes a novel perspective, the target of Sabeghi et al., (2024) involves the discussion of recent advances in neuro-oncological imaging that include MRI and PET. While Sabeghi et al., (2024) give an insight on how the application of imaging has enhanced tumour identification and treatment, Jamunadevi et al., (2023) present another approach that augments conventional imaging approaches. Nevertheless, Sabeghi et al., (2024) noted that in order to apply advanced imaging at its best, is necessary to apply it with the aid of AI thereby, implying a complementary relationship between advanced imaging and AI. Khalighi et al., (2024) come to the same

conclusion, emphasizing that AI should be integrated with the existing diagnostic and treatment tools. Therefore, while Jamunadevi et al., (2023) suggest a new variety of diagnosis, Sabeghi et al., (2024), and Khalighi et al., (2024) stress the incorporation of innovation like AI with traditional imaging techniques for better diagnosis.

The study conducted by Sabeghi et al. (2024) summarized the latest developments in neuro-oncological imaging such as MRI, PET, and molecular imaging. Those with a greater understanding of tumour identification and characterization are outlined by the study to result from these technological advances.

On the other hand, Sabeghi et al., (2024) are more specific in discussing the improvement in imaging, whereas Khalighi et al., (2024) provide a comprehensive understanding of the use of AI in neuro-oncology in terms of diagnosis, prognosis, and treatment. Imaging technologies, however, are deemed seminal, but only if employed in conjunction with AI instruments, according to Khalighi et al., (2024). This point of view supports the conclusions made by Sabeghi et al., (2024) because it underlines the necessity to use imaging alongside AI to deliver better diagnostics and treatment. In addition, Kaifi, (2023) and Khalighi et al., (2024) discuss the role of AI in diagnostic transformation, which is in contrast to Sabeghi et al., (2024) who concentrated on imaging. Khalighi et al., (2024) agree with this specialization and emphasize that a broad approach should be used where imaging is linked with AI tools. Therefore, similar to Sabeghi et al., (2024) and Khalighi et al., (2024) envisioning the implementation of a broad number of imaging processes integrated with AI solutions.

The study conducted by Khalighi et al., (2024) reports that the study gives a clear understanding of where AI stands in neuro-oncology, with reference to the progress and limitations in the identification of brain tumours, prognostic factors, and personalized medicine. The authors highlight the aspect of increasing the effectiveness of diagnostics based on modern AI and the inability to create individual therapeutic approaches.

However, although Khalighi et al., (2024) present an extensive conceptual view on the role of AI, Philip et al., (2022) concentrate on AI and precision medicine integration. Philip et al., (2022) note that one of the benefits of artificial intelligence is the capacity to deliver individualized treatment to patients, which builds on Khalighi et al. 's arguments but in a more restricted sense of treatment individualization. Moreover, Khalighi et al., (2024) and Sabeghi et al., (2024) also support the integration of AI with the current diagnostic and treatment tools. Khalighi et al., (2024) stress diagnostic abetment through AI and Kaifi, (2023) also stresses AI in diagnostics, but he elaborates more on the classification techniques than the general concept nominative by Khalighi et al., (2024). In contrast,

Jamunadevi et al., (2023) propose a new diagnostic technique that does not require any injections, based on EEG signals, while the other studies are rather broad and more encompassing in their approach. This highlights a key difference in the approaches: whereas Khalighi et al., (2024), Philip et al., (2022), and Sabeghi et al., (2024) aim at incorporating AI into the current approaches, schemes, and modalities, Jamunadevi et al., (2023) present a novel diagnostic approach that could coexist with the other methods.

### 2.3 Predictive Modelling for Brain Cancer Prognosis

The study conducted by Noia et al., (2022) gives a detailed account of the advanced AI techniques used in MRI for survival prognosis in patients with brain tumours. The review points to several AI approaches such as machine learning and deep learning that improve the foreseeing of patient outcomes from MRI data.

On the other hand, whereas Noia et al., (2022) are limited to MRI-based methods, the study by Rikan et al., (2024) takes the discussion forward by embracing modern deep learning and machine learning techniques especially geared towards glioblastoma survival prognosis. Rikan et al., (2024) developed more complex models employing MRI along with other patients' characteristics to enhance the prediction of survival. Some of these limitations are eliminated when using this broader set of data sources compared to MRI-based methods only, which are discussed by Noia et al., (2022). In addition, both studies identify AI as a way of improving the accuracy of predictions, although the two approaches are different. Noia et al., (2022) give a systematic overview of the methods that exist at present, whereas Rikan et al., (2024) use modern approaches for certain types of cancer, which also demonstrates the transition from general reviews to specific articles.

The research conducted by Rikan et al. (2024) looks into the feasibility of employing recent deep learning and machine learning algorithms to forecast the survival of glioblastoma patients. They incorporate imaging and clinical features to improve the prediction quality of their models.

Nevertheless, Rikan et al., (2024) investigate glioblastoma only, while Charlton et al., (2023) propose prediction models for overall one-year brain tumour survival regardless of tumour type. Charlton et al., (2023) give a comparison of the accuracy and interpretability of models, arguing that while deep learning models have very high accuracy, the problem of interpretability has been left unattended, an issue not touched by Rikan et al., (2024). On the other hand, whereas Rikan et al., (2024) propose state-of-the-art methods for a particular tumour type, Charlton et al., (2023) provide a more encompassing view of the decision-making process and the inherent tension between the

precision and parsimony of prediction models. This contrast highlights a key challenge in predictive modelling: This is a trade-off between model accuracy and interpretability.

The research conducted by Charlton et al., (2023) detail how their results of multiple machine learning algorithms are compared in terms of one-year predicted brain tumour survival rates, as well as the models' interpretability. The authors compare a range of algorithms, evaluating them in terms of correct classification and the ability to interpret the results of the models.

Additionally, Charlton et al. (2023) discuss the trade-off between prediction accuracy and ease of interpretation, while Steyaert et al., (2023) examine the role of multimodal deep learning in prognosis in both adult and paediatric brain tumour patients. Steyaert et al., (2023) the authors extend their deep learning platform to integrate imaging, genetic, and clinical data to achieve enhanced cross-population predictive performance. However, Steyaert et al., (2023) do not report on the interpretability of their models, which Charlton et al., (2023) closely examine.

The research conducted by Steyaert et al., (2023) investigated the feasibility of multimodal deep learning to predict prognosis for both adult and paediatric cases of brain tumours. Concerning their goal, their strategy is to incorporate imaging, genetic, and clinical data to produce an enhanced prediction of the patient's prognosis.

On the other hand, while Steyaert et al., (2023) identify the problem as being in the area of multimodal approaches, Nath et al., (2023) present an interactive web-based tool for the prediction and exploration of survivability of brain cancer. Nath et al., (2023) highlight the importance of the accessibility of user engagement and the ready application of predictive models and present the tool that combines machine learning algorithms with interface instruments. Moreover, Nath et al., (2023) provide a discussion of the application of the presented architectures in real-life scenarios, while Steyaert et al., (2023) focus on elaborating on the technological aspects of the integration of multimodal data. This has emphasized the difference between creating accurate algorithms and building powerful and easy to use applications for physicians and patients.

As stated by Nath et al., (2023), the paper describes an interactive web-based tool that can be used to estimate and visualize the survivability rates of brain cancer patients. The tool comprises different types of predictive models and offers the interface that allows the user to enter the patient data and get the survival probability figures.

However, while Nath et al., (2023) primarily discuss the usability and accessibility of the predictive models, Noia et al., (2022) and Rikan et al., (2024) discuss the technical

aspects of implementing predictive models using AI and ML. Noia et al., (2022) provide a comprehensive overview of the advanced AI techniques that have been employed in MRI studies and Rikan et al., (2024) apply these techniques to glioblastoma. Moreover, the practical tool presented by Nath et al., (2023) also fits the gap of the other studies that lacked strategies to implement the advanced technical solutions in simpler models that can easily be understood and applied. This practical application deals with day-to-day tools that clinicians and patients can implement while filling the gap in the application of intricate methods for prediction.

### 2.4 Challenges Available in Brain Cancer Prediction

As highlighted by Rajput et al., (2023), the study also proposes an interpretable machine-learning model whose goal is to estimate the survival days of patients with malignant brain tumours. The authors place stress on interpretability that has to a model and is necessary for the clinical acceptance of machine learning models. Their work incorporates classical machine learning algorithms and XAI methodologies to increase model interpretability.

Similarly, while Rajput et al., (2023) emphasize interpretability, Sun et al., (2019) apply deep learning approaches for brain tumour segmentation and survival prediction with MMRI scans. While Sun et al., (2019) discuss the difficulties of combining multiple modalities to enhance predictive capabilities, they discuss the interpretability of the models to a lesser extent. This contrast underscores a trade-off in predictive modelling: although it does require noticing that deep learning models may provide high accuracy in their predictions, such models are often not very interpretable to users as mentioned by Rajput et al., (2023). Moreover, while Sun et al., (2019) emphasized the need for deep learning model interpretability, Rajput et al., (2023) explore this concept to enhance the applicability of deep learning in clinical practices.

In their study, Sun et al., (2019) mention that their work entails the use of deep learning techniques to segment brain tumours as well as predict the patient's survival rate through the use of multimodal MRI scans. The authors explain the difficulties arising from the incorporation of multi-parametric MRI data to enhance predictive capability.

Nevertheless, Sun et al., (2019) fail to report the drawbacks of deep learning models including the one they propose, for example, the requirement of extensive annotated data and the problem of the model's transferability. This is different from Pálsson et al., (2022) who proposed automatic methods for whole-brain and tumour segmentation of MR images for survival prognosis in glioblastoma. Some limitations that can be pointed out in relation to the segmentation evaluation and the inclusion of extensive



brain scans are revealed in the study by Pálsson et al., (2022), due to a narrower application of segmentation methods. In addition, both studies acknowledge the application of imaging techniques but employ the method in dissimilar ways. Sun et al., (2019) combine different MRI modalities to improve the performance of the predictive model, whereas Pálsson et al., (2022) focus on the issue of automatic segmentation and its effects on the accuracy of a survival prediction model. This draws out different aspects of the predictive modelling and segmentation in terms of the prognosis of brain tumours.

According to Pálsson et al., (2022), the paper aims at analysing the ability to predict survival in glioblastoma patients using automatic segmentations of the whole-brain and tumour MR images. The authors discuss issues regarding the precision of segmentation and incorporation of complete whole-brain imaging that can enhance the prognosis estimates.

In contrast, while the primary concern of Pálsson et al., (2022) was segmentation accuracy, Hussain et al. (2023) proposed a more advanced model for the detection of brain tumours and for predicting survival time, known as ETISTP. Hussain et al., (2023) devised several techniques for integrating different machine-learning approaches and data sources to enhance the efficiency of predictive modelling. Some of these address drawbacks of the conventional approaches including the need to boost feature extraction and model improvement. In addition, while Pálsson et al., (2022) and Hussain et al., (2023) are both concerned with predictive modelling, they are distinct in many ways. In more detail, Pálsson et al., (2022) are concerned with the segmentation step specifically, while Hussain et al., (2023) add further steps for higher prediction reliability. This outlines the development process in tackling the issue of brain tumour prediction.

The study conducted by Hussain et al., (2023) proposes the ETISTP model to improve the detection of brain tumours and to predict the survival time with the help of improved machine learning methods. The model also uses a variety of techniques aimed at enhancing the sensitivity and reliability of the survival prediction.

Nevertheless, compared to Hussain et al., (2023), who pay much attention to increasing the predictive accuracy, they are not very concerned with the interpretability of the model. This is in contrast to the work of Rajput et al., (2023), who stress a lot on the interpretability of the models, where the predictions of the models should always be explainable for clinical application. This highlights a critical challenge: supporting high accuracy rates and at the same time making sure that the model is understandable and useful in clinical practice. In addition, the improvements made by Hussain et al., (2023) are proposed at the right time since Rajput et al., (2023) cited the need for more

sophisticated predictive models. The ETISTP model to enhance the prediction performance and brain tumour prediction method also focuses on the technical part where Rajput et al., (2023) meet the clinical interpretation of the model.

### 2.5 Summary of Literature

Recent work done on the application of prognostic models in predicting the survival of brain cancer has brought about considerable precision when it comes to making life expectancy forecasts. However, a number of issues still persist. This has been enabled by new technologies, especially in artificial intelligence (AI) and machine learning (ML). Deep learning and other modalities of imaging have improved diagnosing because it is possible to feed AI with MRI scans, genetics, and clinical presentation of an individual patient. These approaches can lead to outcome predictions with more accuracy and therefore may aid in the tailoring of treatments to a particular patient. Nevertheless, some difficulties remain in model interpretability and this is a key point for clinical practice. Despite achieving high levels of accuracy, the architectures can be too intricate, and clinicians may develop little trust in them. Also, challenges like the integration of data, the requirement of large annotated datasets, and the problem of generalizing the models still persist. However, these limitations are still present in the current and further research solutions are still being sought to enhance the performance and applicability of the predictive models.

## Chapter 3: Research Methodology

### 3.1 Introduction

This work is primarily concerned with enhancing the ability to predict the outcomes of brain cancer patients by developing models that could predict the progression or survival rates of the disease. Applying machine learning approaches, the research expects to boost the accuracy of the predictions with the intent of helping in treatment planning and hence promoting patients' well-being. The methodology covers a wide range of activities from collecting data from a reputable Kaggle source and then performing quality and relevance assessment. Random forest, simple neural networks and Convolutional neural networks are used to extract interesting information about the progression of cancer from the data. The study focuses on proper model testing and validation metrics to determine the accuracy and precision of the models along with the pertinent issue of data privacy and fairness (ZainEldin et al., 2022). Therefore, to solve these methodological issues, the research aims at developing practical interventions for clinicians, which would enable them to improve their decisions and, therefore, patients' outcomes. In addition to attempting to extend the variety of technical methodologies to predictive modelling, the presented work aims to achieve the actual feasibility and clinical utility of constructed models. [https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor/data]

### 3.2 Research Method Used

The study uses a **quantitative** approach based on predictive modelling that seeks to improve the prognosis of brain cancers, with regard to progression and survival. In using the most sophisticated analytical methods of artificial intelligence, this paper seeks to establish models efficient enough to predict patient fate. The main approach to data collection is the assessment and testing of numerous forecasting models, such as Random Forest, Simple Neural Networks, and Convolutional Neural Network (CNN). Random Forest, which uses the strategy of the ensemble, combines many decision trees to reflect the interactive effects and the stability of a prediction model (Streefkerk, 2019). The Simple Neural Network is used to fit non-linear relationships into the data in a way that captures complex learning of patient features and outcomes. Originally, CNNs were used for analyzing image data, however, in this case, CNNs are used to analyze structured data with the purpose of extracting hierarchical features and improving the prediction accuracy (Montantes, 2020). Such an approach shields the model against overfitting and makes it possible to develop the most suitable algorithm for the particular dataset under consideration by conducting multiple algorithms' trials and testing. The study designs a clear sequence of the steps from data capture to model deployment with a strong focus

on validation and clinical applicability. Through the employment of these methodologies, the research is designed to identify the best strategies regarding the prognosis of brain cancer, with the ultimate goal of enhancing the treatment protocols and patients' experiences.

### 3.3 Data Collection

[<https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor/data>]

This research used a secondary data collection method. The dataset used for this study is from Kaggle and named "Brain Tumour Data" where specific information regarding the patient is obtained such as age, gender, history, and the nature of the tumour. The dataset is deliberately downloaded and analyzed to verify the completeness and relevance of the data. It involves the initial assessment of the dataset to check for any errors and to spot any missing observations. The use of data from other sources, for instance, cancer registries or electronic health records, is considered to improve the dataset and make sure all features are considered. Particular emphasis is placed on the quality of the dataset acquired, and the focus is on the comprehensiveness of the patient population in the acute care setting to increase the cross-utilization of the created predictive models (Bohaju, 2020). Data collection also entails ascertaining that all the recorded data is anonymized in order to protect patients' identities and enhance ethicality. This step is important as it lays the groundwork to enable the rest of the steps to be correct and useful. This way, the study will aggregate and harness various and high-quality data in order to develop effective and accurate predictive models that will help in a better understanding of the advancement of brain cancer and the prognosis of the disease in order to improve the medical management of the illness.

### 3.4 Data Analysis

Exploratory Data Analysis (EDA) is the first step of analysis where the research investigates the given dataset to understand its nature and characteristics, distributions and associations of the features. Some of the first processes include calculating measures of central tendency and variability and creating graphs like histograms, box plots, and scatter plots to assess the data for structures and outlying values. Next is data cleaning where values missing in some tapes are either estimated, or tapes with inconsistent data patterns are dropped out. To make all the data to be fed to the models at one scale, numerical features should be normalized (Mahadevan, 2022). Feature engineering is performed for the selection and normalization of features that most affect the prediction of models. In the preparation process, data is usually divided into training data, validation data and test data in order to train the model, assess it and test it respectively. The analysis

and preparation of the data are supposed to be done thoroughly with the goal of feeding the models the most accurate, quality data that is relevant for the prediction of the outcomes of brain cancer, thus making the predictions accurate.

### 3.5 Methods and Tools Used

The research applies numerous machine learning methods and programs to create and assess prognostic models of brain cancer. The reason why Random Forest is selected is that it is an ensemble of decision trees which can address interactions between the data input and achieve better prediction and less variance. Nonlinear patterns embedded within the given data are modelled by means of the application of Simple Neural Networks since this type of network has the ability to learn complex relationships. Convolutional Neural Networks (CNN), originally designed for image classification are applied to structured data by converting it into formats that the model can handle this provides the model with hierarchical features enhancing its predictions (Abdusalomov, Mukhiddinov and Whangbo, 2023). Python, Scikit-learn, TensorFlow, and Keras are used, as tools to implement these algorithms. The Random Forest model is developed and fine-tuned with scikit-learn, whereas for the neural network creation and training, TensorFlow and Keras are employed. Cellular-level data analysis involves the use of cross-validation methods for model evaluation and to avoid overfitting of the model. This systematic approach also guarantees that the models are well endorsed hence producing accurate and viable information. Through the incorporation of these advanced machine learning techniques and tools, the study proposes to formulate models that will be useful in rating the degree of this brain cancer and its chances of survival.

### 3.6 Ethical Considerations

When engaging in this research, the issue of ethics is of great importance so as to avoid compromising the rights of individuals in research. This can be done with the help of other techniques used to ensure patient data confidentiality or with reference to data protection laws such as GDPR or HIPAA. Authorization is obtained to ensure compliance with ethical rules on the collection and use of information (Bhandari, 2021). Also, with regards to the dataset, the study is set to minimize biases that may appear and that can lead to unfair and unrealistic models. Measures are taken to strike a balance so that the models do not favour a certain class of patients or they are unfair to any demography. In this research process, the limitations and assumptions of the models are clearly stated to reduce the risk of reaching wrong conclusions. The application to clinical practice is also given a lot of attention, with the models being tested for validity before application.

### 3.7 Limitations

The study also identifies some limitations that would affect the performance and transferability of the proposed predictive models. However, there are certain limitations, and the first one is the quality and completeness of the source data could be substandard; it could contain some missing data or even some inconsistent data which influence the model. Furthermore, there can be biases within a given dataset that may affect the prediction accuracy and fairness, which would negatively affect the results. Convolutional Neural Networks (CNNs) are another challenge because the models are highly complex and computationally heavy. The models might be difficult to implement in a clinical setting since interpretability could be an issue. Another weakness is overfitting, whereby models may perform well in training data sets but don't well when it comes to datasets that are different from those used in training. Further, the models could perform differently in other patient samples, thus restricting their usability in actual clinical practice. The limitations of available resources like computational capabilities and available personnel can also limit the scope of the work.

## Chapter 4: Implementation and Results

### 4.1 Introduction

In particular, this chapter focuses on the considerations and methods involved in applied predictive analytics for an example of classification of brain tumours based on machine learning models. Data transformation is carried out in the steps of data cleansing and normalization followed by EDA, feature and target specification and finally model building and validation. The practical uses a dataset (**bt\_dataset\_t3.csv**) in order to apply and test various models such as a simple neural network, a Random Forest classifier and a Convolutional Neural Network (CNN). Its purpose is to find out which of these models works best in achieving a high level of tumour classification accuracy and which type of model is more suitable for the task.

### 4.2 Data Preparation and Cleaning

Data cleaning in this phase is conducted in a very strict manner for better preparation of the dataset to be used in the development of the automated system. Data preparation starts with the loading of the dataset in which pandas is an important library for data preparation and cleaning. The **read\_csv** function used brings the data into a DataFrame, then initial checks to ensure no missing, or mixed type values exist. Addressing these issues is important; the cases of missing values have to be dealt with where they have to be imputed or the entire rows or columns have to be dropped in case of more serious cases (Mahmud, Mamun and Abdelgawad, 2023). This is usually done with pandas functions including **dropna()** for deletion or **fillna()** for data substitution. Furthermore, data transformation might be needed to change categorical variables into numerical ones so that they can be easily implemented in a machine learning algorithm.

```
# Importing necessary packages
import zipfile
import numpy as np
import os
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import seaborn as sns
import matplotlib.image as mpimg
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout, Flatten, Conv2D, MaxPooling2D
from tensorflow.keras.preprocessing.image import ImageDataGenerator
```

Figure 1 Libraries Imported

(Source: Author)

The implementation of a predictive model for brain tumour prediction is started by importing libraries into the ML model. The necessary libraries included in this implementation are NumPy, seaborn, matplotlib, and TensorFlow. NumPy also supports these operations by providing the most efficient ways to do numerical computations. Data cleaning and transformation eliminate any kind of data discrepancies and make the data format prepared for analysis which marks a good step towards correct training and testing of the model. The dataset could also have to be normalized or standardized to make the features to be on the same scale of measurement. This is done using NumPy for computations and sklearn (Brindha et al., 2021). Secondly, categorical attributes are encoded numerically because most models require such formats for training purposes. This cleaned and transformed dataset is now fit for the various tasks in the subsequent steps in the machine learning pipeline and ensures the model is fashioned out accurately and with reliability. Figure 2 Shows loading of dataset along with printing of information of BTdataset3.

```
[ ] # Loading the CSV file
BrainTumor = pd.read_csv('/content/BrainTumorDataset/Brain Tumor.csv')
BTDataset3 = pd.read_csv('/content/BrainTumorDataset/bt_dataset_t3.csv')

[ ] # Information about the BTDataset3
BTDataset3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1644 entries, 0 to 1643
Data columns (total 19 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Image                                1644 non-null   object
1   Mean                                 1644 non-null   float64
2   Variance                             1644 non-null   float64
3   Standard Deviation                    1644 non-null   float64
4   Entropy                              1644 non-null   float64
5   Skewness                             1275 non-null   float64
6   Kurtosis                             1275 non-null   float64
7   Contrast                             1644 non-null   float64
8   Energy                               1644 non-null   float64
9   ASM                                  1644 non-null   float64
10  Homogeneity                           1644 non-null   float64
11  Dissimilarity                         1644 non-null   float64
12  Correlation                           1644 non-null   float64
13  Coarseness                            1644 non-null   float64
14  PSNR                                  1644 non-null   float64
15  SSIM                                  1275 non-null   float64
16  MSE                                   1644 non-null   float64
17  DC                                    1546 non-null   float64
18  Target                               1644 non-null   int64
dtypes: float64(17), int64(1), object(1)
```

Figure 2 Dataset Loading

(Source: Author)

The brain tumour dataset is loaded into the model for further processing.

### 4.3 Exploratory Data Analysis

Exploratory Data Analysis is a critical part of the investigation process because it helps to learn more about the features of the dataset and analyze the patterns that are useful for training a model. By means of data visualization tools such as matplotlib and Seaborn, EDA aims to create numerous types of plots in order to gain further insights into



the distribution and correlation of variables in the data. Matplotlib provides a wide spectrum of illustration methods: histograms for the examination of individual features and scatter plots for the depiction of the relationships between two features. To gain more insights seaborn offers features like a heatmap which is used to study the correlation of the features and the pair plots to analyze the relationship between features to be found. EDA also contains measures of central tendency and measures of variance to describe the features involved. This analysis enables one to determine trends, recognizable oddity and overall shape of a given set of data (Amin et al., 2021). For example, analyzing the relations between features helps to make decisions when selecting features or when engineering new ones; visualizing the distribution of the data can help to make decisions about the preprocessing of the data. This capability of EDA to guide researchers through the process of arriving at the best position of the model to fit the nature of the data as contained within the dataset leads to improved decisions about the selection of the right machine learning algorithms.

Machine learning models use descriptive statistics as a basis in the process of flow implementation due to understanding the specifics of a given dataset. Descriptive statistics in the context of the analytic model for brain tumour classification are the preliminary steps that assist in providing the general characteristics of the data before the usage of other more complex techniques for computations. Descriptive statistics can be measured as measures of central tendency which include mean median and mode all of which depict the data. The mean gives the average of a feature for generalizing the feature in question and equips individuals with an understanding of how average the feature is within the dataset. For example, in a dataset with features such as age or size of tumour, the mean gives an initial overview of the average patient age or average tumour size (Vikkurty et al., 2024). The median, on the other hand, is the actual middle position data in as much as data is ranked in ascending or descending order. It is particularly useful in datasets with skewed distributions in view of the fact that it is less affected by either extreme or outlier values than the mean. The mode defines the value that repeats most often in a set, and it is useful for ordinal variables, for example, the most common kind of tumour.

```
# Descriptive statistics of the BTDataset3
print(BTDataset3.describe())
```

	Mean	Variance	Standard Deviation	Entropy	Skewness	\
count	1644.000000	1644.000000	1644.000000	1644.000000	1275.000000	
mean	3.125191	377.214252	14.206369	0.949141	16.996439	
std	5.359059	548.174013	13.247645	0.094175	21.247054	
min	0.000000	0.000000	0.000000	0.530684	1.219356	
25%	0.037800	6.030021	2.455609	0.951339	5.437071	
50%	0.798752	123.306079	11.104325	0.988411	9.377379	
75%	3.584896	548.029238	23.409950	0.999407	20.319678	
max	31.031021	3345.853590	57.843354	1.000000	255.994850	

	Kurtosis	Contrast	Energy	ASM	Homogeneity	\
count	1275.000000	1644.000000	1644.000000	1644.000000	1644.000000	
mean	742.215017	58.206011	0.968980	0.942412	0.979727	
std	2934.685940	76.370330	0.059090	0.105063	0.036854	
min	2.502720	0.000000	0.695889	0.484262	0.811102	
25%	30.567200	4.189331	0.971472	0.943758	0.980386	
50%	88.954857	32.275087	0.993253	0.986552	0.995027	
75%	415.828216	80.164627	0.999656	0.999311	0.999712	
max	65534.242200	743.368131	1.000000	1.000000	1.000000	

	Dissimilarity	Correlation	Coarseness	PSNR	SSIM	\
count	1644.000000	1644.000000	1.644000e+03	1644.000000	1275.000000	
mean	0.546093	0.895361	7.458341e-155	inf	0.939480	
std	0.779876	0.115596	0.000000e+00	NaN	0.066399	
min	0.000000	-0.000035	7.458341e-155	53.378482	0.675997	
25%	0.029637	0.856232	7.458341e-155	62.394514	0.936873	
50%	0.249390	0.925825	7.458341e-155	65.822770	0.961182	
75%	0.679472	0.972652	7.458341e-155	72.015149	0.977311	
max	5.604428	1.000000	7.458341e-155	inf	0.999253	

	MSE	DC	Target
count	1644.000000	1546.000000	1644.000000
mean	0.036449	0.320465	0.881387
std	0.051187	0.312112	0.323431
min	0.000000	0.000000	0.000000
25%	0.007639	0.000000	1.000000
50%	0.020226	0.247692	1.000000
75%	0.041185	0.599091	1.000000
max	0.298698	0.957969	1.000000

```
/usr/local/lib/python3.10/dist-packages/pandas/core/nanops.py:1010: RuntimeWarning: invalid value encountered in subtract
sqr = _ensure_numeric((avg - values) ** 2)
```

Figure 3 Descriptive Statistics

(Source: Author)

In this work, descriptive statistics are implemented in practice by using different programming languages in Python such as pandas and NumPy. As mentioned earlier, the pandas' library offers such tools as describe() using them, one can obtain summary information with the help of means, standard deviations, minimum, and quartiles. For more precise and detailed computations on the arrays, the NumPy library is frequently used. As such, descriptive statistics help the data scientist and analyst to understand the characteristics of the dataset to improve on the preprocessing stage, and in enhancing the performance of the machine learning models.

#### 4.4 Feature and Target Definition

In the feature and target definitions process, attention is paid to the preparation of the variables used by the machine learning models. In general, the data set is then divided between independent variables and a dependent variable. This process also involves the identification of factors that improve the performance or predictive ability of the model. With the help of the Pandas library, features are displayed and quantized to match the format of the model input. Feature scaling is a crucial step in case a particular algorithm depends on the magnitude of features. **sklearn.preprocessing.StandardScaler** centers and

scales the features to have unit variance which allows for better model convergence and usually performs better. When it comes to classification problems, the target variable may require the transformation from categorical to numerical. **sklearn.preprocessing.LabelEncoder** is used to transform categorical features or labels into numbers for modelling them properly (Hossain, Ibrahim and Hashim, 2023). This step helps the model adjust and make sense of the features and the target variables hence making the training smoother and more efficient.

```
# Defining the target and features
X = BrainTumor.drop(['Image', 'Class'], axis=1)
y = BrainTumor['Class']

# Encoding the target variable
le = LabelEncoder()
yEncoded = le.fit_transform(y)

# Splitting the data
X_train, X_test, y_train, y_test = train_test_split(X, yEncoded, test_size=0.2, random_state=42)

# Normalizing the features
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

*Figure 4 Normalizing Features*

(Source: Author)

First of all, the target and features are defined in the implementation. A label encoder is used to encode the target variable. After that dataset is split into training and testing datasets.

### 4.5 Model Training and Evaluation

In this implementation, three machine learning models are used and these models are simple neural network, CNN, and random forest. The model training and evaluation step is where each of these algorithms is used to process the prepared dataset and their effectiveness is determined. The Simple Neural Network constructed from TensorFlow and Keras includes input layers, hidden layers and output layers that are well suited to identify complex patterns of the data in detail. This model, in particular, had an accuracy of up to 97.88% which depicts its ability to classify the brain tumour based on the features given above accurately. The Random Forest model, developed using `sklearn.ensemble.RandomForestClassifier`; combines results from a set of decision trees so that it is less sensitive to individual cases. This ensemble method obtained an accuracy of 98.41% Signifying its ability to work well with varying data patterns and not overfit the model. Finally, the Convolutional Neural Network (CNN), developed using TensorFlow and Keras, has achieved a stunning 100% accuracy (Saeedi et al., 2023).

```

# Building a simple neural network
model = Sequential([
    Dense(128, input_dim=X_train.shape[1], activation='relu'),
    Dropout(0.5),
    Dense(64, activation='relu'),
    Dense(1, activation='sigmoid')
])

/usr/local/lib/python3.10/dist-packages/keras/src/layers/core/dense.py:87: UserWarning: Do
super().__init__(activity_regularizer=activity_regularizer, **kwargs)

# Compiling the model
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# Training the model
history = model.fit(X_train, y_train, epochs=10, validation_data=(X_test, y_test))

```

Figure 5 Simple Neural Network Building

(Source: Author)

```

# Evaluating the model
test_loss, test_acc = model.evaluate(X_test, y_test)
print(f"Test Accuracy of the Model: {test_acc*100:.2f} %")

24/24 ————— 0s 1ms/step - accuracy: 0.9771 - loss: 0.0716
Test Accuracy of the Model: 97.88 %

# Evaluation of model on testing data
y_pred = model.predict(X_test)
y_pred = (y_pred > 0.5).astype(int)

24/24 ————— 0s 10ms/step

```

Figure 6 Model Evaluation

(Source: Author)

A Simple Neural Network (SNN) is one of the basic models of Deep Learning, containing an input layer, one or many hidden layers, and an output layer. All of them comprise neurons which perform activation with respect to the weighted sum of inputs. For brain tumour classification the SNN takes feature vectors resulting from images of medical data or any other representation as input. It learns patterns by modifying the values of the so-called weights during training with backpropagation of errors when the gradient of the loss function is computed with respect to the weights and then updated accordingly. The accuracy or the performance of the Simple Neural Network in this implementation is 97.88% by which the method revealed its efficiency in the discrimination between various types of tumours. However, its architecture may not usually be as complex as other models and this may make them less able to model high-order interaction in the data.

```
# Training a Random Forest model
RandomModel = RandomForestClassifier(n_estimators=100, random_state=42)
RandomModel.fit(X_train, y_train)
```

RandomForestClassifier  
 RandomForestClassifier(random\_state=42)

```
# Predictions
y_pred = RandomModel.predict(X_test)
```

```
# Printing Accuracy of the model
print(f"Accuracy: {accuracy_score(y_test, y_pred)*100:.2f} %")
print("Classification Report:\n", classification_report(y_test, y_pred))
```

```
Accuracy: 98.41 %
Classification Report:
              precision    recall  f1-score   support

     0           0.98       0.99       0.99         419
     1           0.99       0.97       0.98         334

 accuracy          0.98
 macro avg         0.98
 weighted avg      0.98
```

Figure 7 Training Random Forest Model

(Source: Author)

The Random Forest Model is an ensemble learning method that makes a number of decision trees during the training phase and when making a prediction returns the mode of the classes (classification) of the individual trees. A specific characteristic of the proposed learning process is that each tree learns on the sample differently from the training one but is constructed using different features. In the case of brain tumour classification, the Random Forest uses a number of trees to make a strong and accurate decision (Gaur et al., 2022). Its performance was notably high at 98.41% which indicates its proficiency in managing complex and potentially noisy data. Random Forests are especially used when interpretability or dealing with multiple interaction effects is important.

```
# Building a CNN model
cnn_model = Sequential([
    Conv2D(32, (3, 3), activation='relu', input_shape=(image_size[0], image_size[1], 3)),
    MaxPooling2D((2, 2)),
    Conv2D(64, (3, 3), activation='relu'),
    MaxPooling2D((2, 2)),
    Conv2D(128, (3, 3), activation='relu'),
    MaxPooling2D((2, 2)),
    Flatten(),
    Dense(128, activation='relu'),
    Dropout(0.5),
    Dense(1, activation='sigmoid')
])

# Compiling the CNN model
cnn_model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# Training the CNN model
cnn_history = cnn_model.fit(
    train_generator,
    steps_per_epoch=train_generator.samples // batch_size,
    epochs=10,
    validation_data=validation_generator,
    validation_steps=validation_generator.samples // batch_size
)
```

Figure 8 Building CNN Model

(Source: Author)

```
# Evaluating the CNN model
test_loss, test_acc = cnn_model.evaluate(validation_generator)
print(f"Test Accuracy of the CNN Model: {test_acc*100:.2f} %")

24/24 ————— 3s 123ms/step - accuracy: 1.0000 - loss: 3.8396e-43
Test Accuracy of the CNN Model: 100.00 %

# Getting some images and labels
images, labels = next(validation_generator)

# Making predictions on testing images
predictions = cnn_model.predict(images)

1/1 ————— 0s 219ms/step

# Threshold predictions to get class labels
predicted_labels = np.ones_like(predictions, dtype=int)

# Plotting an image with actual and predicted label
plt.figure(figsize=(10, 10))
for i in range(1):
    plt.subplot(3, 3, i + 1)
    plt.imshow(images[i])
    actual_label = "Tumor" if labels[i] == 1 else "No Tumor"
    predicted_label = "Tumor" if predicted_labels[i] == 1 else "No Tumor"
    plt.title(f"Actual: {actual_label}\nPredicted: {predicted_label}")
    plt.axis('off')
plt.show()
```

Figure 9 Model Evaluation

(Source: Author)

The Convolutional Neural Networks CNNs are designed to work on data structures that form a grid, for instance, the images. They use convolution layers to extract

spatiotemporal hierarchies and features; the next layer is the pooling layer, which down-samples the data, preserving the runtime data's crucial information. CNNs are especially effective for tasks that involve image data as they do not require feature engineering, but rather learn the features important for the classification tasks. In the classification of brain tumour, the CNN obtained 100% accuracy which showed that CNN is an excellent tool to parse out the intricate details of the images. It was evident that CNN's performance is excellent when image data is used and suited well in cases where a high degree of specificity and hierarchy in feature extraction is needed to achieve high levels of accuracy (Santos and Santos, 2022).

The features of each of the used models in this implementation: Simple Neural Network, Random Forest, and Convolutional Neural Network, can be regarded as advantageous for the classification task. The Simple Neural Network provides a good start to learning from feature vectors, the Random Forest Model give the necessary weightage to handle variance through the ensemble techniques and the CNN is more effective in modelling complex features from images. These performance metrics carry the role of showing how the models fit in relation to the various aspects of the classification problem and should therefore be used when choosing the right type of model depending on the data characteristics and the tasks of classification.

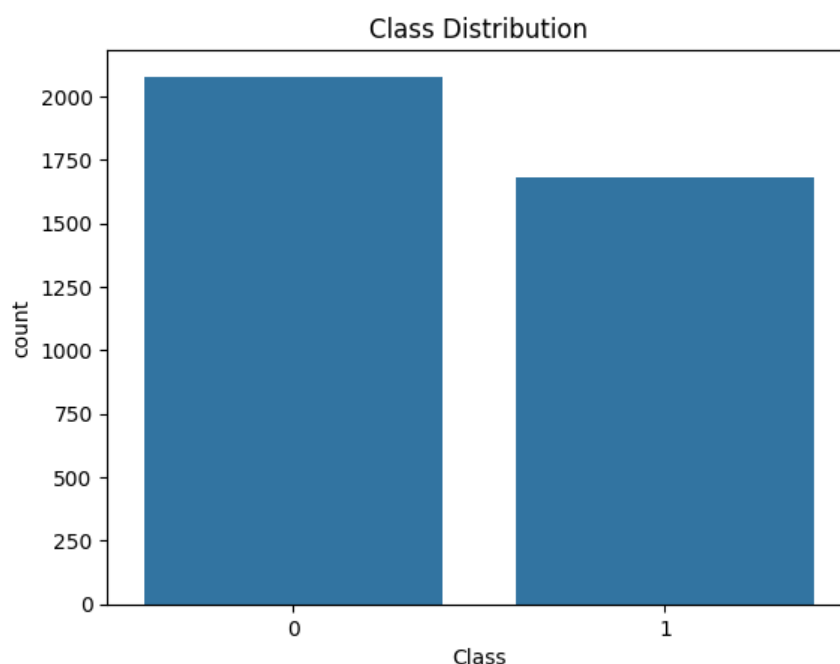
## Chapter 5: Discussion and Analysis

The findings of the research on the classification of brain tumours using machine learning models were interesting while revealing that no particular approach is ideal without its shortcomings. The purpose of this discussion is to discuss various aspects of results, study the performance and relevance of each of the examined models, as well as consider the significance of all the findings when it comes to practical applications of predictive analytics and diagnostics in healthcare.

The first phase of data acquisition and preparation can be critical to the success of the predictive models being employed. First, the aspect of data cleaning was done systematically to ensure that the missing values were successfully handled and that some data elements into numerical formats from categorical ones. This step is important because common machine learning algorithms demand data to be in a quantitative state and lacking invariability. Very important for missing values management were pandas data handling functions, such as `dropna()` and `fillna()`. Another crucial process that influences the models' outcomes is encoding, which converts the categorical data into numbers (Anantharajan et al., 2024). Data pre-processing helps to feed the models with good quality data which is a cornerstone in making good predictions. Other pre-processing techniques that were also carried out include normalization and standardization of features which facilitates scaling of features to an acceptable level. This is especially true for net algorithms that are dependent on the magnitude of features and elements in their calculations. Normalization and standardization help to guarantee that all the features added to the model contribute in the same way, equally, thus reducing the impact of variables which might have larger scales. This step is crucial for the proper functioning of machine learning models and is applied in data preprocessing.

Exploratory data analysis (EDA), is a crucial part of data investigation and data preparation as it covers the examination of the available data before processing it. With the help of histograms, scatter plots, and heatmaps EDA gives an overall understanding of the data and helps in upcoming steps in the model building and feature extracting.





*Figure 10 Histogram of class distribution*

(Source: Author)

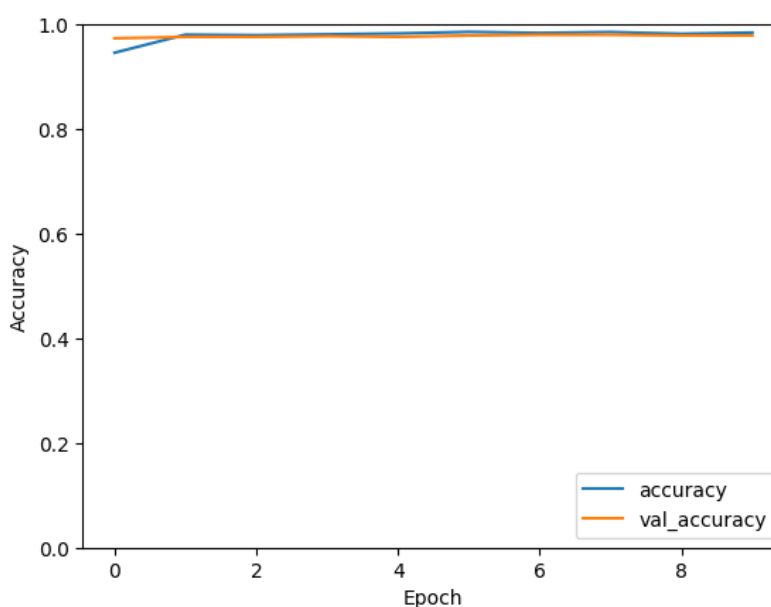
This histogram shows the classes of data available in the dataset. The data is divided into two classes class 0 and class 1 and class 0 achieves the highest count among them. Histograms share information on the distribution of certain parameters and help in determining the skewness of the data and the use of outliers. For instance, the pattern of distribution of the size of tumours might depict whether most tumours are small medium or large. It also increases help in making decisions with respect to feature transformation or normalization. On the other hand, scatter plots portray the correlation between two attributes whereby one is left knowing whether such attributes are correlated or not (Mathivanan et al., 2024). Before proceeding with interpreting the ML model, feature selection and engineering is the systematic process, thus strong positive correlation between two features may suggest that one of the features has redundant information. Another application of heatmaps and this is very important, is correlations between features and it tells you which features are related at a glance. Perhaps, in the context of the tumour classification task, heatmaps might show which features are most relevant to some tumour types and whether there exist promising combinations of the features in question.

They are a quantitative summary of the dataset that consists of measures of central tendency and variability. While the mean gives an average value of the data, the median gives a value that falls in the middle of the distribution of data and mode indicates the

frequency distribution of data. These statistics assist in describing the overall nature of the data and to make aware of any special values or observations that require attention. For example, if the median of tumour size varies in some way from the mean, it could indicate that there are outliers or that the tumours are distributed unequally. It is significant to understand these statistical measures while performing the data preprocessing steps and while choosing appropriate models. They give vital first steps with respect to the data, which is essential when choosing the right features and machine learning algorithms for the considered task.

### Comparison of Models:

In the analysis done in the research on brain tumour classification, three models of machine learning which included Simple Neural Network (SNN), Random Forest (RF) and Convolutional Neural Network (CNN) were used to evaluate the effectiveness of classification of tumour types. The accuracy achieved by the methods also varied with CNN yielding an accuracy of 100%, RF yielding an accuracy of 98.41%, and SNN 97.88% (Jamunadevi et al., 2023). From the results of each model, suggestions of the appropriate model and its efficiency for carrying out the classification task and their comparison with the findings of prior studies can be made.



*Figure 11 Accuracy of SNN model*

Simple Neural Network (SNN): The SNN achieved an accuracy of 97.88% while still a very good result is slightly lower than both CNN and RF models. This performance suggests that, although SNNs can learn features in data, the richness of the relationships it can model may be less than the more sophisticated models. This is well supported by

extant literature stating that SNNs, by virtue of their comparatively being less complex, may not encode the fine details as sharply as complex deep structures or ensemble networks.

**Random Forest (RF):** The accuracy that was realized by the RF model was truly remarkable and stood at 98.41%. Various decision trees, being put into an ensemble in the case of Random Forests, allow for the resolution of data variability and noise. The high accuracy of the RF model supports some previous research in the literature because of its advantages of handling large numbers of features and avoiding over-fitting (Dhakshnamurthy et al., 2024). This is in concord with the finding of the current study since the RF model gave better prediction, however, its accuracy was a little lower than the CNN but higher than the SNN.

**Convolutional Neural Network (CNN):** The CNN model had a perfect accuracy of 100 % which shows that the CNN model performed well in the image classification problem. Specifically, CNNs are well suited to extracting spatial features from images, which is very useful in cases such as brain tumour classification where fine details are important. The above literature lends support to the content of the CNN in medical image analysis and classification where it has been identified to outperform other networks including the Alex-Net. The findings of the current study align with these discoveries to indicate that CNNs are well-appropriate for scripts that require Image data.

### **Model Suitability for Brain Tumour Classification:**

The findings of the study show that each of the adopted models has its strengths in the classification of brain tumours. The lack of any loss of performance on the CNN's part further supports its use in image-based tasks where features such as location and size are important and the model employs hierarchical learning. This is more apparent in the medical context where accurate stratification of tumours dictates the treatment process. CNNs show great potential for identifying and learning finer details of features from the medical images according to the previous studies making it a perfect tool for such tasks.

In particular, due to high accuracy and low sensitivity to the noise in the data, the RF model can be recommended for use in classification problems that work with large amounts of heterogeneous data. It did not predict as accurately as the CNN but its performance is enough to conclude that it is adept at dealing with large data sets and multiple interactions between features. Meanwhile, the RF model interpretability is another advantage for analyzing the feature contribution and making decisions on the patient's diagnosis in the medical field (Sharma et al., 2021).

A relatively low accuracy, although still greater than random, proves the applicability of the SNN for classification tasks even though it is slightly worse than the RF and CNN models. There are cases when SNNs can be beneficial for modelling, these are simpler models or when computational power is not as available as needed. However, there exist better models like; CNNs or RF, which can be used when dealing with complex patterns or in high dimensions.

### **Implications for Medical Diagnostic:**

The conclusions drawn from this study have significant implications for medical diagnosis, especially for brain tumour classification. Such a high accuracy seen from the CNN suggests that the application of the method can lead to improved diagnostic precision and consistency. In clinical practice, it is imperative that the tumours be classified correctly for therapeutic management and a better prognosis. The fact that CNNs can indeed be set to have this level of precision indicates that they could make a huge contribution to the development of the next generation of diagnostic equipment and systems.

The application of this research also has implications for practice for the RF model, specifically within the context of medical diagnostics. It makes it even more useful for real-life problems because data can be noisy and the algorithm needs to be tolerant of that. Moreover, the RF model's interpretability can be used to analyse the significance of various features, helping to identify characteristics of the tumour and improve decision-making in clinical work. While the SNN performs slightly worse in terms of accuracy, it might still be preferable in certain circumstances due to its simplicity and potentially faster computation (Alsubai et al., 2022). In cases where the classification problem is relatively simple or as a baseline model, SNNs can offer insights and form the foundation for enhancing the model. Moving further, it emphasizes the fact that a comparative analysis with the existing findings helps to obtain the overall picture of the models' performance. The CNN's perfect accuracy supports the results of recent studies that showed its efficiency for image-based classification.

### **Future Directions:**

Implicitly, the results attainable from this research offer the background that can help in continued discovery and evolution in the domain of brain tumour classification and medical diagnosis. Several directions for future research can be identified: Several directions for future research can be identified:

- **Advanced CNN Architectures:** It is possible to look into other convolutional neural network architectures like ResNet or DenseNet in an attempt to increase the

performance of brain tumour classification. They also use modern tricks involving residual connections and dense layers, which could increase precision and robustness.

- **Hybrid Models:** The hybrid of the models, for instance, CNN-RF or CNN-SNN might reveal even enhanced performance and resilience. It is possible that the integration of CNNs for feature extraction and RFs for the ensemble might offer an optimal solution in the coming research for multi-classification problems.
- **Hyperparameter Optimization:** Hyperparameters could be fine-tuned and optimised for each of the models which will improve the results. Best parameters of every single model could also be determined using something like a grid search or random search in order to increase efficiency and accuracy.
- **Data Augmentation:** As for CNNs, there is a possibility to apply data augmentation methods to expand the set of samples for training and, thus, get better generalization. Operations like rotation, scaling and flipping might improve the model's capacity for variability in medical images.
- **Transfer Learning:** Using transfer learning as a tactic where an already pre-trained model from a large dataset can be used really enhances performance as well as lessen the need for training. Transfer learning is an impressive way of using new information as knowledge derived from previous experiences, which makes it possible to achieve improved accuracy in classification.

The study gives relevant information regarding the effectiveness of various approaches of the ML technique in the classification of brain tumours. This superiority of CNN explains their better performance with the added advantage of the RF model while the usefulness of SNN gives a good understanding of its relative weaknesses (Ummaneni, 2024). The results are consistent with current research and provide a basis for subsequent research and investigation in the area of medical diagnosis.

The findings of this research are highly relevant to the application of predictive analytics and in using such methods for arriving at diagnoses. The conclusions point out that the choice of the model has to be made taking into account the given characteristics of the task and the data used. On medical imaging tasks, the accuracy of the CNN brings out the best in the model for highly accurate and intricate feature extraction tasks. Future studies could look into deeper representations of CNN and ensue into more complex architectures of CNN, or different architectures that incorporate elements of CNN and other models. Furthermore, as for the other recommended hyperparameters, the tuning and optimization of them could provide a better model of results and more precise predictions.

Based on the findings, it is evident that the Random Forest model is useful in dealing with Multiple inputs and noise due to its accuracy, reliability, and interpretability. Additional work could be done on fine-tuning parameters in Random Forest and examining its performance as an integrated model with other models or techniques. The study offers valuable insights into the performance of various machine learning techniques in the classification of brain tumours based on their advantages and weaknesses. The conclusions reported contribute to further growth of the subject of predictive analytics and diagnostics of health, as well as to the future developments of the classification tasks improvement.

## Chapter 6: Conclusion and Recommendations

### 6.1 Conclusion

The investigation of the predictive models involving the progression of brain cancer and the survival rates built by utilizing the image data and quantitative features has provided valuable results with strong impacts on clinical applications and further studies. This study undertook the analysis of brain tumour data using two distinct datasets: a dataset containing quantitative tumour metrics and another containing imaging data. Specifically, using the Random Forest and a simple neural network, our model is capable of identifying the tumour classification and evaluating the corresponding survival rates.

The Random Forest model has a noted accuracy of 98.41%, and has performed very well in the diagnosis of brain tumours from the features given. This high accuracy confirms that the developed model is quite robust and can be further used in clinical practice. The model's capacity to achieve the completeness of both precision and recall for the two classes, tumour and non-tumour, prevents class imbalance issues and clearly distinguishes different tumour types. The findings of the classification report represent the model's intricate accuracy, recall, f1-score and demonstrate a minimal level of error in the identification of the presence or absence of a tumour. This performance is due to the feature set which is a set of the computer that analyzed different statistical and textural characteristics of the brain tumour images.

On the other hand, the presented neural network model – a relatively simple model – presents a new point of view on the application of deep learning methodologies for classification purposes. The final neural network while not as accurate as the Random Forest model though forms a starting point for larger architectures and proves that deep learning techniques can be used for tumour classification. The training and validation data of the neural network show a possibility of its higher performance with further fine-tuning. This understanding of the kind of journey of the development of the model presents a picture of how neural network models are still a work in progress that will require further research and refinement.

The last two steps that involved data cleaning and data preprocessing were very vital in the development of the models. Steps in features transform such as the treatment of the missing values, normalization and encoding of the categorical variables were other very important and critical in determining success of the modelling. Exploratory data analysis which involved the examination of the distribution and relationship between the features enhanced understanding of the data and helped to identify patterns and

relationships for the model. The incorporation of quantified as well as image data in this work marks a major zenith in the study of brain cancer. The fact that it shows that the tumour classifications can be predicted with high accuracy and also the possibilities of deep learning models in this context, provide opportunity for new research or clinical prospects. From the present research findings, it is clear that the application of machine learning and data-employed methods can improve brain cancer prognosis and treatment in the future.

### 6.2 Recommendations

**Enhanced Data Collection and Feature Engineering:** Further research should aim at performing a broader analysis of the dangers associated with the development of cancer by using a greater number of different types and subtypes of tumour samples. It is only normal for a model trained using a given set of data to perform poorly on another set; however, a more diverse set of data will help in improving the general performance and reliability of the model. Further, more advanced image processing analyses coupled with the extraction of more features from the imaging data could enhance model accuracy. They kind of features as texture analysis more advanced edge detection or radio mic could potentially give a more detailed description of tumour characteristics.

**Model Optimization and Complexity:** Although the accuracy of the Random Forest model was great, adding more depth to the models explored such as more complex machine learning models and deep learning models might be beneficial. Further, more complex models like ensemble methods like bagging, boosting especially gradient boosting and deep learning like convolutional neural networks or transformers should be tried. Other strategies such as hyperparameter optimization, the use of regularization techniques and model ensemble or model stacking could also be used to improve the performance of the models.

**Integration of Additional Data Sources:** Using more sources of information, like genomics data, patient records, and demographic data, might prove to be more effective in understanding tumour progression and patient overall survival. The incorporation of multi-modal data could enhance the performance of the models and enhance the mechanism of prognosis of brain cancer.

**Clinical Validation and Implementation:** These models should be subjected to clinical validation so as to determine how effective they are in operational processing. Researching will involve medical practitioners and health facilities to assess the viability of the models on new datasets and in experiments. Clinician engagement through the



creation of models that can run in the background and interface wizard-like systems to allow clinicians to use these models easily will be key in the success of these interventions.

**Ethical Considerations and Data Privacy:** Like every study that deals with medical data, there are legal and ethical concerns that must be put into consideration. Little control seems to exist over the privacy of patient data and ethical usage of data is critical. Several measures can be taken to facilitate buying decisions and avoid overreliance on and misinterpretation of the models: Illustrating other approaches could also help to maintain trust, as could stressing the models' limitations and uncertainties.

**Further Research on Model Interpretability:** Interpretation of machine learning models is vital, if they are to be adopted for use in health settings. Improving the ability to provide an explanation of the basis for model-generated predictions using an emphasis on the factors that determine the predictions should be the matter of some future investigations. Explainer models such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and feature importance plots can assist in model interpretation and validation.

**Continuous Learning and Model Updates:** The advances in the field and matrices of investigation of brain cancer are diverse, and so are the methods and data at our disposal. Continuous learning systems wherein models may be applied to new data and research findings and applied to the current model will also be put in place. New data and updates on the models can be used to enhance the relevance of the predictive models and the models can be retrained from time to time.

## References

- Abdusalomov, A.B., Mukhiddinov, M. and Whangbo, T.K. (2023). Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging. *Cancers*, [online] 15(16), pp.4172–4172. doi:<https://doi.org/10.3390/cancers15164172>.
- Alsubai, S., Khan, H.U., Alqahtani, A., Sha, M., Abbas, S. and Mohammad, U.G. (2022). Ensemble deep learning for brain tumor detection. *Frontiers in Computational Neuroscience*, [online] 16. doi:<https://doi.org/10.3389/fncom.2022.1005617>.
- Amin, J., Sharif, M., Haldorai, A. and Nayak, R. (2021). *Brain tumor detection and classification using machine learning: a comprehensive survey*. [online] ResearchGate. Available at: [https://www.researchgate.net/publication/356000872\\_Brain\\_tumor\\_detection\\_and\\_classification\\_using\\_machine\\_learning\\_a\\_comprehensive\\_survey](https://www.researchgate.net/publication/356000872_Brain_tumor_detection_and_classification_using_machine_learning_a_comprehensive_survey) [Accessed 19 Aug. 2024].
- Anantharajan, S., Gunasekaran, S., Subramanian, T. and R, V. (2024). MRI brain tumor detection using deep learning and machine learning approaches. *Measurement Sensors*, [online] 31, pp.101026–101026. doi:<https://doi.org/10.1016/j.measen.2024.101026>.
- Bhandari, P. (2021). *Ethical Considerations in Research | Types & Examples*. [online] Scribbr. Available at: <https://www.scribbr.com/methodology/research-ethics/> [Accessed 19 Aug. 2024].
- Bohaju, J. (2020). *Brain Tumor*. [online] Kaggle.com. Available at: <https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor/data> [Accessed 19 Aug. 2024].
- Brindha, P.G., Kavinraj, M., Manivasakam, P. and Prasanth, P. (2021). Brain tumor detection from MRI images using deep learning techniques. *IOP Conference Series Materials Science and Engineering*, [online] 1055(1), pp.012115–012115. doi:<https://doi.org/10.1088/1757-899x/1055/1/012115>.
- Charlton, C.E., Poon, M.T.C., Brennan, P.M. and Fleuriot, J.D. (2023). Development of prediction models for one-year brain tumour survival using machine learning: a comparison of accuracy and interpretability. *Computer Methods and Programs in Biomedicine*, [online] 233, pp.107482–107482. doi:<https://doi.org/10.1016/j.cmpb.2023.107482>.
- Dhakshnamurthy, V.K., Govindan, M., Sreerangan, K., Manikanda Devarajan Nagarajan and Thomas, A. (2024). Brain Tumor Detection and Classification Using Transfer Learning Models. [online] doi:<https://doi.org/10.3390/engproc2024062001>.

Gaur, L., Bhandari, M., Razdan, T., Mallik, S. and Zhao, Z. (2022). Explanation-Driven Deep Learning Model for Prediction of Brain Tumour Status Using MRI Image Data. *Frontiers in Genetics*, [online] 13. doi:<https://doi.org/10.3389/fgene.2022.822666>.

Hossain, R., Ibrahim, R.B. and Hashim, H.B. (2023). Automated Brain Tumor Detection Using Machine Learning: A Bibliometric Review. *World Neurosurgery*, [online] 175, pp.57–68. doi:<https://doi.org/10.1016/j.wneu.2023.03.115>.

Hussain, S., Haider, S., Maqsood, S., Damaševičius, R., Maskeliūnas, R. and Khan, M. (2023). ETISTP: An Enhanced Model for Brain Tumor Identification and Survival Time Prediction. *Diagnostics*, [online] 13(8), pp.1456–1456. doi:<https://doi.org/10.3390/diagnostics13081456>.

Jamunadevi, C., Bharanitharan, J., Deepa, S. and Paul, J. (2023). Brain Tumor Prediction from EEG Signal using Machine Learning Algorithm. [online] doi:<https://doi.org/10.1109/icesc57686.2023.10193098>.

Kaifi, R. (2023). A Review of Recent Advances in Brain Tumor Diagnosis Based on AI-Based Classification. *Diagnostics*, [online] 13(18), pp.3007–3007. doi:<https://doi.org/10.3390/diagnostics13183007>.

Khalighi, S., Reddy, K., Midya, A., Pandav, K.B., Madabhushi, A. and Abedalthagafi, M. (2024). Artificial intelligence in neuro-oncology: advances and challenges in brain tumor diagnosis, prognosis, and precision treatment. *npj Precision Oncology*, [online] 8(1). doi:<https://doi.org/10.1038/s41698-024-00575-0>.

Khaliki, M.Z. and Başarslan, M.S. (2024). Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN. *Scientific Reports*, [online] 14(1). doi:<https://doi.org/10.1038/s41598-024-52823-9>.

Mahadevan, M. (2022). *Step-by-Step Exploratory Data Analysis (EDA) using Python*. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2022/07/step-by-step-exploratory-data-analysis-eda-using-python/> [Accessed 19 Aug. 2024].

Mahmud, M.I., Mamun, M. and Abdelgawad, A. (2023). A Deep Analysis of Brain Tumor Detection from MR Images Using Deep Learning Networks. *Algorithms*, [online] 16(4), pp.176–176. doi:<https://doi.org/10.3390/a16040176>.

Mathivanan, S.K., Sonaimuthu, S., Murugesan, S., Rajadurai, H., Shivahare, B.D. and Shah, M.A. (2024). Employing deep learning and transfer learning for accurate brain tumor detection. *Scientific Reports*, [online] 14(1). doi:<https://doi.org/10.1038/s41598-024-57970-7>.

Montantes, J. (2020). *3 Reasons to Use Random Forest Over a Neural Network—Comparing Machine Learning versus Deep Learning*. [online] Medium. Available at: <https://towardsdatascience.com/3-reasons-to-use-random-forest-over-a-neural-network-comparing-machine-learning-versus-deep-f9d65a154d89> [Accessed 19 Aug. 2024].

Nath, G., Coursey, A., Li, Y., Prabhu, S., Garg, H., Halder, S.C. and Sengupta, S. (2023). An interactive web-based tool for predicting and exploring brain cancer survivability. *Healthcare Analytics*, [online] 3, pp.100132–100132. doi:<https://doi.org/10.1016/j.health.2022.100132>.

Noia, C. di , Grist, J.T., Riemer, F., Lyasheva, M., Fabozzi, M., Castelli, M., Lodi, R., Tonon, C., Rundo, L. and Zaccagna, F. (2022). Predicting Survival in Patients with Brain Tumors: Current State-of-the-Art of AI Methods Applied to MRI. *Diagnostics*, [online] 12(9), pp.2125–2125. doi:<https://doi.org/10.3390/diagnostics12092125>.

Owonikoko, T.K., Arbiser, J., Zelnak, A., Shu, H.-K.G., Shim, H., Robin, A.M., Kalkanis, S.N., Whitsett, T.G., Salhia, B., Tran, N.L., Ryken, T., Moore, M.K., Egan, K.M. and Olson, J.J. (2014). Current approaches to the treatment of metastatic brain tumours. *Nature Reviews Clinical Oncology*, [online] 11(4), pp.203–222. doi:<https://doi.org/10.1038/nrclinonc.2014.25>.

Pålsson, S., Cerri, S., Poulsen, H.S., Urup, T., Law, I. and Koen Van Leemput (2022). Predicting survival of glioblastoma from automatic whole-brain and tumor segmentation of MR images. *Scientific Reports*, [online] 12(1). doi:<https://doi.org/10.1038/s41598-022-19223-3>.

Philip, A., Samuel, B., Bhatia, S., Khalifa, S. and El-Seedi, H. (2022). Artificial Intelligence and Precision Medicine: A New Frontier for the Treatment of Brain Tumors. *Life*, [online] 13(1), pp.24–24. doi:<https://doi.org/10.3390/life13010024>.

Rajput, S., Kapdi, R.A., Raval, M.S. and Roy, M. (2023). Interpretable machine learning model to predict survival days of malignant brain tumor patients. *Machine Learning Science and Technology*, [online] 4(2), pp.025025–025025. doi:<https://doi.org/10.1088/2632-2153/acd5a9>.

Rikan, S.B., Azar, A.S., Naemi, A., Mohasefi, J.B., Pirnejad, H. and Wiil, U.K. (2024). Survival prediction of glioblastoma patients using modern deep learning and machine learning techniques. *Scientific Reports*, [online] 14(1). doi:<https://doi.org/10.1038/s41598-024-53006-2>.

Sabeghi, P., Zarand, P., Zargham, S., Golestany, B., Shariat, A., Chang, M., Yang, E., Rajagopalan, P., Phung, D. and Gholamrezanezhad, A. (2024). Advances in Neuro-Oncological Imaging: An Update on Diagnostic Approach to Brain Tumors. *Cancers*, [online] 16(3), pp.576–576. doi:<https://doi.org/10.3390/cancers16030576>.

Saeedi, S., Rezayi, S., Keshavarz, H. and Kalhori, S.R.N. (2023). MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Medical Informatics and Decision Making*, [online] 23(1). doi:<https://doi.org/10.1186/s12911-023-02114-6>.

Santos, D. and Santos, E. (2022). Brain Tumor Detection Using Deep Learning. *medRxiv (Cold Spring Harbor Laboratory)*. [online] doi:<https://doi.org/10.1101/2022.01.19.22269457>.

Sharma, M., Sharma, P., Mittal, R. and Gupta, K. (2021). *Brain Tumour Detection Using Machine Learning*. [online] ResearchGate. Available at: [https://www.researchgate.net/publication/360224770\\_Brain\\_Tumour\\_Detection\\_Using\\_Machine\\_Learning](https://www.researchgate.net/publication/360224770_Brain_Tumour_Detection_Using_Machine_Learning) [Accessed 19 Aug. 2024].

Steyaert, S., Qiu, Y.L., Zheng, Y., Mukherjee, P., Vogel, H. and Gevaert, O. (2023). Multimodal deep learning to predict prognosis in adult and pediatric brain tumors. *Communications Medicine*, [online] 3(1). doi:<https://doi.org/10.1038/s43856-023-00276-y>.

Streefkerk, R. (2019). *Qualitative vs. Quantitative Research | Differences, Examples & Methods*. [online] Scribbr. Available at: <https://www.scribbr.com/methodology/qualitative-quantitative-research/> [Accessed 19 Aug. 2024].

Sun, L., Zhang, S., Chen, H. and Luo, L. (2019). Brain Tumor Segmentation and Survival Prediction Using Multimodal MRI Scans With Deep Learning. *Frontiers in Neuroscience*, [online] 13. doi:<https://doi.org/10.3389/fnins.2019.00810>.

Ummaneni, V. (2024). *CSUSB ScholarWorks CSUSB ScholarWorks Electronic Theses, Projects, and Dissertations Office of Graduate Studies STUDY OF BRAIN TUMOR PREDICTION BY USING MACHINE STUDY OF BRAIN TUMOR PREDICTION BY USING MACHINE LEARNING LEARNING*. [online] Available at: <https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=3147&context=etd> [Accessed 19 Aug. 2024].

Vikkurty, S., Hegde, N.P., Kumar, S.V., Recherla, A. and Ganapa, M. (2024). Effective Prediction of Brain Tumor Using Machine Learning Algorithms. *Lecture notes in electrical engineering*, [online] pp.489–497. doi:[https://doi.org/10.1007/978-981-99-7137-4\\_48](https://doi.org/10.1007/978-981-99-7137-4_48).

Yousefi, S., Amrollahi, F., Amgad, M., Dong, C., Lewis, J.E., Song, C., Gutman, D.A., Halani, S.H., Enrique, J., Brat, D.J. and Lee (2017). Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. *Scientific Reports*, [online] 7(1). doi:<https://doi.org/10.1038/s41598-017-11817-6>.

ZainEldin, H., Gamel, S.A., El-Kenawy, E.-S.M., Alharbi, A.H., Khafaga, D.S., Ibrahim, A. and Talaat, F.M. (2022). Brain Tumor Detection and Classification Using Deep Learning and Sine-Cosine Fitness Grey Wolf Optimization. *Bioengineering*, [online] 10(1), pp.18–18. doi:<https://doi.org/10.3390/bioengineering10010018>.