Informatics Institute of Technology

A brain tumor detection and classification system using Image Segmentation and Machine Learning

Module ID: 5COSC009C

Module: Software Development Group Project

Module Code: 5COSC009C.2

Module Leader: Mr. Banuka Athuraliya

Group Name: Team Thambapanni

Name	IIT ID	UoW ID
Rasula Yadithya	20191118	W1790364
Sithira Panditha	20191106	W1790355
K.C.P.L. Amarasinghe	2019744	W1790192
Manula Haputhanthrie	20191046	W1790305
Ravindu Kodithuwakku	20191039	W1790300
Udul Sree Halambage	20191149	W1790400

Table of Contents

Table of Figures

List of Tables

1. Chapter 1 – Introduction

1.1. Chapter Overview

This chapter will provide an overview of the project. It is inclusive of the project background, the research gap, the scope of the project, the aim and the proposed features.

1.2. Project Background

The identification and segmentation of brain tumors is an extremely difficult task; prone to errors and dependent on the experience of experts in some instances. While early identification is imperative, such a process may take some time, and vary depending on the physician in question. The brightness and contrast of the screen may also result in varying results and misdiagnosis. Certain similarities between brain tumors as well as normal tissue also makes providing correct diagnosis arduous. (Nasim et al., 2019)

This is an initiative to use Machine Learning and Image Recognition to automate the process of tumor detection and classification.

1.2.1. Introduction to the problem

A brain tumor is an abnormal mass of tissue present in the brain. While more than 150 types of brain tumors exist, they are mainly grouped as,

1. Primary tumor
2. Metastatic tumor

A primary tumor is a tumor which starts within the brain itself. A metastatic tumor begins somewhere else in the body and progresses towards the brain. They are grouped as **benign** or **malignant** depending on their severity. Benign tumors grow slowly and are comparatively harmless, but can be dangerous if located in a vital area. Malignant tumors spread more quickly and are life threatening. (American Association of Neurological Surgeons, no date)

The brain is unable to accommodate for the growing mass, and this results in displaced brain tissue and a blockage in fluids around the brain. This increases the intracranial pressure and can cause the ventricles to enlarge. This swelling causes many of the symptoms seen. (Warnick, McPherson, Gozal, 2

The correct treatment for these different kinds depends on the type of tumor. And therefore, the early and accurate identification of the tumor is the first step in treatment.

And while MRI scans are one of the two most frequently used testing methods, the detection process and time may change due to different circumstances. And the results may depend on even on minute factors like the brightness and contrast of the display screen. (Md. Ariful Islam et al., 2020)

An automated system would provide the means to identify the presence of a tumor, and also to identify the size, shape and location as well, while being significantly more accurate and quicker.

1.2.2. Examples in the problem

1. Delayed diagnosis of patient
2. Complications in the patient's health
3. Insufficient time to take action after prognosis

1.2.3. Attempted Solutions of the Competitors

Many existing brain tumor detection applications are not publicly available. A selected few were explained within research papers but none were present to be tested personally.

'Prometheus' as explained in https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3159221/ and, 'BraTumIA' as shown in https://www.nitrc.org/projects/bratumia both provide insights into our project, but are not publicly available. (Vijayakumar and Gharpure, 2011) (NITRC: BraTumIA, no date)

The research papers we found proposed similar ideas, but do not have any proper implementations. Sometimes existing methods can be classically divided into region-based and contour-based methods. Sometimes this may happen due to low accuracy. So, the given table represents the feature comparison of proposed projects with other research papers.

	Proposed System	(Isselmou, Zhang and Xu, 2016)	(Lavanyadevi, R., et al., 2017)	(Dr Samir Kumar, 2011)
Find out whether brain is present or not by processing MRI images	✔	✔	✔	✔
If positive, finding out whether it is in primary stage or not	✔	-	✔	-
Predict the type of the brain tumor (ex: glioma)	✔	-	-	-
Find out the exact position of the tumor (ex: frontal lobe)	✔	-	-	-
Available as web application	✔	-	-	-

Table 1: Feature Comparison Chart

1.2.4. Problem Definition

Inaccuracies when diagnosing MRI scans due to human error as well as excessive time being taken to reach a final conclusion prove to be ineffective.

1.2.5. Research Questions

This research focuses on solving the following questions.

RQ1: Can a CNN and image segmentation identify the presence of a brain tumor?

RQ2: Can image segmentation differentiate between the varying types of brain tumors?

RQ3: How does this research overcome the problems of the current implemented strategies?

1.3. Aim

To identify and classify a brain tumor by using machine learning and processing MR images.

1.4. Scope

The scope of the project is defined as,

1.4.1. In-Scope

The area that will be focused within the project,

1. Using MRI scans to detect the existence of a tumor
2. The tumor will be defined as,
    1. Benign
    2. Malignant
3. The type of tumor will be identified according to three types as,
    1. Meningioma
    2. Glioma
    3. Pituitary Tumor
4. Identify the location of the tumor within the brain

1.4.2. Out of Scope

The factors that are excluded from the project,

1. The tumor will not be categorized according to its grade. (i.e., Grade I, Grade II etc.)

1.5. Objectives

1.5.1. Research Objectives

1. To identify the presence and type of a brain tumor using ML and CNN
2. To identify the most accurate CNN architecture for MRI tumor detection

1.5.2. Academic Objectives

1. To learn various architectures and approaches to ML

2. To improve team work and management skills
3. To enhance critical thinking and problem-solving skills

1.5.3. Operational Objectives

Objective 1 – Data Collection

- Gathering data from publicly available medical sites

Objective 2 – Revision of existing work

- Identifying previously used approaches and methodologies in related works
- Evaluating existing solutions
- Identifying the most appropriate approach for this research
- Submitting the literature review

Objective 3 – Design

- Identifying the project requirements
- Identifying a suitable design approach
- Prepare design document
- Submission of design document

Objective 4 – Implementation

- Development of prototype using aforementioned approaches and technologies

Objective 5 – Testing

- Testing the accuracy of the prototype
- Reporting test results

1.6. Proposed features

1. Identify presence/absence of tumor
    - The preliminary step would recognize if a tumor is present within the brain.
2. Identify location of tumor

1.7. Resource Requirements

1.7.1. Hardware Requirements

- Desktop with i7 processor and 16GB RAM to train the model
- SLT Fiber Router

1.7.2. Software Requirements

- Python – For data manipulation
- HTM5, CS3, JavaScript, Angular – For development of web application
- Git – For version control
- GitHub – For storing source code
- Word – For documentation

1.8. Chapter Summary

This chapter provided an introduction to the problem domain. The aim, objectives and features of the proposed system were also identified.

2. Chapter 2 – Literature Review

2.1. Chapter Overview

This chapter will focus on the problem domain and currently existing methods and techniques in place, to better understand their pros and cons, to find the most effective implementation method for this system.

2.2. Segmentation techniques used in related research

Different techniques have been used in various researches to find the most suitable method of image segmentation in MRI scans. Edge detection is the most widely used approach and uses detection of abrupt local changes in the intensity of an image to identify edges. (Aslam, Khan and Beg, 2015)

The following section will discuss the most common methods used.

2.2.1. Sobel edge detection technique

The Sobel operator uses a differential operator which utilized two 3x3 kernels for the x and y directions.

The magnitude of each gradient is obtained by,

$$G = |G_x| + |G_y|$$

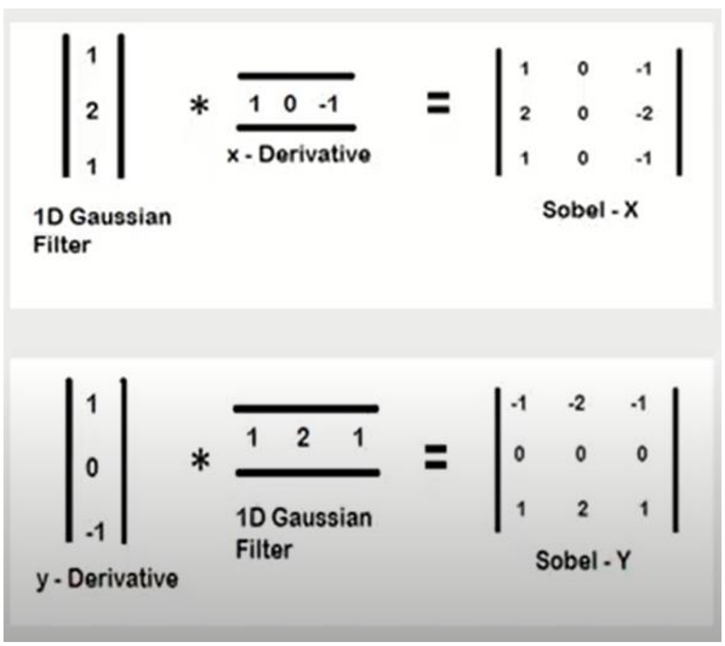


Figure 1: Sobel Matrix formation

Its main advantage is the simplicity it offers due to the approximate gradient calculation. (Chandwadkar et al., no date)

However, edge detection using the Sobel technique produces inaccurate results with the increase of noise. This is due to the smoothing effect (Gaussian smoothing) which is implemented to make the image less sensitive to the noise which is present. And while this does provide a smoother image, it affects the edge detection accuracy.

This makes the Sobel operator 'adequate' for numerous applications, but not for use cases where thin and smooth edge detection is vital. Therefore, the implementation of the Sobel edge detection technique would not provide accurate results if noise is present in the MRI scan.

2.2.2. Edge detection by traditional Canny algorithm

The Canny operator is an improvement of the traditional single threshold method like the Sobel operator which is sensitive to noise. It is a step-by-step process.

1. Apply Gaussian smoothing and reduce noise
2. Compute the gradient magnitude
3. Non-maximum suppression
4. Edge tracking by hysteresis (Stosic and Petar, 2018)

Here, hysteresis uses two threshold, namely upper and lower. A pixel gradient higher than the upper gradient will be marked an edge while anything lower than the lower gradient will be removed. If the gradient is in between, then only the pixel that is connected above the upper threshold is marked as an edge. (Kim, 2013)

However, the complex computation behind the Canny algorithm makes it time consuming and unsuitable to obtain real time results.

2.2.3. Input Cascade Convolutional neural network

This is a deep learning-based technique. This is one of the different types of CNN architecture. This method is 40-fold faster than the other CNNs. There are 2 streams in the Input Cascade CNN.

1. Local features - This is accessible for 7 * 7 fields
2. Global features - This is accessible for 13 * 13 fields

In the first step, getting the local features & the highest and lowest intensities are suppressed by using an algorithm. Then the data is taken from each input channel and normalized by using mathematical techniques. After these two steps, this system applies the post-processing step to remove the noise from the segmented image.

In this post-processing system, the algorithm which is known as the "Connected components labeling algorithm" is used to remove the flat blobs that may appear in tumor regions due to the bright corners of the brain or skull.

So, as a result of the above process, this Input Cascade CNN technique is able to implement a segmented area of the tumor.

2.2.4. Watershed Segmentation

This segmentation method is a gradient-based segmentation technique. It considers the image's gradient map as a gradient map for relief. It segments the magnetic resonance image (MRI) as a dam. In this system, the brain image has been categorized into a normal brain image and an abnormal brain image. In less execution time, this process will correctly classify the brain image. The average filter provides the required smooth image for the step of skull stripping. But the skull tissues can leave in the brain image because of a drawback of the skull stripping method. It can be overcome using morphological operation.

Then, the MRI brain tumor image is effectively segmented by the marker-controlled watershed segmentation process. This process is reducing the over-segmentation issue and provides the required segmentation map for tumor detection. Detection of the tumor by morphological operation provides the exact result of the tumor area. The equation of the tumor area in this research is given below. It is acceptable for any shape of the tumor area. Also, by this system the tumor location can be displayed effectively.

$$Tumor\ Area = A \times total\ number\ of\ pixels\ in\ the\ tumor\ region$$

$$A = V \times H$$

Where,
A = the area of each pixel
H = horizontal dimension of the image
V = vertical dimension of the image
H = 1/horizontal resolution of the image
V = 1/vertical resolution of the image

2.2.5. K - Means Clustering

K- means clustering method is one of the easiest ways to cluster brain MRI data. K groups are initially classified by the data k groups in the cluster. K initial center is identified randomly. The object is allocated to the centers near them. In each center, the means of all objects are marked and classified as new centers. this process repeated until all objects are converged in a cluster. Evaluation of K-means and FCM have been modelled on T1 contrast axial plane MR images for segmentation of brain tumor with histogram guided initialization of cluster. (Nimeesha and Gowda, 2013). An experiment conducted has shown that FCM identifies the vasogenic edema and the white matter as a single tissue class and gray matter and necrotic focus as a single class as well. FCM only finds 3 tissue classes but K means identifies all 6 classes. So, when comparing K means and FCM, K means' ability to cluster the tissue regions are comparatively higher than FCM.

2.2.6. Brain tumor segmentation based on a hybrid clustering technique

There are some image segmentation systems which use K-means algorithms or Fuzzy C-means algorithms. The K-mean algorithm is fast and simple to run on large datasets but this K-means algorithm causes a problem called incomplete detection of tumor when it comes to a malignant tumor. But the Fuzzy C-means algorithm can retain more information of the image and detect the tumor cells compared to the K-means. These systems are much sensitive to noise and outliers of the images and they take some time to execute the process.

In this hybrid clustering technique, they take the benefit of both of the algorithms. This system consists of four stages: Pre-processing, Clustering, Tumor extraction and contouring, validation stages. According to the experiments the hybrid clustering method (KIFCM - kernel-distance - based intuitionistic fuzzy c-means clustering) can detect a tumor that cannot be detected by Fuzzy C-means with less execution time.

Pre-Processing stage:
In this stage It improves the image quality and removes the noise by using several initial processing procedures on the image. Since medical images should be of minimum noise and maximum quality, they use two different techniques; De-noising and Skull removal.

DE-Noising:
Usually, MRI images are disturbed by Gaussian noise and Poisson noise. To remove noise the use a Median Filter. Median Filtering is a nonlinear filter that is used as an effective method used to remove noise while preserving edges. It works by moving pixel by pixel through the image, replacing each value with the median value of neighboring pixels.

Skull removal:
Image background doesn’t contain any useful information but it increases the processing time. Therefore, removing the background Skull and other unnecessary structures can decrease the amount of memory and increase the processing speed. Skull removal is done by using BSE algorithm (Brain Surface Extraction). The BSE algorithm is used only with MRI images and it removes the irregularities, detects edges in the images and performs morphological erosions and brain isolation.

Clustering Stage:
By de-noising the MRI image and removing skulls, the images are fed to KIFCM technique by initializing cluster numbers k, max iterations, and termination parameters. The cluster centers are calculated by:

$$MU = (1:k) * m / (k + 1)$$

where MU is the initial means that can be calculated due to k. k is the number of clusters and m is defined as:

$$m = max\ (MRI\ image) + 1$$

Extraction and contouring stage:
In this stage, there are two methods in use; Segmentation and active contour level set method.

Thresholding segmentation:
This is an Intensity-based segmentation. It used to extract the object from the background. The segmented image which obtained by thresholding, has advantages of smaller storage space and fast processing speed. The output of this stage is segmenting images with a dark background and lighting tumor areas.

Active contour by level set:
Active contours have been used for image segmentation and boundary tracking. The basic idea is to start with initial boundary shapes represented in a form of closed curves(contours) and modify them by applying expansion or shrink operations according to the constraints. The clustering images are entered to the binarization process using an inverse thresholding method with iteration equals 3. The noise of the image is removed by using the median filter that eliminates the small regions that are far away from the tumor cluster. This step is the post processing step of the system.

Validation Stage:

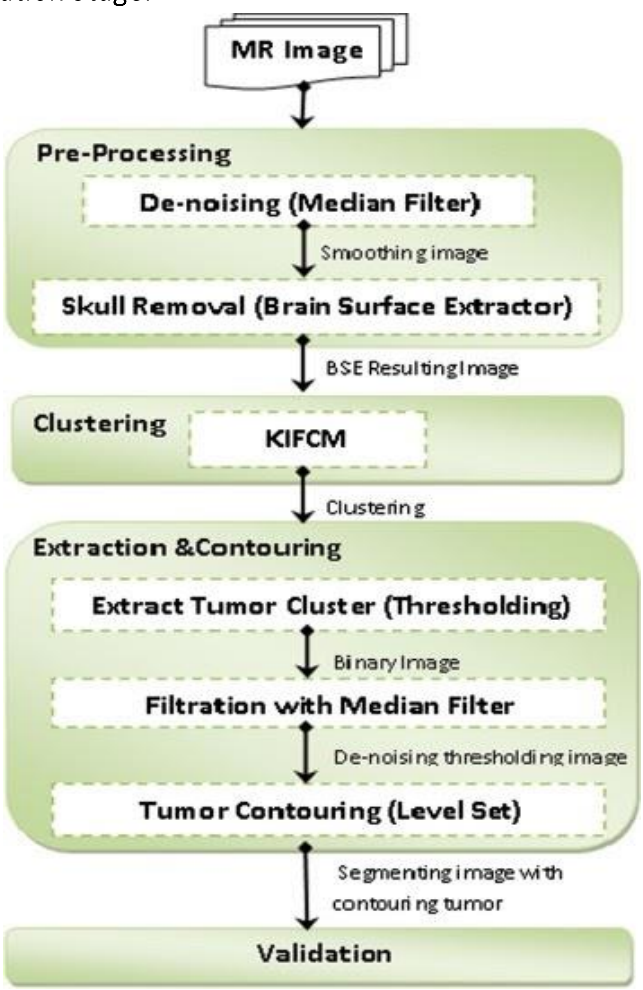


Figure 2: Hybrid Clustering technique

2.2.7. Region Growing Segmentation

This is a region-based extraction method. It selects regions with similar pixels. It firstly selects initial points know as seed points. Then the surrounding pixels of said seed points are examined to see whether they should be added to the region. They are added if they are similar to the seed point. This process is repeated until no more seeds can be added. The region growing segmentation method has been improved upon in Lina et al. (2013), where fuzzy edge technique has been used to identify the initial seed points. The main pros of the region growing segmentation method is that regions with similar pixels are generated. (Angulaskhmi and Priya, 2017) However the major downside would be the initial seed points being sensitive to noise.

2.2.8. Morphological Based Algorithm

This is based on the morphology of features of the image. This method is mainly used to get information from the image based on the representation of the shape. Erosion and Dilation are basic operations. Dilation is used to dilate the size of the brain tumor. This method includes several steps to extract tumors from the MRI images, it includes resampling of image, enhancement of image, histogram application, color plane extraction, and advanced morphological operations to extract brain tumor region. Morphological operations are mainly used to filter to remove low frequency pixels and boundary pixels. This method can calculate the area of the tumor and also the length of the tumor.

2.2.9. Genetic Algorithm

The Genetic Algorithm (GA) is mainly based on natural evolution. GA's natural evolution is based on a search method that optimizes the structure it creates. The chromosomes are used in GA to classify the population of individuals. Population of individuals is updated by using mutation, cross over and selection operators so the population of individuals is updating iteratively. To optimize the fitness function used for assessing each population. (Chandra and Rao, 2016)

The GA algorithm is used via evaluation criteria to optimize the segmentation results of brain tumor from the MRI image. In this method k-means algorithm clusters are used as the initial population. Centers that are clustered are evaluated by a fitness function. Using different selection factors, such as crossover and mutation the weaker chromosomes are then replaced by stronger ones. GA's primary benefit is its high performance in complicated search problems. (Holland, 1992)

Genetic algorithms are randomized, the techniques of search and optimization are motivated by the concepts of evolution and natural genetics, with a large number of implicit parallelism degrees. GA's execute searches in complex ecosystems, broad and multimodal and provide near-optimal solutions for the goal or stress function of a problem with optimization.

In GA's, the search space parameters are encoded in the form of string(chromosomes). A set of such strings is known as a population. Initially a random population representing various points in the search space. An objective and fitness function are associated with each string that represents the degree of goodness of the string. Relying on the principle of the survival of the fittest, picking and selecting a few of the strings. A number of copies are allocated to each one that goes into the pool for mating.

2.3. Research Gap

By integrating software and technology many scientists and engineers try to save the lives of the people by identifying diseases at very early stages. Diagnosis of tumors by professionals is a time-consuming task and as a result, people are not able to take treatment immediately.

Research has shown the ability to find brain tumors using image recognition using methods such as explained in section 2.2. However current systems are limited to identifying whether the tumor is either malignant or benign. It does not classify according to the tumor type. (i.e., glioma or meningioma)
Thus, the proposed solution is a system stating the type of brain tumor and location within the brain. This would help with a far more effective diagnosis of the tumor.

Image segmentation and edge detection would be the main techniques used to identify the type of the tumor.

2.4. Research on Approaches and Techniques

Data augmentation would be used on the dataset to increase the number of MRI scans and to make training the neural network more reliable. The data would then be preprocessed and resized to the same size before being fed to the neural network.
The data would be split as,

1. Data for training
2. Data for validation
3. Data for testing

2.4.1. Selection of a transfer learning model

The purpose of this research is not to reinvent the wheel. Therefore, a pre-trained module will be selected to apply transfer learning to decrease the training time of the neural network. While dozens of pre-trained modules in this regard exist, three of the most common would be,

1. VGG
2. GoogLeNet
3. Residual Network

The selected module would be VGG-16.

VGG-16

This is a deep convolutional neural network. Due to extensive training on the ImageNet database, it gives accurate results even when the data sets are small. The network consists of 16 convolutional layers and a 3 x 3 receptive field. It also has five 2 x 2 Max pooling layers. However, a major downside would be that training a network from scratch would be extremely slow. (Theckedath and Sedamkar, 2020)

2.5. Chapter Summary

This chapter gave an overview of existing methods, technologies and their used techniques and architectures. The most popular edge detection techniques were evaluated in detail. The next chapter will focus on the project management aspect of the project and will give an in-depth analysis of the methods used.

3. Chapter 3 – Project Management

3.1. Chapter Overview

This chapter focuses on the research, software development and project management methods which were considered when managing this project.

3.2. Software Development Methodologies

The models defined for Software Development are followed to make the planning and implementation of any program simpler and straightforward. A few major models would be looked at to select the most viable one.

3.2.1. Waterfall Model

The Waterfall Model consists of several phases. These phases are followed sequentially, where progress is seen as flowing steadily downwards, similar to a cascading waterfall. A phase only begins when the previous phase has been completed and the desired targets have been achieved. Therefore, phases do not overlap. This allows for better control and clear-cut phases within the development project. Additional advantages include this being a simple method to understand, presence of well-defined milestones and proper documentation throughout the development. But it falls short for the implementation in any research related project due to the fact that it does not allow any revision. This method works best for smaller projects where all the requirements have already been identified and noted. It is not suitable for a project with a possibility of change, and especially a project which is research driven.

3.2.2. Rapid Application Development Model

The Rapid Application Development (RAD) model is focused on providing the software as quickly as possible. It has minimal planning. This would require a team of highly skilled individuals, which would not be the case in a research driven team who will be learning as they go.

3.2.3. Iterative Model

This model focuses on designing the complete system in iterations, focusing on a subset on each turn, and enhancing it till the desired model is reached. Every iteration will take this through the requirements, design, implementation and testing phases. This model would be suitable for projects where the major requirements have been understood, but functional requirements have room to be added or changed. Since this is an iterative process, a functioning prototype will be available early on. This makes it easier to improve and find flaws. This would also allow for parallel development, where multiple parts could be worked on at once. And since working models would be tested periodically, it will provide results constantly. Risks are also identified during development and can be fixed during the next iteration. However, if some requirements are not present during the initial stages, like a design or selected architecture, the project cannot proceed forward. This would be suitable for a research-based project where the basic requirements have already been noted.

Figure 2: Iterative Model

3.2.4. Selecting a software development model

Considering the above-mentioned methods, the iterative method was selected as most suitable for the project at hand. The Waterfall Model was disregarded due to its linear nature, inability to have a working model early on and it being difficult to make changes to the software. The RAD model was dismissed due to the high skill level required as well as lack of proper documentation which is needed in a research driven project. Therefore, the model which was selected is the iterative model, which provides a working model early on, is flexible to change and additional requirements, and allows the programmer to learn while developing.

3.3. Programming Methodologies

When a program is developed as a solution to a problem, they might end up being huge and complex. Therefore, the development process of the program should be controlled. These different approaches to analyzing and developing a software are known as programming methodologies.

3.3.1. Procedural Programming

The solution is broken into smaller blocks of code, where one block is aimed at performing one task. All these blocks combined form the entire program. This method is only suitable for programs which are quite simple. It would not be suitable for a software with many functional requirements.

3.3.2. Object Oriented Programming

Here the solution focuses on data as entities and not as functions. How these entities interact with each other and contribute to solving the problem should be thought of.

3.3.3. Functional Programming

The solution is broken into smaller functional units, where one unit performs a specific task. These units are also self-sufficient. These units are brought together to make the final solution.

3.3.4. Selecting a programming method

While all the methods were considered, the selected approach was Object Oriented Programming due to several key reasons.

1. It provides reusability of classes otherwise unavailable in other methods.
2. More secure due to data abstraction and encapsulation provided by OOP
3. Provides easier troubleshooting

3.4. Project Management Methodologies

A project management methodology is a framework that allows you to optimally manage the project. There are several types of methods to consider.

3.4.1. Waterfall Methodology

Similar to the Waterfall model discussed in software development methodologies, this is also a method where tasks are completed in a linear and sequential fashion. It has clearly defined steps as,

1. Requirements
2. Analysis
3. Design
4. Construction
5. Testing
6. Deployment and maintenance

Since this follows a linear method, all the parts of the project should be clearly defined before the start. It is suitable for projects where the end goal is defined, and the ideas and implementation will not change in between. It is not suitable for projects which are liable to change, or where adaptation would be required to accommodate additional requirements, or if rigorous testing would be needed.

3.4.2. Scrum Methodology

This is a form of agile project management, which is a more iterative model which allows the project to change as it progresses. Here the work to be done is broken down into short cycles known as “sprints”. All the work to be completed is present within a backlog. The tasks are prioritized by the project manager.

3.4.3. Kanban Methodology

This is another agile framework where tasks are visually represented on a Kanban board. The board represents stage of the process like 'To Do', 'In Progress' and 'Done'. Work is taken from the backlog and according to the team's availability and completed. This method is suitable to always have a visual representation of the progress.

3.4.4. Lean Methodology

This is a project management methodology which focuses on minimizing waste. It highlights in reducing wastefulness, unevenness and overburden. This method is suitable for projects trying to reduce costs and add value to the customer.

3.4.5. Selecting a project management methodology

Considering the above-mentioned methods, the methodology that was selected was the Scrum method. The division of the tasks into more manageable parts, the prioritization of tasks and the effective use of time make this method suitable.

3.5. Risks and mitigations

Description		Probability	Risk Level	Mitigation
R1	Insufficient knowledge and expertise in development techniques and methods	High	High	Consulting domain experts Engaging in more lessons and online tutorials
R2	Loss of data	Medium	High	Saving all related data on the cloud Backing up every day
R3	Problems with implementation	High	High	Focusing on the most important requirements that provide core functionality and get them fixed first
R4	Failing to meet deadline	High	High	Putting in extra work Refining the project scope
R5	Getting sick or personal matters	Low	High	Dividing the work

3.6. Activity Schedule

Date	Activity	Time Frame
20/10/2020	Finalize project idea	3 weeks
27/10/2020	Submit initial project proposal	1 week
10/11/2020	Literature Review initial submission	2 weeks
22/11/2020	Submit Introduction chapter	1 week
22/11/2020	Literature Review final submission	1 month
01/12/2020	Present draft SRS	1 month
08/12/2020	Present draft project report	2 months

Table 2: Activity Schedule

3.7. Work breakdown structure

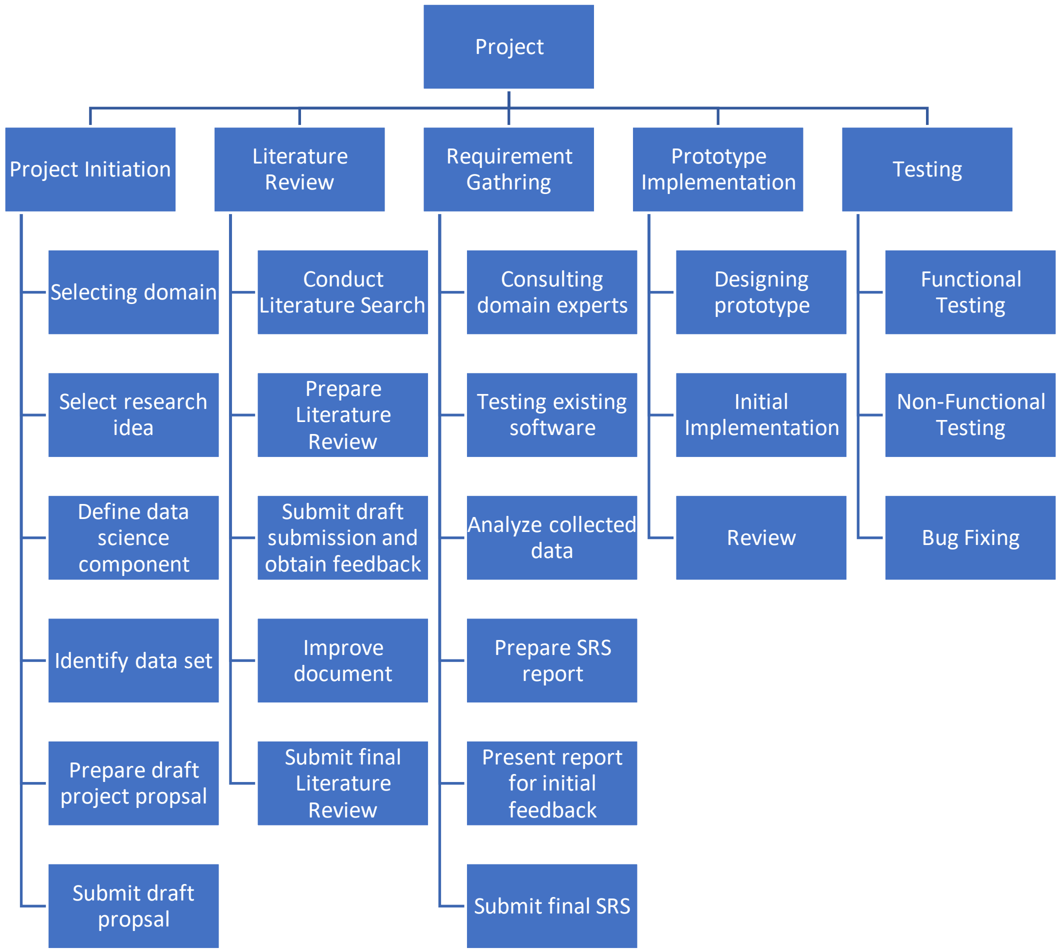


Figure 3: Work breakdown structure

3.8. Gantt chart diagram

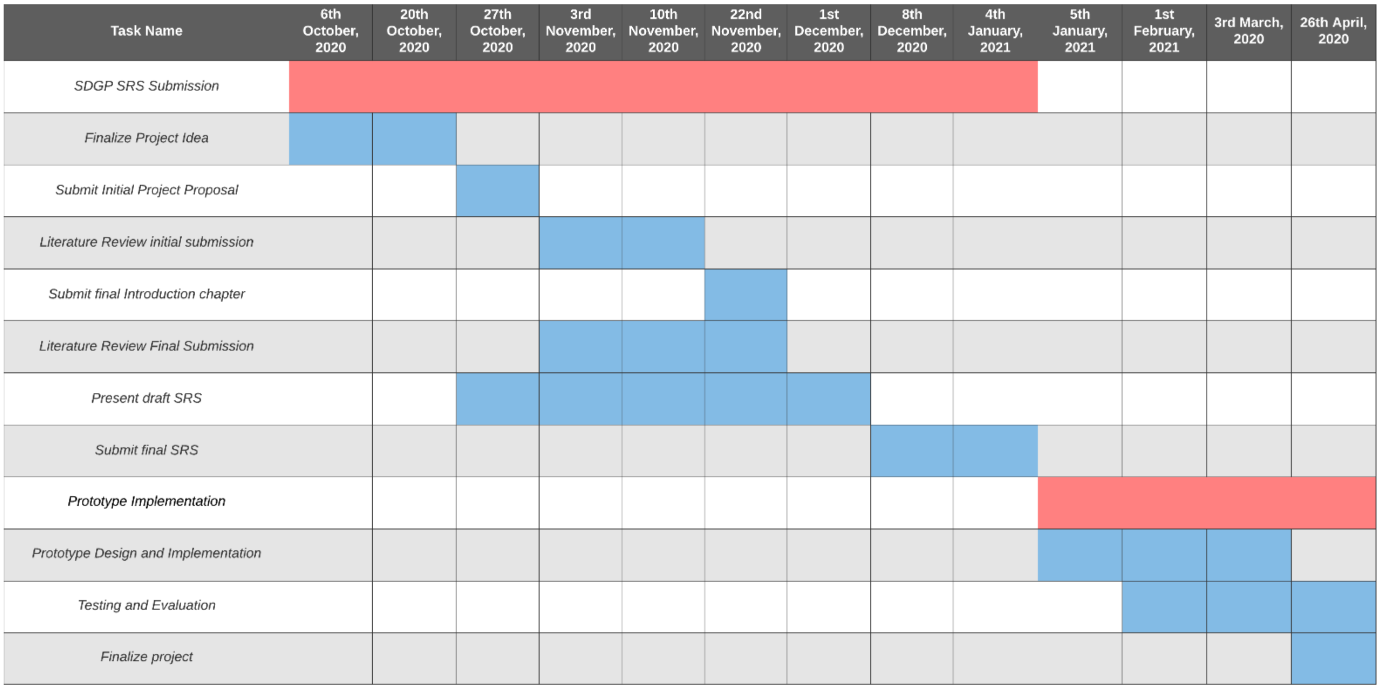


Figure 4: Gantt chart

3.9. Chapter Summary

This chapter gave an overview of the existing software development models and the model which was selected. It went over the risk involved, the activity schedule, the work breakdown structure and Gannt chart. The next chapter which is the System Requirements Specification, will focus on the requirements of the project.

4. Chapter 4 – System Requirements Specifications

4.1. Chapter Overview

The following chapter elaborates on the system requirements specifications of this project. It focuses on the stakeholders, requirement gathering techniques with the functional and non-functional requirements used.

4.2. Stakeholder Analysis

4.2.1. Onion Model

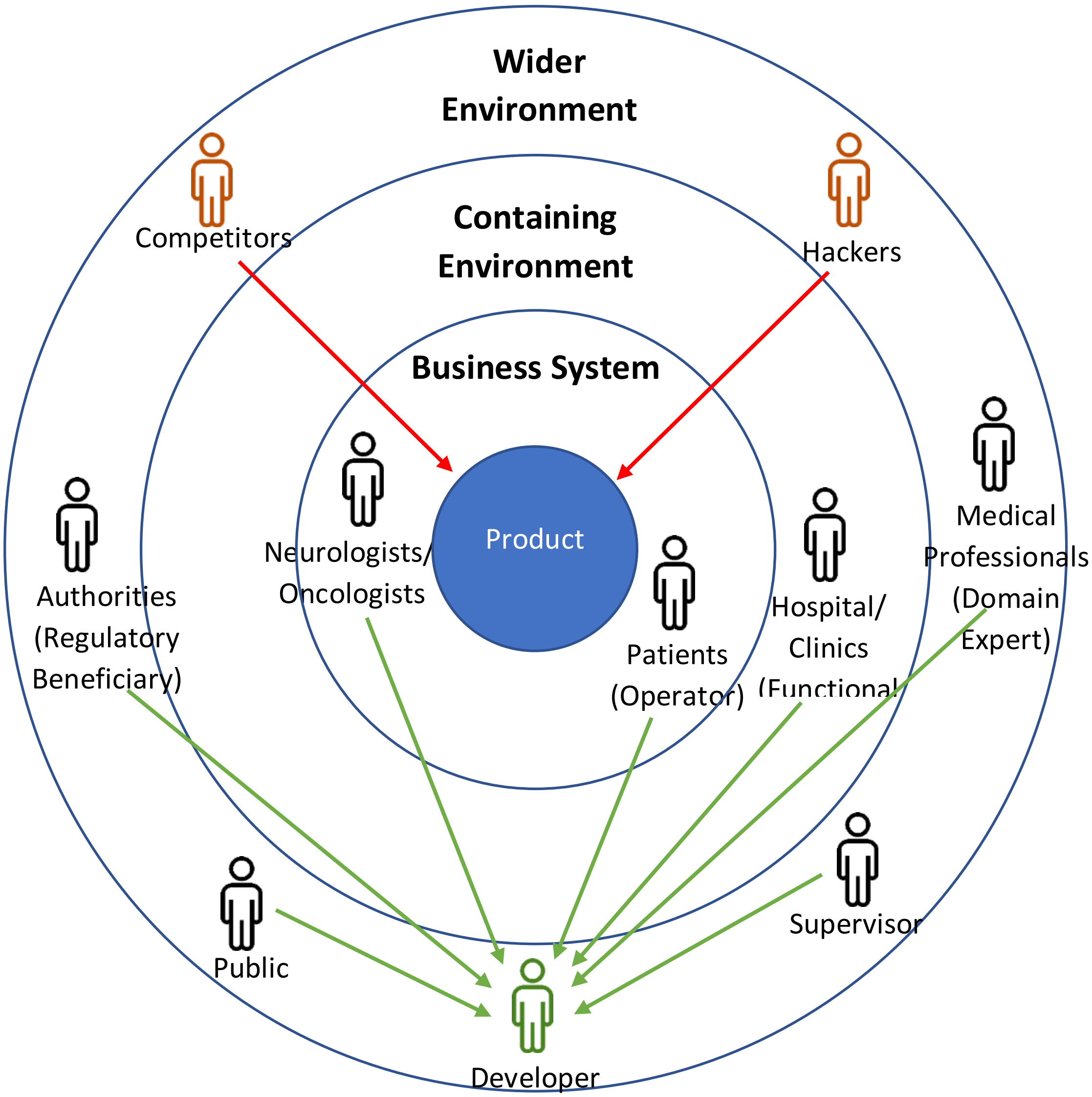


Figure 5: Onion Model

4.2.2. Stakeholder Descriptions

Stakeholder	**Viewpoint**
Functional Beneficiary	
Neurologists, Oncologists	Will utilize the system to get accurate classification of brain tumors
Patients	Will utilize the system to check the results of the MRI scan
Financial Beneficiary	
Hospitals, Clinics	An increase in revenue due to correct and quicker diagnosis of brain tumor patients
Negative Stakeholders	
Competitors	Will try to provide a similar system with even better features
Hackers	Will try to hack into the system, change data, and try to stop the functionality of the system
Regulatory	
Authorities	Making sure the system works as expected and that readings are accurate
Experts	
Medical Professionals	Provides insight into the domain as they have more knowledge regarding it

Table 3: Stakeholder Descriptions

4.3. Requirement Gathering

Requirement gathering is used to collect the data and knowledge which will prove useful to the successful implementation of the proposed system.

4.3.1. Techniques for requirements gathering

1. Brainstorming
   This is a collection of likeminded individuals specializing in the domain, colleagues or even friends to bounce off ideas which could be used to improve the system and provide additional functionality. It could provide insights into previously unthought fallacies and gaps within the project.

2. Literature Review
   Similar systems which have already been implemented within this domain will be researched on and compared to. This will allow to better understand the functionality of existing systems and the features offered. It will provide the knowledge of the most suitable techniques to use when it comes to implementation

3. Questionnaires

   A questionnaire is a series of questions usually directed at a target audience or domain experts. They will provide many responses within a short period, and provide knowledge on features expected by clients and users. Although this would not be as effective as a direct conversation or interview, it does provide valuable insights.

4. Interviews with domain experts

   An interview is a more open ended and personalized method of obtaining answers. And although the answers would be subjective from person to person, they would provide a more detailed and elaborate look into the domain than a questionnaire. Interviews would be of two types as,

   a. Formal interviews
   b. Informal interviews

   While both these methods would be personal, a formal interview will provide the opportunity of a face-to-face interaction. An informal interview, preferably over a video call or phone call would provide an interaction with people who are far away, or unable to meet directly, especially given times where social interaction should be kept at a minimum.

5. Observation

   Observing how the current systems work, as well as the domain and its users, can garner information which can be used to improve the system.

4.3.2. Execution of requirement gathering techniques

4.3.2.1. Literature review on existing systems

By conducting an extensive literature review on brain tumor segmentation and classification techniques, edge detection and image segmentation, the following points were noted.

- Most proposed systems are not publicly available, and none allow direct interaction to the patient.
- Existing systems do not specify the location of the tumor and are limited to stating whether the brain is tumorous or not.

4.3.2.2. Interviews

An informal interview was conducted to get an idea on existing systems in place.
Interviewee: Dr. S.T. Liyanage

Aims:

- To understand the existing procedures in place for brain tumor diagnosis
- To understand the procedures in place for submitting to an MRI scan
- To get an idea of how long an MRI scan takes to reveal a diagnosis
- To recognize how important an MRI is to diagnose a tumor

Findings:

- MRI scans provide a detailed structure of the brain, but are only referred if a CT scan proves suspicious
- MRI scans are very limited in Sri Lanka and are only allowed by referral from a consultant
- The main issue for a delay lies in the minimal number of machines and hence a lengthy waiting list being present
- An MRI is a must to diagnose brain tumors currently

4.4. Models

4.4.1. Use Case Diagram

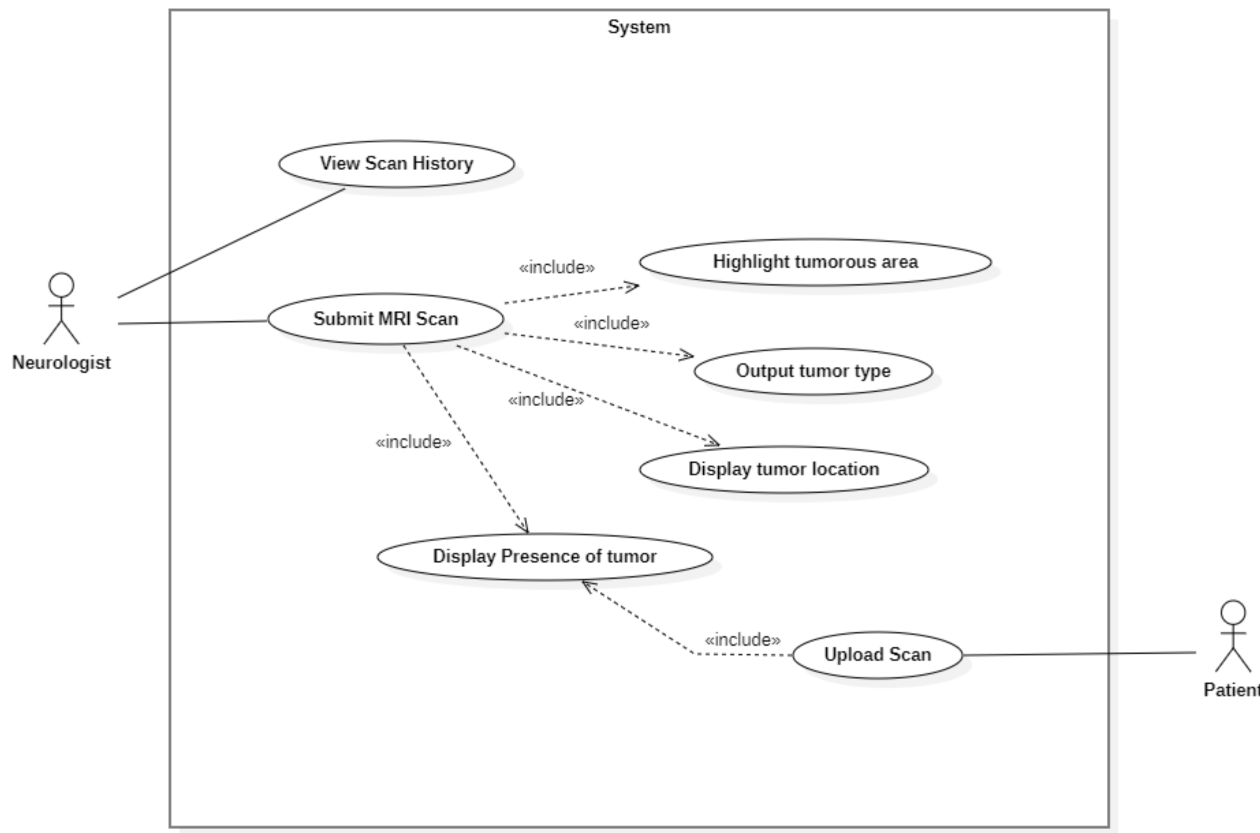


Figure 6: Use Case Diagram of proposed implementation

The neurologist can submit an MRI scan to be analyzed. They can then obtain the results of the scan and see if a tumor is present and where it is.
A patient has the ability to upload a scan for the neurologist to view.

4.4.2. Use Case Description

Use case descriptions related to two use cases will be discussed. They are,

1. Submit MRI scan
2. View Scan History
3. Upload scan

Use Case	Submit MRI Scan	
Description	An MRI scan is submitted to be analyzed by the system	
Actors	Neurologist	
Pre-Conditions	An MRI scan in a suitable format read by the application	
Extending	-	
Including	1. Highlight tumorous areas 2. Output tumor type 3. Display tumor location	
Main Flow	*Admin* 1. User opens application 2. User selects 'Submit Scan' option 3. User browses file to submit	*Device* 4. System checks whether file is of a valid format 5. Accepts file
Alternative Flow	4.a	INVALID FILE FORMAT .1: System prompts to resubmit a file
Exceptions	Cancel the scan	
Post-Conditions	If successful displays scan results, else loads home screen	

Table 4: Submit MRI scan Use Case Description

Use Case	View Scan History	
Description	A list of previous scans with their results are displayed	
Actors	Neurologist	
Pre-Conditions	Connected to the database	
Extending	-	
Including	-	
Main Flow	*Admin* 1. User opens application 2. User selects 'View History'	*Device* 3. System checks for previous scans 4. System prints results
Alternative Flow	3.a	NO HISTORY FOUND .1: System alerts no scans were found
Exceptions	-	
Post-Conditions	If successful displays history	

Use Case	**Upload Scan**	
Description	An MRI scan is submitted to be analyzed by the system by the patient	
Actors	Patient	
Pre-Conditions	An MRI scan in a suitable format	
Extending	-	
Including	-	
Main Flow	*Admin* 6. User opens application 7. User selects ‘Submit Scan as Patient’ option 8. User browses file to submit	*Device* 9. System checks whether file is of a valid format 10. Accepts file
Alternative Flow	4.a	INVALID FILE FORMAT .1: System prompts to resubmit a file
Exceptions	Cancel the scan	
Post-Conditions	If successful displays scan results, else loads home screen	

Table 5: Upload Scan Use Case Description

4.4.3. Domain Model

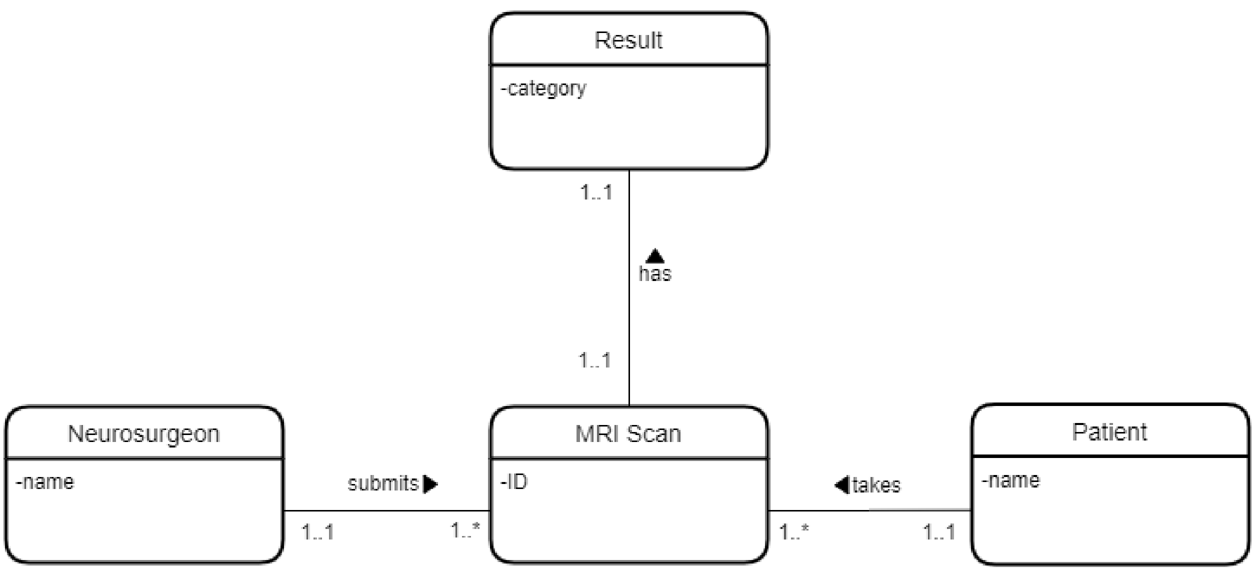


Figure 7: Domain model for proposed system

A neurosurgeon can submit one or more MRI scans. A scan can only be linked to one patient. But a patient can take more than one scan. A scan can only have one result.

4.5. Functional requirements

Requirement		Priority	Description
FR1	Obtain MRI scan and pre-process	Critical	The system should be able to obtain the desired part of an inputted image
FR2	Categorize according to tumor	Critical	Should have the ability to identify a tumor and categorize according to the type
FR3	Display tumor location	Critical	The system should be able to state the location the tumor is present
FR4	Identify tumorous areas and highlight	Desirable	An outline of the tumor and a highlighted region should be available
FR5	Search for previous scans and results	Luxury	An extensive scan history should be displayed
FR6	Patient shares uploaded MRI scan with doctor	Luxury	A patient who uploads an image from home has the ability of sharing it with the doctor via the application

Table 6: Table of Functional Requirements

4.6. Non-functional requirements

Requirement		Priority	Description
NF1	Accuracy	Critical	The system should be at least 97.87% accurate when identifying the brain tumor and its type
NF2	Performance	Critical	The system should work without any performance issues and provide the required output within a short time
NF3	Reliability	Critical	The system should be reliable and run without crashing
NF4	Usability	Desirable	The system should be easy to use and understand

Table 7: Table of Non-Functional Requirements

4.7. Chapter Summary

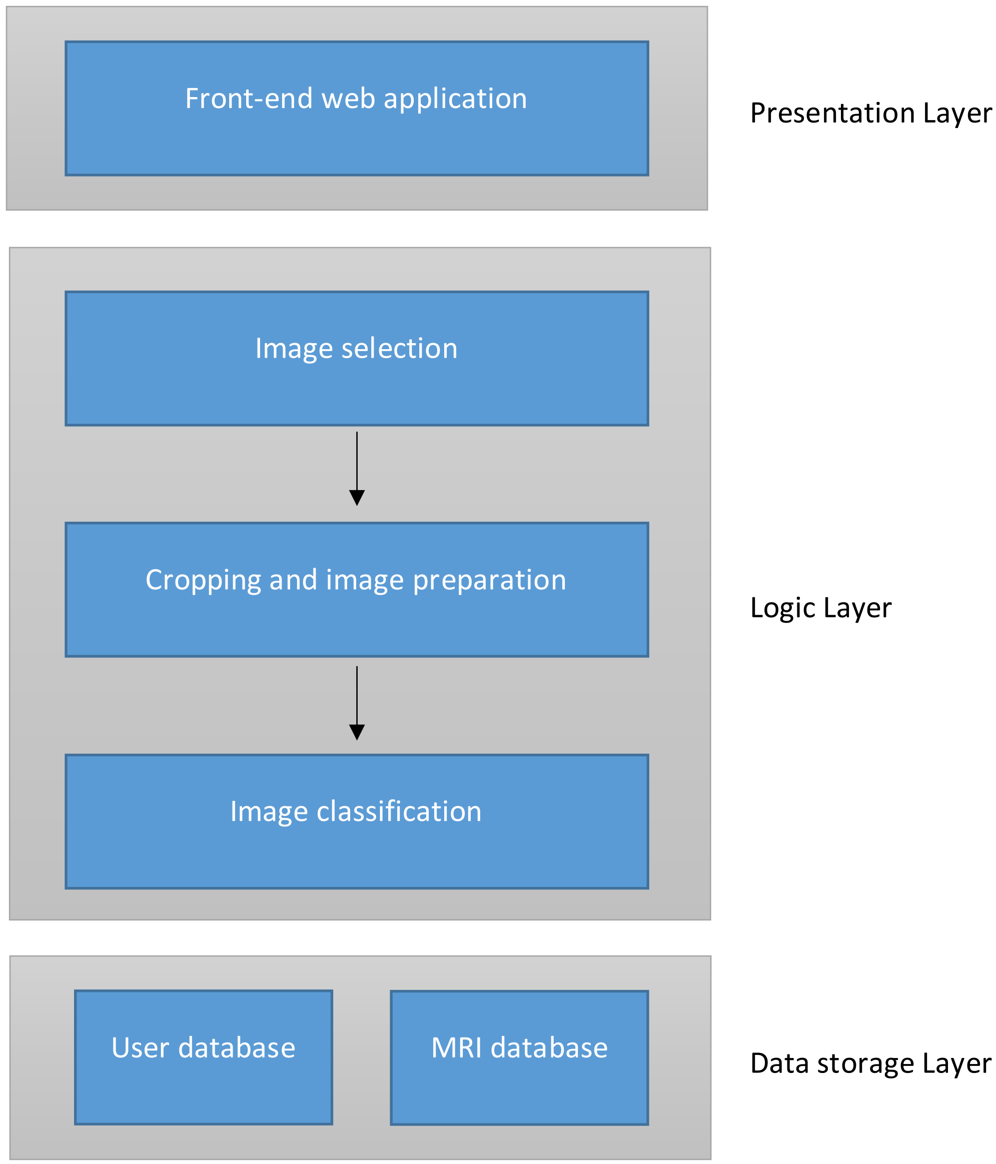
This chapter focused on the stakeholders of the project, requirement gathering techniques as well as the functional and non-functional requirements that are proposed.

5. Chapter 5 – Design

5.1. Chapter Overview

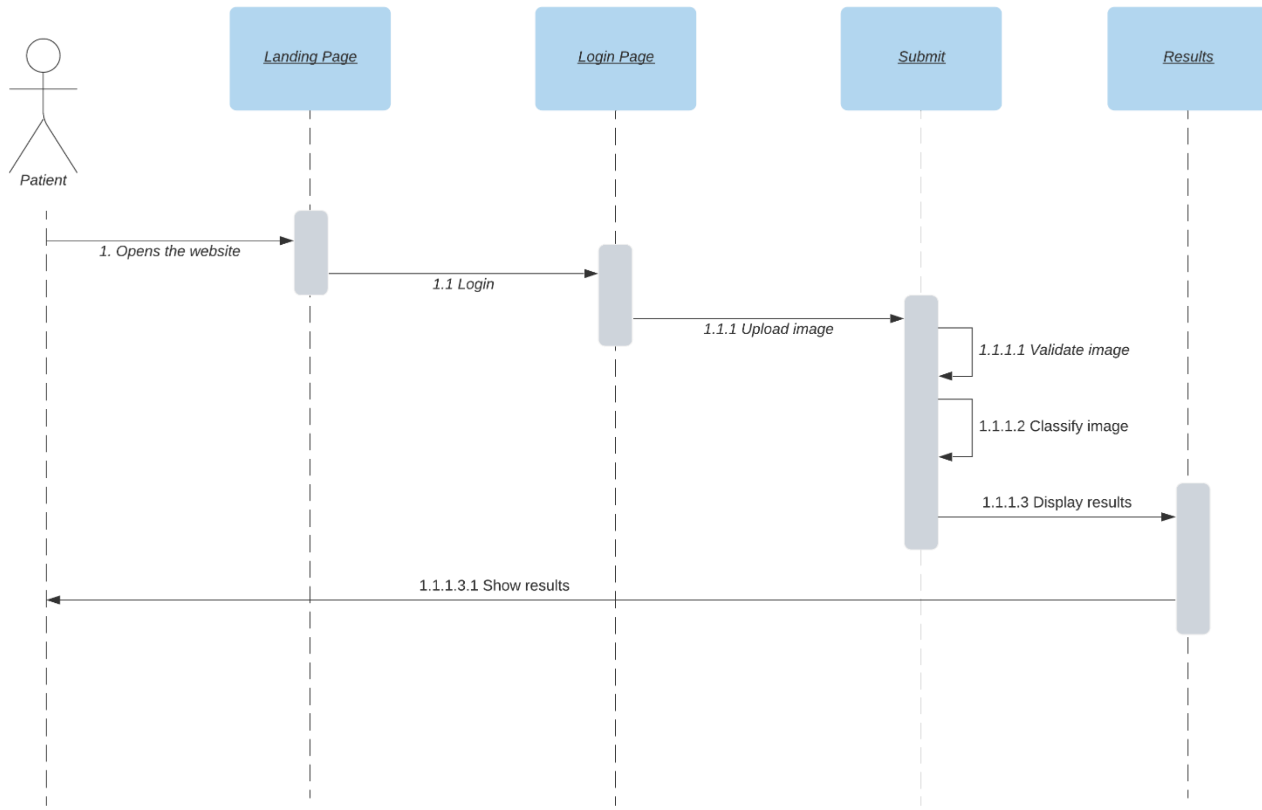
This chapter will focus on the design aspect of the proposed system. Apart from presenting a wireframe of the system thought out so far, it will also focus on discussing the high-level architecture diagram, the class diagram as well as sequence diagrams.

5.2. High-Level architecture Diagram

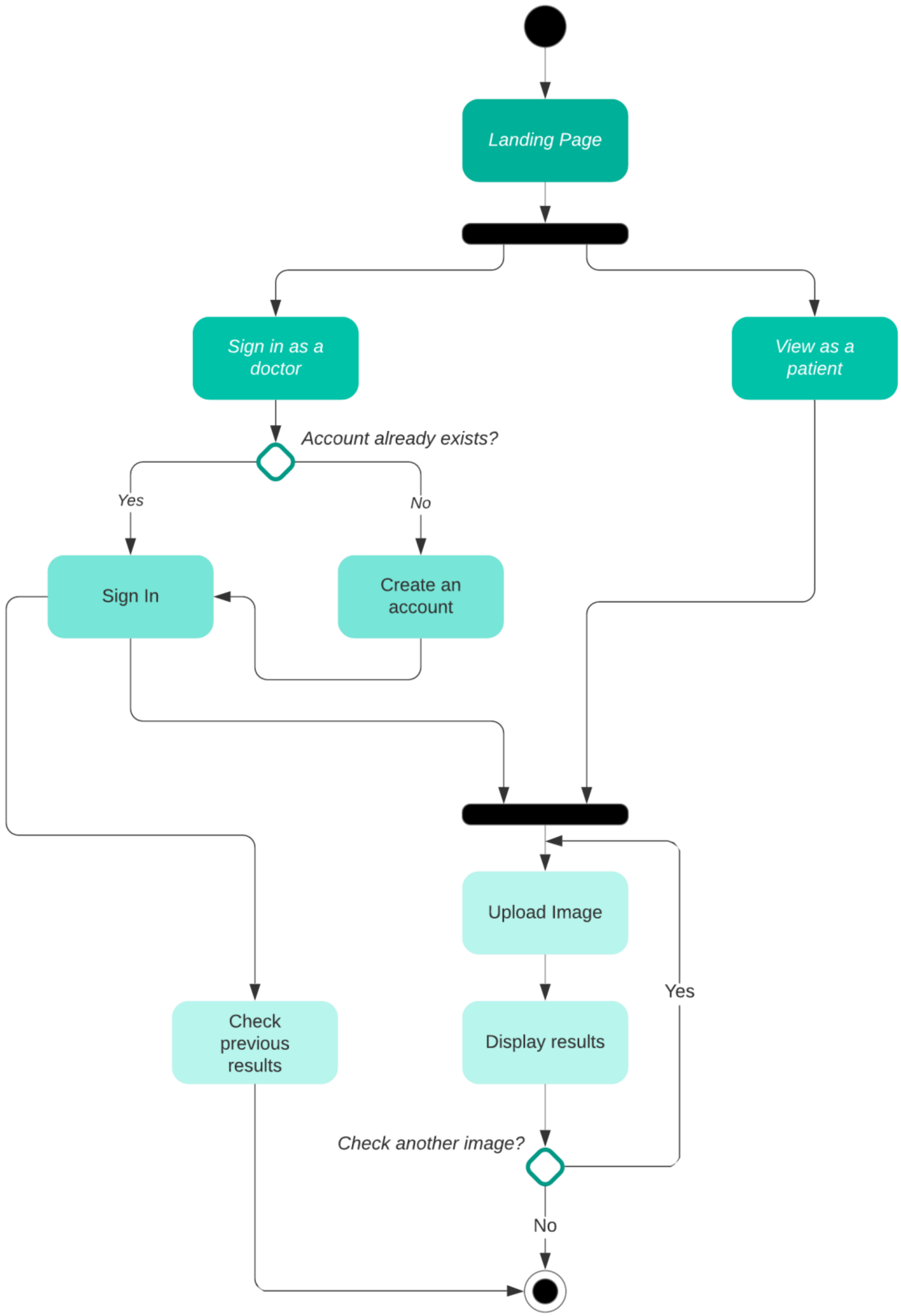


5.3. Class Diagram

5.4. Sequence Diagram



5.5. Activity Diagram



5.6. Wireframes

6. Chapter 6 – Conclusion

6.1. Chapter Overview

In the previous chapter, the system requirements specifications of the research were discussed. This chapter would focus on the dataset used for the research, the issues and challenges which arose, as well as the expected implementation of the system.

6.2. Dataset

Several data sets were considered suitable for this system. However, one set had to be selected to train and test the model.

The criteria measured in selecting the most appropriate data set were,

1. Number of images available
2. Types of brain tumors covered within the set

Out of all the considered sets, only one had a significant number of images, spread across three distinct types of brain tumors. This set was selected to train and test the model.

It contains 3064 T1-weighted enhanced images from 233 patients. It covers three kinds of brain tumor. The images are split among the three types as,

1. Meningioma (708 slices)
2. Glioma (1426 slices)
3. Pituitary tumor (930 slices)

The data is organized within a MATLAB data format with the following fields.

cjdata.label: 1 for meningioma, 2 for glioma, 3 for pituitary tumor

cjdata.PID: patient ID

cjdata.image: image data

cjdata.tumorBorder: a vector storing the coordinates of discrete points of the tumor border

cjdata.tumorMask: a binary image with 1s indicating the tumor region

The above-mentioned data set can be obtained from (Jun Cheng et al., 2017)

6.3. Legal, social, ethical and professional issues

6.3.1. Legal

When obtaining the data set from figshare.com, priority had to be given to make sure no legal restrictions were broken, and that the use of the dataset happened within the terms of use and conditions provided by Figshare. Accordingly, the proprietary markings including trademarks or copyright notices were not removed, suppressed or modified in anyway in compliance with the terms of use of Figshare.

Software licensing was also focused on in regards to legal implications. No pirated versions of any software were used. Community editions were used when available, while alternatives were found for software which would other be unavailable freely. Trial versions and evaluation versions were also used as to avoid any legal misuse. Accordingly, StarUML's unregistered evaluation version was used, while Google Docs, which is a free web-based suite was used for documentation.

6.3.2. Social

The current implementation of the software only allows English, and as such would provide difficulty of use to any person unfamiliar to English. Apart from that, the social implications on the medical field are a bare minimum.

6.3.3. Ethical

Ethical clearance was obtained in the interviews by the participating individuals before any further interactions. Their consent was obtained to record the conversations and to transcript the following dialogue which ensued. Furthermore, the dataset was examined to make sure no personal or confidential data was present.

6.4. Plans for Future Implementation

Future implementations of the project will consist of increasing the accuracy of diagnosis, reducing the time taken for a successful diagnosis and also the implementation of a mobile application. Recommendation of health tips to follow and doctors to consult will also be present in future versions.

Member	Contribution
Rasula Yadithya	1.2.1 Introduction to the problem 1.2.2 Examples in the problem 2.1 Chapter Overview 2.2.1 Sobel Edge Detection Technique 2.2.2 Canny Edge Detection Technique 3.9 Gantt chart diagram 4.3.2 Execution of requirement gathering techniques 5.2 Dataset
Sithira Panditha	1.5 Objectives 1.8 Chapter Summary 3.5 Testing Methodologies 3.8 Work breakdown structure 4.4 Models
K.C.P.L. Amarasinghe	1.3 Aim 1.6. Proposed features 3.4 Project Management Methodologies 4.1 Chapter Overview 4.5 Functional Requirements 5.4 Plans for Implementation
Manula Haputhanthrie	1.4 Scope 1.7 Resource Requirements 3.6 Risks and Mitigations 4.2 Stakeholder Analysis 4.6 Non-functional Requirements
Ravindu Kodithuwakku	1.2.5 Research questions 3. Chapter Overview 3.3 Programming Methodologies 3.7 Activity Schedule 4.3 Requirement Gathering 5.3.3 Ethical
Udul Sree Halambage	1.2.4 Attempted solutions of the competitors 3.2 Software Development Methodologies 5.1 Chapter overview 5.3.1 Legal 5.3.2 Social

References

1. American Association of Neurological Surgeons, no date. Brain Tumors. *American Association of Neurological Surgeons.* Available from https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Brain-Tumors [Accessed 23rd October 2020]
2. Available from https://biomedpharmajournal.org/vol13no3/accurate-classification-and-detection-of-brain-cancer-cells-in-mri-and-ct-images-using-nano-contrast-agents/ [Accessed on 23 October 2020]
3. Bahadure, Nilesh Bhaskarrao, et al. “Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM.” *International Journal of Biomedical Imaging*, 6 Mar. 2017. Available from https://www.hindawi.com/journals/ijbi/2017/9749108/ [Accessed 27 October 2020]
4. Bhattacharyya, Debnath, and Tai-hoon Kim. “Brain Tumor Detection Using MRI Image Analysis.” *Ubiquitous Computing and Multimedia Applications*, edited by Tai-hoon Kim et al., Springer, 2011, pp. 307–14. *Springer Link*, doi:10.1007/978-3-642-20998-7_38.
5. Dina Aboul Dahab, Samy S. A. Ghoniemy, Gamal M. Selim (2012). International Journal of Image Processing and Visual Communication ISSN (Online)2319-1724: Volume 1, Issue 2, October 2012
6. Md. Ariful Islam, Tanjima Akhter, Amena Begum, Md. Rakib Hasan, Farzana Sultana Rafi (2020). International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-8, June 2020
7. Nasim, M.A.A. (2019). Brain Tumor Detection Using Convolutional Neural Network. In: Hossain, Tonmoy & Shishir, Fairuz & Ashraf, Mohsena & Nasim, Md Abdullah & Shah, Faisal. 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT-2019). At: Dhaka, Bangladesh. December 2019: Researchgate, 1
8. Nitrc.org. 2020. *NITRC: Bratumia (Brain Tumor Image Analysis): Tool/Resource Info*. [online] Available at: https://www.nitrc.org/projects/bratumia [Accessed 3 November 2020].
9. Spine, M., 2020. *Brain Tumors: Overview Of Types, Diagnosis, Treatment Options | Cincinnati, OH Mayfield Brain & Spine*. [online] Mayfieldclinic.com. Available from https://mayfieldclinic.com/pe-braintumor.htm [Accessed 3 November 2020].
10. T. Ruba, et al. “Accurate Classification and Detection of Brain Cancer Cells in MRI and CT Images Using Nano Contrast Agents.” *Biomedical and Pharmacology Journal*, vol. 13, no. 3, Sept. 2020, pp. 1227–37.
11. Theckedath, D., Sedamkar, R.R. Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks. *SN wwwwwwCOMPUT. SCI.* **1,** 79 (2020). https://doi.org/10.1007/s42979-020-0114-9
12. Verneda Lights [Updated on June 6, 2017] “Brain Tumor: Types, Risk Factors, and Symptoms.” *Healthline*, 16 July 2012. Available from https://www.healthline.com/health/brain-tumor [Accessed 24 October 2020]

Bibliography

1. Dr Samir Kumar. (2011) "Detection of Brain Tumor-A Proposed Method." *Journal of Global Research in Computer Sciences*, vol. 2, no. 1, Jan. 1970, pp. 55–63. Available from https://www.rroij.com/open-access/detection-of-brain-tumora-proposed-method-55-63.php?aid=37003 [Accessed 26 October 2020]
2. https://ieeexplore.ieee.org/document/8191888/references#references [Accessed 25 October 2020]
3. Isselmou, A., Zhang, S. and Xu, G. (2016) A Novel Approach for Brain Tumor Detection Using MRI Images. *Journal of Biomedical Science and Engineering*, 9, 44-52. Available from https://www.scirp.org/journal/paperinformation.aspx?paperid=70753 [Accessed 25 October 2020]
4. Jun Cheng et al., (2017). brain tumor dataset. *figshare.* Available from https://figshare.com/articles/brain_tumor_dataset/1512427 [Accessed on 11th November 2020]
5. Lavanyadevi, R., et al. "Brain Tumor Classification and Segmentation in MRI Images Using PNN." *2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE)*, 2017, pp. 1–6. *IEEE Xplore*, Available from
6. Vijayakumar, C. and Gharpure, D., 2020. Development of Image-Processing Software for Automatic Segmentation of Brain Tumors in MR Images. Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3159221/ [Accessed 25th October 2020]