## LOGIC OF PROBLEM

**Analysis and Classification of Magnetic Resonance Images of Brain Tumors using Supervised Classification Algorithms.**

**Team 1:**

**Archit Jain**

**Ayush Arora**

**Siddharth Chauhan**

**Statement of the problem:-** With the growing demands of information technology and the e-health care sector, it is now possible to provide state of the art facilities to the patients suffering from brain tumors. However, the problem of classifying brain tumors as benign or malignant is still primarily dependent on the traditional method of biopsy. According to the World Health Organization(WHO) and American Brain Tumor Association(ABTA) around 70,000 Americans are living with brain tumors out of which 69.8% are benign, and 30.2% are malignant. Brain tumors are graded on a scale of 1 to 4, where benign tumors are graded from 1-2 and malignant tumors from 3-4. The problem addressed here is to do an in-depth analysis of images obtained from Magnetic Resonance Imaging(MRI) and classifying brain tumors as benign or malignant. As classifier algorithms Naive Bayes, J48, jRip and k-NN are applied to classify the tumor, the accuracy of each algorithm is calculated and analyzed against each other.

**The key question at issue is:**

The treatment procedure for benign tumors (grade 1-grade2) and malignant tumors(grade3-grade4) are considerably different where one can be treated with surgical excursions other require radiotherapy, chemotherapy or both. If grade 1 -grade 2 is left untreated,

there are high chances of it later turning into grade3-grade4 glioma. The key question here is how can we precisely identify whether the brain tumor being analyzed is normal(benign) or abnormal(malignant)? How do the image quality and image contrast obtained from MRI contribute in determining the desired outcomes? Based on the given data, can we predict whether the tumor can be classified as benign or malignant? Also, the noise present in the image data can result in low accuracy, which is also a significant concern for future predictions.

**The purpose of reasoning through the issue is:** A report by the American Cancer Society shows that about 71% of the detected brain cancer cases resulted in the death of the patient within the year which is significantly higher than any other cancer. An automated system to detect the cancerous nature of a tumor would mean the availability of the results at a higher proficiency, faster rates, and lower prices. It would also mean that the overall errors caused by human mistakes would be limited. Achieving such a system might be a long way ahead in the future and might need much effort from the researchers' community. But this system is something that would bring a significant impact.

**Information needed:** A good insight is needed about the anatomy of the brain and of the abnormal tissue growths that result in tumors. A thorough differentiation between the types of tumors and their further classification into primary or secondary tumors would provide a better understanding of the working of the project. Information on Magnetic Resonance Imaging(MRI) properties would also come in handy. Image Processing stages, namely preprocessing, segmentation provides a path for extraction of the tumor from the brain. Understanding the statistical functions of the images used in the creation of the matrix is mandatory. A thorough

study of the various supervised classification algorithms (ANN, Random Forest and K-nn) would enable us to analyze the prediction results and errors better. Wide range information on Image Preprocessing techniques is needed to extract the features from the extracted tumors. These techniques include the addition of filters, removal of noise, enhancement of contrast, etc. Also, knowledge of Edge–Detection algorithms such as Canny Edge Detection, Devernay Edge Detection vs the Sub- Pixel Detection is required for better evaluation. The selection of Image features for the application was extensively done. Programming background in MATLAB is relevant for the development of the Feature Extraction from the image recognition part of the project.

**Assumptions made:** The training dataset collected from the source is reliable, and the labels in the data are verified. Another assumption made is that the 13 features considered in the application for the segmented tumor images are ample to derive a definite matrix for the application. It is assumed that such predictions could someday be used to make a life-threatening decision. Also, it is assumed that the amount of data available for analysis is sufficient to train a model to derive decisions for the test dataset.

**Concepts guiding the reasoning:** Neoplastic diseases cause abnormal growth of tissues within the human body. These growths turn into tumors. The tumors either could be non-toxic to the body or could be cancerous. The non-cancerous tumors are called Benign Tumors and aren't generally aggressive. The tumors cancerous in nature are called Malignant and are highly aggressive. Magnetic Resonance Imaging(MRI) is a useful imaging technique used to capture the

tumors within the human body. The MRI uses a series of rotating magnets around the body which creates a magnetic field and interacts with the nuclei within the body, in turn, generating an image. The image captured from the body part, in this case, the brain is preprocessed, and the tumor is segmented. The tumor now acts as the region of interest for the application and calculated statistical features with the provided labels act as the train data set for the machine learning model.

**Implications of reading:** The system in consideration once constructed with an acceptable level of accuracy would mean automation of a long and delicate task, hence reducing the human effort and errors. On the other hand, a false positive or a true negative of the system might result in the loss of human life. Therefore complete reliability on such a system for the application might not be the best option in the near future.

**Point of view:** From a patient's health point of view a system that predicts the malignancy of the tumor would mean that the biopsy could be avoided which eventually means that the whole process of detection of cancer becomes faster and the chances of human error are reduced. From the doctors' point of view, a system that derives from decisions as fragile as a patient's life should be highly accurate. So doctors' might be reluctant to use the system until it can be completely relied upon. Statisticians would want to check how properties of image vary with the nature of the tumor to increase the size of the feature matrix. Researchers in the domain would want to collect a dataset as large as possible so as to build an accurate model for the application.

**Inferences made / Conclusions derived:** In conclusion, the desired systems should have high accuracy which can be achieved only through a large amount of training data and through

fine-tuning the model to increase recall to the maximum. An early prediction of the tumor if its cancerous can help patients seek the right treatment faster. A false negative can prove to be deadly in this application as compared to a false positive case. As fields of data analytics are evolving, applications like this can be made widely available at a low investment, unlike costly lab equipment to make predictions with improved accuracy and higher reliability.

**Summary:** The proper analysis of the Magnetic Resonance Images generated and already labelled as cancerous or non-cancerous based on the results from the biopsy results suggests a set of distinguishing features between the two. The segmentation of the tumors, followed by the calculation of these features creates a matrix. This matrix used with the adequately supervised classification algorithms (Naive Bayes, J48, jRip and K-nn) would make a model that could distinguish between the types of cancers for the new MR Images. The basic necessity of this system is a large dataset to increase its accuracy. Once achieved, this system would be a small victory against the never-ending battle between cancer and man-kind.

**Summary of Updation in the Logic of Problem**

Wide range information on Image Preprocessing techniques is needed to extract the features from the extracted tumors. These techniques include the addition of filters, removal of noise, enhancement of contrast, etc. Also, knowledge of Edge–Detection algorithms such as Canny Edge Detection, Devernay Edge Detection vs the Sub- Pixel Detection is required for better evaluation. The selection of Image features for the application was extensively done. Programming background in MATLAB is relevant for the development of the Feature Extraction

from the image recognition part of the project. Further Information on Random Forest Algorithm and Artificial Neural Networks was collected as these algorithms provide better accuracy.