

# Research Paper: Machine Learning Application for Brain Tumor Localization in the Repository of Molecular Brain Neoplasia Data (REMBRANDT)

Machine Learning for Health Care (Heinz 95-845) - Lukas Mohs

## 1. Abstract

Within this paper, I want to present the outcome of the combined application of *Computer Vision* (CV) and *Machine Learning* (ML) to the *Repository of Molecular Brain Neoplasia Data* (REMBRANDT) dataset. This dataset contains pre-surgical *magnetic resonance* (MR) multi-sequence images of 130 patients. Based on a visual analysis and the application of several ML algorithms, I tried to predict the location of the brain tumor and compare it to the evaluation of three Radiologists, which defined the affected brain part and the potential impact on the brain functionality.

## 2. Detailed Description

### 2.1 What is your proposed analysis?

The publicly available dataset, which includes extensive meta information, aggregates vertical and horizontal brain scans of different cancer patients. The proposed approach for this study tries to develop a supporting technology for radiologists to identify the brain tumor in each image sequence. This is realized by the application of CV algorithms that extract prognostic image features. These features will then be taken as the input for three Machine Learning algorithms: *Logistic Regression*, *Decision Trees* and *Naive Bayes*, which will be trained to classify the location of the brain tumor.

### 2.2 Why is your proposed analysis important?

The early detection of cancer greatly increases the chances for successful treatment. Two components can support the early detection of cancer:

- Education to promote early diagnosis and
- Screening

(World-Health-Organization)

The application of CV and ML could scale the screening opportunities for a high number of patients due to automatization. In case the localization based on the proposed, fairly simple model performs well, the applicability of these technologies in the field of cancer prevention could be proven. Then, further investigation and development could be started in order to advance the applicability and reliability of the model.

### 2.3 How will your analysis contribute to existing work? Provide references.

In the past, CV has been applied to several fields in order to identify objects or patterns within images. *Image Recognition* and ML usually form the underlying basis for CV, which can be understood as *understanding the image*. CV libraries like *OpenCV* and ML libraries like *Tensorflow* are both developed with a general purpose, which improves their overall performance but limits their applicability to niche areas. Since tumors can vary heavily in their appearance given different body parts and scan methods, such a general purpose tool would perform poorly. Therefore, just a few approaches have been undertaken to utilize this combination of technologies. One example was provided by N. Nandha Gopal and M. Karnan by applying clustering algorithms such as *Fuzzy C Means* to diagnose brain tumor (Nandha Gopal).

### 2.4 Describe the data. Y outcome(s), U treatment, V covariates, W population as applicable.

The REMBRANDT dataset is described as "a cancer clinical genomics database", which includes a Web-based data mining and analysis platform that tries to connect clinical information and genomic characterization data (Subha Madhavan and Buetow). The MRI was undertaken prior to surgery and other treatment, therefore the outcome can be seen as the positioning of the tumor within the brain. The provided metadata uses the following measure for tumor localization within the brain: *Temporal, Insular, Parietal, Occipital, Brainstem* and *Cerebellum*. These *labels* are used for the ML implementation.

### 2.5 What evaluation measures are appropriate for the analysis? Which measures will you use?

One core challenge of this study is given by the creation of the prognostic image features. To train a ML algorithm some features must be extracted from the images to be able to calculate the probability of the tumor location. These probabilities indicate how likely it is for a tumor to be in one of the labelled brain regions.

### 2.6 What study design, pre-processing, and machine learning methods do you intend to use? Justify that the analysis is of appropriate size for a course project.

First, a selection of the right image series must be undertaken, which determines, whether horizontal or vertical MRI scans will be used for the analysis. Afterwards, these images might be preprocessed to maximize the CV performance. Based on the CV outcome, *Logistic Regression, Decision Trees* and *Naive Bayes* algorithms are trained on a training dataset and tested again a test dataset.

### 2.7 What are possible limitations of the study?

Since the field of prediction is chosen to be very narrow, the developed model will not be applicable to other types of tumors without modifications. Furthermore, the final model

will also not be too sophisticated since it can be seen as a novel approach for automated tumor localization and might just provide the basis for further investigation.

## References

M. Nandha Gopal, N. ; Karnan. Diagnose brain tumor through mri using image processing clustering algorithms such as fuzzy c means along with intelligent optimization techniques.

Yuri Kotliarov Himanso Sahni Howard A. Fine Subha Madhavan, Jean-Claude Zenklusen and Kenneth Buetow. Rembrandt: Helping personalized medicine become a reality through integrative translational research.

World-Health-Organization. Early detection of cancer. URL "<http://www.who.int/cancer/detection/en/>".