Precision medicine and quantitative imaging in glioblastoma

ELMED219 (Artificial Intelligence and Computational Medicine), January 3-28, 2022

https://github.com/MMIV-ML/ELMED219-2022

Team #4

1 Research plan

1.1 A brief background to the field

For many years clinicians have tried to find a better treatment for glioblastomas but to this day this disease is still incurable. Glioblastomas are the most aggressive, malignant brain tumors that arise within the brain and spinal cord [1]. These tumors often affect older people but can occur in all ages [1]. The prognosis of glioblastoma is less than 2 years and on average of 3 months without treatment [2].

Glioblastomas are grade four cancers that arise from glia cells called astrocytes, which are supporting cells within the neural tissue [1]. When these cells become cancerous, they earn the ability to proliferate uncontrollably and spread out to other regions. The difference between glioblastomas and other cancer types, which makes them so lethal, is the diffused growth pattern. These tumor cells grow invasively through the cortex and neural networks [2]. Such a dispersed growth pattern makes it extremely hard for surgeons to remove the tumor without damaging important parts of the brain.

Brain cancers often have nonspecific symptoms during early stages. As the disease reaches later stages more specific symptoms appear such as seizures and neural failure [1]. Finding the cancer late is an additional reason why treatment is so difficult.

Diagnosis primarily starts off with a biopsy and an MRI scan, an imaging technique that is used to view the organs and tissues of the human body. It uses a magnetic field and radio waves to change the spin of hydrogen nuclei (protons) in water molecules to produce anatomical images [3]. MRI is a standard protocol for viewing the brain tissue and is used to determine treatment options for the patient. Using MRI, the glioblastoma can be located and its biochemical changes can be identified [4]. Tumor location and whether the tumor has necrosis or edema are important to detect because of their effect on prognosis [4].

The prognosis of glioblastoma is less than 2 years and the patient is only provided life prolonging treatment. The primary treatment is surgery. The aim is to remove as much as possible of the tumor without affecting crucial parts of the brain. After surgery, radiation and chemotherapy is used to kill remaining cancerous cells. However, due to the diffused growth pattern there are often cancerous cells left behind, which causes the cancer to return.

For many years there has been ongoing research on glioblastomas to improve early diagnosis, segmentation, prognosis and predicting overall survival (OS). Improvement in these areas is needed to increase the

quality of life for these patients. There is still a long way to go, however machine learning (ML) has the potential to make significant improvements.

1.2 Objectives and expected impact

Improving OS predictions is an interesting and important part of glioblastoma research. If the clinician could provide a high-probability prognosis of OS, this could help the patient choose the right treatment for their unique glioblastoma. [5]

The goal of this project is to give a high probability prognosis of OS for glioblastoma patients, based on tumor location and other features such as age and number of disconnected tumor cores, see Table 1. We want to make a program that can sort the patients into short-term, mid-term and long-term OS using ML techniques. Our long-term goal is to help clinicians aid their patients into making the best possible choices of treatment.

In this study we will use different ML techniques for the segmentation, extraction of features and OS prediction. Using machine learning for this task is essential for both the patient and clinician. The clinician can spend more time consulting with the patient and evaluating the treatment. It will also allow effective treatment to start earlier, which is crucial for this group of patients, as OS is typically less than 2 years.

If the clinician has a highly probable OS estimation for the patient, treatment can be selected accordingly. For instance, if estimated OS is low, the treatment can be focused on relieving pain and other symptoms. If estimated OS is high, the treatment can be more aggressive to ensure a longer OS. If OS is classified as mid-term, the patient might want to control the tumor, but not go trough all the side effects related to extending OS [5].

1.3 Material and methods

This study proposes to use the Deep Learning (DL) method nnU - Net as the segmentation method to extract the MRI features. To predict the OS we will use the method from McKinley et al. [6] and further develop the model. The model consists of an ensemble of least squares linear regression and random forest classifier.

nnU-NET

DL is an order of ML methods grounded in artificial intelligence that make use of experience to gain knowledge. The DL - based method nnU – Net is a segmentation method which configures itself without the use of expert knowledge for any given biomedical dataset [7]. This particular architecture is able to deliver a complete automated pipeline that includes pre – processing, data augmentation and post – processing of MRI segmentation.

The nnU – Net pipeline works by creating a ‘dataset fingerprint’ to extract the image features [7]. The fingerprint consists of data – dependent hyperparameters, determined utilizing a set of heuristic rules. The fingerprint then concludes on fixed parameters (loss function, architecture template and training procedure) and rule – based parameters (image resampling, intensity normalization, patch size, batch size). This step allows the fingerprint to use the unit architecture to identify the optimal configuration out of three possible; 2D, 3D and 3D cascade [7]. nnU – net will then empirically determine the group selection of the configurations and whether post – processing is needed; this is known as the empirical parameters. The optimal configuration will be used for the prediction of OS. In the study by Isensee et. al [8] nnU – net was able to achieve an accuracy of 90.6% on their validation set, showcasing that it is a robust segmentation method.

Linear Regression

The first model used for OS prediction is linear regression with least squares method. Linear regression is a supervised ML technique using linear input data and fitting the data to form a linear combination. This can be demonstrated in 2D as points on an xy coordinate system and trying to find a linear line to best fit

the points. When several features are used as inputs, the linear regression model increases in dimension and forms a hyperplane instead of a line. The resulting formula (1) is [9]:

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \tag{1}$$

Where β_0 is the bias coefficient [10], p is the number of different features, X_j is quantitative inputs and β_j is the unknown coefficients.

To determine the quality of the fit, different methods for measuring errors are used. In this project we will use the least squares method. It is a simple method using the squared distance from the datapoint to the regression hyperplane in the y-direction and summed over all datapoints. This can be expanded to include our set of training data and accompanying coefficients from formula (1) as shown in formula (2) [9]:

$$RSS(\beta) = \sum_{i=1}^{N} (y_i - f(x_i))^2 = \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta j)^2 \tag{2}$$

Where N is the total number of datapoints, y_i is the label and the rest are the same as in formula (1). The coefficients are then chosen to minimise the sum of squares.

Random Forest

The second model used for OS is a random forest classifier. To build a random forest, firstly a bootstrap sample is made. This is a subset of the original dataset where the different datapoints can be collected multiple times, but the resulting sample is still the same size as the original. From this sample a tree is built. To form the nodes, a predetermined number of features are chosen at random from all the features and the best split is determined. This is repeated until minimum node size is reached. The number of random features and node size should be at minimum the square root of all the features. The whole process is then repeated until desired number of trees are made. To make a prediction the majority vote of all the subtrees is chosen.

The final classifier for OS prediction is an ensemble of the two previous models classifying the survival time into short-($<$10 months), mid-(10-15 months) and long-survivors ($>$15 months). The result of the linear regression model is the prediction unless the random forest classifier disagrees with more than 50% probability. This is based on the model in McKinley et al. [6], however the different hyperparameters will be adjusted to how the new features impact the model.

1.4 Evaluation

For a more robust evaluation cross validation with five folds will be used. This involves combining the training and validation sets before splitting it into five equal parts. The model will then run for five rounds, varying which parts is used as validation while the rest is used for training. The validation for the model is obtained by averaging the validation from the five rounds.

The confusion matrix will be used for evaluation. It gives the number of true positives and negatives, as well as false positives and negatives. This gives a good overview of which classes the model has problems distinguishing. Moreover, from the confusion matrix one can find the accuracy, precision, recall and the F1-score [11]. We will look at the accuracy and the F1-score, since accuracy is widely used and F1-score works well on unbalanced data, which can be the case on small datasets. Lastly, Spearman rank correlation will be used since we are studying how new features affect the model. This looks at the correlation between two variables and rank it from -1 to 1, where negative values indicate a negative relationship, positive values a positive relationship and zero indicates no correlation. Furthermore, it includes the p-value where values over 0.05 indicate that the correlation is not statistically significant [12].

References

[1] Mayo Clinic. Glioblastoma. [Internet]. 2020 April 04 [cited 2022 January 22] https://www.mayoclinic.org/diseases-conditions/glioblastoma/cdc-20350148

[2] TedxTalks. Fighting Glioblastoma—Dr. Christopher Duma, M.D.—TedxCollegeoftheCanyons. [Internet]. 2016 May 06 [cited 2022 January 22] https://www.youtube.com/watch?v=hErXkeIadsY&t=759s

[3] Rockall A, Hatrick A, Armstrong P, Wastie M. Diagnostic Imaging. 7th ed. Wiley-Blackwell:2013.

[4] Wu CX, Lin GS, Lin ZX, Zhang JD, Liu SY, Zhou CF. Peritumoral edema shown by MRI predicts poor clinical outcome in glioblastoma. World Journal Surgicl Oncology. 2015;13. https://doi.org/10.1186/s12957-015-0496-7

[5] Mayo Clinic. Cancer treatment decisions: 5 steps to help you decide [Internet]. 2021 June 18 [cited 2022 January 22] https://www.mayoclinic.org/diseases-conditions/cancer/in-depth/cancer-treatment/art-20047350

[6] McKinley R, Rebsamen M, Dätwyler K, Meier R, Radojewski P, Wiest R. Uncertainty-Driven Refinement of Tumor-Core Segmentation Using 3D-to-2D Networks with Label Uncertainty. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Cham: Springer International Publishing; 2021. p. 401–11.

[7] Isensee, F., Jaeger, P.F., Kohl, S.A.A. et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation [Internet]. Nat Methods 2021;18:203–211. [cited 2022 January 23]. Available from: https://www.nature.com/articles/s41592-020-01008-z

[8] Isensee F., Jäger P.F., Full P.M., Vollmuth P., Maier-Hein K.H. nnU-Net for Brain Tumor Segmentation [Internet]. In: Crimi A., Bakas S. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Cham, Springer; 2021. 118-132. [cited 2022 January 23]. Available from: https://link.springer.com/chapter/10.1007%2F978-3-030-72087-2_11

[9] Hastie TJ, Tibshirani RJ, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer Science + Business Media; 2009. XXII, 745.

[10] Brownlee J. Linear Regression for Machine Learning [Internet]. Machine Learning Mastery; 2020 August 15 [cited 2022 January 21]. Available from: https://machinelearningmastery.com/linear-regression-for-machine-learning/

[11] Ayuya C. Evaluating Machine Learning Model Performance [Internet]. Section; 2020 November 26 [cited 2022 January 21] Available from: https://www.section.io/engineering-education/evaluating-ml-model-performance/

[12] Zach How to Calculate Spearman Rank Correlation on Python [internet]. Statology; 2020 December 7 [cited 2022 Januray 22] Available from: https://www.statology.org/spearman-correlation-python/

2 Data management plan and ethical considerations

2.1 Description of generated data and code

We plan to use the BraTS dataset from 2020 [1, 2, 3]. The data includes MRI T1, T1Gd, T2 and T2-Flair, age and survival data.

Our model will try to predict overall survival and will only include subjects with "Gross Tumor Resection" and high grade gliomas. The output should be number of survival days, and we are going to validate our findings with the accuracy of the classification into three classes: long-survivors ($>$15 months), short survivors ($<$ 10 months) and mid-survivors (between 10 and 15 months).

Based on McKinley et. al [4], who won the survival challenge in BraTS 2020, we plan to construct our model around age, disconnected number of cores and number of disconnected whole tumors. These image-derived features are extracted from a nnU-net model. In addition, the nnU-net model will include features such as edema and necrosis. The nnU-Net model from Isensee et. al, who won the segmentation challenge in BraTS 2020, will be the foundation of our nnU-net model. [5] To complete our model, a brain region description of the tumor will be included. We intend to use the model from Chato et. al [6] as a framework.

A combined linear regression model and a random forest classification model will be used to predict survival time. The basis for the model is the linear regression model, but the classification model can overwrite the regression model in case of a probability of 50% or more. This model design is inspired by McKinley et. al [4]. Prior to the prediction model, the dataset will be divided into a training (70%), validation (15%) and test set (15%). Planned features and pipeline are shown in Table 1 and Figure 1.

Category	**Feature**
Patient	Age
	Survival days
Image	Disconnected tumor cores
	Disconnected whole tumor regions
	Necrosis
	Edema
	Sub-region location

Table 1: Features

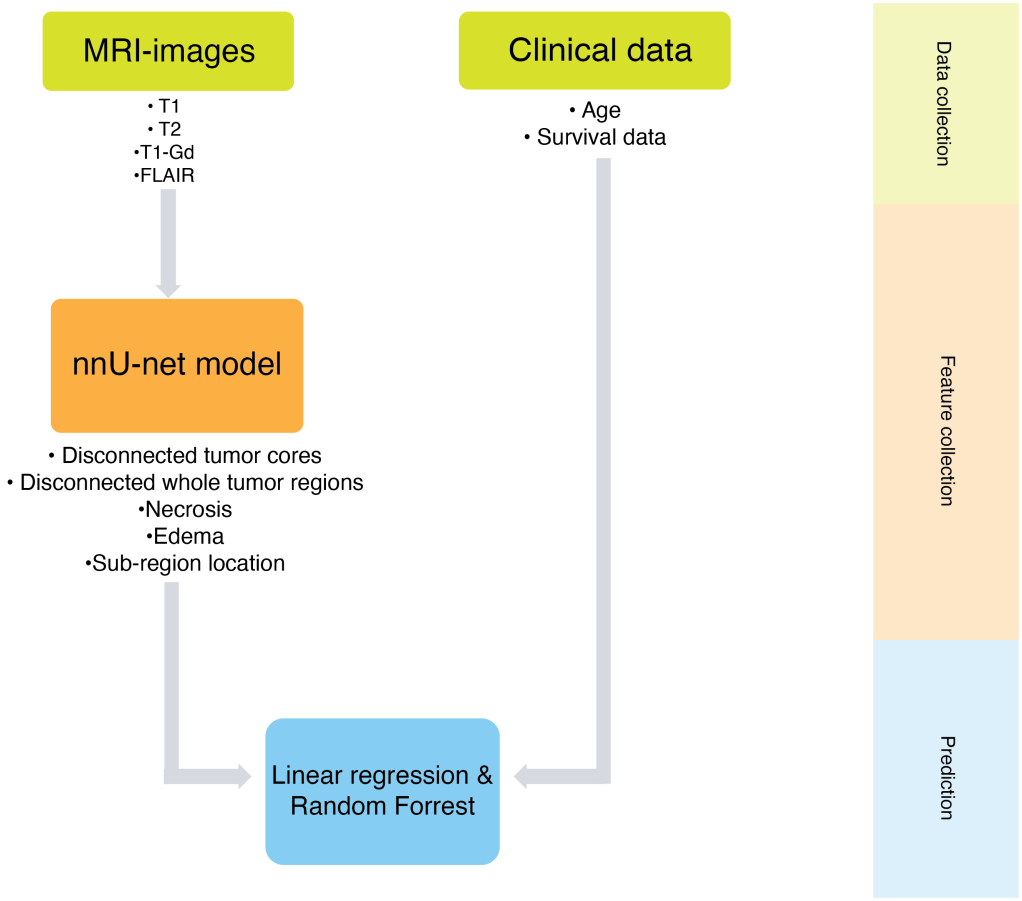

Figure 1: The overview of the process

2.2 Sharing of data and code

The code will be publicly available on GitHub accessible for future research. The data is already public, and the requirement to access the dataset is through an account in CBICA's Image Processing Portal.

2.3 Ethical considerations

Machine Learning and Artificial Intelligence can pose different ethical challenges [7]. These issues are often due to the divisive desire of patient privacy and sharing of data. On one hand, confidential patient information should only be available to a limited amount of people. On the other hand, research data should be available to other researchers for replication and validation. Dilemmas like these are important to discuss before releasing AI solutions into the world of health care.

We access patient data from the BraTS 2020 dataset. Consent has already been given and can be accessed by anyone. The data itself uses pseudo-identifiers to connect all the data to the same person but cannot be traced back to the person itself [1]. Therefore, there is no action needed to uphold the anonymity and no more ethical problems arise on this front.

As previously mentioned, the code will be publicly available on GitHub. The code itself does not contain any sensitive information. However, the results after using it on medical data could show some bias. This depends on the dataset used to train the algorithm and could be clarified when accessing the code. For that reason, it is important to use enough data when training the algorithm so that it becomes more generalizable. However, in this field it is a continuous challenge to obtain large amount of data.

References

[1] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et. al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans Med Imaging. 2015 Oct;34(10):1993-2024
https://doi.org/10.1109/TMI.2014.2377694

[2] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data. 2017 Sep 5;4:170117
https://doi.org/10.1038/sdata.2017.117

[3] Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Cromi A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint. 2018 arXiv:1811.02629.
https://arxiv.org/abs/1811.02629

[4] McKinley R, Rebsamen M, Daetwyler K, Meier R, Radojewski P, Wiest R. Uncertainty-driven refinement of tumor-core segmentation using 3D-to-2D networks with label uncertainty. arXiv preprint. 2020. arXiv:2012.06436.
https://arxiv.org/abs/2012.06436

[5] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021 Feb;18(2):203-211.
https://doi.org/10.1038/s41592-020-01008-z

[6] Chato L, Latifi S. Machine Learning and Radiomic Features to Predict Overall Survival Time for Glioblastoma Patients. J Pers Med. 2021 Dec 9;11(12):1336
https://doi.org/10.3390/jpm11121336

[7] Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 2020;368:l6927.
https://doi.org/10.1136/bmj.l6927