TM10007: Machine Learning
Assignment guidelines

March 15, 2020

1 Introduction

This guide describes how to conduct and report on the final research project for the TM10007 Machine learning course. In this project, you will use one of the provided medical datasets to train and evaluate different classification methods.

2 Deliverables

You will have to hand in a written report similar to a paper, shortly giving an introduction to the problem, and describing your methods, results, and a discussion. The report should have no more than 5 pages, excluding references and appendices, and should contain the following:

1. An introduction concerning the (clinical) problem to be solved.
2. A description of the methods applied, with proper motivation, containing for example:
    - A description of the dataset
    - Preprocessing of the data
    - Classifiers (at least two)
    - Experimental and evaluation setup
    - Statistics

    You methods should be sufficiently detailed for another person to replicate your study.
3. A results section, providing a comprehensive overview of the results. When possible, try to illustrate your results with plots or tables.
4. A discussion, containing
    - an interpretation of the results and overall conclusion

- limitations of your study
- the clinical relevance of your study
- recommendations for future research

We do not expect you to perform a statistical analysis of the results, but we do expect you to perform the necessary experiments to reach a conclusion on the most promising method and expected performance. If possible, try to relate to your knowledge of the used methods.

5. A reflection of the team process, with an individual contribution for every team member describing their personal contribution and their opinion on the teamwork. This should be no more than 200 words per person.

Note that your code is also a deliverable, see the last section of this handout.

3 Grading

When grading the report, we will take the following aspects into account:

- Is a proper motivation supplied for the methodological choices?
- Was a proper experimental setup applied and described in the report?
- Are the results interpreted well, with respect to the different methods?
- Is the clinical relevance of the results discussed?
- Are the conclusions supported by quantitative results and figures?

Note that although a proper analysis and smart choice of methodology will lead to a classifier with high performance, a better performance does not automatically lead to a higher grade.

4 Deadlines

The code and report have to be handed in by April 19th 2020 (23:59 at the latest). The code has to be uploaded to your groups Github reposity and tagged with the tag *final*. Note that the code should be runnable using Google colab.

5 Datasets

All datasets contain quantitative medical image features, on which you are going to apply machine learning to find a relation between these and clinical outcome. This field of research is also known as radiomics.

For each dataset, you will have assess to a set of these quantitative imaging features, on which you have to perform a binary classification problem.

5.1 Prediction of tumor grade in brain cancer [1]

Gliomas are the most common form of brain tumors. The prognosis and treatment is mostly dependent on the tumor grade, which is defined by histological analysis. A rough categorization is that of low-grade (II/III) and high-grade (IV) glioma, the latter of which are also called glioblastoma. For the latter category the median survival is approximately 15 months, while low-grade glioma have a 10-year survival rate of approximately 47%.

The grade is known to affect the presentation of the tumor on magnetic resonance imaging (MRI), but a tissue sample is still needed to get a proper diagnosis. If a diagnosis could be made purely based on imaging, this might eliminate the need for a biopsy in some cases. Note, however, that a tumor resection by craniotomy is the generally preferred treatment for glioma, which automatically results in a tissue diagnosis.

The aim of this study is therefore to predict the tumor grade of low grade glioma's (high or low) based features extracted from a combination of four MRI images: T2-weighted, T2-weighted FLAIR and T1-weighted before and after injection of contrast agent. A good performance on this dataset would be above 80% mean accuracy.

5.2 Distinguishing Alzheimer patients from healthy controls [2]

Dementia is a major problem worldwide, affecting over 36 million. In patients with dementia, the brain shrinks at a higher rate than in healthy people. Dementia consists of several underlying diseases, of which Alzheimer can be seen as one of the extreme forms. As everyone's brain shrinks with age, abnormal shrinkage, i.e. dementia, can be difficult from normal shrinkage.

Recent studies have tried to use quantitative MRI in order to distinguish between patients with normal and abnormal growth. Especially in specific regions such as the hippocampus, abnormal, heterogeneous growth patterns may be observed. To this end, the Alzheimer's Disease Neuroimaging Initiative (ADNI) consortium has build a MRI database of almost 1000 persons, including both patients with Alzheimer and healthy controls.

The aim of this study is therefore to distinguish between Alzheimer patients and healthy controls based features extracted from T1-weighted MRI. A good performance on this dataset would be above 75% mean accuracy.

5.3 Predicting tumor stage in head and neck cancer [3]

The annual incidence of head and neck (H&N) cancer is around 550.000 cases worldwide per year, with a mortality rate of 300.000 per year. This cancer comes in various forms, and therefore a wide spread in the prognosis and survival of H&N tumors. One of the most important clinical biomarkers is the tumor stage or so called T-stage. This biomarker is used to judge the malignancy of the tumor, and is thus closely related to survival.

While T-stage is highly related with volume, the molecular profile of the tumor does also play a role in the malignancy. However, quantification of the molecular profile can only be obtained after a biopsy or resection. While imaging, mostly computed tomography (CT) may be used to quantify T-stage, manual rating is observer dependent, subjective, and challenging.

The aim of this study is therefore to predict the T-stage (high/low) in patients with H&N cancer based on features extracted from CT. A good performance on this dataset would be above 70% mean accuracy.

References

[1] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4:170117, 2017.

[2] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*, 74(3):201–9, January 2010.

[3] Hugo J. W. L. Aerts, Emmanuel Rios Velazquez, Ralph T. H. Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Cavalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, Frank Hoebers, Michelle M. Rietbergen, C. René Leemans, Andre Dekker, John Quackenbush, Robert J. Gillies, and Philippe Lambin. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 5, June 2014.