Combining deep learning and radiomics-based machine learning to optimize predictions on medical images

Summary

In 2040 the estimation of patients with cancer reaches 28 million people, an increase of almost 50% compared to the figures from 2020 (GLOBACAN 2020). This will make cancer even more of a burden for society and for healthcare. Moreover, lack of clinicians is already a worldwide issue, making the demand for tools to reduce their workload very high; hence the need to keep improving and developing clinical decision support systems, which is the focus of this thesis.

This thesis consists of two parts, both aiming to explore the combine value of feature-based models and deep learning models for medical image analysis in cancer. The first part investigated the combination of the predictions obtained with feature-based and deep learning models, potentially leading to more accurate and robust frameworks. The second part of this thesis explored the use of feature-based models to augment the predictions of deep-learning models.

Part 1: Comparing and combining deep learning and feature-based machine learning

Radiomics and deep learning are two machine-learning methods which can be used to classify medical images and promising results using these methods have been reported in the literature. In **Chapter 2** we explored the pros and cons of those two methods: compared to deep learning, radiomics can perform better on small datasets (which is usually the case when analyzing medical imaging datasets), but cannot be used to segment data. Furthermore, radiomics requires input from clinicians/experts who need to identify and delineate the data/region of interest by hand and from data scientists who select features based on a preset list. We hypothesized in this thesis that both methods yield complementary information and using them in combination improves the results.

In **Chapter 3**, we analyzed the capability of mass spectrometry imaging (MSI) data and data from haematoxylin & eosin (H&E) stained tissue sections for automatic identification of patients with Barrett's oesophagus and prediction of progression in patients with low grade dysplasia. Due to the differences in the acquisition method of these two datasets, the datasets needed to be homogenized. This study showed that the model based on H&E data alone was better at identifying tissue type and the models based on MSI data alone were more suitable for predicting dysplasia grade for all patients and disease progression in patients with low grade dysplasia. In **Chapter 4**, we compared and combined a radiomics-based model and a deep learning based model aiming to predict adverse radiation effects (ARE) in a dataset of pre-treatment brain MRI images containing metastasis. In this case, the most efficient pre-processing method was selected independently for the two different models. We observed that the best results on the external dataset were obtained when combining the predictions of the radiomics-based model and the deep learning model. This study suggests that the predictions of the two models could be used in combination and improve the classification of ARE/none-ARE lesions within brain-MRI compared to using those models independently.

Part 2: Using feature-based models to augment deep learning predictions

In **Chapter 5**, we evaluated wether the use of a machine-learning model based on handcrafted-features in addition to a conventional U-Net model can improve performance when trying to find digital H&E stained slides containing lymph node (LN) and subsequently segment them using a large H&E dataset from patients with oesophageal cancer. We compared our results to the conventional U-Net model approach and found that the accuracy of our model was better than conventional U-net model approaches. Moreover, our method allowed us to obtain a likelihood score per potential LN found, allowing the creation of an "uncertain" class, for which the model cannot provide a prediction whether the candidate contour is a LN or not. The addition of an "uncertain" class allowed us to identify slides, which (for sure) require manual quality control more specifically instead of quality controlling of a random set of few slides. Six percent of the images from the external dataset were classified in the uncertain category and thus would need to be quality checked by a pathologist/expert.

Finally, in **Chapter 6**, we implemented a deep learning model which uses pre-processed data of contrast enhanced mammograms containing a suspicious mass, returning predictions on mass location, contour and a label differentiating between "benign" and "malignant". In parallel, we implemented a radiomics-based model on the contours made by the radiologist to predict the malignancy of the masses, comparing and combining the predictions obtained with the different models. We also implemented a radiomics model based on the predicted contours obtained after prediction by the deep learning model and compared and combined the score obtained there. We observed that for both scenarios (ground truth contours and predicted contours), the combination of radiomics and deep learning results obtained the best performance.

In every study presented in this thesis, we implemented a reproducible workflow by establishing a common pre-processing strategy and validating the models either with cross-validation or with external validation datasets provided by another institute. We conclude from the studies presented in this thesis that using feature-based models (e.g. radiomics) in combination with deep learning models leads to stronger predictions as different but complementary information is extracted and processed by the two methods. Future studies are needed whether decision support models can be further improved by also including patient characteristics as a non-image based dataset in the final model.