Linköping University | Department of Computer and Information Science
Master's thesis, 30 ECTS | Datateknik
2017 | LIU-IDA/LITH-EX-A--2017/001--SE

A very very long title

– with a subtitle

En himla bra svensk titel

Dávid Hrabovszki

Supervisor : Anders Eklund
Examiner : Annika Tillander

Linköpings universitet
SE–581 83 Linköping
+46 13 28 10 00 , www.liu.se

Upphovsrätt

Copyright

Abstract

The abstract resides in file `Abstract.tex`. Here you should write a short summary of your work.

Acknowledgments

Acknowledgments.tex

Contents

Abstract **iii**

Acknowledgments **iv**

Contents **v**

1 Introduction **1**
1.1 Motivation 1
1.2 Aim 2
1.3 Research questions 2
1.4 Delimitations 2

2 Related work **3**

3 Method **5**

4 Results **6**

5 Discussion **7**
5.1 Results 7
5.2 Method 7
5.3 The work in a wider context 8

6 Conclusion **9**

Bibliography **10**

1 Introduction

1.1 Motivation

It is estimated that around 300,000 new brain and nervous system cancer cases occurred in 2020 worldwide. Around 250,000 deaths occurred from this type of cancer in the same year ([20]). The World Health Organization classifies tumors into grades based on their malignancy, where grade I is the least malignant and grade IV is the most malignant ([12]). Grade II and III cancers are called Lower Grade Gliomas (LGG), and grade IV cancers are called Glioblastoma or Glioblastome Multiforme (GBM) ([16]).

It is important to diagnose cancer types correctly, because treatment options and survival expectancy depend largely on how malignant a tumor is and what characteristics it has. There are histological differences between different types, which helps the expert pathologist in the decision making. Grade I lesions have the possibility of cure after surgery alone, grade II tumors are more infiltrative, can progress to higher grades, and often recur, and grade III is reserved for cancer that has some evidence of malignancy. The treatment of grade III lesions usually include radiation and chemotherapy. Grade IV tumors are malignant, active, necrosis-prone (death of the tissue), progress quickly and often cause fatality ([13]).

Histology is a branch of biology concerned with biological tissues, and its aim is to discover structures and patterns of cells and intercellular substances. Histologists examine tissue samples that have been removed from the living body through surgery or biopsy. These samples are processed and stained with chemical dyes to make the structures more visible. They are then cut into very thin slices that can be placed under an optical microscope and examined further. [3]

Pathology is a medical branch, where experts aim to determine the causes of diseases [4], and histopathology connects the two fields, by studying the diseases of tissues under a microscope. Histopathologists make diagnoses based on tissues and help clinicians in the decision making process. Specifically, they often provide diagnostic services for cancer, by reporting its malignancy, grade and possible treatment options [18].

With the advance of technology, it is now possible to scan, save, analyze and share tissue images using virtual microscopy. This technology scans a complete microscope slide and creates a single high resolution file called Whole Slide Image (WSI). These files take up significant storage and require specific software to view and manipulate them, because they are stored in special file formats [14].

In this paper, Whole Slide Images from The Cancer Genome Atlas (TCGA) are used. The dataset is publicly available, and contains tissues from GBM and LGG brain cancer types from many different clinics. There are 860 examples of GBM and 844 examples of LGG available as Diagnostic Slides. The images are labeled as a whole, therefore no pixel-wise annotation is available. The files can be more than 3 GB in size, and their resolution is ofter higher, than 100,000 x 100,000. This is why they are saved in a special format (svs), that allows for storage of such large files.

The images were scanned at multiple different resulutions, which can be separately obtained thanks to the special file format. Not all images have the same highest magnification level, however. All LGG classified scans were recorded at 40x magnification (0.25 μm/pixel), while 77% of GBM scans have only 20x magnification (0.5 μm/pixel) available. In order to analyze them together, all images need to be obtained at the same level.

Since the images are so large, it is impossible to process them as a whole, therefore patches or tiles are extracted from them, that are easier to handle for a neural network.

1.2 Aim

The aim of this thesis is to classify two different types of brain tumor (GBM and LGG) from Whole Slide histology images using Deep Learning.

1.3 Research questions

This paper intends to find the answer to following complex research question:

What is the best approach with Deep Learning for GBM vs LGG classification using histology images without pixel level annotation?

The competing approaches are compared using statistical methods. There are several challenges regarding the research question, one of which is that the slide images are large in size, therefore they need to be divided into patches. They also come from different sources, so they must be normalized. There is no annotation available on a pixel level, so it is possible that patches in a slide are cancer-free. The patches need to be combined to slide level in the prediction phase, which is not a straight-forward task.

1.4 Delimitations

This is where the main delimitations are described. For example, this could be that one has focused the study on a specific application domain or target user group. In the normal case, the delimitations need not be justified.

2 Related work

Numerous researchers have applied computer algorithms for histology image analysis, and they can be split into two groups. One requires the extraction of hand-crafted features that expert pathologists would recognize from the slide images, the other does not. The second group can afford to not explicitly obtain these features, because they use deep convolutional networks to automatically do that within the model. The first group therefore uses more traditional machine learning algorithms that accept well defined features. This thesis belongs in the second group, and the author aims to achieve higher accuracy scores with deep learning than the researchers did in the first group.

This is not an easy task, because some researchers achieved very good results without using deep learning. [2] extracted coarse features of shape, color and texture from the image patches focusing on cell nuclei, reduced the dimensionality of the features, and created clusters representing similar patterns. Fine features were extracted from a select few patches from each cluster, then an Elastic Net was used on these patches to make the diagnostic decision of classifying a slide into GBM or LGG. Extracting fine features was a very computationally expensive task, this is why only a smaller number of representative patches were a part of this step. Slide level aggregation was done by weighted voting of the predictions of the patches. All Whole Slide Images came from the TCGA data repository, and were resized to 20x magnification level using bicubic interpolation, and the patches were 1024 x 1024 pixels. They achieved an accuracy of 93.1% on a dataset containing 302 brain cancer cases.

[17] approached the problem similarly, by using a more traditional machine learning model, the support vector machine. They extracted features from the texture, intensity and morphology along with several clinical measures, and trained the SVM model on them. The creation of features required knowledge about what differentiates GBM from LGG, such as microvascular proliferation, mitotic activity and necrosis. The validation accuracy was 75.12% on images obtained from TCGA.

There is a lot more literature in this field that utilized the power of deep learning for image analysis. [11] presented the three best performing methods from the 21st International Medical Image Computing and Computer Assisted Intervention (MICCAI 2018) conference for classification of oligodendroglioma and astrocytoma (which are two subclasses of LGG) patients. They all used a combination of radiographic and histologic image dataset, where the histologic images were obtained from TCGA, but they processed the two types of images separately. The three methods achieved accuracy scores of 90%, 80% and 75% respectively.

The best one, [1] applied several preprocessing steps on the images, including region of interest detection, stain normalization and patch extraction (224 x 224). They trained an autoencoder with convolutional layers to extract features from each patch, then used these features to identify patches that can potentially contain tumor regions using anomaly detection (Isolation Forest), where tumor was considered the anomaly. The DenseNet-161 network, pretrained on ImageNet, was then trained on these anomaly patches only, and the final prediction was done according to majority voting.

The second best approach, [15] argues that since only whole slide level annotation is available, but the training is done on patches, this is a weakly-supervised learning problem. To tackle this, they incorporated a Multiple Instance Learning (MIL) framework into the CNN architecture, which helps combine the patch predictions to slide level intelligently. The preprocessing steps were similar to [1], but they used 256 x 256 patches and a more simple histogram equalization for color normalization. Here a pretrained DenseNet-169 model was used with dropout regularization, which produced an average slide level score from all sampled patches for that slide. They concluded that the dropout technique did not improve the accuracy significantly.

The third best solution (only described in [11]), used a different approach. They identified tissue characteristics that differentiate the two classes, such as necrosis, cell density, cell shape and blood vessels. The images were then partitioned into 512 x 512 patches, and fed into a VGG16 CNN network with data augmentation to tackle class imbalance, and dropout and batch normalization layers to reduce overfitting.

A mixture of traditional machine learning and deep learning approaches exists, when the CNNs are only used for automatic feature extraction, but the classification is done by another machine learning algorithm. [22] used a pretrained AlexNet CNN for extracting 4096 features, some of which revealed biological insights, and then employed a SVM. They achieved 97.5% accuracy and concluded that these CNN features are significantly more powerful than expert-designed features.

[5] conducted a very extensive research, where they used a deep learning approach to classify whether a slide image has cancer in it or not. They tested their methods on very large datasets of different types of cancer, and different slide preparation methods. The datasets were similar to TCGA in a way that they were also not labeled at pixel level, therefore the authors presented different Multiple Instance Learning approaches to tackle this weakly supervised problem in the form of slide aggregation models. These models included logistic regression, random forest, and recurrent neural networks that were trained on the validation set to avoid overfitting. They showed by statistical comparisons that fully supervised learning models based on curated datasets do not generalize well to real world data, where detailed annotation is not available. Even though the authors did not use brain cancer data, some very useful findings and methods can be applied to the TCGA brain tumor dataset, including the statistical comparison of different models.

Other papers have experimented with deep learning for digital pathology, from which this research can benefit in questions such as optimal patch size, architecture, data augmentation methods, preprocessing steps, and slide level aggregation techniques [9], [6], [19], [7], [10], [21], [8].

3 Method

In this chapter, the method is described in a way which shows how the work was actually carried out. The description must be precise and well thought through. Consider the scientific term replicability. Replicability means that someone reading a scientific report should be able to follow the method description and then carry out the same study and check whether the results obtained are similar. Achieving replicability is not always relevant, but precision and clarity is.

Sometimes the work is separated into different parts, e.g. pre-study, implementation and evaluation. In such cases it is recommended that the method chapter is structured accordingly with suitable named sub-headings.

4 Results

This chapter presents the results. Note that the results are presented factually, striving for objectivity as far as possible. The results shall not be analyzed, discussed or evaluated. This is left for the discussion chapter.

In case the method chapter has been divided into subheadings such as pre-study, implementation and evaluation, the result chapter should have the same sub-headings. This gives a clear structure and makes the chapter easier to write.

In case results are presented from a process (e.g. an implementation process), the main decisions made during the process must be clearly presented and justified. Normally, alternative attempts, etc, have already been described in the theory chapter, making it possible to refer to it as part of the justification.

5 Discussion

This chapter contains the following sub-headings.

5.1 Results

Are there anything in the results that stand out and need be analyzed and commented on? How do the results relate to the material covered in the theory chapter? What does the theory imply about the meaning of the results? For example, what does it mean that a certain system got a certain numeric value in a usability evaluation; how good or bad is it? Is there something in the results that is unexpected based on the literature review, or is everything as one would theoretically expect?

5.2 Method

This is where the applied method is discussed and criticized. Taking a self-critical stance to the method used is an important part of the scientific approach.

A study is rarely perfect. There are almost always things one could have done differently if the study could be repeated or with extra resources. Go through the most important limitations with your method and discuss potential consequences for the results. Connect back to the method theory presented in the theory chapter. Refer explicitly to relevant sources.

The discussion shall also demonstrate an awareness of methodological concepts such as replicability, reliability, and validity. The concept of replicability has already been discussed in the Method chapter (3). Reliability is a term for whether one can expect to get the same results if a study is repeated with the same method. A study with a high degree of reliability has a large probability of leading to similar results if repeated. The concept of validity is, somewhat simplified, concerned with whether a performed measurement actually measures what one thinks is being measured. A study with a high degree of validity thus has a high level of credibility. A discussion of these concepts must be transferred to the actual context of the study.

The method discussion shall also contain a paragraph of source criticism. This is where the authors' point of view on the use and selection of sources is described.

In certain contexts it may be the case that the most relevant information for the study is not to be found in scientific literature but rather with individual software developers and open

source projects. It must then be clearly stated that efforts have been made to gain access to this information, e.g. by direct communication with developers and/or through discussion forums, etc. Efforts must also be made to indicate the lack of relevant research literature. The precise manner of such investigations must be clearly specified in a method section. The paragraph on source criticism must critically discuss these approaches.

Usually however, there are always relevant related research. If not about the actual research questions, there is certainly important information about the domain under study.

5.3 The work in a wider context

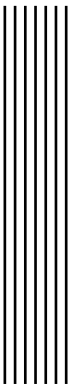
There must be a section discussing ethical and societal aspects related to the work. This is important for the authors to demonstrate a professional maturity and also for achieving the education goals. If the work, for some reason, completely lacks a connection to ethical or societal aspects this must be explicitly stated and justified in the section Delimitations in the introduction chapter.

In the discussion chapter, one must explicitly refer to sources relevant to the discussion.

6 Conclusion

This chapter contains a summarization of the purpose and the research questions. To what extent has the aim been achieved, and what are the answers to the research questions?

The consequences for the target audience (and possibly for researchers and practitioners) must also be described. There should be a section on future work where ideas for continued work are described. If the conclusion chapter contains such a section, the ideas described therein must be concrete and well thought through.



Bibliography

[1] Aditya Bagari, Ashish Kumar, Avinash Kori, Mahendra Khened, and Ganapathy Krishnamurthi. "A Combined Radio-Histological Approach for Classification of Low Grade Gliomas". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Ed. by Alessandro Crimi, Spyridon Bakas, Hugo Kuijf, Farahani Keyvan, Mauricio Reyes, and Theo van Walsum. Cham: Springer International Publishing, 2019, pp. 416–427. ISBN: 978-3-030-11723-8.

[2] Jocelyn Barker, Assaf Hoogi, Adrien Depeursinge, and Daniel L. Rubin. "Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles". In: *Medical Image Analysis* 30 (2016), pp. 60–71. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2015.12.002. URL: https://www.sciencedirect.com/science/article/pii/S1361841515001838.

[3] T. Editors of Encyclopaedia Britannica. "Histology". In: *Encyclopedia Britannica* (Oct. 2013).

[4] T. Editors of Encyclopaedia Britannica. "Pathology". In: *Encyclopedia Britannica* (Nov. 2014).

[5] Gabriele Campanella, Matthew Hanna, Luke Geneslaw, Allen Miraflor, Vitor Silva, Klaus Busam, Edi Brogi, Victor Reuter, David Klimstra, and Thomas Fuchs. "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images". In: *Nature Medicine* 25 (Aug. 2019), p. 1. DOI: 10.1038/s41591-019-0508-1.

[6] Mehmet Ertosun and Daniel Rubin. "Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of convolutional neural networks". In: *AMIA Annu Symp Proc* 2015 (Nov. 2015), pp. 1899–1908.

[7] Azam Hamidinekoo, Tomasz Pieciak, Maryam Afzali, Otar Akanyeti, and Yinyin Yuan. "Glioma Classification Using Multimodal Radiology and Histology Data". In: Nov. 2020.

[8] Le Hou, Dimitris Samaras, Tahsin Kurc, Yi Gao, James Davis, and Joel Saltz. "Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification". In: vol. 2016. June 2016. DOI: 10.1109/CVPR.2016.266.

[9] Andrew. Janowczyk and Anant. Madabhushi. "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases". In: *Journal of Pathology Informatics* 7.1 (2016), p. 29. DOI: 10.4103/2153-3539.186902.

[10] Justin Ker, Yeqi Bai, Hwei Yee Lee, Jai Rao, and Lipo Wang. "Automated brain histology classification using machine learning". In: *Journal of Clinical Neuroscience* 66 (2019), pp. 239–245. ISSN: 0967-5868. DOI: https://doi.org/10.1016/j.jocn.2019.05.019. URL: https://www.sciencedirect.com/science/article/pii/S0967586819306563.

[11] Tahsin Kurc, Spyridon Bakas, Xuhua Ren, Aditya Bagari, Alexandre Momeni, Yue Huang, Lichi Zhang, Ashish Kumar, Marc Thibault, Qi Qi, Qian Wang, Avinash Kori, Olivier Gevaert, Yunlong Zhang, Dinggang Shen, Mahendra Khened, Xinghao Ding, Ganapathy Krishnamurthi, Jayashree Kalpathy-Cramer, James Davis, Tianhao Zhao, Rajarsi Gupta, Joel Saltz, and Keyvan Farahani. "Segmentation and Classification in Digital Pathology for Glioma Research: Challenges and Deep Learning Approaches". In: *Frontiers in Neuroscience* 14 (2020), p. 27. ISSN: 1662-453X. DOI: 10.3389/fnins.2020.00027. URL: https://www.frontiersin.org/article/10.3389/fnins.2020.00027.

[12] Louis. In: ().

[13] Louis. In: (). DOI: https://doi.org/10.3322/caac.21551.

[14] mbfbioscience. In: ().

[15] Alexandre Momeni, Marc Thibault, and Olivier Gevaert. "Dropout-Enabled Ensemble Learning for Multi-scale Biomedical Data". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Ed. by Alessandro Crimi, Spyridon Bakas, Hugo Kuijf, Farahani Keyvan, Mauricio Reyes, and Theo van Walsum. Cham: Springer International Publishing, 2019, pp. 407–415. ISBN: 978-3-030-11723-8.

[16] Ostrom. In: (). DOI: doi:10.1093/neuonc/noy131.

[17] Saima Rathore, Tamim Niazi, Muhammad Aksam Iftikhar, and Ahmad Chaddad. "Glioma Grading via Analysis of Digital Pathology Images Using Machine Learning". In: *Cancers* 12.3 (2020). ISSN: 2072-6694. DOI: 10.3390/cancers12030578. URL: https://www.mdpi.com/2072-6694/12/3/578.

[18] rcpath. In: ().

[19] Amin Shirazi, Eric Fornaciari, Narjes Bagherian, Lisa Ebert, Barbara Koszyca, and Guillermo Gomez. "DeepSurvNet: deep survival convolutional network for brain cancer survival rate classification based on histopathological images". In: *Medical & Biological Engineering & Computing* 58 (Mar. 2020). DOI: 10.1007/s11517-020-02147-3.

[20] Sung. In: (). DOI: doi:10.3322/caac.21660.

[21] Xi Wang, Hao Chen, Caixia Gan, Huangjing Lin, Qi Dou, Efstratios Tsougenis, Qitao Huang, Muyan Cai, and Pheng-Ann Heng. "Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis". In: *IEEE Transactions on Cybernetics* PP (Sept. 2019), pp. 1–13. DOI: 10.1109/TCYB.2019.2935141.

[22] Y. Xu, Zhipeng Jia, L. Wang, Yuqing Ai, F. Zhang, Maode Lai, and E. Chang. "Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features". In: *BMC Bioinformatics* 18 (2017).