Automatic identification and classification of brain tumours

Jonathan Rittmo
University of Gothenburg
Gothenburg, Sweden

May 26, 2021

Abstract

Medical image analysis is an increasingly important concept for aiding patient diagnosis. But it is a tedious and time consuming task relying on subjective (albeit expert) knowledge of clinicians and technicians. Decision support systems relying on standardised analysis pipelines and classification could greatly aid and speed up these processes. Brain tumours are one of the more important diseases when it comes to early classification and treatment selection. In this project I propose an analysis pipeline for automatic segmentation unhealthy tissue, based on a dataset of 3064 MR images from tumour patients with 3 different kinds of tumours. The pipeline consists of three parts: preprocessing, segmentation and classification. The aim of the project is to see whether automatic segmentation using a residual neural net can improve classification of extracted features using support vector machines, compared to classification of features extracted from full images. Results indicate that the proposed segmentation yields a better accuracy in the classification part than using full images. However, the improvement is only slight. The small dataset is a limitation of the study and future research with more data should be performed to solidify any potential effects.

***K*eywords** Brain tumor classificaton · Image segmentation · Support vector machines · ResNet

1 Introduction

Decision support systems that facilitate clinical diagnoses are becoming more and more important in health care systems. In this project I will focus on classification of the brain tumour types: meningioma, glioma and pituitary tumour. A tumour is a lump of cells characterised by its uncontrolled growth. These lumps can be either malignant or benign where the malignant tumours are recognised through their heterogenous shapes. All of these tumour types can be malignant, but meningioma and pituitary tumours are generally benign. Glioma on the other hand make up 80% of all malignant tumours in adults (Chen et al., 2012). Detection of malignant brain tumours at an early stage is essential

for treatment selection and the survival rate of the patient and hence have become an important issue in clinical image analysis. So the issue of segmentation of MR images have been considered thoroughly. There are many exisiting techniques, where support vector machines (SVM) and neural nets are some of the more popular choices of models (e.g. Alfonse & Salem, 2016; Damodharan & Raghavan, 2015; Guo et al., 2011; Torheim et al., 2014).

Segmentation is the process of clustering the pixels of an image into k classes dependent on some shared properties. For the removal of the skull in MR images one must segment out the background which in the present dataset, due to the quality and uniform background, could be done with simple thresholding. Extracting tumour tissue on the other hand requires pixels in images to be classified either as tumour tissue or healthy tissue (Abdel-Maksoud et al., 2015; Ain et al., 2014). This is a more difficult task since the intensities of the tissue (healthy or unhealthy) pixels are more alike than tissue and background. And even though SVMs are popular for this kind of task I will use a residual neural net (He et al., 2016) for the segmentation and instead SVM for the tumour type classification task.

In my proposed analysis pipeline the essential parts involves: removal of non-nervous tissue through thresholding, segmentation of infected tissue using residual neural nets (ResNet), feature extraction (first and second order statistics) from the segmented regions and classification of these features into the categories "meningioma," "glioma" and "pituitary." The proposed pipeline is summarised in 1.

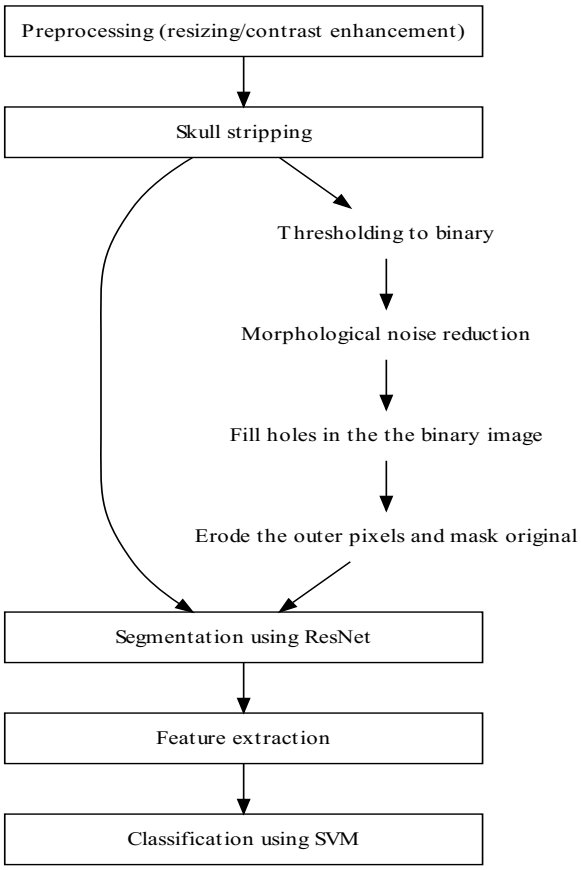

Figure 1: Proposed analysis pipeline for automatic classificaton of tumours from MR images.

2 Analysis

2.1 Data

The data for this project was 3064 T1 weighted 512x512 magnetic resonance images from 233 patients with three types of tumours: meningioma, glioma and pituitary tumours. These were resized to 256x256 to speed up computation using nearest-neighbour interpolation. The data has been used in Cheng et al. (2015) and Cheng et al. (2016) and contains ground truth masks of tumour regions generated by manual delineation of tumour regions. There are two problems with this dataset: first the data is grouped on patient level with varying number of observations per patient and second it is imbalanced with the respect to number of images and type of tumour. The grouped data would probably not be problematic if this was purely a segmentation task but because we want to classify each tumour, it is important to ensure that no patient appears both in train and test data simultaneously. It should also be noted that the dataset is not really imbalanced with respect to classes if you take the grouping into account, this can be seen in Figure 2 where there is a quite hefty skew between the classes when we look at the observation level but just a small skew disfavouring pituitary tumours.

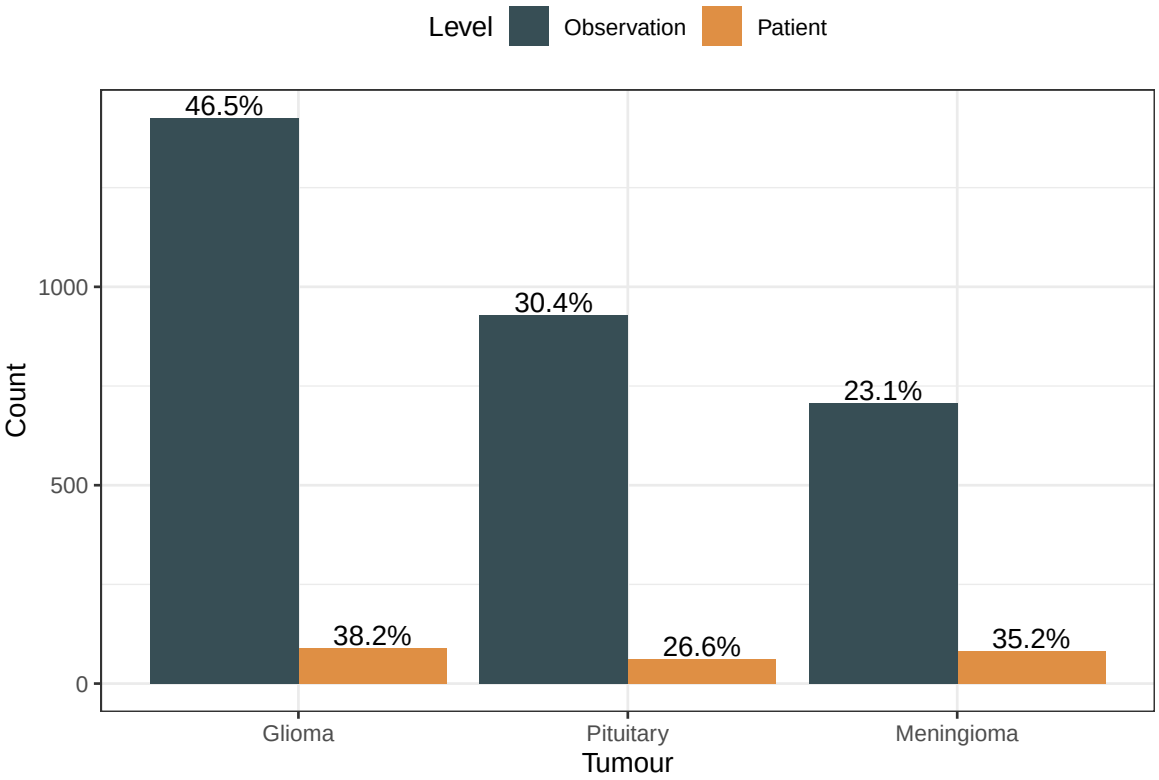

Figure 2: Imbalance of dataset.

When partitioning the data into train and test sets for the segmentation task no consideration was taken to the classes of tumours, but the data was partitioned with consideration to patient ID. For the classification task on the other hand the data was partitioned on patient and stratified with regards to tumour class. A sample of the original images from the dataset can be seen in Figure 3.

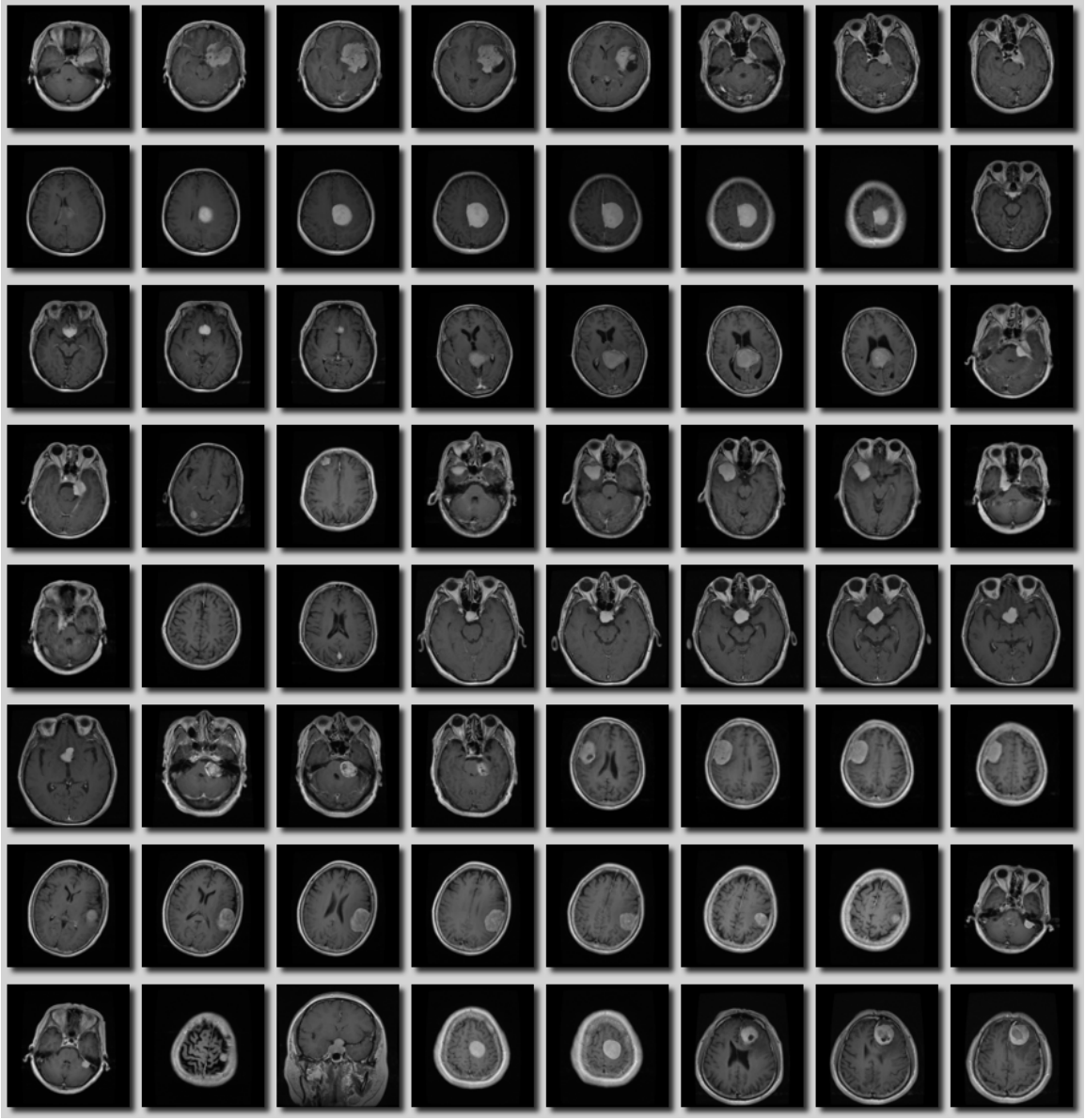

Figure 3: Example images from the original dataset.

2.2 Preprocessing

One of the most important parts of medical image analysis of brain tissue is the removal of non-nervous tissue such as the skull. This is usually referred to as skull stripping and several techniques are available to do this. Morphology based skull stripping is one of the most commonly used techniques and has proved to work well (Benson & Lajish, 2014). Hence it was the technique that I used for this project as well. The overall processing pipeline can be seen in Figure 1. The images were first normalised to a scale in the range [0, 1] and then I used contrast-limited adaptive histogram equalization (CLAHE) to enhance contrasts of the images. Simple histogram equalisation computes a single histogram for the entire image, but in contrast CLAHE computes several histograms for different sections of the image (Zuiderveld, 1994). In that way it improves local contrasts and edges – suitable for tumour detection.

After this enhancement simple thresholding was very effective to segment out any tissue from the background and create a binary mask due to the quality of the dataset and the clear intensity differences. Area opening was then applied to the mask to remove potential noise, i.e. removing connected objects with fewer than 10 pixels and any holes in the main object (i.e. the skull) was filled. The outer pixels were then eroded away using a disk shaped structuring element with a 15 pixel radius. This mask was then applied to the gray-scale image to remove the skull. Results of these operations on the images in 3 can be seen in Figure 4.

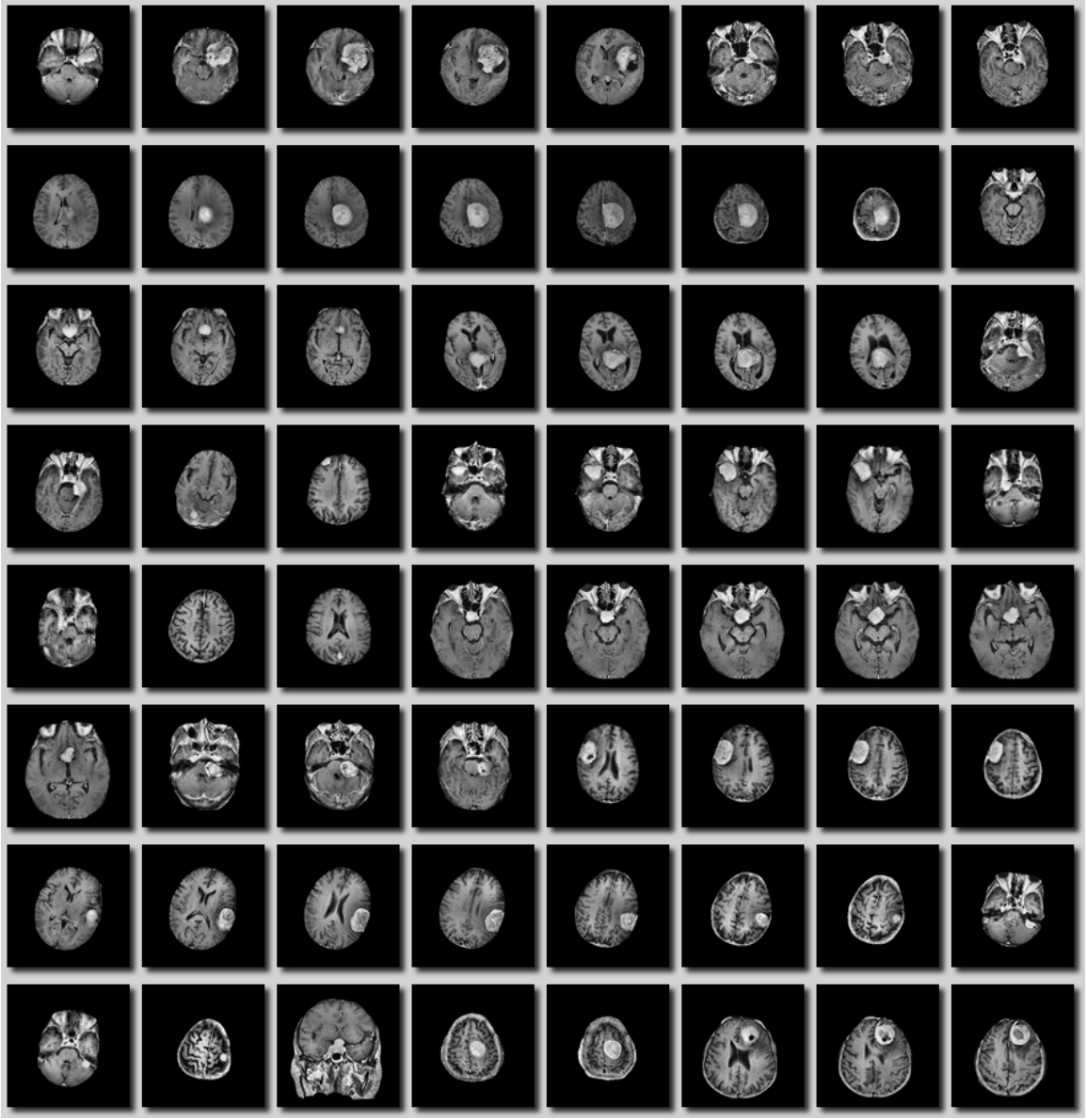

Figure 4: Skull stripped and contrast enhanced example images.

2.3 Segmentation

The segmentation was conducted by training a residual neural net (He et al., 2016) on the ground truth images so that each pixel in every image was given the label "tumour," "normal" or "background."

The ground truth for tumour has been manually delineated by professionals (Cheng et al., 2015) and the "normal" pixels was segmented out by simple thresholding. Residual neural nets have been used extensively in semantic segmentation tasks and proven effective in tumour detection (e.g. Zeineldin et al., 2020). I created a DeepLab v3+ convolutional neural network based on ResNet-18 which is an 18 layer deep pre-trained residual neural net, in Matlab. DeepLab v3+ is a network used for semantic segmentation and basing it on a predefined network adds a few additional layers to it (i.e. to ResNet-18) that are set up for the specific pixel classification and image input. The data was partitioned 60/20/20 for training, validation and testing in such a way that no patient appeared in either set simultaneously.

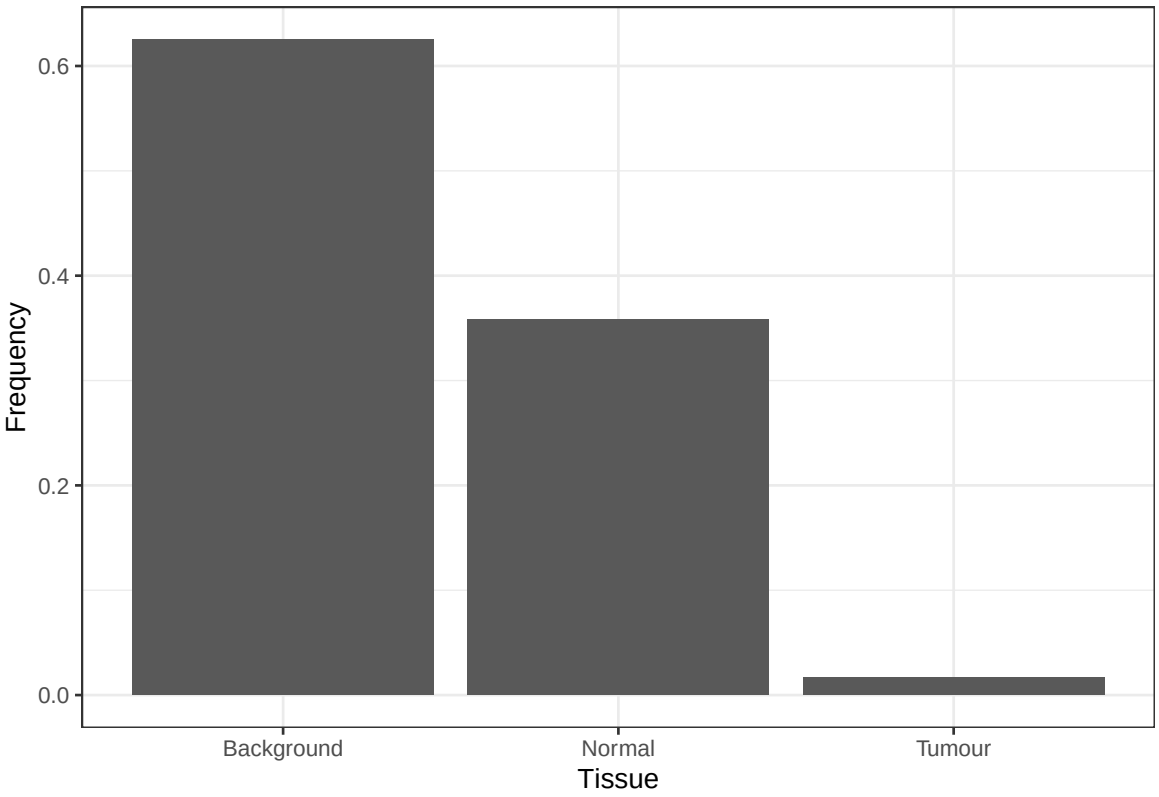

Figure 5: Pixel frequency of the three classes.

Due to both background and normal tissue pixels being overrepresented in the images as seen in Figure 5 I balanced the classes by weighting them relative to the frequency of the normal tissue. So the weight of "tumour" was $0.3583/0.0166 = 21.6205$ and so on. This effectively punish the network more for misclassifying a "tumour" pixel than "background" or "normal." The network was trained using stochastic gradient descent with momentum optimisation with weight decay (L2 regularisation) and early stopping to avoid over-fitting. Some evaluation metrics of the network can be seen in Table 1.

IoU refers to the correctly classified pixels divided by the sum of false positives and false negatives and the mean BF score refers to how well the predicted boundary of an object aligns with the ground truth boundary. And as can be seen neither of these scores are very good for the "tumour" class, while the accuracy still is (moderatly) high. This indicates that the tumour regions predicted by the network probably are much larger than the ground truth tumour. In Figure 6 the first 64

Table 1: Evaluation metrics of semantic segmenation on test data

	Accuracy	IoU	MeanBFScore
Tumour	0.840	0.386	0.221
Normal	0.943	0.924	0.804
Background	0.992	0.989	0.982

segmented images in the test set are shown where the yellow represents ground truth tumour and the non-coloured areas represent the predicted tumour region. As noted it seems like the network often classify much larger areas than the ground truth.

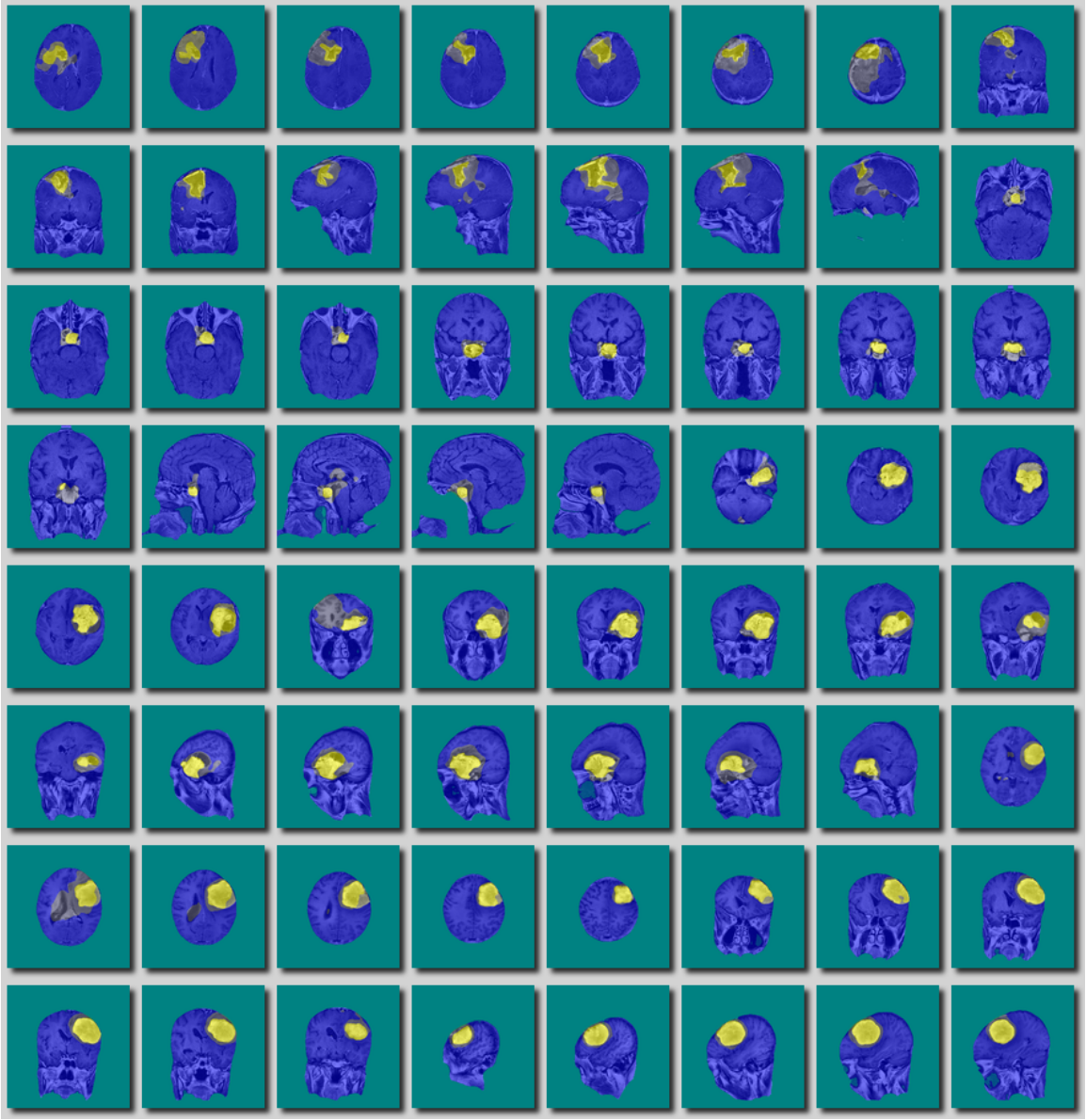

Figure 6: Segmented images from test set.

Since I am going to use the predicted tumour areas as masks for feature extraction the irrelevant false positive pixels (i.e. the larger areas classified as tumour) might affect later classification of tumour type. However, Cheng et al. (2015) show that using the surrounding tissue of tumours can aid classification due to the fact that they often grow in different areas of the brain and hence I went on with the analysis.

2.4 Tumour classification

For classifying type of tumour the segmented tumour areas were used as masks to hide any irrelevant information when extracting features (eight of the 3064 tumours were not detected by the network at all so no mask was used for these images). Preferably classification of tumour would have been done only on the images in the test dataset from the segmentation task. However, due to the small dataset, the predicted tumour areas from the training images were used as well. It is common to use features based on Wavelet transformation when classifying magnetic resonance images. However, Aggarwal & K. Agrawal (2012) found that first and second order statistical features yielded better results. Hence, the extracted first order features were: mean, variance, skewness and kurtosis. In addition Hu's seven invariant moments (Hu, 1962) were extracted to represent the shape of the tumours. Unfortunately these moments are mostly appropriate if we have a very clean segmentation of the objects of interest which, as noted and displayed in Figure 6, is not the case for many of the images in this dataset. However, due to the discriminatory power of these moments when objects are segmented properly these moments they were included anyway.

The extracted second order features were correlation, energy and homogeneity. Contrast is also a commonly used second order feature but since it is basically the same thing as variance it was left out of the analysis. These features are calculated from a normalised gray-level spatial dependence matrix, which is a matrix with counts of how often a pixel with grey-level intensity I_1 occurs horizontally adjacent to a pixel with grey-level intensity I_2. In this analysis the grey-levels of the image were binned to 8 values so that the gray-level spatial dependence matrix was of size 8x8. Correlation is a measure of the linear dependency between a pixel and its neighbour, for all the pixels in the image. Energy is the summation of the squared elements in the normalised spatial dependence matrix and homogeneity is a measure of the closeness to the diagonal of the distribution of the elements in the spatial dependence matrix. That is, if no pixel of grey-level intensity I_1 appeared adjacent to any pixel of grey-level intensity I_2 the spatial dependence matrix would be a diagonal matrix and the homogeneity would be 1.

So, these features were extracted from the areas predicted by the segmentation model and then fed into a support vector machine (SVM) with a Gaussian kernel. SVM was the model of choice because it has been used extensively for these types of classification problems (Aggarwal & K. Agrawal, 2012; e.g. Bahadure et al., 2017) with good results. After training the model had a ten fold cross validation error of 0.1877 and a test error of 0.1794 indicating that the model can generalise to some degree even if a better accuracy would have been desirable. To have something to compare against I also trained the same model on features extracted from the full images (i.e. the skull stripped images without tumour masking). This model got a ten fold cross validation error of 0.1914 and a test error of 0.2292. So the segmentation seems to improve performance somewhat even if the improvement is slight. To get an intuition of the possibly better discriminatory power when using segmented images one can observe Figure 7 where the features have been reduced to the first two principal components. Interestingly we can see that the most separable classes are "glioma" and "pituitary"

which is consistent with the mentioned fact that pituitary tumours generally are benign, gliomas make up ~80% of the malignant tumours in adults and malignant tumours are characterised by their heterogenous shapes.

We do however not know the malignancy of the tumours in this dataset. But if we merge the meningiomas and pituitary tumours and call this class "generally benign tumours" and call the gliomas "generally malignant tumours" and train a model with these labels we achieve a ten fold cross validation and test error 0.1050 and 0.0774 on the segmented images compared to a ten fold cross validation and test error of 0.1520 and 0.2170 respectively. Further indicating that segmentation of the images, at least to predict malignancy, is helpful.

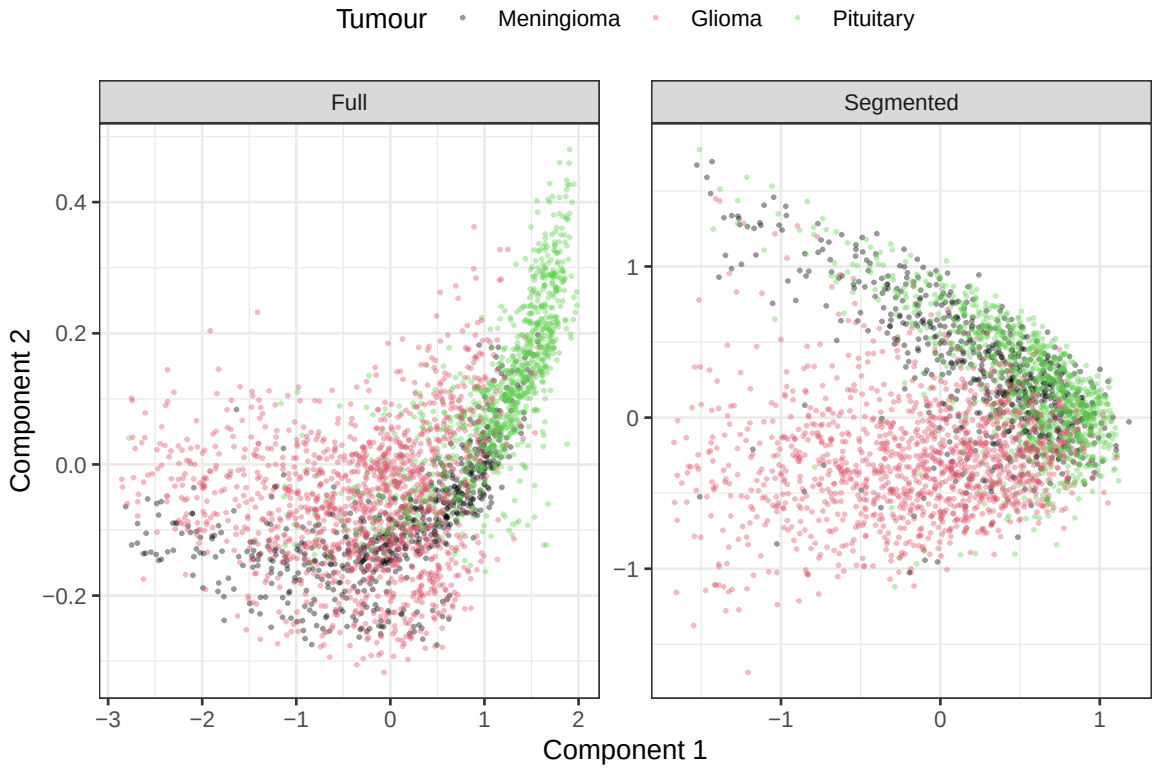

Figure 7: Dimension reduced feature space from segmented and full images.

Table 2: Evaluation of class-wise predictions on segmented test data.

	Meningioma	Glioma	Pituitary
Sensitivity	0.706	0.889	0.779
Specificity	0.895	0.913	0.926
Pos Pred Value	0.622	0.914	0.814
Neg Pred Value	0.925	0.887	0.910
Precision	0.622	0.914	0.814
Recall	0.706	0.889	0.779
Prevalence	0.197	0.510	0.293
Detection Rate	0.139	0.454	0.228
Detection Prevalence	0.223	0.497	0.280
Balanced Accuracy	0.800	0.901	0.853

As can be seen the feature separability is somewhat increased for the segmented images even though the effect might be slight. Furthermore, Figure 7 seems to indicate that Meningioma is the most difficult class to discriminate. Looking at class-wise prediction performance on test data in Table 2, this is obviously the case. The measures are calculated by comparing each class to the remaining classes and their definitions can be find in appendix A. As can be seen Meningioma has the lowest scores on all measures except for the negative predictive value (i.e. the probability that a tumour is not Meningioma if it has been classified as anything else). However, both the positive and negative predictive values are heavily dependent on the prevalence (base rate) of the classes which we can see is imbalanced in the test set. The reason for this imbalance even though the the training and testing set were stratified on tumour class is probably that the data also were partitioned on patient and each patient had a varying number of images. Therefore exactly equal ratios of the classes in test and training data were difficult to achieve.

3 Discussion and conclusion

This project has looked at methodology to automatically classify MR images from brain cancer patients into three different tumour types. This analysis was done in three major steps: preprocessing, segmentation and classification. The preprocessing, with skull stripping being the most important operation to automatise, was in general successful. The segmentation on the other hand had a more ambiguous result. For some images it worked great but overall it had a high false positive rate. Extracting shape features based on dirty segmentation is generally not a good idea and as we saw for the classification part of the project the result for segmented images was only marginally better than using the full. Another limitation of this project was the small dataset used. Preferably, the tumour classification should only have been performed on the test set from the segmented images to fully evaluate performance of this analysis pipeline but due to the limitations of such a small dataset I chose to perform the classification on the segmented images from the train set as well. Hence, the results of the study might not be fully generalisable and more research is encouraged. All in all, however, the slight classification improvement of the segmented images shows that using automatic segmentation pipelines could aid the classification and diagnosis of brain tumours.

References

Abdel-Maksoud, E., Elmogy, M., & Al-Awadi, R. (2015). Brain tumor segmentation based on a hybrid clustering technique. *Egyptian Informatics Journal*, *16*(1), 71–81. https://doi.org/10.1016/j.eij.2015.01.003

Aggarwal, N., & K. Agrawal, R. (2012). First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images. *Journal of Signal and Information Processing*, *03*(02), 146–153. https://doi.org/10.4236/jsip.2012.32019

Ain, Q., Jaffar, M. A., & Choi, T.-S. (2014). Fuzzy anisotropic diffusion based segmentation and texture based ensemble classification of brain tumor. *Applied Soft Computing*, *21*, 330–340. https://doi.org/10.1016/j.asoc.2014.03.019

Alfonse, M., & Salem, A.-B. M. (2016). *An Automatic Classification of Brain Tumors through MRI Using Support Vector Machine*. *40*(03), 11.

Bahadure, N. B., Ray, A. K., & Thethi, H. P. (2017). Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM. *International Journal of Biomedical Imaging*, *2017*, 1–12. https://doi.org/10.1155/2017/9749108

Benson, C. C., & Lajish, V. L. (2014). Morphology Based Enhancement and Skull Stripping of MRI Brain Images. *2014 International Conference on Intelligent Computing Applications*, 254–257. https://doi.org/10.1109/ICICA.2014.61

Chen, J., McKay, R. M., & Parada, L. F. (2012). Malignant Glioma: Lessons from Genomics, Mouse Models, and Stem Cells. *Cell*, *149*(1), 36–47. https://doi.org/10.1016/j.cell.2012.03.009

Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., Wang, Z., & Feng, Q. (2015). Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition. *PLOS ONE*, 13.

Cheng, J., Yang, W., Huang, M., Huang, W., Jiang, J., Zhou, Y., Yang, R., Zhao, J., Feng, Y., Feng, Q., & Chen, W. (2016). Retrieval of Brain Tumors by Adaptive Spatial Pooling and Fisher Vector Representation. *PLOS ONE*, 15.

Damodharan, S., & Raghavan, D. (2015). *Combining Tissue Segmentation and Neural Network for Brain Tumor Detection*. *12*, 11.

Guo, L., Zhao, L., Wu, Y., Li, Y., Xu, G., & Yan, Q. (2011). Tumor Detection in MR Images Using One-Class Immune Feature Weighted SVMs. *IEEE Transactions on Magnetics*, *47*(10), 3849–3852. https://doi.org/10.1109/TMAG.2011.2158520

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, *8*(2), 179–187. https://doi.org/10.1109/TIT.1962.1057692

Torheim, T., Malinen, E., Kvaal, K., Lyng, H., Indahl, U. G., Andersen, E. K. F., & Futsæther, C. M. (2014). Classification of Dynamic Contrast Enhanced MR Images of Cervical Cancers Using Texture Analysis and Support Vector Machines. *IEEE Transactions on Medical Imaging*, *33*(8), 1648–1656. https://doi.org/10.1109/TMI.2014.2321024

Zeineldin, R. A., Karar, M. E., Coburger, J., Wirtz, C. R., & Burgert, O. (2020). DeepSeg: Deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images. *International Journal of Computer Assisted Radiology and Surgery*, *15*(6), 909–920. https://doi.org/10.1007/s11548-020-02186-z

Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In *Graphics gems IV* (pp. 474–485). Academic Press Professional, Inc.

A Appendix A

Given a two-class classification problem and the confusion matrix:

Table 3: Confusion matrix for a two-class problem.

		Predicted class	
		Class 1	Class 2
True class	Class 1	A	B
	Class 2	C	D

Then the measurements in Table 2 are defined as:

$$\text{Sensitivity} = \frac{A}{(A+C)}$$

$$\text{Specificity} = \frac{D}{(B+D)}$$

$$\text{Prevalence} = \frac{(A+C)}{(A+B+C+D)}$$

$$\text{PPV} = \frac{\text{sensitivity} * \text{prevalence}}{(\text{sensitivity} * \text{prevalence}) + [(1-\text{specificity})(1-\text{prevalence})]}$$

$$\text{NPV} = \frac{\text{specificity} * (1-\text{prevalence})}{[(1-\text{sensitivity}) * \text{prevalence}] + [(\text{specificity})(1-\text{prevalence})]}$$

$$\text{Detection Rate} = \frac{A}{A+B+C+D}$$

$$\text{Detection Prevalence} = \frac{A+B}{A+B+C+D}$$

$$\text{Balanced Accuracy} = \frac{\text{sensitivity} + \text{specificity}}{2}$$

$$\text{Precision} = \frac{A}{A+B}$$

$$\text{Recall} = \frac{A}{A+C}$$

However, note that since the classification in this project was a multi-class problem the measurements are calculated by comparing one class to the all the other.