8DC00 Medical Image Analysis

Project 2 – Computer Aided Diagnosis

Report

29-10-2021

Group 3

Aiik Biermans – 1241616

Brigitte van der Geest – 1464027

Pauline Haulez – 1462245

Willem Schellekens - 1636308

# Index

[Index 2](#_Toc85622611)

[1. Introduction 3](#_Toc85622612)

[2. Methods 3](#_Toc85622613)

[3. Results 3](#_Toc85622614)

[4. Discussion 3](#_Toc85622615)

[References 3](#_Toc85622616)

[Appendix 3](#_Toc85622617)

# 1. Introduction

Computer aided diagnosis (CAD) has become standard practice for the detection of, for example, breast cancer at many screening sites and hospitals. An example of this is a CAD network which analyzes images of cell tissue and aids in the classification of these cells into malignant or benign cells. CAD is developing rapidly and has become one of the major research areas of medical imaging. (Doi, 2007)  
The size of the cell nuclei of cancer cells in the breast can be a useful feature to give a prognosis of the aggressiveness of the tumor. The bigger the nuclei the worse the prognosis. Automating the measurement of the cell nuclei saves much time in pathology workflow. The automatic classification also opens up possibilities for large scale analysis, while taking a fraction of the time when compared to manual analysis.

The goal of this project is to create a program that automates the cell nuclei measurement and classifies these cells to the large nuclei class or the small nuclei class. A prediction model for the size of the nuclei is given with the use of a linear regression model, where the calculated error of the fit of the model is evaluated. The classification is done with the use of a logistic regression model and evaluated with the outcoming accuracy. Multiple variations of parameters will be tried to find the best performing model, which is the model with the maximum accuracy and minimum loss.

Furthermore, this report will end with a comparison of these linear and logistic regression models to deep neural networks as described in the paper of Graham et al. (2019)

# 2. Methods

The used dataset, the computer-aided diagnosis methods of linear and logistic regression, and the evaluations of these methods are elaborated below.

## 2.1 Dataset

The dataset provided consists of RGB images of nuclei with size 24x24 pixels retrieved from Veta et al. (2015). The dataset involved 39 slides from patients with invasive breast cancer. The representative tumor regions of size 1x1 mm were selected by a pathologist. In the tumor regions of a size of 1x1 mm, which is a size of 4000x4000 in pixels, around 100 nuclei were manually segmented.

The dataset with 39 slides are divided in 21 slides for the training dataset with 2191 segmented nuclei. The other 18 slides are used as test dataset with 2073 segmented nuclei.

## 2.2 Linear regression

Linear regression is a method to estimate the association between a continuous dependent variable and an independent variable by fitting a linear line to the observed data. The equation of linear regression line is shown in formula 2.1. X is the independent variable and Y is the dependent variable. Due to the fitted linear line with a specific equation, for new data the outcome can be predicted.

(2.1)

In this project, the X are the RGB images with size of 24x24 pixels of the training dataset. Each image gives an input of 24x24x3 features. The output, the dependent variable Y, is the size of the nuclei. The 300 smallest and the 300 largest nuclei are picked and visualized. Then, the resulted equation from the fitted linear line of the training data set is used to predict the areas of the nuclei in the test dataset.

## 2.3 Logistic regression

Three independent datasets are used for training, validating, and testing of the logistic regression model for classifying nuclei. The datasets consist of nuclei images with corresponding ground truths. Nuclei are either classified as small or large . The trained logistic regression model uses the sigmoid function (formula 2.2) to make a prediction (p) of the chance whether a nucleus is classified as small (p<0.5) or large (p>0.5). Training of the model is done by minimizing the training and validation losses. The function loss is given by formula 2.3.

The weights are adjusted to obtain the minimum for the loss function. For the training of the model, several starting parameters are used. These parameters consist of the learning rate (mu), batch size, total number of epochs and theta. The learning rate is the step size in gradient descent that determines the magnitude of in- or decrease of the model parameters. A good initial value for theta ensures the loss on the first epoch is already close to the minimum. Training of the model stop once the total number of epochs, or iterations, is reached. The values of the initial parameters are found by trial and error. The effects of the changing starting parameters are visualized by using starting parameters above and below current values. For increased accuracy, a variable learning rate is used, resulting in a smaller learning rate once the validation loss increases. To limit the number of epochs and time needed, a stopping criterium is set. As the validation loss is the most stable, the training loop is terminated once the differences in validation loss between the last four epochs is beneath a certain threshold. The last four epochs are needed, to avoid outliers triggering a premature termination of training the model. The effect of the stopping criteria threshold is also compared. Lastly, the effect of the training set size is considered. The model is trained with a reduced training dataset with factor 0.5% of the original size (14607) to a new training dataset of size 73.

## 2.4 Methods of evaluation

Linear regression evaluation

To evaluate the result of the linear regression, the distance of every actual data point to the fitted linear regression line is calculated with the formula 2.2. The result is called the error of the linear regression model. A smaller number of the error means a better approximation of the linear regression model.

(2.4)

The effect on the error of a smaller training dataset is also evaluated. For this evaluation only the first 5000 segmented nuclei are used instead of the full training dataset with 21910 input samples.

Logistic regression evaluation

Evaluation of the results of logistic regression was done by determining the accuracy of the predictions by comparing the predicted results of an independent test set with the ground truths. This prediction is done with formula 2.3, and compares the ground truth (*y)* with the predicted classification (*ŷ)*. (source <https://scikit-learn.org/stable/modules/model_evaluation.html#accuracy-score>)

The result of this accuracy score is the fraction of correct predictions. In addition, a comparison is made of the prediction certainties of the model to the ground truths. A histogram is plotted, indicating the possibilities of the individual nuclei in the test set belonging to a certain classification.

# 3. Results

The results of the linear and logistic regressions and their evaluations are mentioned below.

## 3.1 Linear regression

The results of the linear regression to predict the area of the nuclei in the test dataset is visualized in figure 1. On the left the result of the linear regression with an input of the full training dataset of 21910 samples is shown. The error of this linear regression is 374.9 pixels to the fourth power. On the right, the result of the regression with an input of only 5000 samples of the training dataset is shown. The error of this linear regression is 744.2 pixels to the fourth power. The red line in both figures is the actual area of the test dataset on both the x axis and y axis. In figure 2 in the appendix, the 300 smallest and the 300 largest nuclei of the training dataset are shown as a result of the linear regression model.

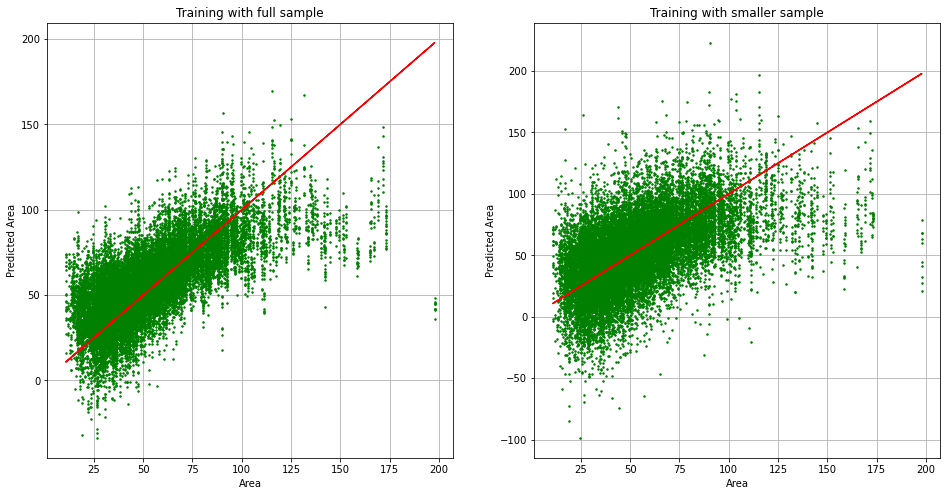


Figure 1 Predicted area versus actual area of the linear regression for the full and small training dataset

## 3.2 Logistic regression

By trial and error, the initial parameters for mu, the batch size, the total number of epochs and theta was determined. The values for the initial parameters are shown in table one.

Table 1 Initial parameters logistic regression

|  |  |
| --- | --- |
| Mu | 0.00003 |
| Batch size | 350 |
| Number of epochs | 1000 |
| Theta | -0.000105 |

Ten models are trained, validated and tested with the same data. In figure 3, for one model, the training loss, validation loss and test accuracy are plotted for every epoch. On average, the stopping criterion terminates the models after 390.2 (±175.3) epochs with a validation loss of 0.42 (±0.015).

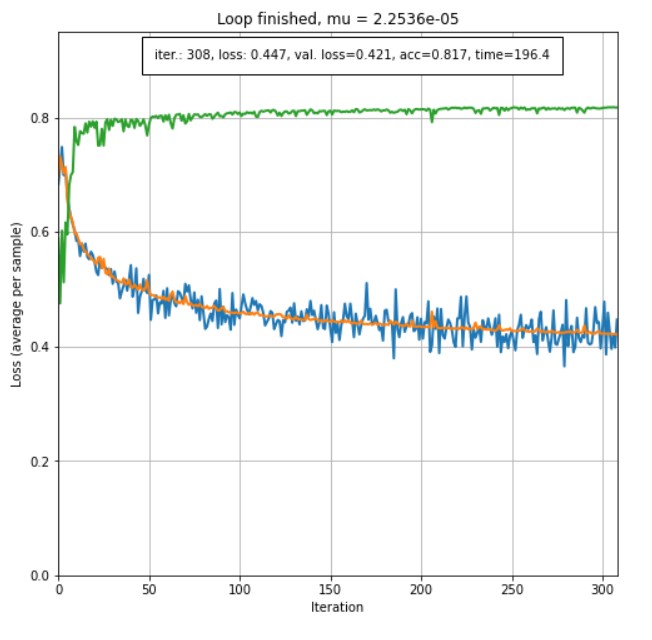


Figure 2: Training loss (blue), validation loss (orange) and accuracy (green).

## 3.3 Variation of hyperparameters

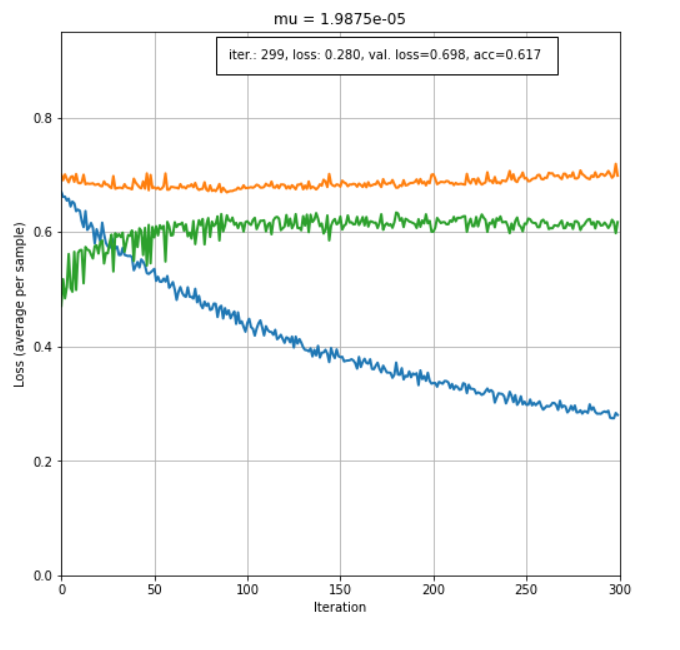
Figures 4-6 show the effect on validation and training loss curves when the learning rate is increased or decreased ten times. Note the axis values are the same in all three figures.

The next three figures in appendix B, show the effect of varying the batch size. The figures do not show the runtime of the optimization for the three different batch sizes.

Lastly, the final two images of appendix B show the effect of varying the initial value for \theta.

In figure 9 of appendix B the effect of lower threshold criterium is shown. This model is trained for 955 epochs, with an accuracy of 0.826.

To test the effect of a much smaller dataset, a model is trained with a small fraction of the original training set. The effect is visualized in figure 4. As the training loss decreases, the validation loss remains around 0.7.

  
Figure 4: Loss curves and accuracy for a smaller training set

## 3.4 Logistic regression evaluation

Ten different models, trained with the same data, predict the classifications of the test set with a mean accuracy of 0.818 (±0.0052). Figure 14 shows the prediction histogram of a finished model.

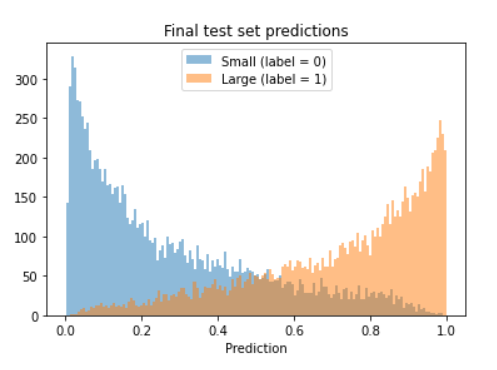


Figure 5: Prediction histogram

# 4. Discussion

The goal of this project was to create a program that automates the cell nuclei measurements and classifies these cells to the large nuclei class or the small nuclei class, because the size of cell nuclei can be useful by the diagnosis of breast cancer.

Firstly, the linear regression model calculated the area of the segmented nuclei. The nuclei were manually segmented, so there are human errors possible due to this segmentation method. The outcome of the linear regression of the full training dataset resulted in a smaller error than the outcome of the model with an input of half the sample size of the training dataset. This model with the full dataset, can be seen as a good prediction method of the size of the nuclei, because the error of 375 pixels to the fourth power is very small in comparison of input of the model (the full-size tumor regions of 4000x4000 pixels).

Figure 1 shows that nuclei with a size bigger than 100 squared pixels are almost always underestimated in the model. This shows the model is not suitable for very large nuclei, with a size larger than 100 squared pixels.

Secondly, the logistic regression model categorized the area of the segmented nuclei into the small or large group. The values of the parameters (learning rate, batch size, number of epochs and theta) are selected by trial and error to minimize the resulted loss function. It always remains unclear when the optimum for the minimalizes loss functions is reached and the parameters are the best possible values. The combinations of all these parameters resulted in a lot of different combinations.

In figures 1-3 of appendix B the variations of the learning rates is evaluated. It shows that when the learning rate is too small, the loss function does not minimize well. The validation and training loss curves hardly appear the windows of the figure, which is caused by the steps of optimization of each iteration being too big. When the learning rate is too big, it takes many more iterations to reach to reach the same value in the loss functions as for a more optimal learning rate around .

In figures 4-6 of appendix B the variations of the batch size are evaluated, and the effect of the size is not very noticeable. The lowest loss for the biggest batch size is expected, as a bigger batch size means there are more values used in the calculations of each iteration, resulting in a better prediction of the parameters for the next iteration. However, it is not expected for the validation loss to look more noisy as the batch size increases, as the prediction for the next iteration should be better. Furthermore, the shorter runtime for the smaller batch time is as expected, but the effect of the runtime is not shown in the figures. This is because of less relevance for this report, as calculation times are not a main point of focus.

In figures 7 and 8 of appendix B the variation of the initial value for theta is evaluated and this effect is very minimal. After about 50 epochs the loss curves start to look very similar.

Figure 9 of appendix B shows the effect of a smaller threshold of the stopping criterium for the training of the model. This model is trained for 955 epochs, resulting in an prediction accuracy of 0.826. With the larger threshold, the model is trained over an average of 390.2 epochs, leading to a final accuracy of 0.818. The large amount of extra epochs needed, results in only a small increase of accuracy.

The result of the logistic regression model with the input of a smaller dataset resulted in a overfitting model, which is not a good model to predict the size of the nuclei on a new dataset. An adequate size of the training dataset is needed for both linear and logistic regression to make accurate predictions.

The accuracy of approximately 81.7% is pretty high, however this means still 18.3% of the aggressiveness of tumors would be over- or underestimated, which indicates that the results should still be checked.

# 5. Reading assignment

The deep neural network used in Graham et al. (Graham et al.) uses a slightly modified implementation of the Preact-ResNet50 residual network (He et al, 2016) followed by three branches, each with its own function. This feature extracting network of 50 layers is then followed by branches that (1.) predicts whether a pixel belongs to the nuclei or background, (2.) predicts the horizontal and vertical gradients of the nuclei, useful for separating clustered nuclei, and (3.) determine the type of each nucleus. Altogether, this is neural network model is much more deep and complex than the models used in this report for the linear and logistic regression.

The optimization of the linear regression model is done by finding the closed-form solution for minimization of the loss function, as opposed to backwards propagation in the neural network, where the minimum is iteratively approached using the training dataset. Still, the same is done for the logistic regression model in this report. Each iteration, the parameters are adjusted until the minimum is reached, after which the minimization is ended when the improvements in the loss function start becoming insignificant. Another difference is that the linear regression model gives a prediction for the size of the nuclei, while the logistic regression network and deep neural network perform a classification in size.

The result of the complexity of the neural network used by Graham et al. is that is generalizes strongly, meaning it can be used well on data it has not been trained on. The result is that the authors claim the network would perform well on additional tissue types, even though the results follow from it being trained on only a single tissue type. This is because the network is shown to segment multiple tissues well. The strong generalization, combined with the ability to effectively segment clustered nuclei while still being computationally quick (11 seconds) means it might prove to be a useful network in the clinic.

# References

Doi, K. (2007). Computer-aided diagnosis in medical imaging: Historical review, current status and future

potential. Computerized Medical Imaging and Graphics, 31(4–5), 198–211.

<https://doi.org/10.1016/j.compmedimag.2007.02.002>

Graham, S., Vu, Q. D., Raza, S. E. A., Azam, A., Tsang, Y. W., Kwak, J. T., & Rajpoot, N. (2019).

Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images.

*Medical Image Analysis*, *58*, 101563. https://doi.org/10.1016/j.media.2019.101563

He, Kaiming, e.a. ‘Mask R-CNN’. arXiv:1703.06870 [cs], januari 2018. arXiv.org, <http://arxiv.org/abs/1703.06870>.

Veta M., van Diest P.J., Pluim J.P.W. (2016). Cutting Out the Middleman: Measuring Nuclear Area in Histopathology Slides Without Segmentation. *Medical Image Computing and Computer-Assisted Intervention;  Lecture Notes in Computer Science, vol 9901*. DOI: https://doi.org/10.1007/978-3-319-46723-8\_73

# Appendix A

Result of the linear regression model; the 300 smallest and 300 largest nuclei

Afbeelding met tekst

Automatisch gegenereerde beschrijving

Figure 1: The 300 smallest and 300 largest nuclei of the full training dataset

# Appendix B

Figures for analysis of hyperparameters

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Calendar  Description automatically generated with medium confidence  Figure 1 Loss curves for | Chart  Description automatically generated  Figure 2 Loss curves for | | | Chart, line chart  Description automatically generated  Figure 3 Loss curves for | |
| Figure 4 Loss curve for a batch size of 128 | | Figure 5 Loss curves for a batch size of 350 | | | Figure 6 Loss curves for a batch size of 572 |
| Chart  Description automatically generated  Figure 7 Loss curves for | | | Chart  Description automatically generated  Figure 8 Loss curves for | | |

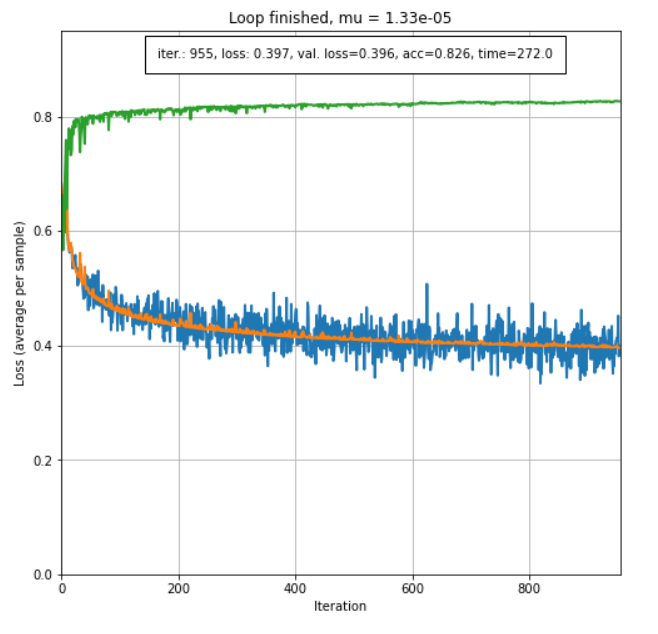


Figure 9: The effect of lower threshold of stopping criteria