8DC00 Medical Image Analysis

Project 2 - CAD

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

CAD, short for computer aided diagnosis, has become a standard part of the detection of for example breast cancer at many screening sites and hospital. The CAD programs gives for example as output to which class the analyzed cells belong to, malignant or benign cells. CAD is developing even more 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 would spare a lot of time for the pathologists.  
The goal of this project is to create a program that automates the cell nuclei measurement and classify these cells to the large nuclei class or the small nuclei class. The automatic cell nuclei measurement will be done with the use of a linear regression model and evaluated with the outcoming error. The classification will be 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 model. This is the model with the highest accuracy and lowest loss.   
Furthermore this report will end with a comparison of these linear- and logistic regression models to deep neural networks that are explained 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 an 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

## 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.

(2.2)

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**

# 3. Results

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

## 2.2 Linear regression

In the following figure 1 the 300 smallest and the 300 largest nuclei of the training dataset are shown.

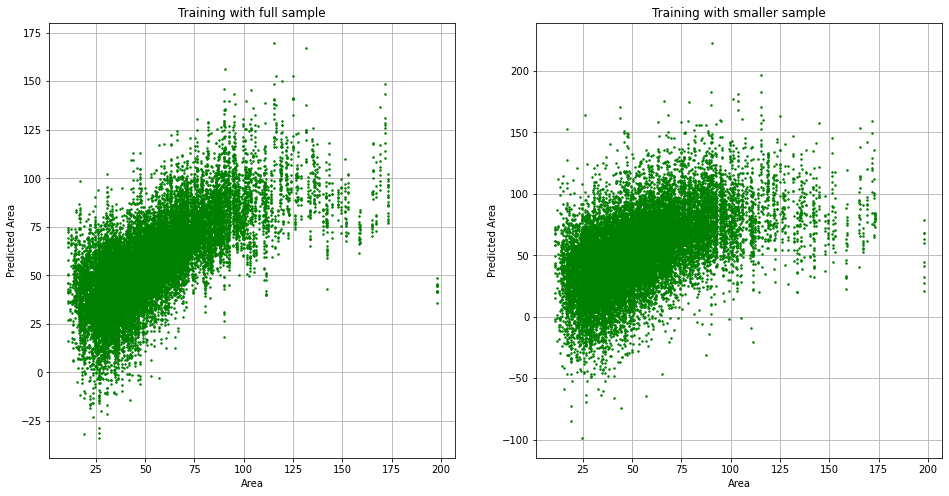
#### Figure 1: The 300 smallest and 300 largest nuclei of the full training dataset

Afbeelding met tekst

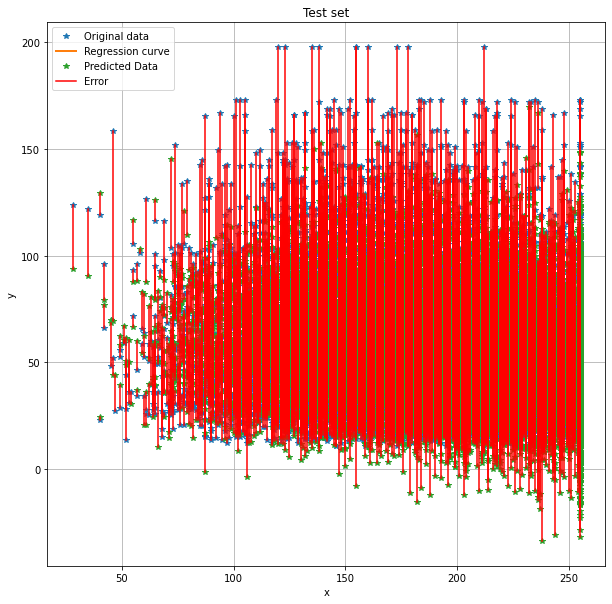
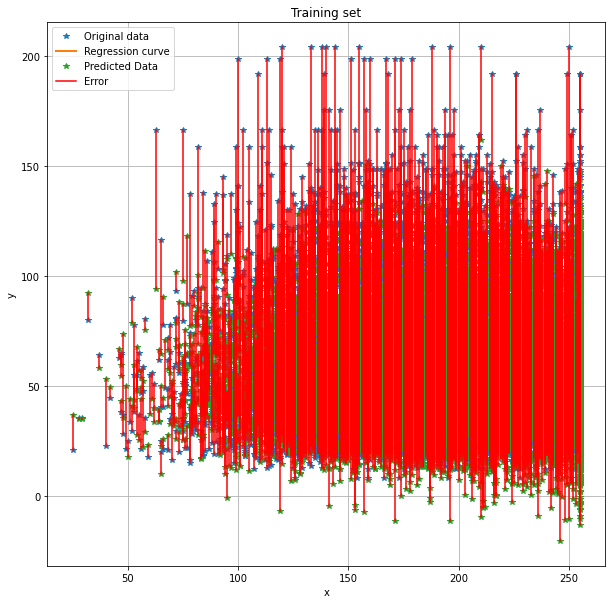
Automatisch gegenereerde beschrijving

Below, figure 2 is shown. The results of the linear regression to predict the area of the nuclei in the test dataset is visualized. 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.

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 with a smaller training dataset is 744.2.

*Figure 2: Predicted area versus actual area of the linear regression for the full and small training dataset  
*

Willen we dit er ook in hebben, zoals in notebook 2.1?



# 4. Discussion

# 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

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