

Lappeenranta University of Technology  
School of Business and Management  
Degree Program in Computer Science

Bachelor's Thesis

**Teemu Huovinen**

**SENSITIVITY OF RETINAL IMAGE SEGMENTATION ON  
GROUND TRUTH ACCURACY**

Examiners: Prof., D.Sc. (Tech.) Lasse Lensu

Supervisor: Prof., D.Sc. (Tech.) Lasse Lensu

# **ABSTRACT**

Lappeenranta University of Technology  
School of Business and Management  
Degree Program in Computer Science

Teemu Huovinen

**Sensitivity of retinal image segmentation on ground truth accuracy**

Bachelor's Thesis

2014

39 pages, 20 figures, 1 table, and 4 appendices.

Examiners: Prof., D.Sc. (Tech.) Lasse Lensu

Keywords: retinal image analysis, image segmentation, sensitivity analysis, ground truth

Diabetes is a growing problem in developed countries, which motivates the research for an automated method for diagnosing and monitoring the disease. One promising approach is detection of diabetic retinopathy in retinal images. The object of this thesis is to study the effects of inaccurate ground truth in exudate segmentation in retinal images, and to compare the performance of different image features and methods. This is done by studying the performance of Naïve-Bayes and Gaussian mixture models (GMM) with Bayesian classification, and also unsupervised edge detection methods with varying degrees of ground truth accuracy. The results confirm the negative effect inaccurate ground truth has on segmentation accuracy. Also the method- and feature-specific performance is explored.

# **TIIVISTELMÄ**

Lappeenrannan teknillinen yliopisto  
LUT Kauppatieteet ja tuotantotalous  
Tietotekniikan koulutusohjelma

Teemu Huovinen

**Silmänpohjakuvien segmentoinnin herkkyyssä asiantuntijatiedon tarkkuudelle**

Kandidaatin työ

2014

39 sivua, 20 kuvaaa, 1 taulukko ja 4 liitettä.

Tarkastajat: Prof., TkT Lasse Lensu

Hakusanat: silmänpohjakuvien analyysi, segmentointi, herkkyyssanalyysi, asiantuntijato

Keywords: retinal image analysis, image segmentation, sensitivity analysis, ground truth

Diabetes on kasvava ongelma kehittyneissä maissa, mikä motivoi automatisoidun sairauuden diagnosointi- ja seurantamenetelmän tutkimista. Yksi lupaava menetelmä on tunnistaa diabeettinen retinopatia silmänpohjakuvista. Tämän tutkielman tavoite on tutkia epätarkan asiantuntijatiedon vaikutuksia eksudaattien segmentointiin silmänpohjakuvissa, sekä vertailla eri piirteiden ja metodien suorituskykyä. Tämä toteutetaan vertailemalla Naïve-Bayesin, Gaussisten sekoitemallien Bayesiläisellä luokittelulla, sekä ohjaamattoimien reunantunnistusmenetelmien suorituskykyä erilaisilla asiantuntijatiedon tarkkuuksilla. Tulokset varmistavat epätarkan asiantuntijatiedon negatiiviset vaikutukset segmentoinnin tarkkuuteen, Myös menetelmä- ja piirrekohtaista suorituskykyä tutkitaan.

## PREFACE

I would like to thank the Machine Vision and Pattern Recognition Laboratory of Lappeenranta University of Technology for giving me this opportunity to carry out my research. I would like to give special thanks to Prof. Lasse Lensu for the endless guidance and support during my summers as an intern.

I would like to thank University of Bristol for their retinal image database and its accurate ground truth data, which was extensively used in this thesis.

Lappeenranta, October 19th, 2014

*Teemu Huovinen*

# CONTENTS

|   |           |
|---|-----------|
| <b>1 INTRODUCTION</b>   | <b>7</b>  |
| 1.1 Background . . . . .  | 7         |
| 1.2 Objectives and Restrictions . . . . .                             | 7         |
| 1.3 Structure of the Thesis . . . . .                                 | 8         |
| <b>2 RETINAL IMAGES AND THEIR SEGMENTATION</b>                        | <b>9</b>  |
| 2.1 Structure of the eye fundus . . . . .                             | 9         |
| 2.2 Optic disc detection . . . . .                                    | 10        |
| 2.3 Blood vessel detection . . . . .                                  | 10        |
| 2.4 Color transform . . . . .   | 12        |
| 2.5 Unsupervised image segmentation methods . . . . .                 | 14        |
| 2.5.1 Kirsch operator . . . . .                                       | 14        |
| 2.5.2 Morphological operations . . . . .                              | 14        |
| 2.6 Supervised methods . . . . .                                      | 15        |
| 2.6.1 Naïve Bayes . . . . .   | 15        |
| 2.6.2 Gaussian Mixture Models . . . . .                               | 16        |
| 2.7 Evaluation metrics . . . . .                                      | 16        |
| <b>3 EXPERIMENTS AND RESULTS</b>                                      | <b>18</b> |
| 3.1 The ground truth . . . . .  | 18        |
| 3.2 Supervised image segmentation methods . . . . .                   | 19        |
| 3.2.1 Test settings . . . . .   | 19        |
| 3.2.2 Results and discussion . . . . .                                | 19        |
| 3.3 Unsupervised methods . . . . .                                    | 20        |
| 3.3.1 Finding the best parameters . . . . .                           | 20        |
| 3.3.2 Results and discussion . . . . .                                | 21        |
| <b>4 DISCUSSION</b>   | <b>25</b> |
| <b>REFERENCES</b>   | <b>25</b> |
| <b>APPENDICES</b>   |           |
| Appendix 1: Example results of unsupervised methods                   |           |
| Appendix 2: Example results of Naïve-Bayes                            |           |
| Appendix 3: Example results of GMM-Bayes                              |           |
| Appendix 4: Supervised methods' performance using gray world pictures |           |

## ABBREVIATIONS

|              |  |
|--------------|--|
| <b>CLAHE</b> | Contrast Limited Adaptive Histogram Equalization |
| <b>GMM</b>   | Gaussian Mixture Model                           |
| <b>pdf</b>   | Probability Density Function                     |
| <b>FJ</b>    | Figuero-Jain                                     |
| <b>EM</b>    | Expectation Maximization                         |
| <b>RGB</b>   | Red-Green-Blue                                   |
| <b>TP</b>    | True Positive                                    |
| <b>TN</b>    | True Negative                                    |
| <b>FP</b>    | False Positive                                   |
| <b>FN</b>    | False Negative                                   |
| <b>FPR</b>   | False Positive Rate                              |
| <b>FNR</b>   | False Negative Rate                              |

# 1 INTRODUCTION

## 1.1 Background

The growing number of diabetes patients and the estimated amount of undiagnosed patients motivate the research for an effective mass screening method for early detection and monitoring of diabetes. The most common complication of diabetes, diabetic retinopathy, causes abnormalities in the eye, and detecting these abnormalities in the eye fundus is a promising mass screening method. [1] When developing a method for detecting these abnormalities, manual segmentation of the objects in the image are used as ground truth to train classifiers and to evaluate the results. In eye fundus image segmentation research, ground truths are usually produced by medical experts of the field (ophthalmologists) marking these abnormalities, such as exudates.

An optimal ground truth would be a pixel-accurate binary representation of the abnormalities, but as ground truths are produced manually, such accuracy is not possible. Because the marking of an accurate ground truth takes a good amount of time and patience, it is often necessary to use inaccurate markings of the abnormalities. Clusters of exudates are marked, rather than each small finding specified separately.

## 1.2 Objectives and Restrictions

This thesis will address two main questions in its research:

- How will inaccurate ground truth affect image features and segmentation methods?
- How will other features than color perform?

This thesis focuses on exudate detection in terms of segmentation performance, and Bristol database is used as it is readily available and includes accurate ground truth information. To enable comparison between image features, color, edge and texture features are used. Blood vessel detection is explored only to create a mask for them. A rough method for optic disk detection is also implemented as a preprocessing step for masking reasons. Structure of the eye fundus is briefly explored.

Both supervised and unsupervised segmentation methods are used. In supervised methods, the ground truth is used to evaluate the results and to label pixels on the training set as either exudate or background.

In unsupervised methods, the ground truth is used to evaluate the performance of the methods on the training set with varying values of method parameters. From these results, the best set of parameters are then selected for each method based on their performance. The method performance is then compared to supervised methods only with these selected parameters.

### **1.3 Structure of the Thesis**

Section 2 takes a look at the different features of eye fundus images, and how they are relevant in this thesis. It also explains the theory behind the applied pre-processing and segmentation methods. Section 3 describes the experiments in detail, and presents the results for each experiment. Section 4 sums up and interprets the results, and discusses the impact of this thesis and possible future research.

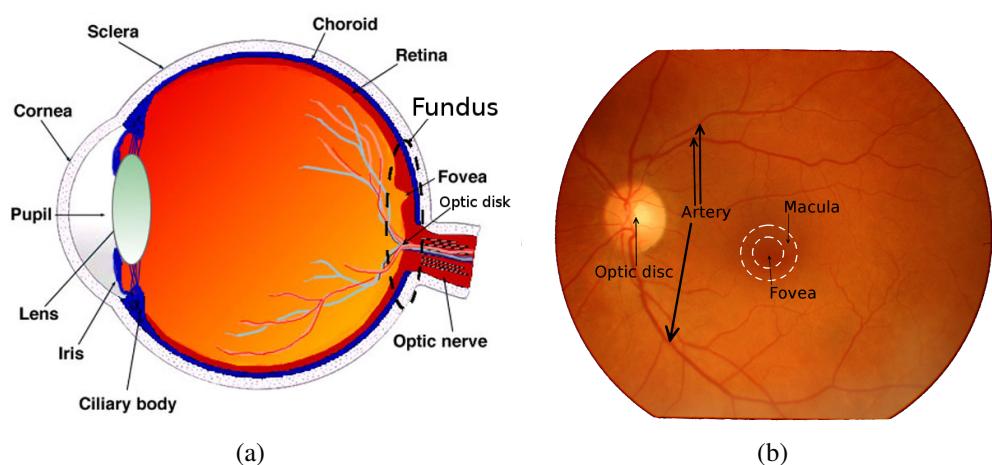
## 2 RETINAL IMAGES AND THEIR SEGMENTATION

### 2.1 Structure of the eye fundus

The function of the human eye on an abstract level can be compared to a camera. Light reflected from an object passes through the cornea, pupil and lens, and is then focused on the retina (the inner part of the eye). The light is absorbed by photoreceptors and the response is then transmitted to the brain via the optic nerve. [2] Rather than covering the eye as a whole, this chapter will focus on the structure of the eye fundus.

The eye fundus is the interior lining of the eyeball, located at the very back of the eye. The location of the eye fundus is illustrated in Figure 1a. Its most noticeable parts are the optic disk, the macula, and the veins. Macula is a highly light-sensitive area in the central region of the retina. Fovea is a round area located in the macula, which is densely populated by cones (cells sensitive to color). The orange background seen in the eye fundus images is formed by the pigment epithelium, which is a barrier between the choroid (vascular layer surrounding the retina) and the photoreceptors. [3]

Optic disc is where the optic nerve and main arteries connect with the eye. There are no light-sensitive cells inside the optic disc, which creates a blind spot in the retina. Veins inside the eye together with the choroid provide nutritional support to the eye. [1] These parts of the eye fundus are illustrated in Figure 1b.



**Figure 1.** Structure of the human eye: (a) cross section of the eye (modified from [3]), (b) structure of the fundus.

## 2.2 Optic disc detection

Optic disc is very similar to exudates in terms of color and intensity, so detection and masking of the optic disk is an important preprocessing step in exudate detection. There are papers dedicated to the localisation of the optic disk [4], and it is also covered in papers concerning the detection of other parts of the eye fundus, such as exudates [5].

The method used in this thesis is based on the brightness of the optic disk, and the vertical blood vessels inside it. The horizontal image gradient is calculated using Sobel operator [6], the result is shown in Figure 2a. The image is then divided into slightly overlapping square areas with a side of 140 pixels. The area with the highest sum of gradients is considered as region of interest, i.e. to hold the optic disk. This is because the dark blood vessels inside the bright optic disk result in a strong horizontal gradient. Images with a “camera glare”, i.e. a high intensity strip in the corner of the eye fundus are problematic, as that area also has a high horizontal gradient. An example region of interest is shown in Figure 2b.

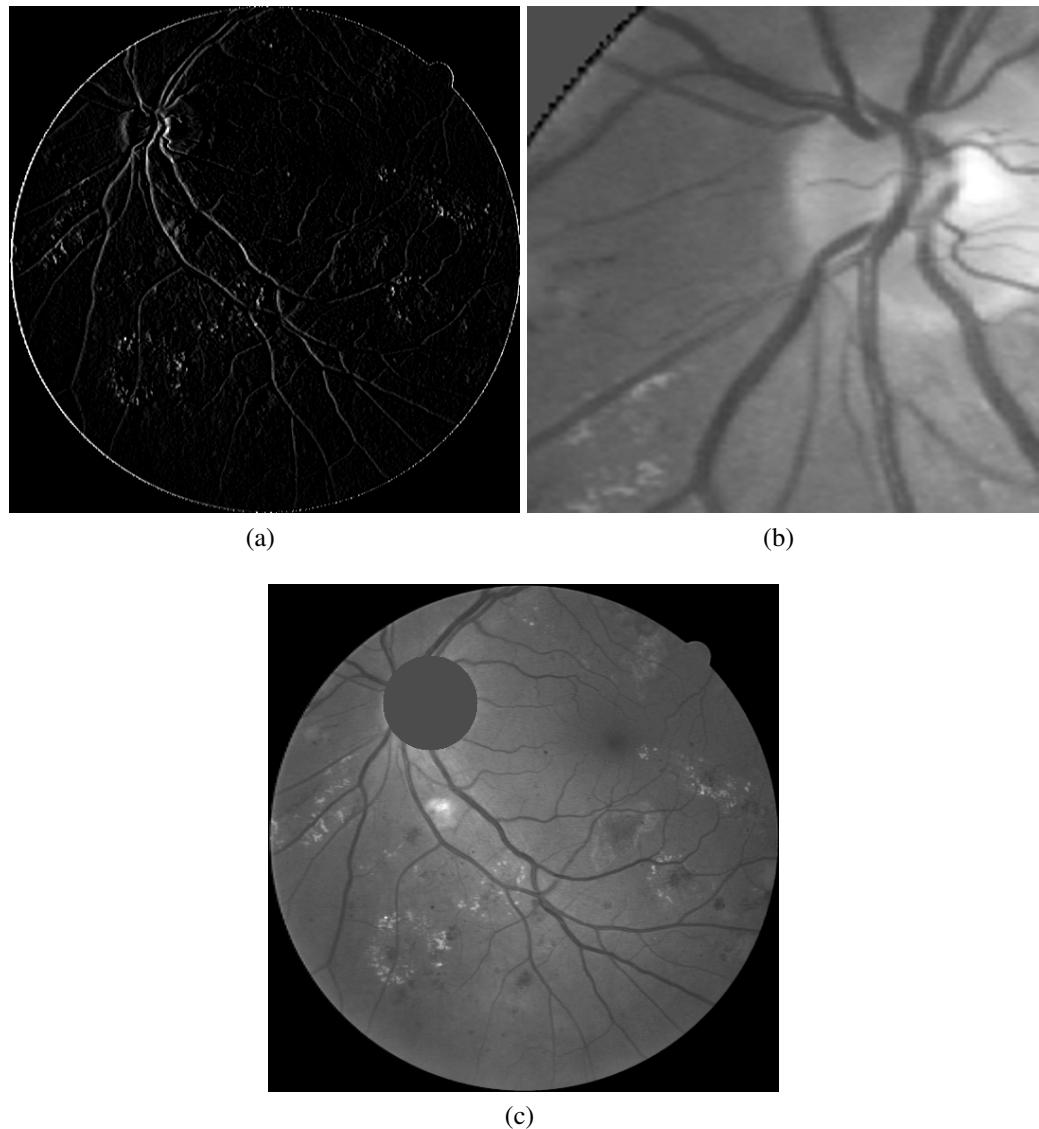
Inside the area of highest sum of horizontal gradients, the pixel with the highest intensity is considered to be inside the optic disk. This pixel is then used as the center of a circle that will mask out the optic disk. The final masking result is shown in Figure 2c.

## 2.3 Blood vessel detection

The issue of blood vessel detection in fundus images has been a popular topic of research [7–9].

In the context of this thesis, however, the purpose of blood vessel detection is to create a mask, and to use that mask to remove false positives from exudate segmentation results. For example, edge detection techniques often highlight the borders of vessels as well as exudates. For this purpose, accurate vessel detection itself is not important and including other dark areas of the image is even beneficial.

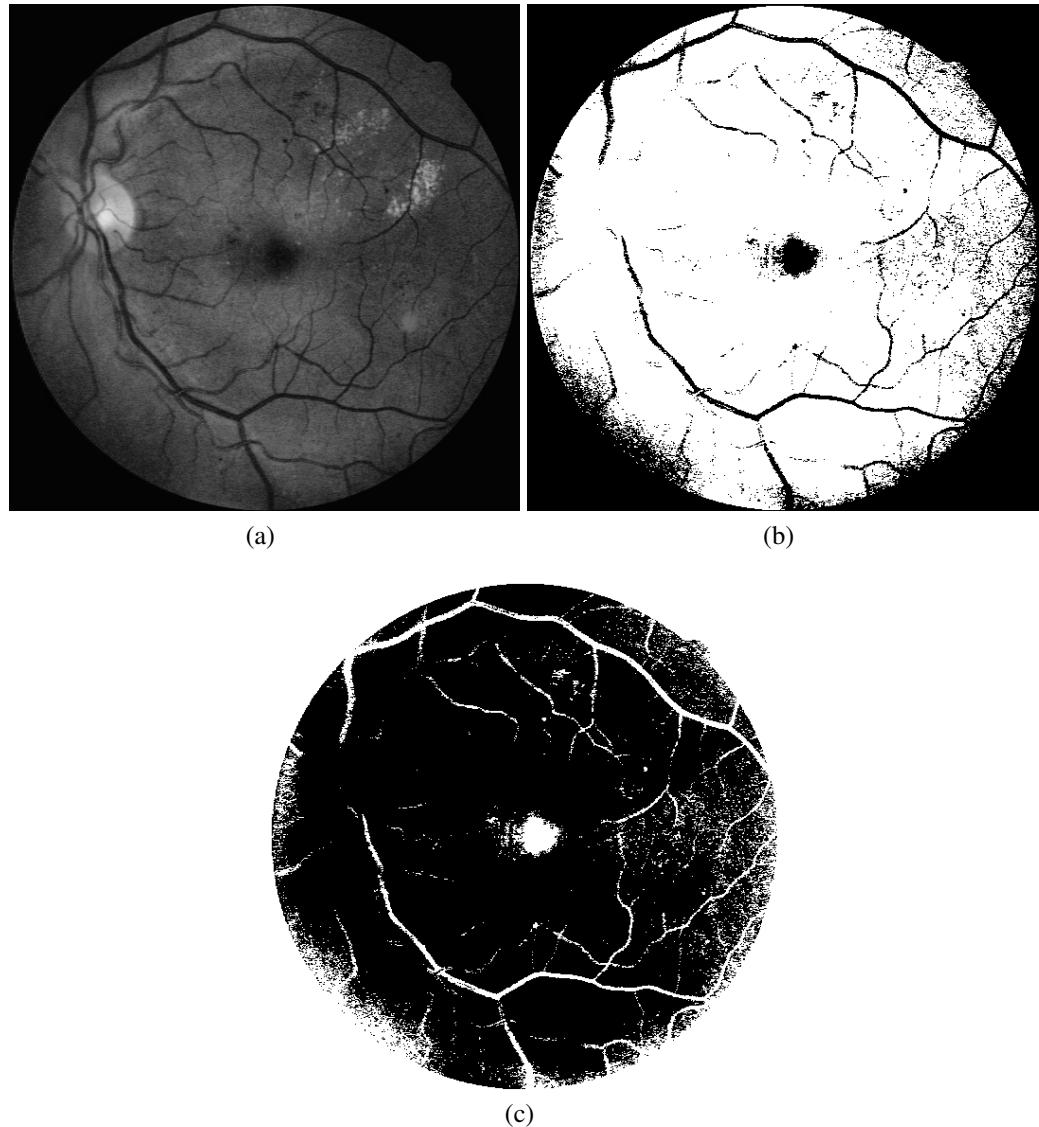
The mask is formed by first using contrast-limited adaptive histogram equalization (CLAHE) [10] to enhance contrast in the green channel of the image, the result for this is shown in Figure 3a. This contrast enhanced image is then thresholded with Otsu’s method [11], which separates the image into foreground and background by minimizing the intra-class variance. This results in all the vessels and other darker areas classified as



**Figure 2.** Locating and masking the optic disk: (a) horizontal gradient, (b) region of interest, (c) optic disk masked out.

background, and all brighter areas classified as foreground. This is shown in Figure 3b. To create a binary mask of the darker areas, we use the complement of this thresholded image. The final version of the mask is shown in Figure 3c.

This method is inadequate for blood vessel detection as it also includes other darker areas of the image, such as the fovea. As a mask, however, it clearly reduces the amount of false positives in exudate segmentation results. It also does not remove true positives, as only the darker areas of the image are included in the mask.



**Figure 3.** Creating the blood vessel mask: (a) CLAHE, (b) thresholding using Otsu’s method, (c) the final mask, complement of thresholded image.

## 2.4 Color transform

The human eye is capable of correcting the effect of varying light sources (illuminants) while perceiving color [12]. In contrast, an image of an object taken under different light sources is perceived differently in terms of color by a computer. This promotes the normalization of color features in image sets where there’s high variation in color properties. In this thesis, the use of color normalization is motivated by the variation of the color of the pigment epithelium layer in the eye fundus.

The illumination and general color properties of the eye fundus images vary quite heavily.

To be able to effectively use color features in teaching a classifier, the variance between images needs to be minimized. This is achieved by estimating the illuminant by applying the gray-world assumption [13]: *the average reflectance in a scene under a neutral light source is achromatic*. While this does not hold true in the case of eye fundus images, the application had positive results in terms of color feature normalization as seen in Figure 4. Normalization was done by multiplying each channel with a coefficient defined as  $K_r = I_{avg}/r_{avg}$ , where  $I_{avg}$  is the mean of all RGB-values in the image, and  $r_{avg}$  is the mean of the specific channel.



**Figure 4.** Color space transform: Original images on the left, adjusted images on the right.

## 2.5 Unsupervised image segmentation methods

### 2.5.1 Kirsch operator

Edge detection operators that only detect gradients in specified directions are called compass operators. Kirsch operator applies a compass operator in eight directions by rotating a mask in  $45^\circ$  shifts. [14] It is defined as:

$$K_1 = \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}, K_2 = \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}, \dots, K_8 = \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}.$$

### 2.5.2 Morphological operations

Mathematical morphology operations use a structuring element to perform an operation on an input image. Most morphology operations are based the two basic morphological operations, dilation and erosion. Top-hat transform is an operation that highlights bright areas in an image, where as bottom-hat (also known as black top-hat) operation highlights the dim areas of an image. [15] These transforms are defined as follows:

$$\text{TopHat}(f, B) = f - (f \circ B) \quad (1)$$

$$\text{BottomHat}(f, B) = (f \bullet B) - f \quad (2)$$

where  $f$  is the input image,  $B$  is the structuring element,  $\circ$  denotes the opening operation and  $\bullet$  denotes the closing operation. Opening and closing are standard operations based on dilation and erosion, and they are defined in [15].

The exudate detection method proposed in [16] was also implemented, except for the optic disc detection, for which the method described in Section 2.2 was used. The method is designed to use the top-hat operation to highlight the exudates, and the bottom-hat operation to highlight and ultimately remove dim areas, e.g. vessels from the results. The method is defined as follows:

$$F(f) = TopHat(f) - BottomHat(f) \quad (3)$$

where  $Tophat(f)$  and  $BottomHat(f)$  denote Equations 1 and 2, respectively. Unlike in [16], thresholding is applied after these operations to enable evaluation by creating a binary representation of the results.

## 2.6 Supervised methods

Color, edge and texture features were used to teach the classifiers of the supervised methods. Color was extracted from the image as pixel-specific RGB-values. Edge was represented with the differential of the maximum and minimum value in a 3-by-3 neighborhood (implemented with *rangefilt*-function in MATLAB [17]). Location-invariant local binary patterns (LBP) [18] were used for texture.

### 2.6.1 Naïve Bayes

Naïve Bayes is a probabilistic classifier that gets its name from the assumption that all features are independent from each other. Each feature contributes to the classification regardless of the absence or presence of other features. As the name also implies, the Naïve Bayes classifier uses the Bayesian theorem as the basis of classification. [19]

The Naïve Bayes classifier calculates the posterior probability for the sample to belong in each of the possible classes based on its features, and then classifies the sample to the class with the highest posterior probability. The classes are modeled with a single probability density function, and in the case of binary pixel classification, the classes are foreground and background. In this thesis, exudates represent the foreground and the rest is considered the background. The posterior probability is calculated with the Bayesian theorem as follows:

$$p(C_1) = \frac{P(C_1)p(\bar{x}|C_1)}{P(C_1)p(\bar{x}|C_1) + P(C_2)p(\bar{x}|C_2)} \quad (4)$$

where  $p(C_1)$  is the posterior probability and  $P(C_1)$  is the prior probability for the sample to belong in class one, and  $p(\bar{x}|C_1)$  denotes the conditional probability of the feature vector having the values of the sample pixel, if it were of class one.

## 2.6.2 Gaussian Mixture Models

A Gaussian mixture model (GMM) is a single probability density function (pdf) that describes a weighted sum of multiple Gaussian pdfs [20]. When using a GMM to model image classes (more specifically, class-conditional pdfs), parameters of the underlying Gaussian pdfs are estimated by fitting the model to training data. After the formation of the model, it is possible to use Bayesian classification to classify pixels to the given classes.

Application, parameter estimation and performance of a classifier using GMMs and Bayesian classification is discussed in [21]. GMMBayes Toolbox [22] is a MATLAB implementation of these methods. For parameter estimation, Expectation Maximization (EM), Figuero-Jain (FJ) and greedy EM algorithms are implemented in the toolbox and described in both [21] and [22]. The main difference between these algorithms is that EM requires a fixed amount of Gaussian components, while FJ estimates the amount (maximum amount is fixed in the implementation). In this thesis, only FJ is used.

## 2.7 Evaluation metrics

In evaluation, correctly classified samples can be split into true positives and true negatives. Errors are then similarly split into false positives and false negatives. In this thesis, true positives are findings correctly classified as exudates in the segmentation results, and true negatives are consequently findings correctly classified as the background. False negatives are classified as background in the results, but are actually positive (classified as exudate) in the ground truth, and false positives are classified as positive in the results, when they are in fact background [23]. These terms are illustrated with an error matrix shown in Table 1.

Sensitivity represents the proportion of samples correctly classified as positive. Its values range from 0 to 1, where 1 means all foreground samples were correctly classified. Specificity in turn represents the proportion of background pixels correctly described as negative. Its values also range from 0 to 1, where 1 means all the background pixels were

**Table 1.** General error matrix

|       |          | Ground truth        |                     |
|-------|----------|---------------------|---------------------|
|       |          | Positive            | Negative            |
| Tests | Positive | True Positive (TP)  | False Positive (FP) |
|       | Negative | False Negative (FN) | True Negative (TN)  |

correctly classified. [23] They are defined as:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (6)$$

The Dice coefficient [24], Jaccard index and F-score all measure set agreement, representing the success of segmentation with a single number. All of these coefficients can have values from the range [0,1], where 0 indicates no similarities, and 1 indicates perfect agreement. The coefficients are defined as follows:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} \quad (7)$$

$$\text{Jaccard} = \frac{|A \cap B|}{|A \cup B|} \quad (8)$$

$$\text{F-score} = \frac{2PR}{P + R} \quad (9)$$

where A is the segmented set, and B is the ground truth, P stands for precision and R stands for recall (sensitivity).

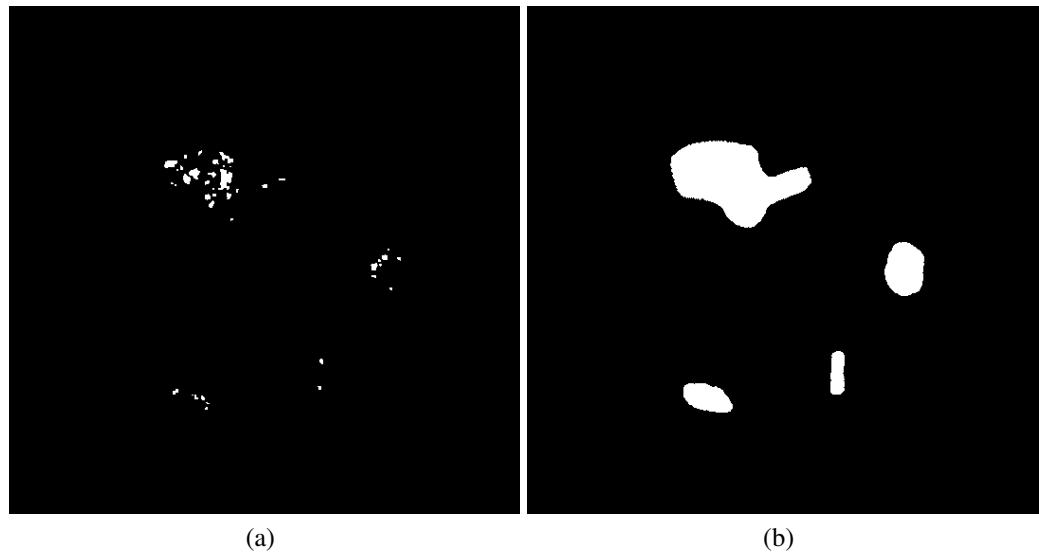
## 3 EXPERIMENTS AND RESULTS

### 3.1 The ground truth

To enable the comparison of results and sensitivity analysis, more inaccurate ground truths were made by hand, using the Bristol ground truth as a basis. Instead of the original color images, markings for the inaccurate ground truth were made on the black-and-white images of the Bristol ground truth. This was to ensure every exudate present in the Bristol ground truth was to be also present in the inaccurate ground truth. Markings were done by grouping clusters of exudates together, and marking single exudates more loosely. This is illustrated in Figure 5.

To create a basis for more comprehensive testing, different stages of inaccurate ground truths were created by dilating the accurate ground truth. Three different stages were created; one, three and five iterations of dilation by a disk-shaped structuring element, with a radius of 1 pixel. The dilation was restricted to stay inside the inaccurate ground truth.

The accuracy of the ground truth has a direct impact on training and performance of supervised methods. In general, the inaccurate ground truth will mean a significant amount of background samples close to exudates will be categorized as positive. This will result in higher variation of feature distributions and an increase of false positives.



**Figure 5.** Ground truths: (a) Original accurate ground truth (b) The most inaccurate ground truth

## 3.2 Supervised image segmentation methods

### 3.2.1 Test settings

Preprocessing was kept to a minimum as only the timestamp was removed from the eye fundus images. As the optic disk's color is very close to the exudates, a mask was generated with the method described in Sec. 2.2. The eye outline was also included in the mask, as it was included in the results by edge detection methods. Vessels were not included in the mask.

The amount of background pixels in the images is considerably higher than the amount of exudate pixels. This was balanced by applying sampling to the training data, using equal amounts of background and exudate pixels. All exudate pixels were used, and background pixels were selected randomly.

Classifiers were trained with color, edge and texture features, and with the combination of the three. The feature extraction methods are described in Sec. 2.6. No contrast enhancement was used, so raw color information from the image was used as the color feature.

The Bristol image set was divided into two parts, using one half for training and the other half for evaluation, and then vice versa. This was done to study the effect the selection of training data has on the results. Classifiers were trained with each level of inaccurate ground truth described in Sec. 3.1. Two different image sets were used; the original Bristol images, and the color-normalized images. The colors were adjusted with the method described in Sec. 2.4, with the objective of reducing the color variation in the images. With this setup, each calculation used a specific ground truth accuracy, image set, and teaching set.

### 3.2.2 Results and discussion

Using inaccurate ground truth was expected to result in more false positives in image segmentation results. This can be seen in the results as an increase in sensitivity and a decrease in specificity with all features and classifiers when the level of inaccuracy in the ground truth increases. This is also visible in the image-specific results shown in Appendices 2 and 3. All image segmentation methods and classifiers behave similarly in this regard, having gradually more and more false positives in their segmentation results as the ground truth is made more inaccurate.

Comparing the sensitivity and specificity of Naïve-Bayes and GMM-Bayes classifiers shown in Figures 6 and 7, a slightly higher performance of color feature GMM-Bayes classifier can be seen. The Naïve-Bayes edge-classifier on the other hand, is outperforming other single-feature trained classifier and even the classifier with all the features included. It is worth noting that the linear progression of Naïve-Bayes' sensitivity and specificity as the ground truth becomes more inaccurate. Based on the specificity of the texture-based classifiers, their performance was below the others. The image-specific results explain this, as the texture-based classifier behave similarly to a random classifier. This could be because of the feature's variation in the training set, but this was not confirmed. It is worth noting that images without exudates are excluded from these results, as they have no other effect than decreasing the average performance. It is worth noting that it is necessary to also include the exudate-free images when developing a practical application.

Color normalization's impact on the results was negligible. It was expected to improve the performance of a classifier using the color feature by reducing the color variance of the exudates and background. However, results with Naïve-Bayes are almost identical between the image sets, and no clear improvement can be seen from the results of GMM-Bayes.

In the Figures 6 and 7, selection of training data is represented via the dotted or filled line. Generally, the selection of training data had a minor impact on the results. Performance of both methods' classifiers using edge feature, however, changed drastically when using the most inaccurate ground truth. This can be seen as a drop in specificity. Selection of training data also had a noticeable effect on the performance of GMM-Bayes classifier using all features. For others, the selection of training data did not appear to have a meaningful effect.

### 3.3 Unsupervised methods

#### 3.3.1 Finding the best parameters

When researching feature sensitivity to the ground truth accuracy, standard segmentation with unsupervised methods is not really useful as it does not use the ground truth. In this thesis, the ground truth is used in finding the best parameters for the unsupervised methods. First, the images used for teaching are segmented with each method and a large set of parameters. The results are then evaluated with the ground truth.

For this purpose, it was desired to have a single value describing the goodness of segmentation results to enable easier comparison and ranking of the results. Different values, such as sensitivity, specificity and precision (defined in Sec. 2.7) were considered, but these were discarded as they only described one aspect of the result, either the amount of samples correctly positive or negative. The used coefficient would have to describe the goodness in a more wholesome way, and for that reason, the Dice coefficient, Jaccard index and F-score were considered (defined in Sec. 2.7).

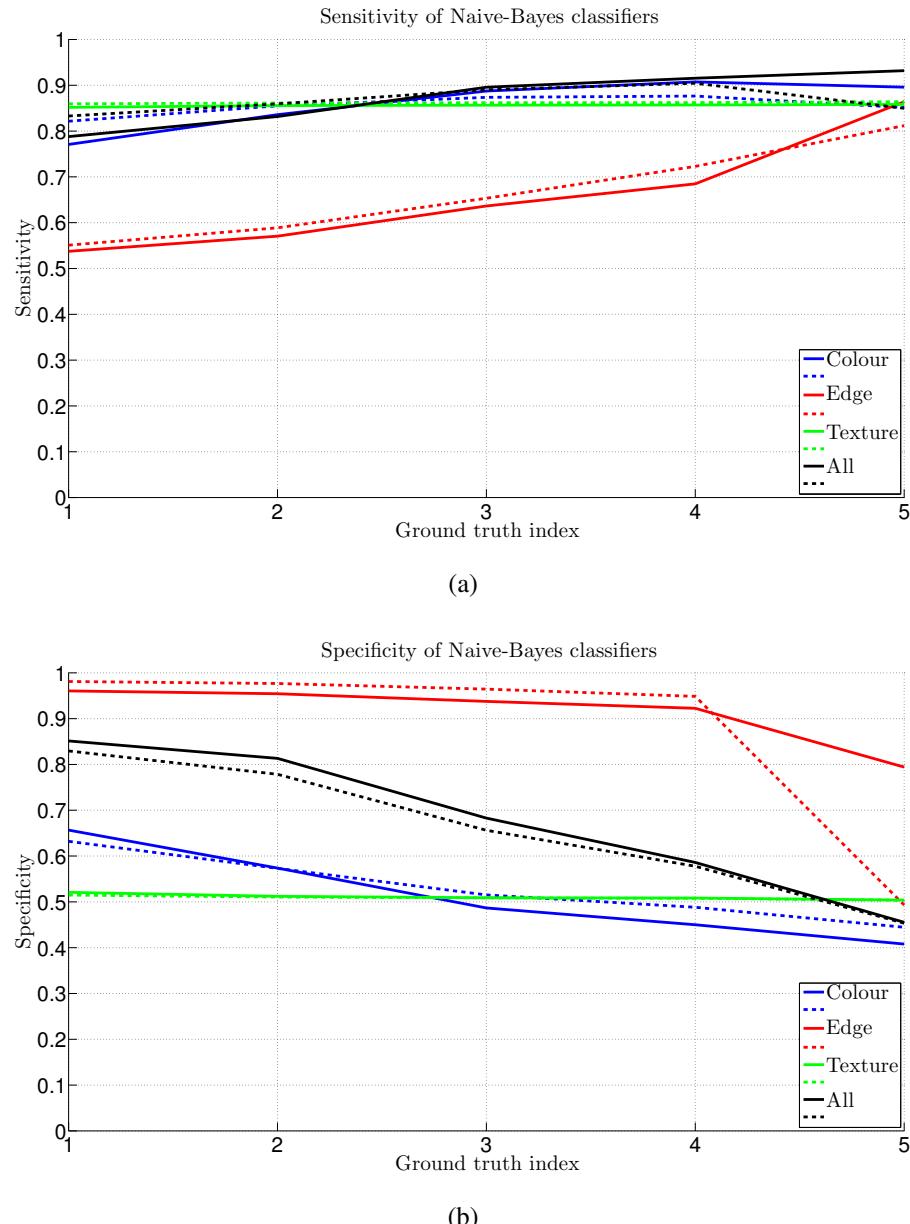
Dice coefficient and F-score values are interchangeable in practice as their values are exactly equal. Jaccard index behaves similarly to the two, though being considerably more critical. The use of Jaccard was discarded because of its criticality, and the choice between Dice and F-score was arbitrary. Dice coefficient was chosen as it is simple to implement.

### 3.3.2 Results and discussion

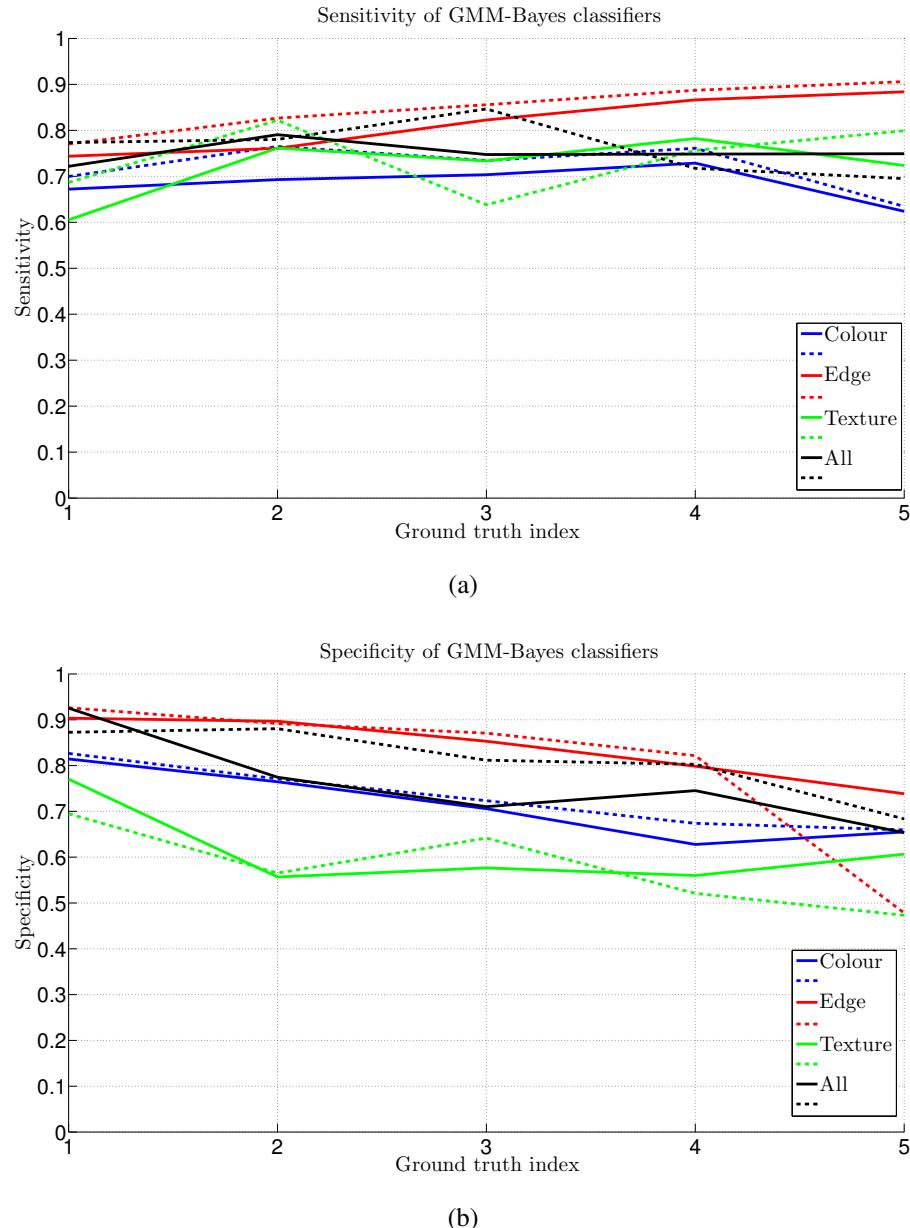
Kirsch and Tophat (described in Sec. 2.5) and the method described in [16], were evaluated. Eadgahi's methods performance is not present in this thesis as it was noticed to be very close to the performance of Tophat.

The sensitivity of the unsupervised methods behaves similarly with the supervised methods, as can be seen in Figure 8a. More exudates are found as the ground truth inaccuracy increases. However, their specificity is significantly more resistant to the inaccurate ground truth, maintaining a high value throughout the different levels of inaccurate ground truth. This can be seen in Figure 8b.

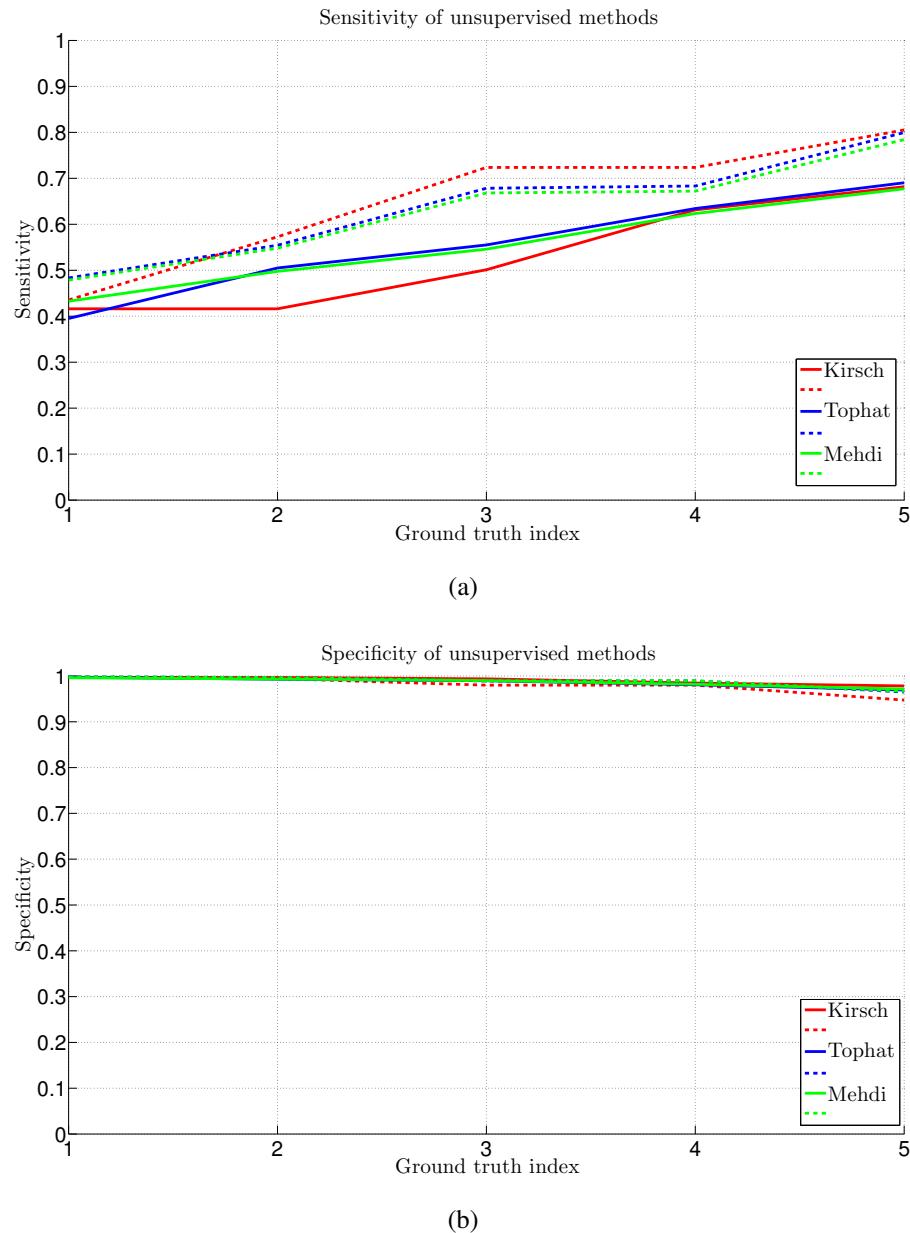
The used mask has a major impact on these results. It was expected that the edge detection methods highlight the blood vessels in the image, so the method described in Sec. 2.3 was used to create a mask of the main arteries in the image. Without this mask, the amount of false positives in the image segmentation results would drastically increase with the inaccurate ground truths.



**Figure 6.** Performance of Naïve-Bayes classifiers: (a) Sensitivity (b) Specificity. Note that these graphs represent the methods' performance over images that have exudates present. The original Bristol ground truth is on the left on the x-axis, and the most inaccurate ground truth is on the right. Solid and dotted lines represent the use of different training/testing sets, as described in Section 3.2.1.



**Figure 7.** Performance of GMM-Bayes classifiers: (a) Sensitivity (b) Specificity. Note that these graphs represent the methods' performance over images that have exudates present. The original Bristol ground truth is on the left on the x-axis, and the most inaccurate ground truth is on the right. Solid and dotted lines represent the use of different training/testing sets, as described in Section 3.2.1.



**Figure 8.** Performance of unsupervised methods: (a) Sensitivity (b) Specificity. Note that these graphs represent the methods' performance over images that have exudates present. The original Bristol ground truth is on the left on the x-axis, and the most inaccurate ground truth is on the right. Solid and dotted lines represent the use of different training/testing sets, as described in Section 3.2.1.

## 4 DISCUSSION

The main goal of this thesis was to study if and how much the spatial accuracy of the ground truth affects the segmentation of retinal images using different image features and segmentation methods. Two supervised methods were compared, Naïve-Bayes and GMM-Bayes. Classifiers were trained with color, edge and texture features. The performance of unsupervised edge detection methods was also evaluated.

The segmentation results of all used methods confirm the fact that accuracy of the used ground truth has a significant impact on the segmentation performance. As for the features, classifiers using the edge feature outperformed color and texture based classifiers. Even though texture seemed promising in initial testing, texture-based classifiers were unable to even partially find exudates from the images.

GMM-Bayes was expected to outperform the simpler Naïve-Bayes, but there was no clear difference in overall performance of these methods. However, Naïve-Bayes' performance with higher levels of inaccuracy in the ground truth was much more predictable, as the segmentation accuracy declined almost linearly as the ground truth became more inaccurate. The performance of GMM-Bayes classifiers, mainly the ones using texture, was less linear. The performance of unsupervised edge detection methods was significantly better than the performance of Naïve- and GMM-Bayes, largely due to the mask used to exclude the blood vessels.

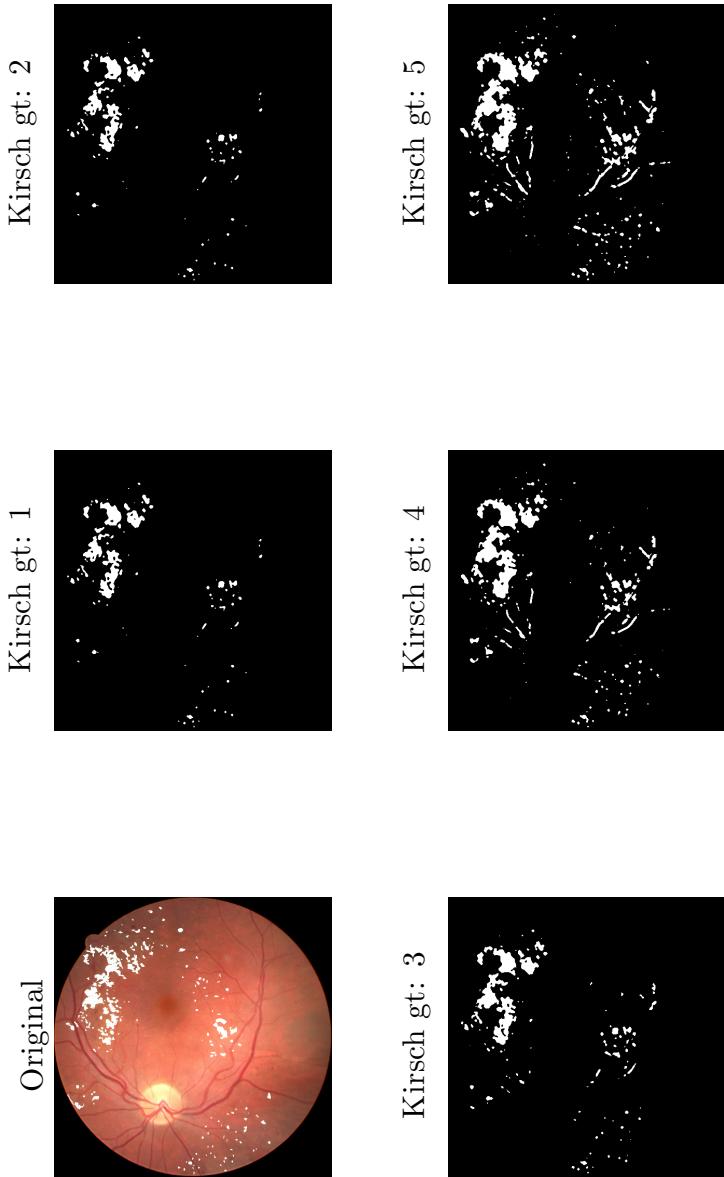
The second goal of the thesis was to evaluate the performance of features other than color. Based on the performance of the unsupervised methods and edge-based classifiers of supervised methods, using the edge feature in exudate segmentation proves to be a valid method. Future work is required in studying the usefulness of texture in this matter.

## REFERENCES

- [1] Tomi Kauppi. *Eye fundus image analysis for automatic detection of diabetic retinopathy*. PhD thesis, 2010.
- [2] Brian A Wandell. *Foundations of vision*. Sinauer Associates, 1995.
- [3] Helga Kolb. Simple Anatomy of the Retina. <http://webvision.med.utah.edu/book/part-i-foundations/simple-anatomy-of-the-retina/>. Accessed: 2014-09-30.
- [4] Sribalamurugan Sekhar, Waleed Al-Nuaimy, and Asoke K Nandi. Automated localisation of retinal optic disk using hough transform. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 1577–1580. IEEE, 2008.
- [5] Thomas Walter, J-C Klein, Pascale Massin, and Ali Erginay. A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina. *Medical Imaging, IEEE Transactions on*, 21(10):1236–1243, 2002.
- [6] Cristian Perra, Francesco Massidda, and Daniele D Giusto. Image blockiness evaluation based on sobel operator. In *ICIP (1)*, pages 389–392, 2005.
- [7] Frédéric Zana and Jean-Claude Klein. A multimodal registration algorithm of eye fundus images using vessels detection and hough transform. *Medical Imaging, IEEE Transactions on*, 18(5):419–428, 1999.
- [8] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram van Ginneken. Ridge-based vessel segmentation in color images of the retina. *Medical Imaging, IEEE Transactions on*, 23(4):501–509, 2004.
- [9] Meindert Niemeijer, Joes Staal, Bram van Ginneken, Marco Loog, and Michael D Abramoff. Comparative study of retinal vessel segmentation methods on a new publicly available database. In *Medical Imaging 2004*, pages 648–656. International Society for Optics and Photonics, 2004.
- [10] Ali M Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, 38(1):35–44, 2004.
- [11] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

- [12] David H Foster. Color constancy. *Vision research*, 51(7):674–700, 2011.
- [13] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980.
- [14] Anil K Jain. *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.
- [15] Xiangzhi Bai, Fugen Zhou, and Bindang Xue. Image enhancement using multi scale image features extracted by top-hat transform. *Optics & Laser Technology*, 44(2):328–336, 2012.
- [16] Mehdi Ghafourian Fakhar Eadgahi and Hamidreza Pourreza. Localization of hard exudates in retinal fundus image by mathematical morphology operations. In *Computer and Knowledge Engineering (ICCKE), 2012 2nd International eConference on*, pages 185–189. IEEE, 2012.
- [17] MATLAB rangefilt-function. <http://se.mathworks.com/help/images/ref/rangefilt.html>. Accessed: 2014-11-18.
- [18] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Gray scale and rotation invariant texture classification with local binary patterns. In *Computer Vision-ECCV 2000*, pages 404–420. Springer, 2000.
- [19] Kevin P Murphy. Naive bayes classifiers. *University of British Columbia*, 2006.
- [20] Douglas Reynolds. Gaussian mixture models. *Encyclopedia of Biometrics*, pages 659–663, 2009.
- [21] Pekka Paalanen, Joni-Kristian Kämäräinen, Jarmo Itonen, and Heikki Kälviäinen. Feature representation and discrimination based on Gaussian mixture model probability densities: practices and algorithms. *Pattern Recognition*, 39(7):1346–1358, 2006.
- [22] Pekka Paalanen. Bayesian classification using Gaussian mixture model and EM estimation: implementations and comparisons. *Information Technology Project*, 2004.
- [23] Tom Fawcett. An introduction to ROC analysis. *Pattern recognition letters*, 27(8):861–874, 2006.
- [24] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

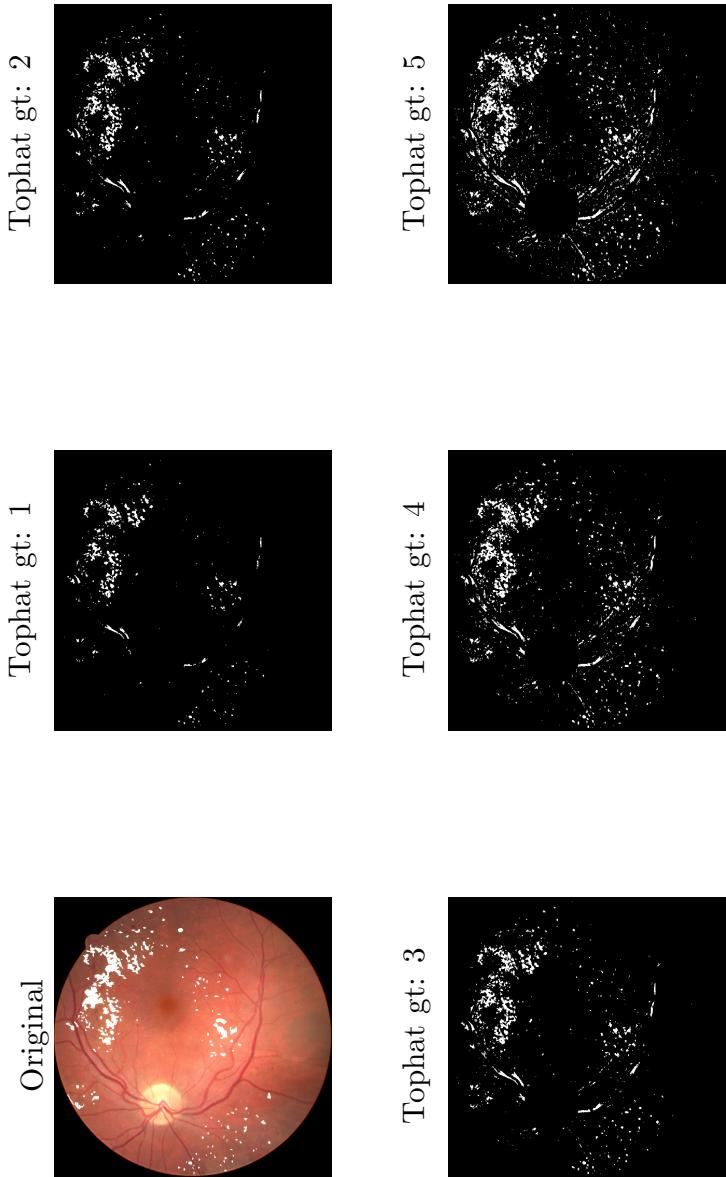
## Appendix 1. Example results of unsupervised methods



**Figure A1.1.** Example results of Kirsch. The original ground truth is shown superimposed on the image of the eye fundus. The segmentation results of the method are shown as a binary image. The results with the accurate ground truth are shown first (1), then the results with the three iterations of inaccurate ground truth (2-4), and lastly the results with the most inaccurate ground truth (5).

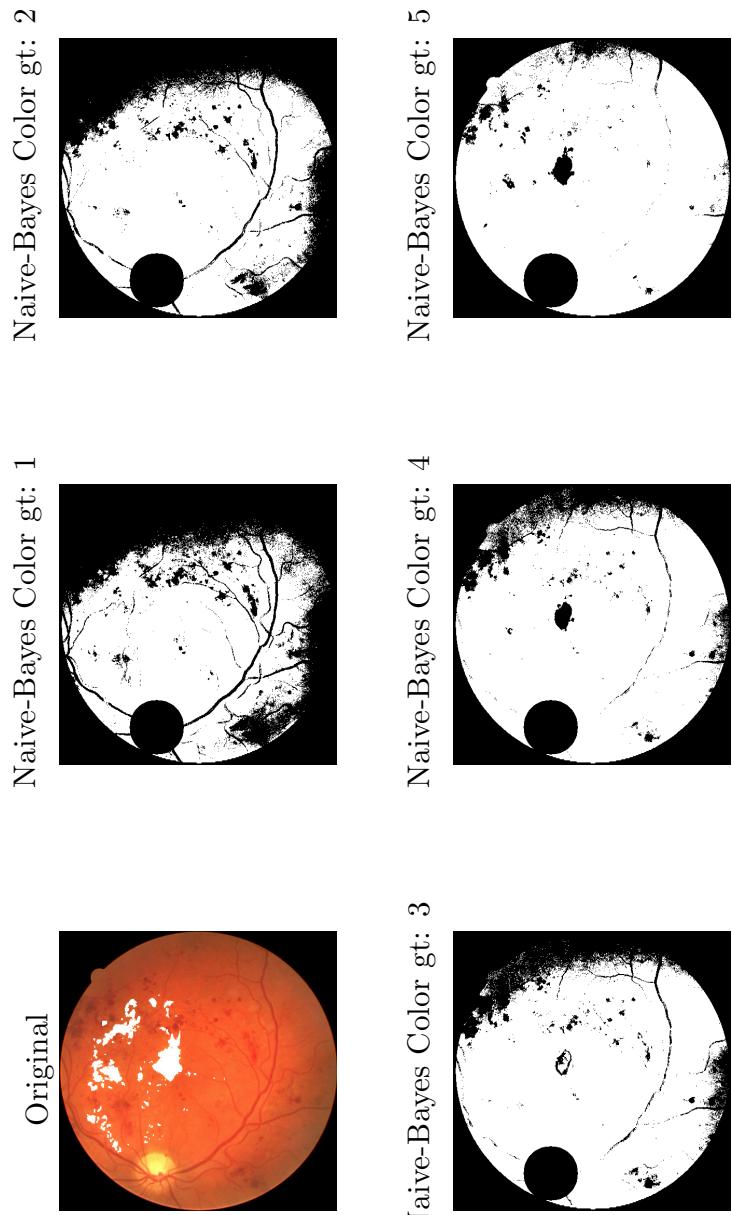
(continues)

## Appendix 1. (continued)



**Figure A1.2.** Example results of Tophat. The original ground truth is shown superimposed on the image of the eye fundus. The segmentation results of the method are shown as a binary image. The results with the accurate ground truth are shown first (1), then the results with the three iterations of inaccurate ground truth (2-4), and lastly the results with the most inaccurate ground truth (5).

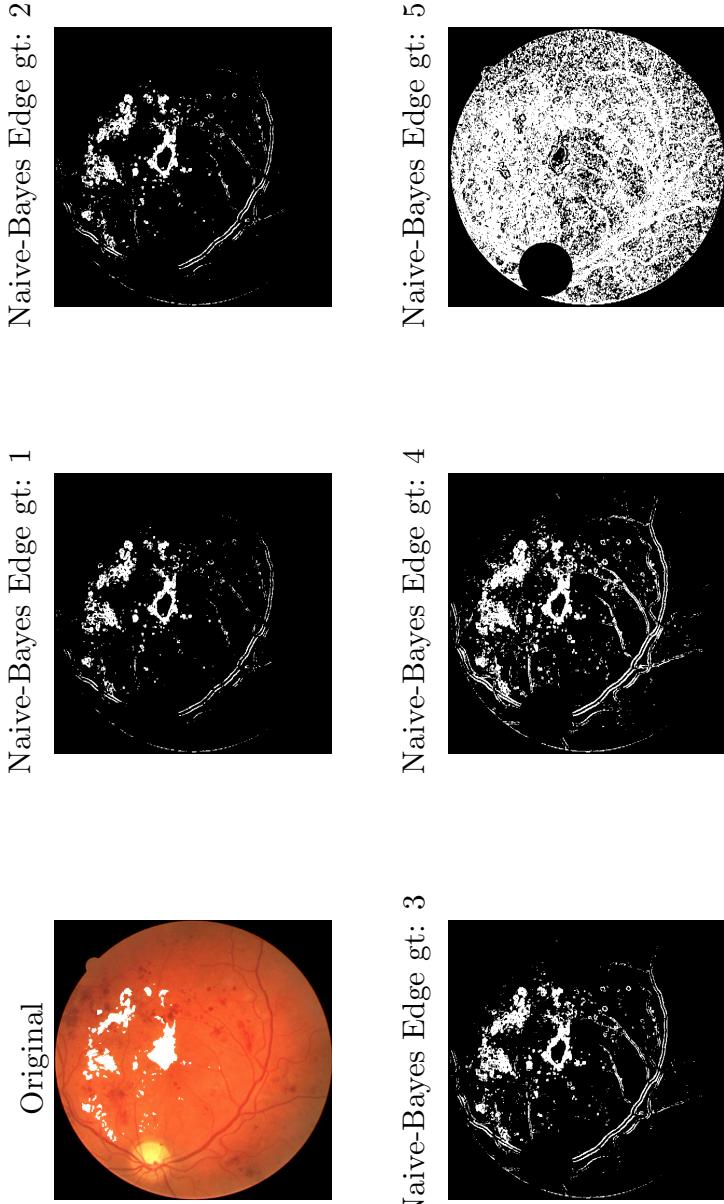
## Appendix 2. Example results of Naïve-Bayes



**Figure A2.1.** Example results of Naïve-Bayes classifier using color. The original ground truth is shown superimposed on the image of the eye fundus. The segmentation results of the method are shown as a binary image. The results with the accurate ground truth are shown first (1), then the results with the three iterations of inaccurate ground truth (2-4), and lastly the results with the most inaccurate ground truth (5).

(continues)

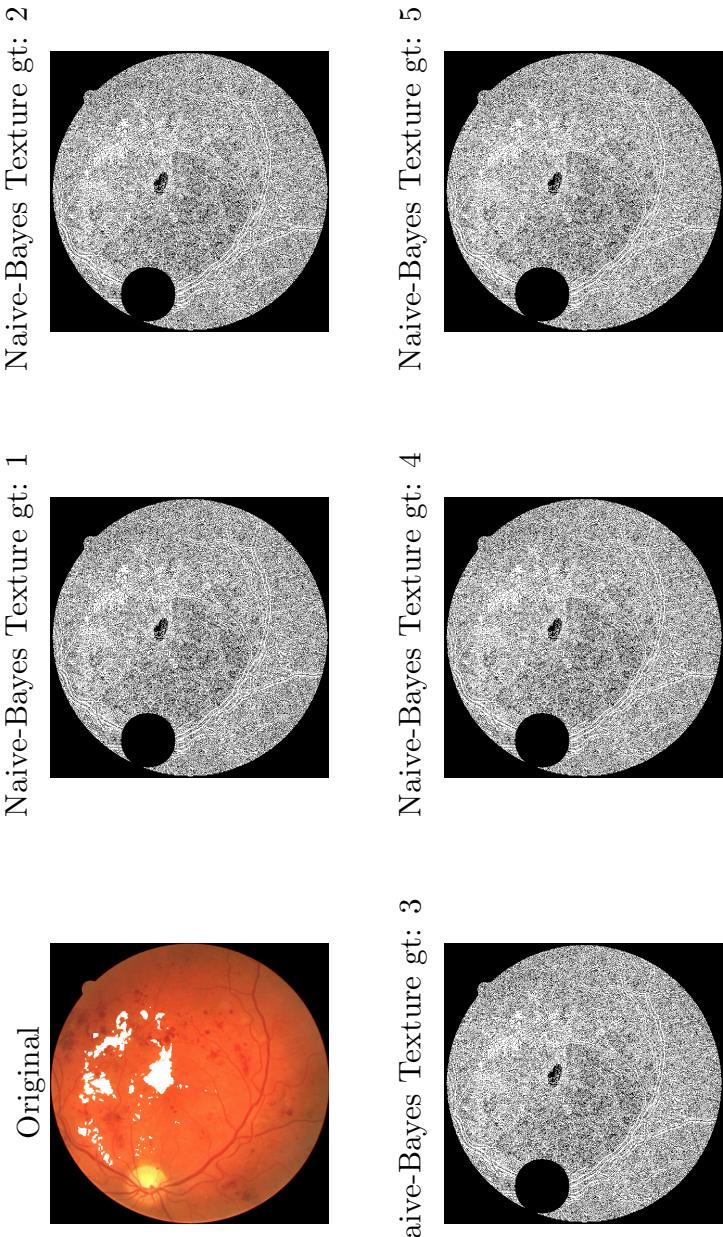
## Appendix 2. (continued)



**Figure A2.2.** Example results of Naïve-Bayes classifier using edge. The original ground truth is shown superimposed on the image of the eye fundus. The segmentation results of the method are shown as a binary image. The results with the accurate ground truth are shown first (1), then the results with the three iterations of inaccurate ground truth (2-4), and lastly the results with the most inaccurate ground truth (5).

(continues)

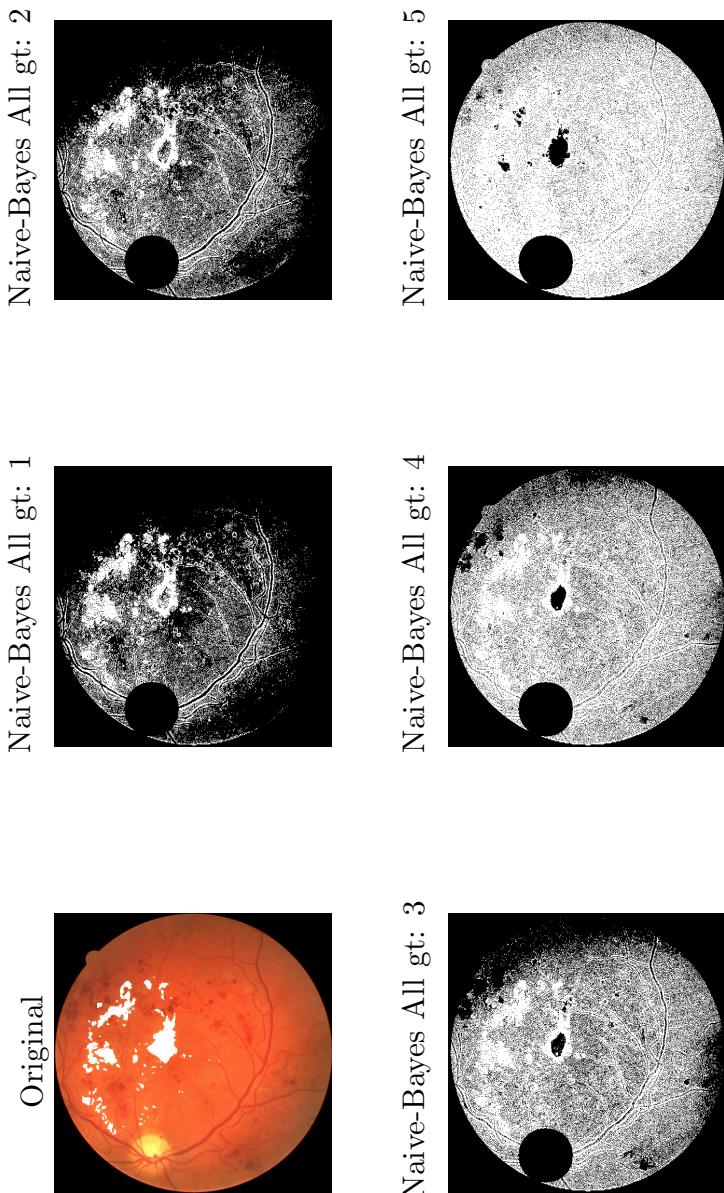
## Appendix 2. (continued)



**Figure A2.3.** Example results of Naïve-Bayes classifier using texture. The original ground truth is shown superimposed on the image of the eye fundus. The segmentation results of the method are shown as a binary image. The results with the accurate ground truth are shown first (1), then the results with the three iterations of inaccurate ground truth (2-4), and lastly the results with the most inaccurate ground truth (5).

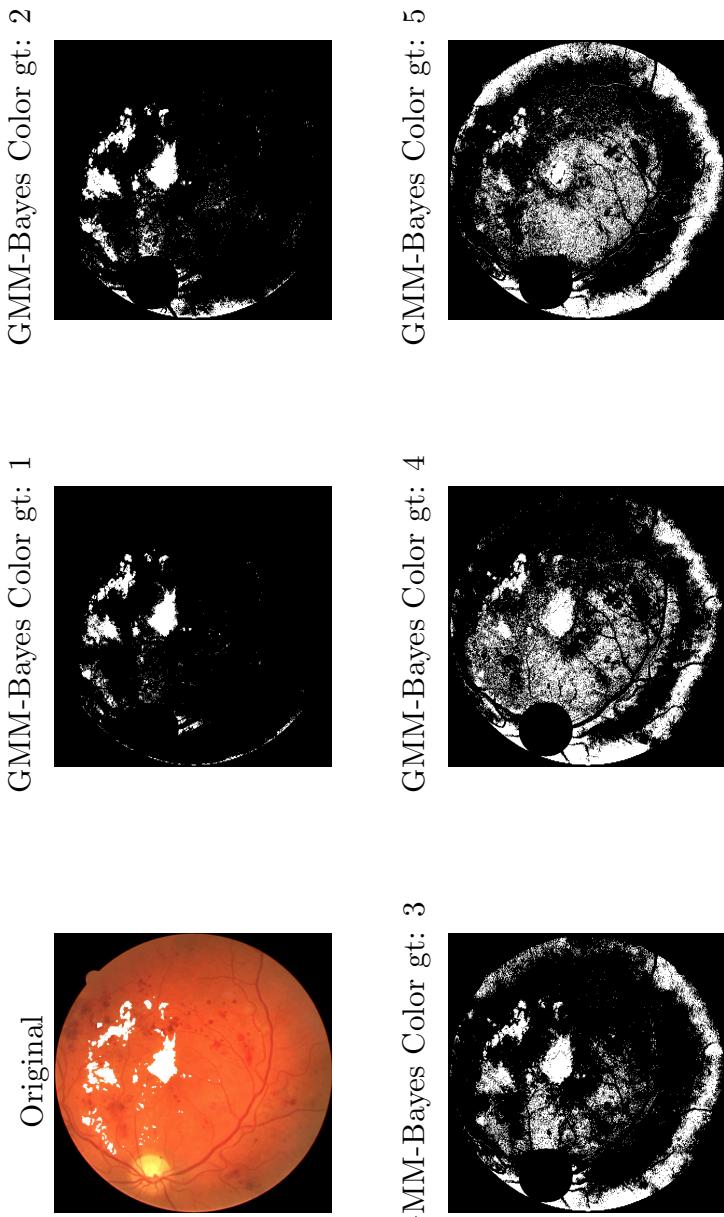
(continues)

## Appendix 2. (continued)



**Figure A2.4.** Example results of Naïve-Bayes classifier using all features. The original ground truth is shown superimposed on the image of the eye fundus. The segmentation results of the method are shown as a binary image. The results with the accurate ground truth are shown first (1), then the results with the three iterations of inaccurate ground truth (2-4), and lastly the results with the most inaccurate ground truth (5).

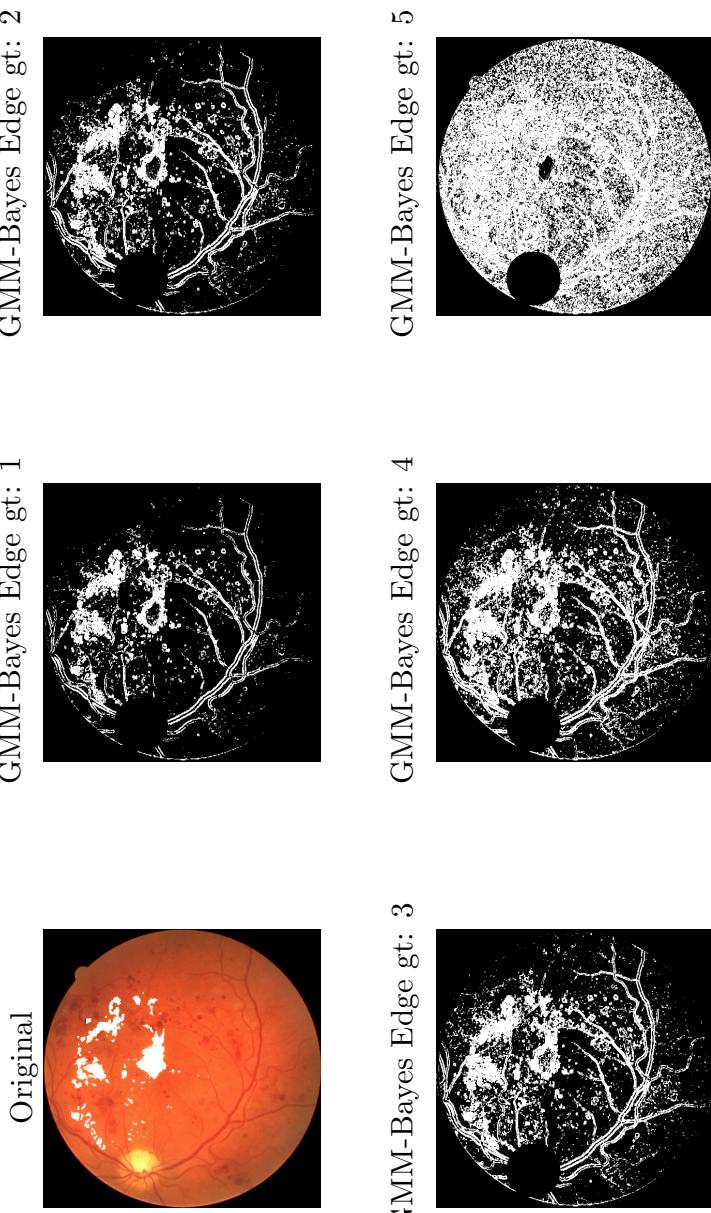
### Appendix 3. Example results of GMM-Bayes



**Figure A3.1.** Example results of GMM-Bayes classifier using color. The original ground truth is shown superimposed on the image of the eye fundus. The segmentation results of the method are shown as a binary image. The results with the accurate ground truth are shown first (1), then the results with the three iterations of inaccurate ground truth (2-4), and lastly the results with the most inaccurate ground truth (5).

(continues)

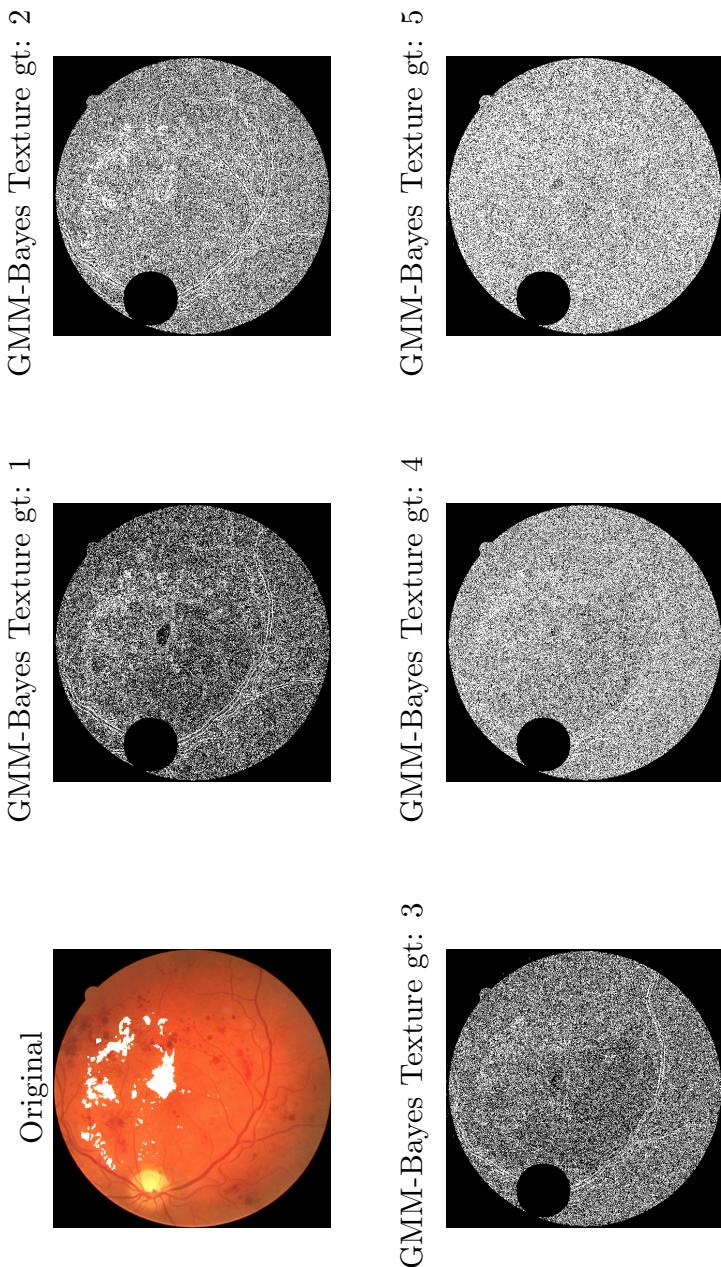
### Appendix 3. (continued)



**Figure A3.2.** Example results of GMM-Bayes classifier using edge. The original ground truth is shown superimposed on the image of the eye fundus. The segmentation results of the method are shown as a binary image. The results with the accurate ground truth are shown first (1), then the results with the three iterations of inaccurate ground truth (2-4), and lastly the results with the most inaccurate ground truth (5).

(continues)

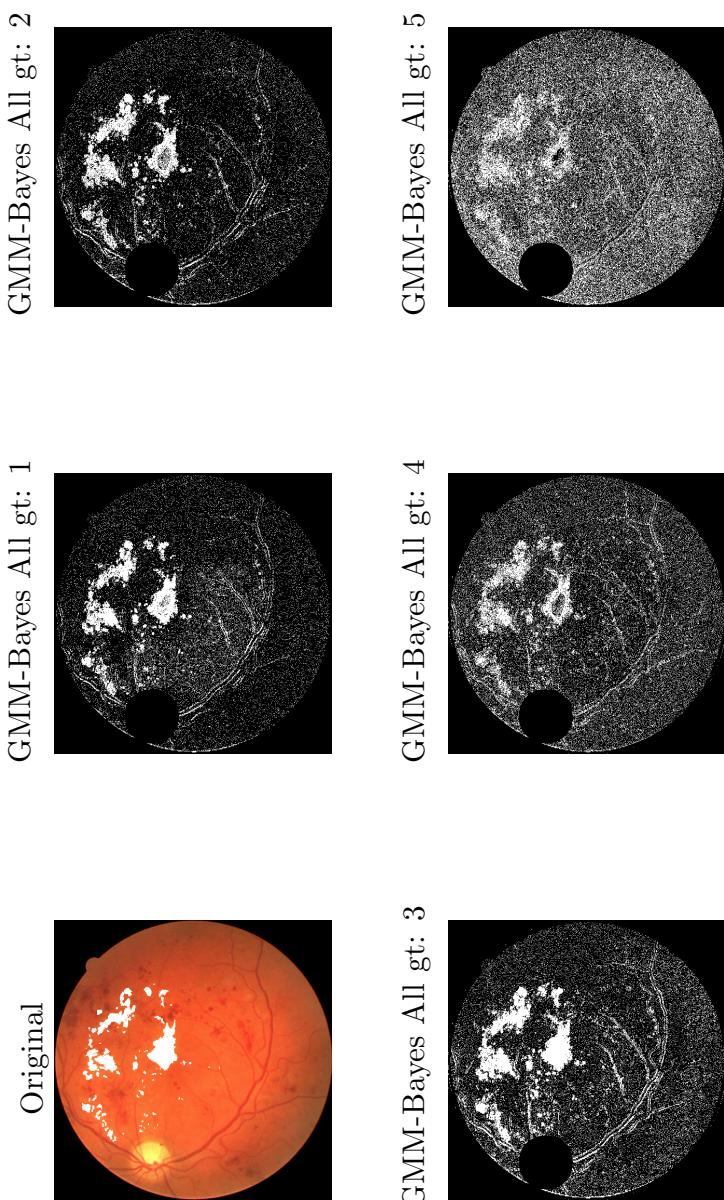
### Appendix 3. (continued)



**Figure A3.3.** Example results of GMM-Bayes classifier using texture. The original ground truth is shown superimposed on the image of the eye fundus. The segmentation results of the method are shown as a binary image. The results with the accurate ground truth are shown first (1), then the results with the three iterations of inaccurate ground truth (2-4), and lastly the results with the most inaccurate ground truth (5).

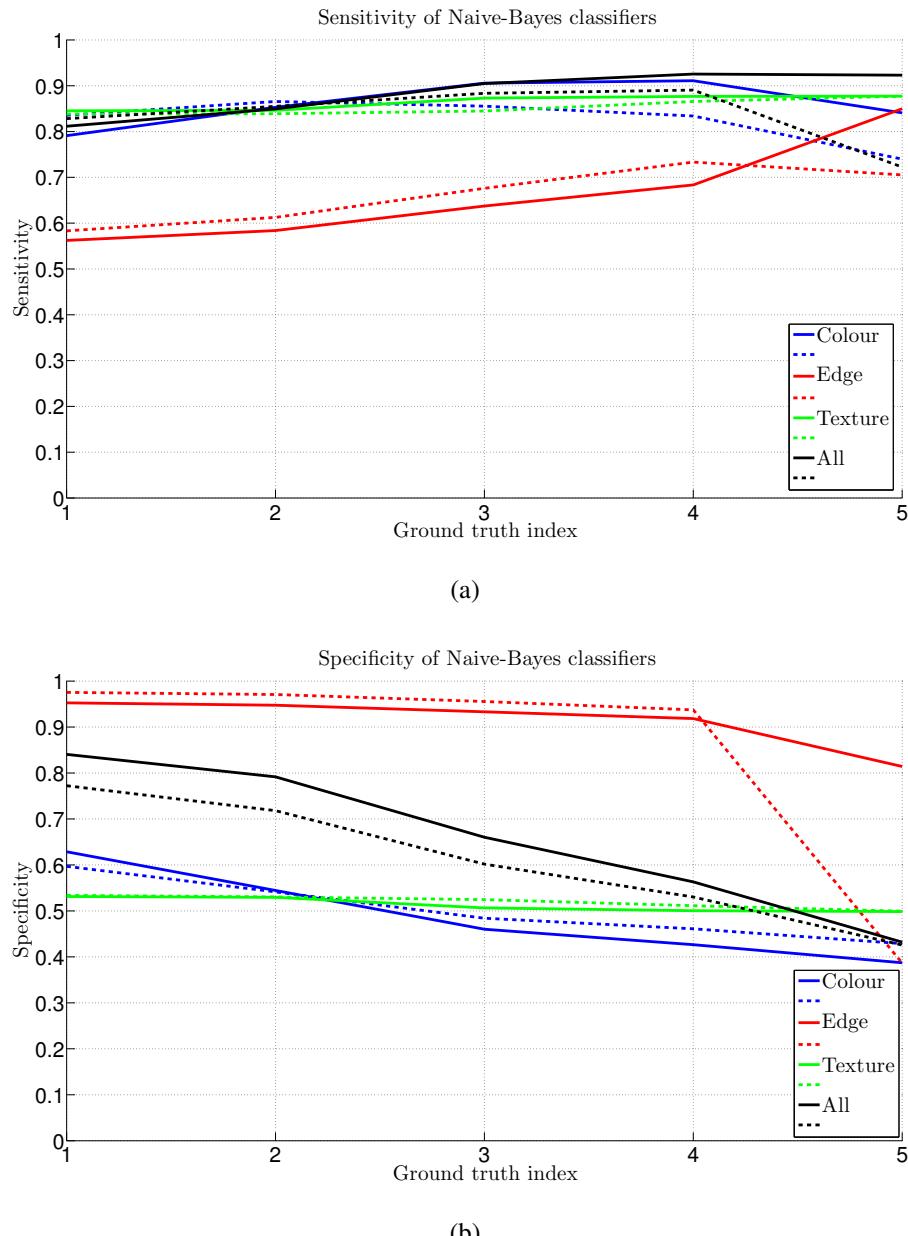
(continues)

### Appendix 3. (continued)



**Figure A3.4.** Example results of GMM-Bayes classifier using all features. The original ground truth is shown superimposed on the image of the eye fundus. The segmentation results of the method are shown as a binary image. The results with the accurate ground truth are shown first (1), then the results with the three iterations of inaccurate ground truth (2-4), and lastly the results with the most inaccurate ground truth (5).

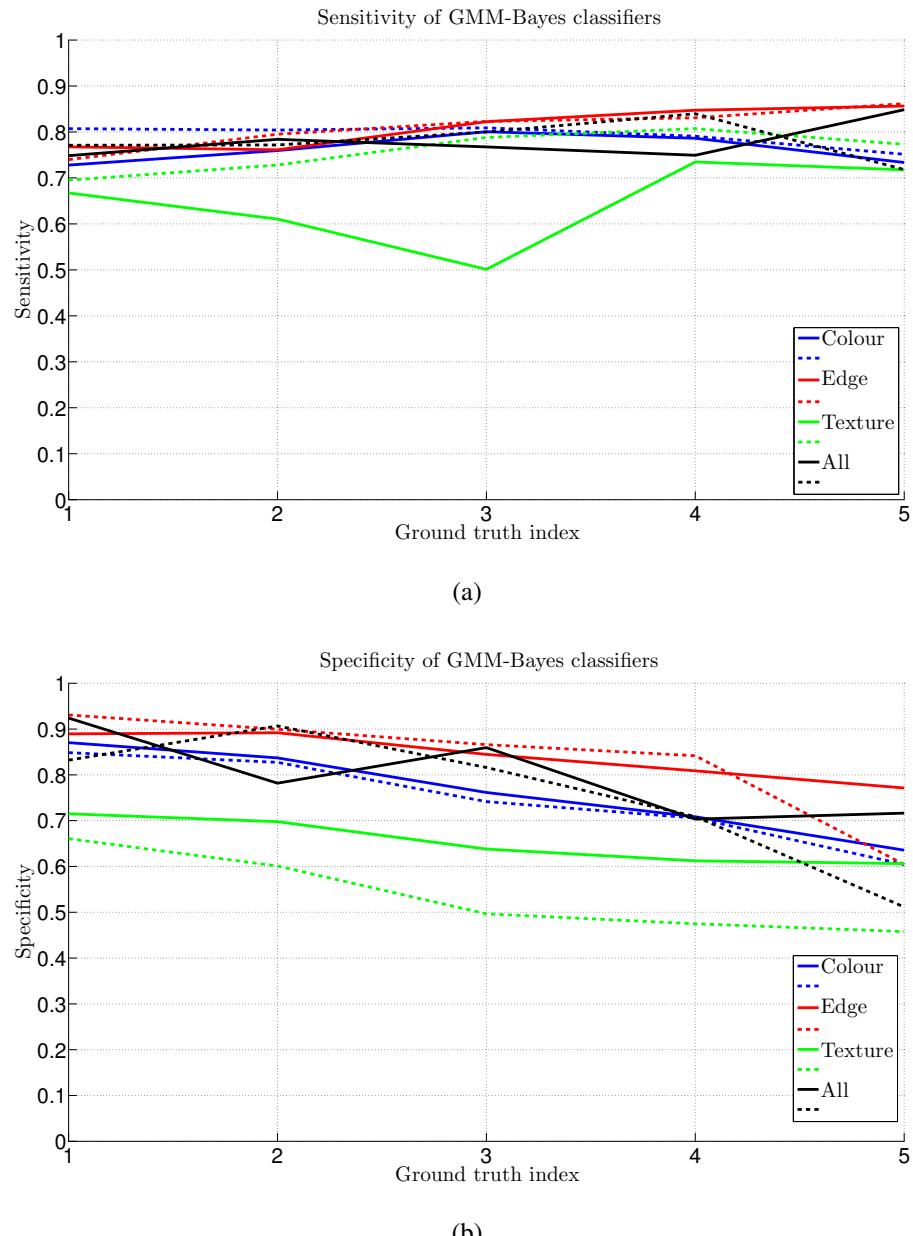
## Appendix 4. Supervised methods' performance using gray world pictures



**Figure A4.1.** Performance of Naïve-Bayes classifiers using gray world images: (a) sensitivity, (b) specificity.

(continues)

## Appendix 4. (continued)



**Figure A4.2.** Performance of GMM-Bayes classifiers using gray world images: (a) sensitivity, (b) specificity.