Objective assessment for tracheomalacia severity in neonates born with an esophageal atresia

Thesis

Nienke Muijen

Nienke Muijen (BSc.)

MSc student Technical Medicine,

Track | Medical Imaging & Intervention

M3 internship | Pediatric surgery department

Wilhelmina Children's Hospital, Utrecht, The Netherlands

January 25th, 2024

Supervisors

Prof. dr. I.A.M.J. Broeders	Chairman, University of Twente
Dr. S.H.A.J. Tytgat	Medical supervisor, Wilhelmina Children's Hospital
J.W. Verweij, MSc.	Medical supervisor, Wilhelmina Children's Hospital
Dr. C.O. Tan	Technical supervisor, University of Twente
J.R. Abbing, MSc.	Technical supervisor, University of Twente
J. de Witte, MSc.	Process supervisor, University of Twente
B. Wermelink, MSc.	External member, University of Twente

UMC Utrecht Wilhelmina Kinderziekenhuis

UNIVERSITY OF TWENTE. | TECHMED CENTRE

Abstract

Goal The goal of this thesis was to explore how the tracheomalacia severity in neonates born with an esophageal atresia could objectively be assessed, using bronchoscopy videos.

Method Three subgoals were established to achieve the goal. The first subgoal aimed to determine the appropriate region of interest and image representation, considering the 'red, green, blue' (RGB), 'hue, saturation, value' (HSV), and gray color-spaces. The second subgoal aimed to select the most suitable segmentation techniques, considering Otsu-thresholding, Multi-Otsu thresholding, K-means clustering, and Canny edge detection. The final subgoal aimed to develop an objective scoring system, using the obtained segmentations. Therefore, the area and anteroposterior : transverse (APT) diameter ratio was calculated for the end-inspiratory and -expiratory phase at the malacia and non-malacia part of the trachea. Five formulas were proposed to relate the APT-ratio outcomes from these frames to the clinically estimated tracheal lumen collapse (TLC). Linear regression was used to assess whether the calculated outcomes could predict the percentage of TLC.

Results The pixel values corresponding to the tracheal lumen exhibit a different range than the cartilage ring and posterior wall in the red-, value- and gray channel. Following dice similarity coefficient (DSC) values were obtained for tracheal lumen segmentation in these channels; Otsu-thresholding: DSC_{red}=0.9567, DSC_{value}=0.9571 & DSC_{gray}=0.9584; Multi-Otsu thresholding: DSC_{red}=0.9674, DSC_{value}=0.9681 & DSC_{gray}=0.9615; K-means clustering: DSC_{red}=0.9650, DSC_{value}=0.9638 & DSC_{gray}=0.9667. Linear regression resulted in the following outcomes; Formula 1: R^2=0.046, $RMSE$=22.12 & p=0.365; Formula 2: R^2=0.064, $RMSE$=21.91 & p=0.283; Formula 3: R^2=0.003, $RMSE$=22.61 & p=0.819; Formula 4: R^2=0.007, $RMSE$=22.56 & p=0.718; Formula 5: R^2=0.002, $RMSE$=22.62 & p=0.840.

Conclusion The tracheal lumen is the ROI that can be segmented from a bronchoscopy video. The highest DSC value (DSC = 0.9681) was obtained for three-threshold Multi-Otsu thresholding in the value channel. However, the examined techniques had all quite similar DSC values. It remains uncertain whether this particular combination is really the best, given the constraints of the limited data. The APT-ratio seems not to be a useful measurement because none of the proposed formula's had a statistically significant linear regression analysis. Further research should focus on how the area could be used as measurement.

Keywords Bronchoscopy; Computer Vision; Esophageal atresia; Tracheal lumen collapse; Tracheomalacia.

Abbreviations

APT	Anteroposterior : transverse
CSA	Cross-sectional area
CSA_{exp}	End-expiratory cross-sectional area
CSA_{insp}	End-inspiratory cross-sectional area
DSC	Dice Similarity Coefficient
EA	Esophageal atresia
FOI	Field of interest
HSV	Hue, saturation, value
RGB	Red, green, blue
ROI	Region of interest
TLC	Tracheal lumen collapse
TM	Tracheomalacia
WKZ	Wilhelmina Children's Hospital
3D	Three-dimensional

List of Figures

List of Tables

Contents

1

General introduction

This chapter provides clinical background information about tracheomalacia, the problem synthesis, and the goal of this thesis.

1.1 Congenital tracheomalacia

Approximately 1 in 2100[1][2][3] children suffer from tracheomalacia[4] (TM), a condition in which the tracheal lumen can (partly) collapse during expiration due to tracheal weakness. TM is one of the most common congenital tracheal malformations. Besides congenital (primary TM), it can also be acquired (secondary TM) for example due to prolonged intubation, external compression, trauma, or inflammation. However, secondary TM will be outside the scope of this thesis.

The typical congenital malformation associated with TM is esophageal atresia[5] (EA). Normally, the esophagus is a direct connection from mouth to stomach, see *Figure 1A*. In patients with EA, the esophagus is interrupted, resulting in a blind upper pouch, see *Figure 1B*. There are different types of EA, but around 90% of the neonates born with EA have a tracheo-esophageal fistula[6]: an abnormal connection between the trachea and the distal esophagus, see *Figure 1C*. This could be explained embryologically because the trachea and esophagus both develop from the foregut. Around 87%[5] of the neonates born with an EA suffer from TM, especially those with a tracheo-esophageal fistula. This fistula is transected during the EA surgery. A suture is placed in this fistula, followed by cutting through the fistula. The distal esophagus is now disconnected from the posterior wall of the trachea. The tension of this posterior wall is also removed, resulting in an improved floppiness of this posterior wall.

1.2 Anatomy

The trachea[8] is located in the thoracic cavity as part of the respiratory tract, which can be divided into an upper- and a lower tract. The trachea is part of the lower respiratory tract, which originates at the larynx and runs down to the carina. Here, the trachea separates into the left- and right main bronchus. The heart, and the aorta are located anterior from the trachea. Posteriorly, the esophagus is directly adjacent to the trachea with a common wall, and thereafter the spine. The trachea consists at the anterior side of 15 to 20 C-shaped, rigid cartilage rings which are connected via intercartilaginous membranes. The pars membranacea is located on the posterior side, which contains the tracheal muscle.[1][9]

1.3 Physiology

Along the respiratory cycle[10][11], the tracheal diameter changes, see *Figure 2.* During expiration, the tracheal lumen narrows due to inward movement of the posterior wall, resulting in acceleration of the airflow and clearance of the mucus. Normally, this narrowing ranges between 10-20%[2][10].

Tracheal compliance[1] is increased in TM, making the trachea more susceptible to tracheal lumen collapse (TLC) along the respiratory cycle. Several factors[10][11][12][13] can contribute to this; The cartilage rings can have a lower intrinsic strength or a flattened shape, and the posterior wall can be too floppy and broader than normal.

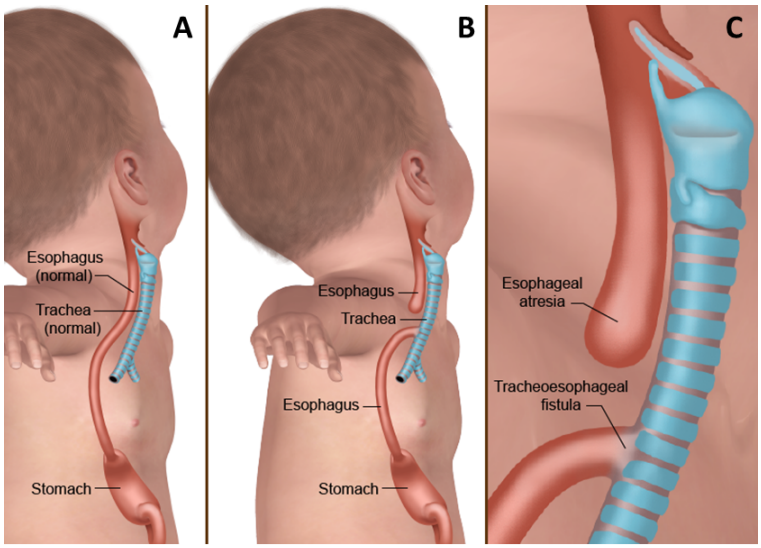

Figure 1: Anatomy visualisation.
A: Normal esophagus and trachea;
B: Esophageal atresia in combination with a tracheo-esophageal fistula;
C: Closer view of tracheo-esophageal fistula.[Adapted from: [7]]

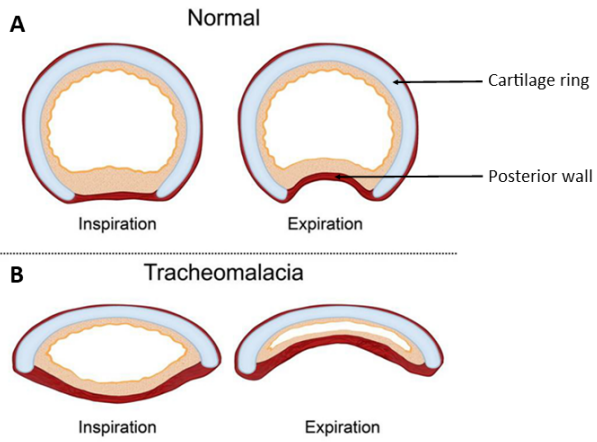

Figure 2: Cross-section of the trachea along the respiratory cycle.[Adapted from: [14]]
A: Normal trachea;
B: Tracheomalacia.

1.4 Symptoms

Common symptoms[1][2][3][4][15] of TM include abnormal breathing sounds (inspiratory stridor, expiratory rhonchi, wheezing), and a barking, ineffective cough. Clearing secretions may be challenging, making these patients susceptible to airway infections. A severe manifestation of TM are 'brief resolved unexplained events'[12]: the neonate stops breathing, turns blue, and becomes unresponsive. In a worst case, the neonate may need a resuscitation to recover.

1.5 Diagnostics

TM is diagnosed[1][2][3][15] before the surgery required for EA repair starts. A bronchoscopy is performed under general anesthesia while the neonate is spontaneously breathing, without the use of positive end-expiratory pressure[4][16]. This procedure provides a cross-sectional view of the trachea, see *Figure 3*. Typical aspects of TM observed in the bronchoscopy videos are a flattened cartilage shape and a (broader) posterior muscle wall, which intrudes into the tracheal lumen. The severity of TM is subjectively assessed during the procedure. In some cases, a bronchoscopy is not possible due to respiratory instability. Those neonates are not checked for TM. Since the neonates are about 3-5 days old during this procedure, they may be fragile. Especially those who are born too early.

There is no universally accepted grading system for TM. According to *Wallis et al*[15], the following (descriptive) classification is used for grading the severity of TM in pediatrics: 50-75% TLC is mild, 75-90 % TLC is moderate, and >90% TLC is severe. Clinical symptoms are not considered in the classification. In the The Wilhelmina Children's Hospital (WKZ, Utrecht, The Netherlands), they use a grading system[5][12] for neonates with EA to assess the severity of TM prior to EA repair; <33% TLC is mild, 33-66% TLC is moderate, and >66% TLC is severe. If the TLC is >33%, surgery seems to be required. This classification considers the increased floppiness of the posterior wall after transection. However, there is insufficient

evidence to establish this classification as a generally accepted grading system.

For three reasons, it is not preferred to do the bronchoscopy after tracheo-esophageal fistula transection. First, the neonate already received muscle relaxation medication. This is not a realistic scenario for TM assessment. Second, the neonate is positioned on the left side. Therefore, it will be challenging to position the neonate on the back, which is required for the bronchoscopic assessment. Last, not all neonates have a fistula so it is essential to check this before the EA repair.

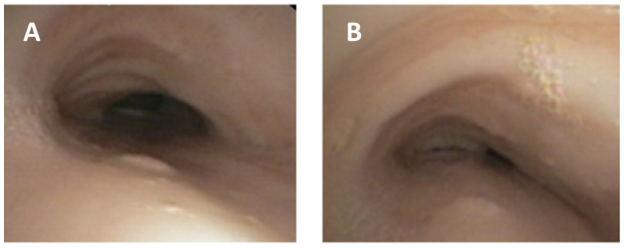

Figure 3: Bronchoscopy video in a tracheomalacia patient.[Adapted from: [1]]
A: Inspiration;
B: Expiration.

1.6 Treatment

Around 16-33%[12] of the neonates with EA and a tracheo-esophageal fistula suffer from symptomatic TM. The gold standard for treatment is in most hospitals an aortopexy[2]. During this surgery, the aorta is sutured to the sternum, resulting in decompression on the anterior side of the trachea. In the WKZ, a thoracoscopic posterior tracheopexy[5][17] is the gold standard. During this surgery, the posterior wall of the trachea is sutured with non-absorbable sutures at one to three places to the spinal ligament, see *Figure 4*. Due to this suturing, the posterior wall is stabalised under traction, preventing collapse of the tracheal lumen.

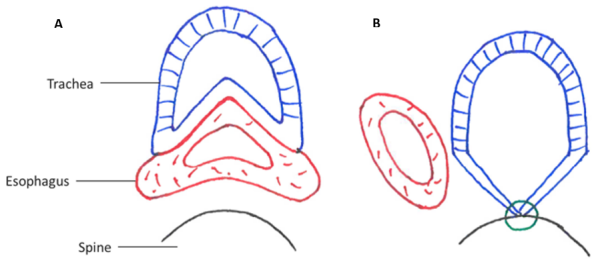

Figure 4: Posterior tracheopexy.[Adapted from: [18]]
A: Anatomy before a posterior tracheopexy;
B: Anatomy after a posterior tracheopexy.

1.7 Impact on life

Not treating or failing to treat TM can result in various consequences. As described in *Paragraph 1.4*, TM can result in 'brief resolved unexplained events' which could be fatal for the neonate. Such events occur particularly when the intrathoracic pressure[1] is increased, for example during feeding, crying, and coughing. Sometimes, tube feeding or a gastrostomy tube (a tube inserted into the stomach via a surgically made opening in the abdomen) is required to feed the neonate. This is disadvantageous because it is crucial for a neonate to learn how to drink and eat. Parents are required to receive training where they learn how to respond in a non-breathing situation, including resuscitation. However, non-breathing events could still be a reason for the child to remain in the hospital for a longer time, causing concerns for the parent(s) and/or relatives. When the TM is not severe enough, the child can go home, but it may raise other concerns, like sending the child to daycare or entrusting a nanny. In severe cases, the neonate may need a tracheostomy (a surgically created tracheal opening at the front of the neck) as temporary measure until a second surgery is performed.[5][12][19]

Neonates who do not undergo surgery will always keep some form of TM because it is not possible to grow out of it. During the years, the symptoms could become less[2] because the trachea grows, and the cartilage matures. Children with TM are susceptible to recurrent infections of the respiratory tract, exhibiting symptoms similar to those of asthma- and croup patients, and they could have apnea. Prolonged treatment with antibiotics and corticosteroids is often required, which is undesirable. About 27%[12] of the patients have permanent lung damage due to these respiratory infections. Additionally, children with severe TM may experience a reduced quality of life, and growth delay.[5][12][19]

1.8 Problem synthesis

As mentioned before, the determination of the TLC percentage relies on a subjective estimation by the otolaryngologist and the pediatric surgeon, which makes it susceptible to inter- and intra-observer variability. To enhance the outcome for the children suffering from TM, it would be helpful if an accurate and objective assessment could be made to determine whether a posterior tracheopexy is warranted or not. Especially for the neonates born with an EA it will be beneficial if this can be done during the the primary surgery for EA repair. This prevents a challenging second surgery due to adhesions from the previous surgery. Besides, it could prevent or limit the worse consequences associated with TM.

1.9 Related work

Several papers have been published on the quantification of TM. *Hysinger et al.*[1] explored the use of self-gated ultrashort echo-time magnetic resonance imaging to quantify TM. They measured the end-inspiratory cross-sectional area (CSA_{insp}) and end-expiratory cross-sectional area (CSA_{exp}), which they used to calculate the percentage change in cross-sectional area (CSA), using *Formula 1*. *Douros et al.*[20] explored the use of helical multi-detector computer tomography To quantify TM. They used a ratio between CSA_{exp} and CSA_{insp} as a measurement. *Ciet et al.*[21] explored the use of spirometer-controlled cine

magnetic resonance imaging to quantify TM. They measured the anteroposterior diameter end-inspiratory and end-expiratory at one point, which they used to calculate the percentage of diameter change. *Ebrahimian et al.*[13] explored automatic segmentation and measurement of the tracheal collapsibility in adult TM patients from a Computed Tomography (CT)-scan. They segmented the trachea from the CT-scan and used a commercial software to measure automatically the diameter and area; both along the trachea and at one single point (at the height of the aortic arch). For quantification, they used the difference between inspiratory and expiratory measurement outcomes, concluding that measurements along the whole trachea are more accurate than at one single point.

$$\Delta\text{CSA} = \frac{\text{CSA}_{insp} \text{ - } \text{CSA}_{exp}}{\text{CSA}_{insp}} \tag{1}$$

Only one study has been published which quantified the tracheal lumen area based on bronchoscopy videos. This study was published in 2005 by *Masters et al.*[22]. They developed a method for measuring the tracheal lumen by outlining the airway lumen using a color histogram. Recently, it has been investigated whether this technique is also suitable for the bronchoscopy videos available at the pediatric surgery department of the WKZ. However, the results were not accurate. Therefore, another approach should be explored.

1.10 Research aim

The main research question of this thesis is: '*How can the percentage of tracheal lumen collapse in neonates born with esophageal atresia objectively be assessed, using bronchoscopy videos?*'. Three subquestions are formulated to answer this question; The first subquestion is: '*What image representation is appropriate for cartilage ring and posterior wall, or tracheal lumen recognition in bronchoscopy videos?*', see *Chapter 2*. The second subquestion is: '*What image segmentation technique is suitable for the segmentation of cartilage ring and posterior wall, or tracheal lumen?*', see *Chapter 3*. The last subquestion is: '*What measurement (a calculation based on the segmentation) is usable for determining the percentage of tracheal lumen collapse?*', see *Chapter 4*.

2

Color-space analysis

In this chapter, various image representations will be explored that could be useful for segmenting a relevant structure for the objective assessment. Histogram analysis will be used to gain insight into the pixel value distribution of the structures in multiple image color-space channels. This type of analysis is known as color-space analysis.

2.1 Introduction

Images can be visualized in different color-spaces to gain a better understanding of their characteristics, and explore various aspects of the image. Commonly used color-spaces are: 'Red, green, blue' (RGB)[23], 'hue, saturation, value' (HSV)[23], and grayscale. RGB and HSV images consist of three channels, which can be separated into individual channels to represent different aspects. The three primary colors, red, green and blue, can individually be explored via the RGB color-space. The HSV color-space represents additional characteristics. Hue represents the pixel color itself, saturation the intensity of each pixel, and value the brightness each pixel. A grayscale image is a one channel image which also represents the brightness. However, the difference between grayscale and value is the formula[24] used for the new pixel value computation, resulting in different outcomes. The gray color space[25][26] is commonly used for image processing tasks related to objects with prominent edges that can easily be discriminated from a background.

This chapter aims to answer the first subquestion of this thesis: *What image representation is appropriate for cartilage ring and posterior wall, or tracheal lumen recognition in bronchoscopy videos?* In order to understand the intensity distribution of colors, and other features of the image, multiple (region-specific)

pixel intensity histograms are used. It is decided to focus on the tracheal lumen, cartilage ring, and posterior wall based on the expectation that these regions could probably provide useful measurements for objective TLC assessment.

2.2 Materials & method

2.2.1 Materials

Three bronchoscopy videos from neonates with EA were used. All patients received treatment in the WKZ in 2022, and were 2 and 10 days old. The algorithm used for the color-space analyses was developed with Python as programming language.

2.2.2 Method

Image pre-processing In *Figure 5*, a schematic overview is given of the image pre-processing method. One end-inspiratory frame was extracted per video. Each frame was converted using binary thresholding. The minimum and maximum, non-zero pixels in the x and y directions were used for cropping the original frames to the field containing the bronchoscopy view. This was defined as the field of interest (FOI).

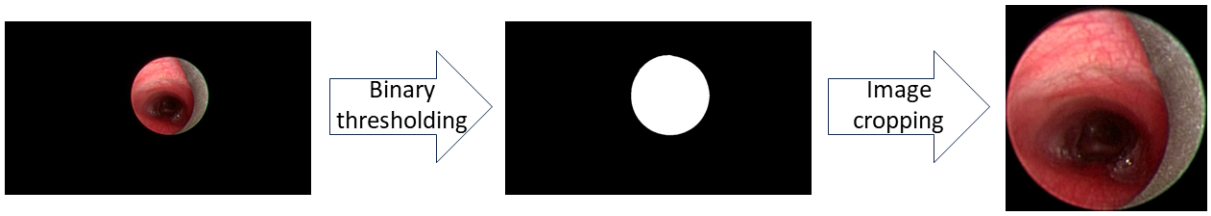

Figure 5: Overview of the image pre-processing steps for image cropping to the field of interest.

Histogram analysis Pixel intensity histograms were extracted from the three cropped frames. The individual channels from the following color-spaces were used: RGB, HSV, and gray. Besides, region-specific pixel intensity histograms were extracted for each color-space channel individual. The following regions of interest (ROI) were manually defined on each frame: cartilage ring, posterior wall, and tracheal lumen. Exactly the same regions were analysed across the different color-spaces. Furthermore, single-channel images were generated for each of the RGB and HSV channels individually, to gain a comprehensive understanding of the general pixel value distribution throughout the whole FOI.

2.3 Results

In *Figures 6, 7 and 8*, the extracted histograms are visualized for each patient. In each sub-figure, the upper histogram represents the pixel intensity distribution of the whole image. The other three histograms represent the region-specific pixel intensity distribution. The x-axis of each histogram represents the pixel value, and the y-axis the number of pixels corresponding to each pixel value. In the RGB histograms, the blue line represents the distribution for the blue channel, the green line for the green channel, and the red

line for the red channel. In the HSV histograms, the magenta line represents the distribution for the hue channel, the turquoise line for the saturation channel, and the yellow line for the value channel. In the gray histograms, the blue line represents the distribution for the gray channel. On the right side of the RGB and HSV plots, the single-channel images are visualized. The color bar represents the corresponding pixel value. In this paragraph, the pixel value ranges extracted from *Figures 6, 7 and 8* are presented per channel, and patient in the following order: cartilage ring/posterior wall/tracheal lumen.

Red channel The following ranges are roughly observed in the red channel: 80-150/60-160/10-100 for patient 1, 210-250/140-245/45-145 for patient 2, and 175-205/105-250/30-75 for patient 3. The cartilage ring, and posterior wall have quite similar ranges in each patient.The tracheal lumen has in general another range than the cartilage ring and posterior wall. The values corresponding to all three ROI's varies among the patients.

Green channel The following ranges are roughly observed in the green channel: 25-75/0-50/0-70 for patient 1, 95-155/30-130/5-90 for patient 2, and 95-140/30-210/0-45 for patient 3. All three ROI's have quite similar ranges in each patient. The values corresponding to all three ROI's varies among the patients.

Blue channel The following ranges are roughly observed in the blue channel: 25-70/0-50/0-60 for patient 1, 125-195/50-170/15-115 for patient 2, and 110-150/45-230/10-50 for patient 3. All three ROI's have quite similar ranges in each patient, except for the tracheal lumen in patient 3 which has another range than the cartilage ring and posterior wall in this patient. The values corresponding to all three ROI's varies among the patients.

Hue channel The following ranges are roughly observed in the hue channel: a peak at 0 & 180/a peak at 0 & 180/0-15 & 170-180 for patient 1, 165-175/165-180/165-180 for patient 2, and 170-180/160-175/a peak at 0 & 160-180 for patient 3. All three ROI's have quite similar ranges in each patient.

Saturation channel The following ranges are roughly observed in the saturation channel: 145-195/180-250/ 95-250 for patient 1, 100-145/115-195/70-240 for patient 2, and 85-125/50-190/95-220 for patient 3. All three ROI's have quite similar ranges in each patient.

Value channel The following ranges are roughly observed in the value channel: 80-145/60-170/20-95 for patient 1, 210-250/ 125-240/45-145 for patient 2, and 180-205/105-250/30-75 for patient 3. The cartilage ring, and posterior wall have quite similar ranges in each patient. The tracheal lumen has in general another range than the cartilage ring and posterior wall. The values corresponding to all three ROI's varies among the patients.

Gray channel The following ranges are roughly observed in the gray channel: 40-95/25-80/0-75 for patient 1, 135-190/ 55-160/15-110 for patient 2, and 120-155/55-225/15-55 for patient 3. The cartilage ring, and posterior wall have quite similar ranges in each patient. The tracheal lumen partly. The values corresponding to all three ROI's varies among the patients.

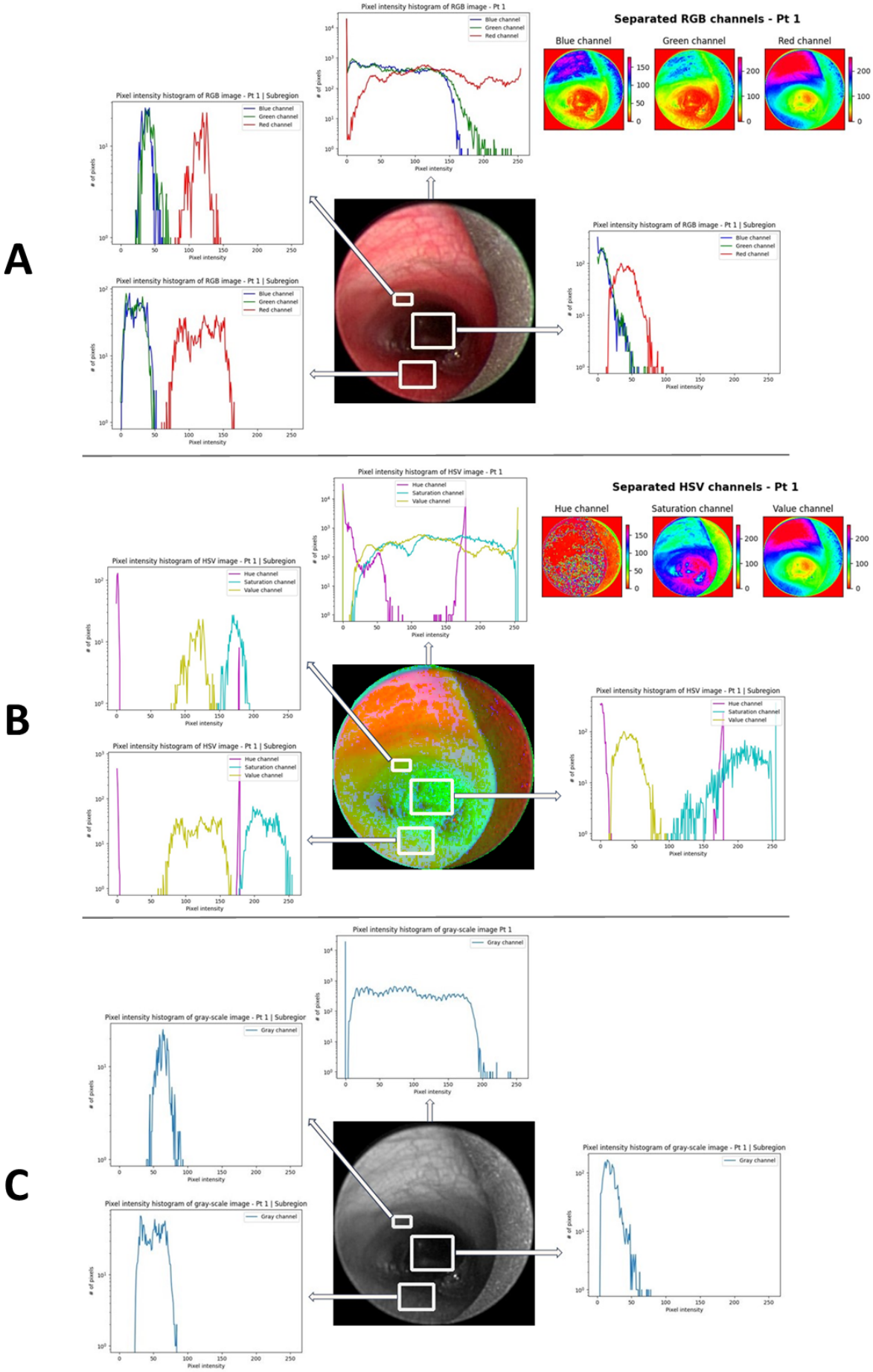

Figure 6: Pixel intensity distribution analysis in multiple color spaces for the same frame of the bronchoscopy video from patient 1.

A: RGB color-space;

B: HSV color-space;

C: Gray color-space.

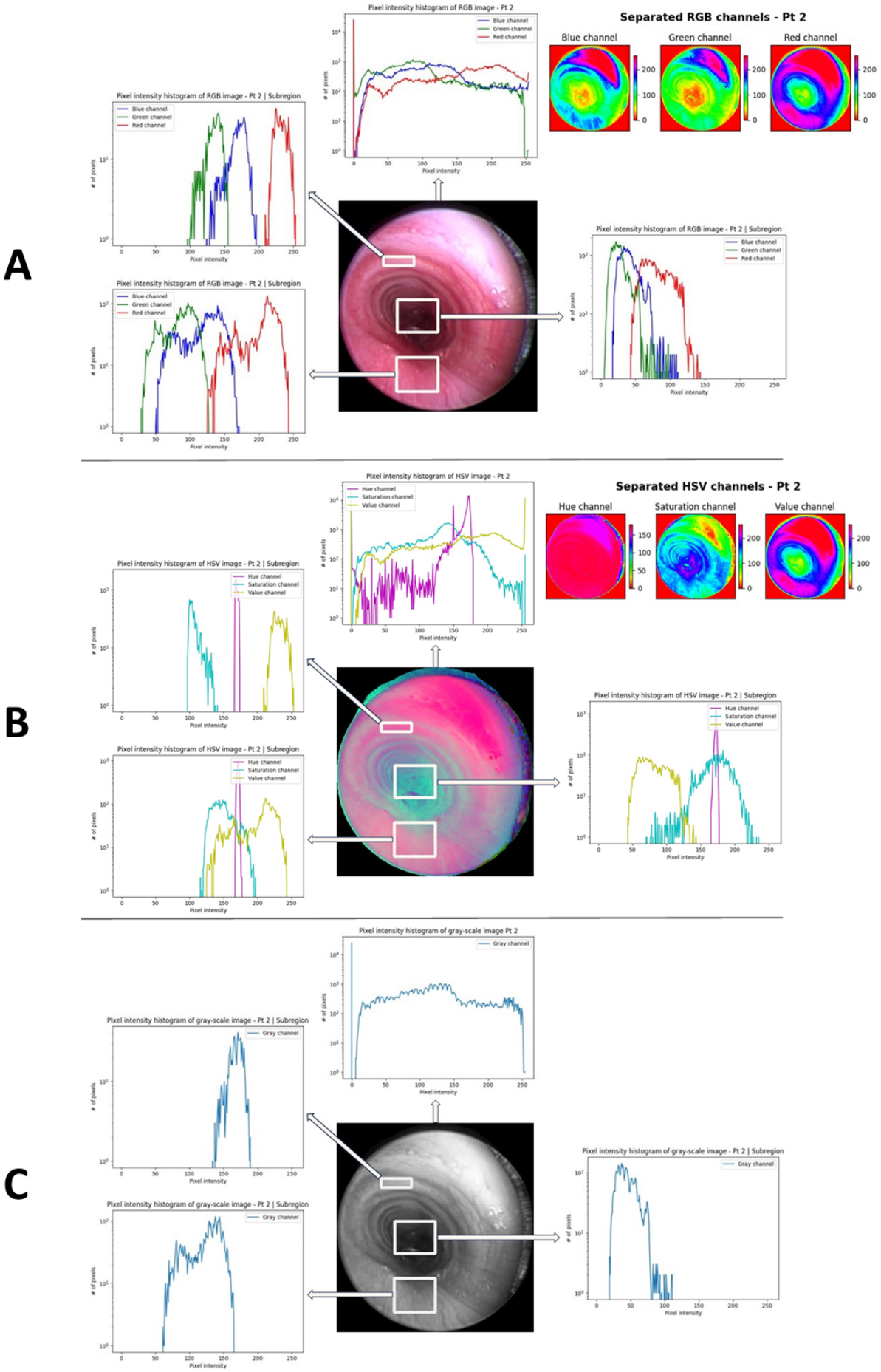

Figure 7: Pixel intensity distribution analysis in multiple color spaces for the same frame of the bronchoscopy video from patient 2.

A: RGB color-space;

B: HSV color-space;

C: Gray color-space.

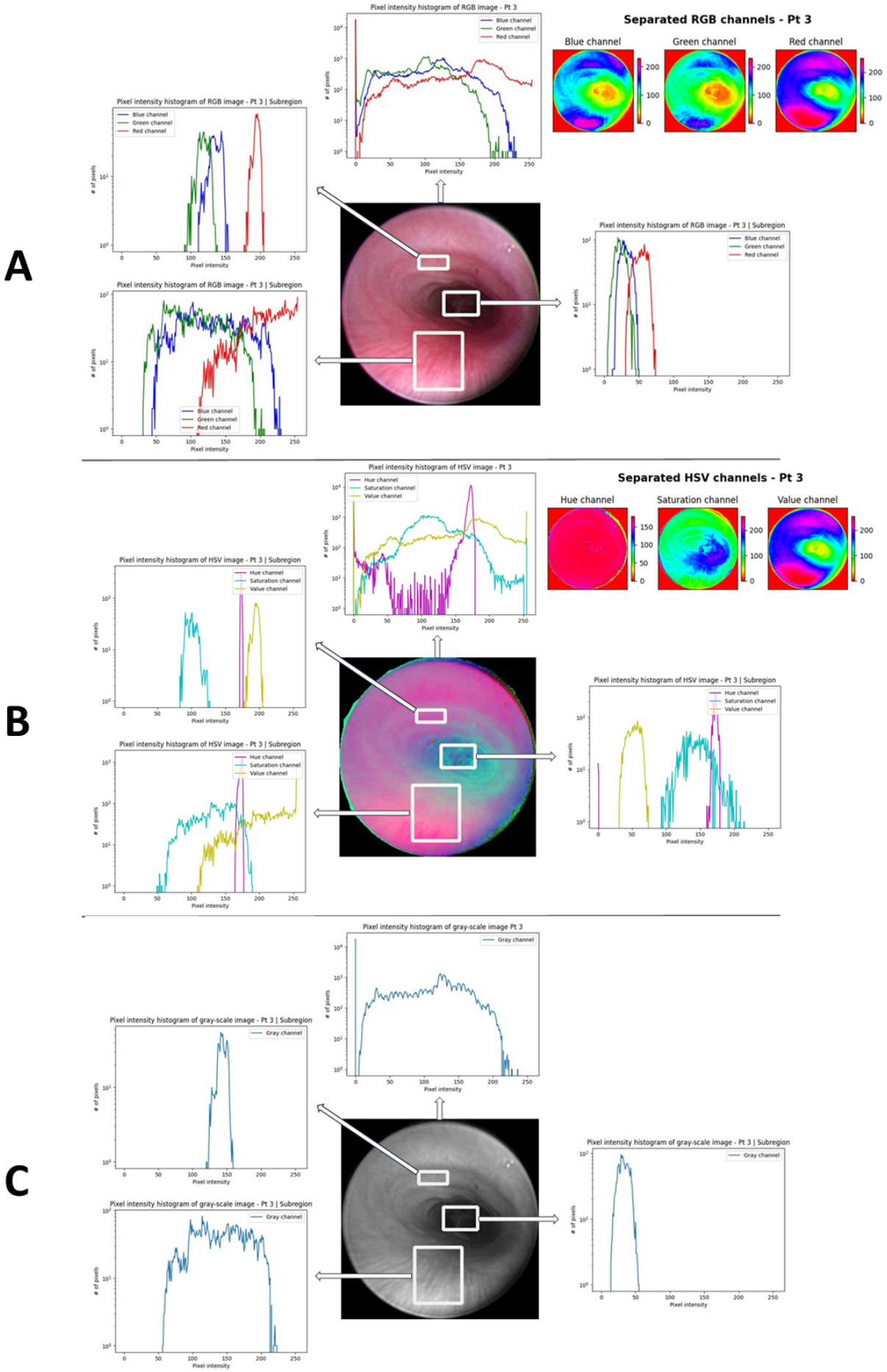

Figure 8: Pixel intensity distribution analysis in multiple color spaces for the same frame of the bronchoscopy video from patient 3.

A: RGB color-space;

B: HSV color-space;

C: Gray color-space.

Single-channel map Considering the single-channel maps for all three patients, the three ROI's can primarily be recognized in the red-, and value channel. In the blue-, green-, and hue channel maps this is overall not possible. In the saturation channel map from patient 2, and 3, it is possible to recognize the cartilage rings separately but these values appear quite similar to those of the tracheal lumen.

2.4 Discussion

The goal of this chapter was to investigate which image color-space and ROI could be appropriate for objective TLC assessment. As mentioned before, the ROI's were the cartilage ring, posterior wall, and tracheal lumen. It is expected that the tracheal lumen could be segmented from the red-, value, and gray channel based on the distinctive pixel value ranges for the three ROI's. An overlap in ranges for the three ROI's was observed for the blue-, green, hue-, and saturation channel, suggesting that they may not be useful. In the single-channel images was observed that the tracheal lumen could only be clearly identified in the red- and value channel images. This aligns with the results from the histogram analysis. For none of the explored channels, it was possible to detect the cartilage ring and posterior wall individually.

It is not possible to correlate the results with other studies. It depends on the goal, application, and the data which image channel could best be used. One study was found were they tried to segment the tracheal lumen in bronchoscopy video's, published by *Masters et al.*[22] in 2005. They used manual adaptation of the color balance, which is not comparable to the investigation in this chapter. No study was found related to cartilage ring and posterior wall segmentation.

There are six limitations in this study. First, one bounding box per ROI was used. When multiple bounding boxes for the same ROI are drawn, it could be possible to check if the pixel distribution is the same at all these points, for example along the whole shape of a cartilage ring. Second, the bounding box of the posterior wall was quite large, instead of just in the continuation of one cartilage ring. Therefore, the histogram was too general. Third, only three frames were used. Maybe, one of the used frames was an outlier. More frames would be recommended for the next time to improve the understanding of the data. Fourth, one frame per video was chosen. How the distribution changes along multiple frames was not considered. Fifth, only end-inspiratory frames were used. In an end-expiratory frame, the depth of field could be changed. Therefore, it could be relevant to consider the pixel distribution of the tracheal lumen relative to the other structures in that stage. Last, the inter-cartilage tissue was outside the scope of this study. However, considering the saturation images, this could probably be used as an outline of the lumen. Therefore, it would be recommended to include this ROI the next time.

Initially, it was uncertain whether pixel intensity or brightness majorly contributed to the dark, red color observed deeper in the tracheal lumen compared to the tracheal lumen near to the bronchoscope. Based on the conclusion that the value channel is usable while the saturation is not, it could be concluded that brightness contributes to this effect. This brightness is likely to be the consequence of the illumination provided by the bronchoscope.In general, the tracheal lumen had a lower pixel value range compared to the other two ROI's. However, in the red-, value-, and gray channels was also some pixel value range overlap between the three ROI's for some patients. This is not expected to be problematic due to the limited

overlap. Another important observation is that the ranges of the tracheal lumen varied among the patients. Therefore, it is recommended to use a segmentation technique based on an automatic principle.

2.5 Conclusion

The red-, value- and gray channel are considered suitable for tracheal lumen segmentation. For none of the used color-spaces, it was possible to detect the cartilage ring and posterior wall separately. The subquestion addressed in this chapter is answered.

3

Tracheal lumen segmentation

In the previous chapter was concluded that the tracheal lumen could probably be segmented from the red-, value-, and gray channels. In this chapter, the appropriate segmentation technique for tracheal lumen segmentation in those channels will be explored. Multiple techniques will be tried, and evaluated, with the use of labeled ground truths to determine the most suitable segmentation technique.

3.1 Introduction

Image segmentation is one of the tasks computer vision[27][28][29] tackles, with the purpose of processing and analyzing images high-detailed on the pixel level. The goal of image segmentation is to divide an image into various regions by assigning all the pixels into a class, based on shared characteristics. In this way, it is possible to process only the ROI's[30] instead of the whole image. Common computer vision segmentation techniques rely on the following principles[30][31][32][33]: thresholding, regions, or boundaries. It depends on the images and purpose which technique would be appropriate. Thresholding is a relatively easy and fast technique. It filters pixels based on the pixel value distribution in the image. However, this technique is susceptible to differences in pixel values between the image frames and patients, and overlapping pixel intensities of ROI's. Region-based image segmentation relies on clustering the pixels that meet the same criteria, for example pixel intensity or texture, and takes the spatial relationship (pixel connectivity) into account for defining multiple regions within an image. However, it could be challenging to define the boundaries in regions with quite similar characteristics while they actually not belong to the same region. Furthermore, it could be too time consuming for complex images as it analyzes individual regions of an image. Boundary-based segmentation relies on abrupt pixel value changes, which works well in images with

limited number of prominent edges. However, it may not be useful in images with a lot of edges, and blurred edges due to noise.

Performance of a technique can be evaluated via evaluation metrics. These compare the segmentation relative to a (manual) labeled ground truth. A commonly overlap-based metric is the Dice Similarity Coefficient[34] (DSC), see *Formula 2*. A DSC value ranges between 0 and 1. A value of 0 indicates no overlap between the segmentation and ground truth, while a value of 1 indicates a full overlap between segmentation and ground truth.

This chapter aims to answer the second subquestion of this thesis: *What image segmentation technique is suitable for the segmentation of cartilage ring and posterior wall, or tracheal lumen?* The following segmentation techniques are explored: Otsu-thresholding, Multi-Otsu thresholding, K-means clustering, and Canny edge detection. The DSC is used as evaluation metrics.

$$\text{DSC} = \frac{2 \times \text{Area of overlap}}{\text{Total area}} \tag{2}$$

3.2 Materials & method

3.2.1 Materials

Six bronchoscopy videos from neonates with EA were used. All patients received treatment in the WKZ in 2022, and their ages ranges between 2 and 10 days old. The algorithms used for the segmentation techniques and calculating the DSC value were developed with Python as programming language. Data labeling was performed using *Label studio*.

3.2.2 Method

Based on the results of *Chapter 2*, it was decided to use the red-, value, and gray channels for tracheal lumen segmentation in this chapter.

Data extraction From each bronchoscopy video, four frames were extracted; An end-inspiratory, and an end-expiratory frame from both the malacia and non-malacia segment of the trachea. It was decided to use these specific frames based on the fact that they correspond to the points along the respiratory cycle where the lumen is at its maximum (end-inspiratory) and minimum (end-expiratory). The frames were extracted from a malacia and non-malacia segment of the trachea in order to establish a patient-specific reference value for what a normal collapse value is for that patient.

Image pre-processing The same method as described in *Paragraph 2.2.2* was used for image pre-processing all the frames. Normalization of the image was not performed.

Segmentation techniques It was decided to include at least one threshold-based, one region-based, and one edge-based segmentation technique for better understanding which type would be useful at all. The focus was placed on relatively easy, and already known techniques. Especially because of the limited amount of time. Algorithms for the following segmentation techniques were developed and individually applied to the three channel representation of all frames: Otsu thresholding, Multi-Otsu thresholding (three thresholds), K-means clustering (with K=3 and K=5), and Canny edge detection. For Multi-Otsu thresholding, various numbers of thresholds were explored by plotting these in the histograms of *Chapter 2*. Using three thresholds resulted in the best thresholds considering the tracheal lumen pixel value range. Gaussian filtering, a Sobel kernel, and hysteresis between 20 and 100 were used for the Canny edge detection. Furthermore, local gradients were extracted using a Scharr kernel (first-order derivative). The segmentation algorithms were applied to each frame and image representation (red-, value-, and gray channel) individually. Several operations were conducted to improve the segmentation mask generated by the segmentation algorithms. Especially to remove the outline of the FOI that remained in the segmentation mask, and remove some structures, unrelated to the tracheal lumen. The used operations will be described in the following paragraph.

Mask post-processing A segmentation mask was generated after applying the segmentation technique. This mask was closed to ensure that regions belonging to the tracheal lumen were not removed. The center of the tracheal lumen was defined in the mask via interactive pixel selection. A circular area was defined around this point, with a radius equal to 0.3 times the image height. This radius was chosen because it was observed that none of the tracheal lumens had a larger radius than 0.3 times the image height. Any structure outside this defined area that did not belong to the tracheal lumen itself was removed, and the contours of the remaining structures were determined. The object with the largest boundary was defined as the tracheal lumen and used as the final segmentation mask. This mask was applied to the corresponding cropped RGB frame, resulting in an image that only contained the tracheal lumen.

Evaluation The ground truths were manually labeled, using the cartilage ring at the light/dark transition in the RGB image as guidance. The lumen enclosed by this cartilage ring was defined as tracheal lumen ground truth. All DSC values were calculated, using *Formula 2*. The mean DSC value and standard deviation were calculated for each combination of segmentation technique and image representation channel. The evaluation included the following techniques: Otsu thresholding, Multi-Otsu thresholding (using the middle threshold value) and K-means clustering (using K = 5). For this last technique, the cluster representing the tracheal lumen was used as the segmentation.

3.3 Results

Otsu thresholding In *Figure 9*, some examples of the results for Otsu thresholding are visualized. A white outline, representing the segmentation, is drawn at the cropped frames. Each column belongs to one input frame. The images in row A belong to using the red channel as input, row B to the value channel, and row C to the gray channel. The aimed goal to segment was the dark, red area deeper in the lumen. It can be observed that it was not always successful to include this whole area in the segmentation, resulting in an undersegmentation. Due to this, the segmentation is not in the same plane relative to the camera, when considering a three-dimensional (3D) space. Furthermore, it can be observed that some segmentations are just a circle, for example in row 'C', the first, and fifth column.

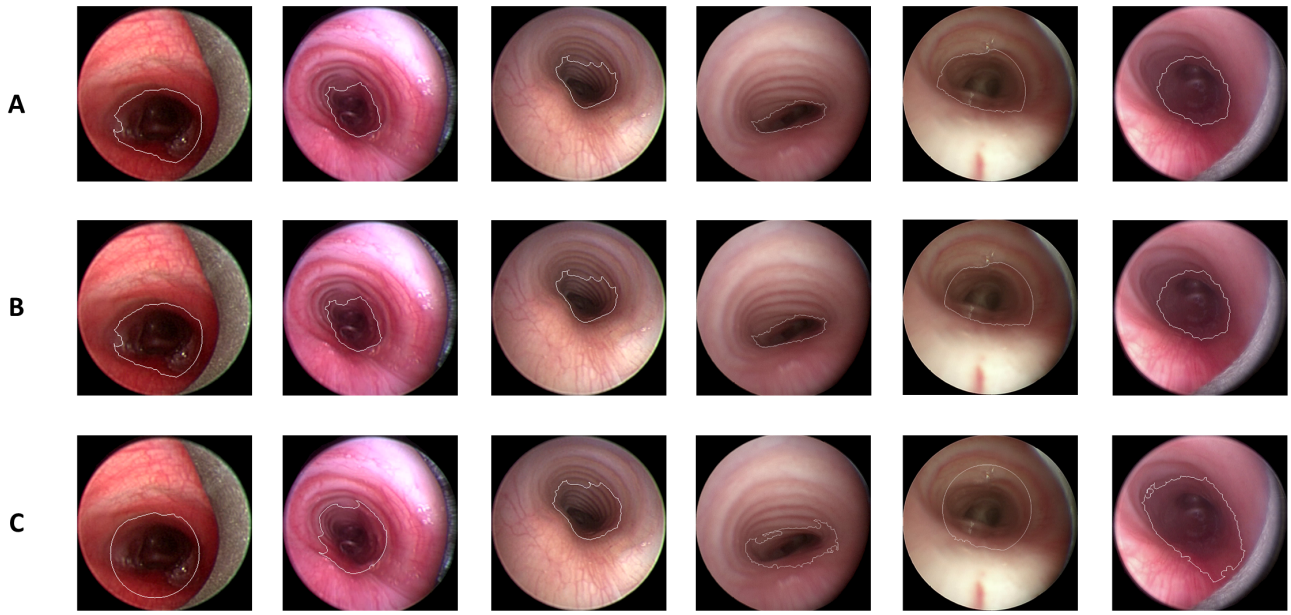

Figure 9: Some examples of using Otsu thresholding including mask post-processing. Each column belongs to one input frame, and the white outline represents the segmentation.
A: Red channel;
B: Value channel;
C: Gray channel.

Multi-Otsu thresholding In *Figure 10*, some examples of the results for the middle value Multi-Otsu thresholding are visualized. A white outline, representing the segmentation, is drawn at the cropped frames. Each column belongs to one input frame. The images in row A belong to using the red channel input, row B to the value channel, and row C to the gray channel. The aimed goal to segment was the dark, red area deeper in the lumen. It can be observed that it was not always successful to include this whole area in the segmentation, resulting in an undersegmentation. Due to this, the segmentation is not in the same plane relative to the camera, when considering a 3D space. Furthermore, it can be observed that some segmentations are just a circle, for example in row 'C', the first column.

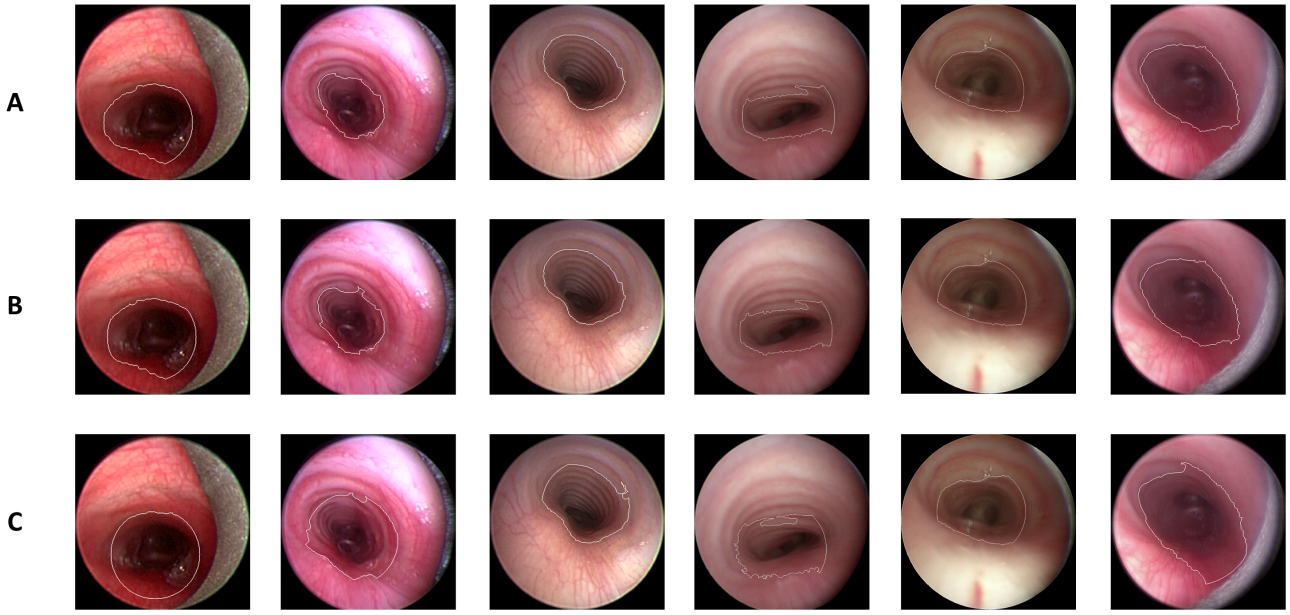

Figure 10: Some examples of using the middle value threshold, based on Multi-Otsu thresholding. These results include mask post-processing. Each column belongs to one input frame, and the white outline represents the segmentation.
A: Red channel;
B: Value channel;
C: Gray channel.

K-means clustering In *Figure 11*, some examples of using the tracheal lumen cluster defined via K-means clustering with K=5 are visualised. A white outline, representing the segmentation, is drawn at the cropped frames. Each column belongs to one input frame. The images in row A belong to using the red channel input, row B to the value channel, and row C to the gray channel. The correct cluster was manually extracted. The aimed goal to segment was the dark, red area deeper in the lumen. It can be observed that it was not always successful to include this whole area in the segmentation, resulting in an undersegmentation. Due to this, the segmentation is not in the same plane relative to the camera, when considering a 3D space.

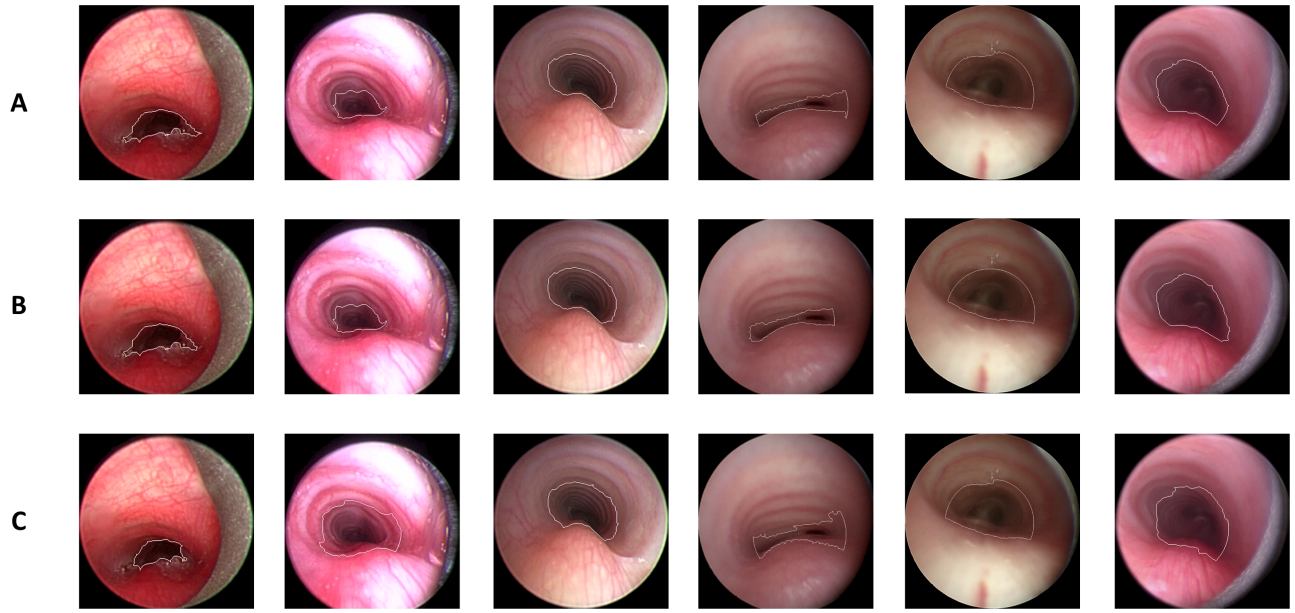

Figure 11: Some examples of using K-means clustering (K = 5), including mask post-processing. Each column belongs to one input frame, and the white outline represents the segmentation.
A: Red channel;
B: Value channel;
C: Gray channel.

Canny edge detection In *Figure 12*, some examples of using Canny edge detection are visualized. Each column belongs to one input frame. The images in row A belong to using the red channel as input, row B to the value channel, and row C to the gray channel. The aimed goal to segment was the same area as aimed in the first, second, and third column of *Figures 9 and 10*. It is in none of the images plot in *Figure 12* possible to recognise an enclosed tracheal lumen.

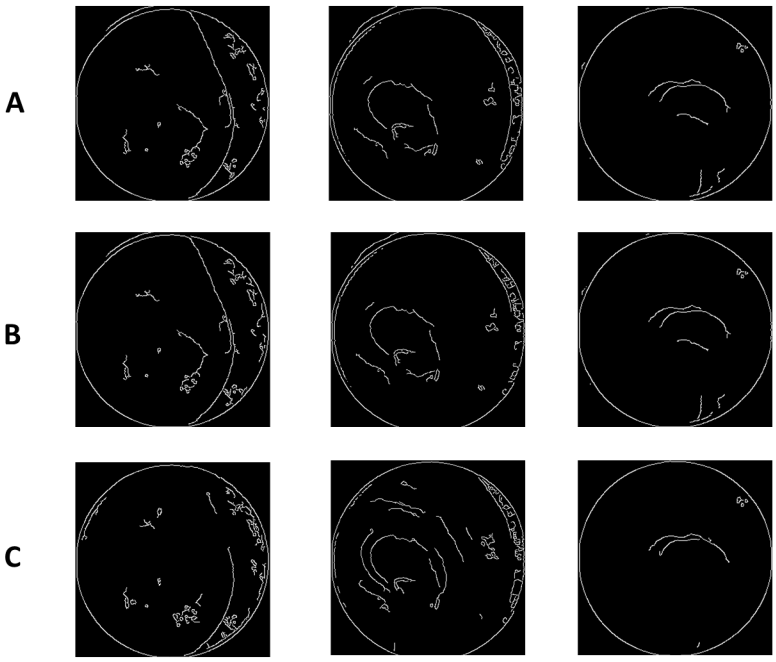

Figure 12: Some examples of the results for Canny edge detection, using different image channels. Each column belongs to one input frame.
A: Red channel;
B: Value channel;
C: Gray channel.

Gradients In *Figures 13, 14*, and *15*, some examples of the corresponding local gradient maps are visualized. These gradients were extracted via a Scharr kernel in the x- and y-direction independently. Each column belongs to one input frame. The images in row A belong to the x-direction gradient, and row B to the y-direction gradient. The aimed goal to segment was the same area as aimed in the first, second, and third column of *Figures 9 and 10*. In all images plot in *Figures 13, 14 and 15*, a considerable amount of gradients were detected in each channel and in both directions. It is in none of these images, it is possible to recognise an enclosed tracheal lumen.

In *Appendix A*, the results for the segmentations without mask post-processing, the effect of all three thresholds determined with Multi-Otsu thresholding, and the effect of K=3 and K=5 clustering can be found.

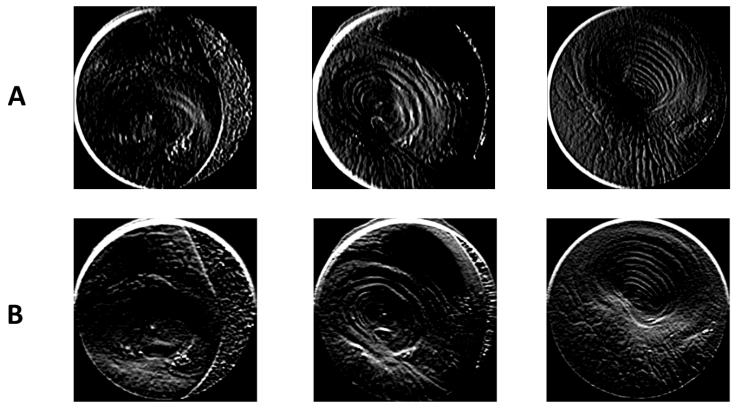

Figure 13: Some examples of Scharr gradients in the red channel. Each column belongs to one input frame.
A: Scharr filter in x-direction;
B: Scharr filter in y-direction.

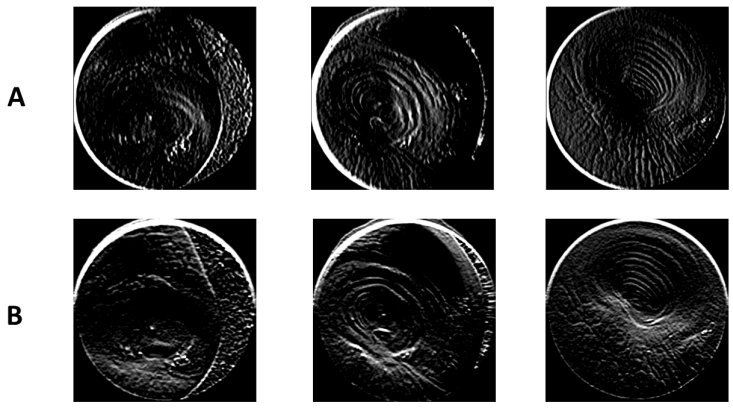

Figure 14: Some examples of Scharr gradients in the value channel. Each column belongs to one input frame.
A: Scharr filter in x-direction;
B: Scharr filter in y-direction.

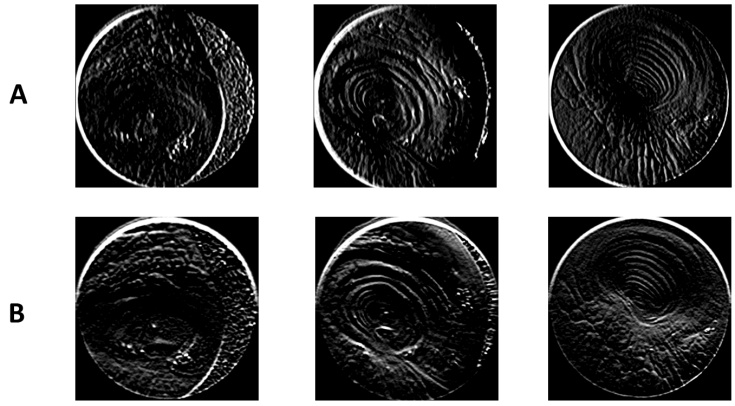

Figure 15: Some examples of Scharr gradients in the gray channel. Each column belongs to one input frame.
A: Scharr filter in x-direction;
B: Scharr filter in y-direction.

DSC value In *Table 1*, the mean DSC value, including the standard deviation, for each technique and channel combination can be found.The DSC values were all quite similar to each other. The lowest DSC value (DSC = 0.9567) was observed for Otsu thresholding in the red channel. The highest DSC value (DSC = 0.9681) was observed for Multi-Otsu thresholding in the value channel.

Table 1: Mean DSC and standard deviation for multiple segmentation technique and image channel combinations (n = 24).

Segmentation technique	**Red channel**	**Value channel**	**Gray channel**
Otsu thresholding	0.9567 ±0.03234	0.9571 ±0.03201	0.9584 ±0.02706
Multi-Otsu thresholding	0.9674 ±0.02191	0.9681 ±0.02207	0.9615 ±0.02032
K-means clustering	0.9650 ±0.02061	0.9638 ±0.02016	0.9667 ±0.02005

3.4 Discussion

The goal of this chapter was to investigate which segmentation technique would be appropriate for tracheal lumen segmentation in the bronchoscopy videos. The highest mean DSC value was achieved with Multi-Otsu thresholding in the value channel. Nonetheless, the results for all techniques were all quite similar. Therefore, it is not possible to claim that this combination is really the best for tracheal lumen segmentation. Edge-based techniques will not be appropriate for tracheal lumen detection because the frames are too complex.

Two relevant studies were found. The first was published in 2005 by *Masters et al.*[22]. They also used color-based tracheal lumen segmentation for objective TM assessment. Considering their results, the images look quite similar to the segmentations in this study; deep into the lumen. They used manual thresholding and did not use ground truths. Therefore, the results of the DSC can not be compared. The second was published in 2023 by *Keuth et al.*[35]. They used a Convolutional Neural Network for tracheal lumen

segmentation based on a depth map instead of a color-based technique. They also used DSC for performance evaluation, but their results are higher than 1 without understanding how they calculated this so it is not possible to compare both results with each other. Furthermore, it can be doubted if it would be beneficial to use such a complex technique while the simpler techniques used in this study results in good results. However, a monocular depth map could probably be used to force the segmentation to have the same distance relative to the bronchoscope. Normally, depth could be estimated by using two camera's. However, this is not possible with the bronchoscope. Pre-trained Deep Learning models are available for monocular depth map estimation, which could probably be used. Considering bronchoscopic (and endoscopic in general) segmentations, Deep Learning gains more attention for segmentation. For now, this is not preferred because it could be overkill. Besides, it requires a lot of data, takes a long time to train the network, and dedicated hardware is required.

There are four limitations in this study. First, a limited number of 24 images was used. More data would be recommended for the next time. Second, the ground truths were manually labeled by one person, making the results less reliable. It is recommended to let multiple persons do the labelling, also a multiple times, to improve the inter- and intra-observer reliability. Third, frames near to the carina were eliminated on purpose. Therefore, the results are not generalisable for the whole trachea. Fourth, the segmentations equal to the mask processing circle are difficult to reproduce exactly the same every time. Hence, the DSC results could be different when the same frames are used multiple times.

It was not expected that thresholding and cluster techniques would perform this well. Thresholding and cluster-based segmentation techniques[30][31][36] are relatively easy and fast, but are susceptible to differences in color values between the image frames and patients, and overlapping pixel intensities of ROI's. The pixel intensity difference between frames and patients was overcome by using automatically determined thresholds and clusters. The pixel intensity overlap between ROI's was minimized by the used image channels, based on the conclusion of *Chapter 2*. All this contributed to the promising results.

As mentioned before, it is not possible to state that Multi-Otsu thresholding in the value channel is really the best to use. However, there was one disadvantage in this study related to K-means clustering which supports the advice to focus on threshold-based segmentation techniques. The cluster corresponding to the tracheal lumen was manually selected each time because the corresponding number belonging to this cluster varied among patients. This manual selection will not be optimal and is time consuming, especially when it is desired to analyse the whole trachea.

Three problems were noticed in the segmentations. First, some segmentations were just a circular shape. The lumen in these frames was still in connection with the FOI border before the mask post-processing was used. Therefore, the circular segmentation shape was just the circular area used for mask processing. A typical aspect of these frames was a lumen close to the border of the frame. It was checked what the result would be when multiple frames before and after the extracted frame were used. The circular segmentation was gone when the same lumen was captured somewhere in the middle of the frame. However, the minimal distance between the lumen boundary and the FOI boundary was not established. Therefore, it would be recommended to set the requirement that the tracheal lumen should be somewhere around the middle of the frame. Second, frames near to the carina were not usable because this resulted in segmentation of one

of the main bronchi. Third, not all segmentations were in a perpendicular plane relative to the camera, considering a 3D space. This phenomenon could probably be assigned to two aspects. The first aspect is the bronchoscope position. The trachea is larger than the bronchoscope, so there can be variation in position and angles. Additionally, the rigid bronchoscope barrel appeared sometimes curved, as seen in the frames with a metal structure around the FOI border. As a result, the light source was not in the middle of the trachea, resulting in an unequal illumination of the left/right and posterior/anterior side of the tracheal lumen. This could be overcome during the procedure by keeping the bronchoscope somewhere in the middle and paying attention to equal illumination. However, this can not be adjusted in the available data. Therefore, this could be relevant to keep in mind for the future. The second aspect is related to the use of the largest object boundary as segmentation. In single-channel images, like those used in *Chapter 2*, it was visually observed that the intercartilaginous tissue sometimes had another pixel value than the tracheal lumen. In the segmentations of this chapter, it was observed that the segmentation boundary was at the transition of the cartilage ring to the intercartilaginous tissue. Therefore, it could be possible that the missing part of the segmentations was not connected to the tracheal lumen (the largest object boundary), so, not included in the segmentation.

The similarity in DSC values could probably be explained by the results from *Chapter 2*; The pixel intensity distributions for the red- and value channels were quite similar, as well as for the gray channel, but in another range. These distributions were utilized in the segmentation techniques for determining the thresholds and clusters. This could probably explain the similarity in DSC values. Some frames had a lower DSC value in all three channels, which could probably be assigned to an inaccurate ground truth.

3.5 Conclusion

Using Multi-Otsu thresholding in the value channel resulted in the highest DSC value (DSC = 0.9681). However, the DSC values were all quite similar. Therefore, it is not possible to claim that this combination is really the best. The subquestion addressed in this chapter is answered, based on limited evidence.

4

Scoring system development

In the previous chapter was concluded that Multi-Otsu thresholding in the value channel resulted in the highest DSC value. In this chapter, the application of the segmentations for an objective tracheomalacia severity assessment will be explored. Two measurements that could potentially be useful (the area and the anteroposterior : transverse diameter ratio) will be calculated based on the segmentations. These measurement outcomes will be used as input for a linear regression model to explore the feasibility of developing a scoring system.

4.1 Introduction

After extracting the tracheal lumen, measurements can be calculated for objective collapse assessment. Two measurement are expected to be meaningful: the tracheal lumen area, and the anteroposterior : transverse (APT) diameter ratio of the tracheal lumen, see *Figure 16*. The area is clinically used, but not calculated. The disadvantage of using the area as objective measurement is that distances relative to the camera should be involved. This is not possible in the bronchoscopy videos. If the bronchoscope remains in the same position during the end-inspiratory and end-expiratory phases, it could be possible to analyze the CSA (cross-sectional area) change along the respiratory cycle using *Formula 1*. However, this is not a measure for TM unless it can be related to the CSA change in a non-malacia tracheal segment. It is assumed that the TLC (tracheal lumen collapse) is normal (no collapse) directly below the vocal cords. The CSA_{insp} and CSA_{exp} should be calculated at two crucial points in the trachea: just below the vocal cords and at the malacia part. However, it is currently unknown how these should be related to each other.

The second potential, and distance-independent measure, could be the APT diameter ratio[1][13]. This ratio could provide insight into the diameter change along the respiratory cycle. It is hypothesized that the real transverse diameter does not change along the respiratory cycle because this diameter relates to the (rigid) cartilage ring. However, the shape of the tracheal lumen is usually asymmetrical. An average of the diameters, measured at various points in the segmentation, should be used for both the anteroposterior and transverse diameter.

This chapter aims to answer the third subquestion of this thesis: *What measurement (a calculation based on the segmentation) is usable for determining the percentage of tracheal lumen collapse?* In order to translate the measurement outcomes into a clinical assessment, the relationship between outcomes and real TLC must be established. Otherwise, it is not known how a calculated measurement should be interpret. To address this relationship, linear regression[37][38] will be used.

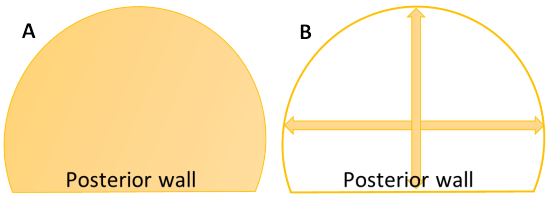

Figure 16: Visualization of the proposed objective measurements.
A: Cross-sectional area;
B: Anteroposterior : transverse diameter.

4.2 Materials & Method

4.2.1 Materials

19 bronchoscopy videos from neonates with EA were used. However, 36 neonates were included, but not all had an available bronchoscopy video or they lacked a determined percentage of TLC in the surgical report. These patients were excluded. All patients received treatment in the WKZ in 2022 and 2023, and their ages ranges between 2 and 10 days old. The algorithm used for segmentation and measurement calculation was developed with Python as programming language. IBM SPSS statistics (version 28.0) was used for linear regression analysis.

4.2.2 Method

Based on the results of *Chapter 3*, it was decided to use Multi-Otsu thresholding in the value channel as segmentation technique in this chapter.

Data extraction From each bronchoscopy video, four frames were extracted; An end-inspiratory and end-expiratory frame from both the malacia and non-malacia segment of the trachea. This is similar to *Paragraph 3.2.2*. The clinically determined TLC was extracted from the surgery report.

Algorithm development An algorithm was build and rendered for each frame individual. The frame was loaded and cropped to the FOI, as described in *Paragraph 2.2.2*. Next, the segmentation was performed using the algorithm developed in *Chapter 3*. Most of the segmentations were not aligned to the horizon due to bronchoscope rotation along the video. It was necessary to do a horizontal alignment, otherwise it would be difficult to calculate the diameters of the segmentation. Two points were interactively defined on the resulted segmentation mask: the left and right corners of the cartilage ring/posterior wall transition, see *Figure 17A*. A vector was defined through these points. The angle between the horizon and this vector was used to rotate the segmentation and align it with the horizon, see *Figure 17B*. As mentioned in *Chapter 3*, not all segmentation were in a perpendicular plane relative to the bronchoscope. This should be compensated to ensure an accurate measurement of the diameters. It was decided to use manual perspective correction over an automatic method because it was expected that compensation for depth distance would be challenging via object movement tracking across multiple frames. A transformation matrix was calculated based on interactive corner selection of a plane in the horizontal aligned segmentation, see *Figure 17C*. This matrix was applied to the segmentation, see *Figure 17D*, and followed by binary thresholding. The area, and APT-ratio were calculated based on this resulting segmentation.

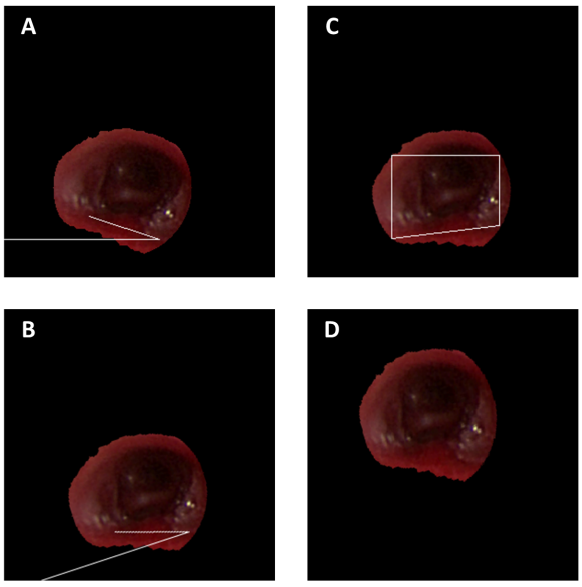

Figure 17: Segmentation position correction.
A: One vector is drawn between the left and right corners of the cartilage ring/posterior wall transition, and one vector equal to the horizon;
B: Horizontal segmentation alignment;
C: Interactive corner selection for a plane in the rotated segmentation;
D: Result after applying the transformation matrix.

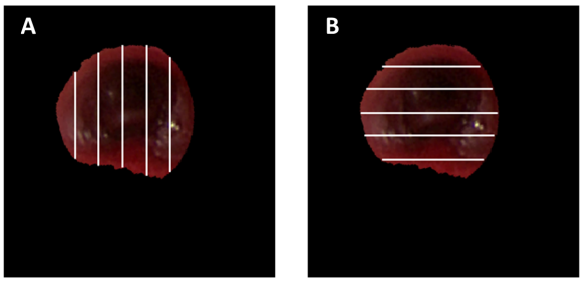

Figure 18: Points where the diameters were automatically measured.
A: Anteroposterior diameters;
B: Transverse diameters.

Measurement calculations The area was calculated by summation of the number of white pixels in the binary image. The diameters were automatically calculated. Therefore, the minimum and maximum x- and y boundary values were extracted from the binary segmentation. Based on these ranges, the diameters were automatically calculated at five points (1/6th, 2/6th, 3/6th, 4/6th, and 5/6th of the ranges), as shown in *Figure 18*. The mean diameter for each direction was calculated, followed by the APT-ratio.

Scoring system Five different formulas were explored to assess how the measurements could be used. The APT-ratios per frame were used as input. One paper, published by *Hysinger et al.*[1] was found were they used the area to calculate the CSA change along the respiratory cycle, see *Formula 1*. This formula takes the correlation between the malacia and non-malacia segment of the trachea not into account. Therefore, self-established formula's were also explored since relevant literature was not found. In general, two types of formula's were established: those comparing the malacia versus the non-malacia segment (see *Formula's 3, 4 and 5*), and those comparing end-inspiratory versus end-expiratory (see *Formula's 6 and 7*). A scatterplot was made for the outcomes of each formula, where the calculated outcomes (independent variable) were plotted against the estimated TLC (dependent variable). Furthermore, linear regression was performed, using ordinary least squares for the fitting. The following descriptive statistics are used for evaluation: the coefficient of determination (R^2), the root mean squared error, and a p-value of 0.05.

The first model corresponds to *Formula 3*. The same formula as used by *Hysinger et al.*[1] is used to calculate the APT change ratio along the respiratory cycle for the malacia and non-malacia segment individually. A ratio between these two values is calculated.

$$\text{TLC} = \frac{\frac{\text{APT M}_{insp} - \text{APT M}_{exp}}{\text{APT M}_{insp}}}{\frac{\text{APT N}_{insp} - \text{APT N}_{exp}}{\text{APT N}_{insp}}} \tag{3}$$

The second model corresponds to *Formula 4*. The APT-ratio difference along the respiratory cycle is calculated for the malacia and non-malacia segment individually. This distinguishes model 2 from model 1 because now the actual value change is used instead of a ratio change itself, and the impact of the end-inspiratory APT value (the denominator in model 1) is now limited. A ratio between these two actual value changes is calculated which should be interpret as how many times smaller the change along the

respiratory cycle is for the malacia segment relative to the non-malacia segment.

$$\text{TLC} = \frac{\text{APT M}_{insp} \text{ - APT M}_{exp}}{\text{APT N}_{insp} \text{ - APT N}_{exp}} \tag{4}$$

The third model corresponds to *Formula 5*. This formula is based on the principle of proportion. Basically, the concept is 'part divided by whole', so it should be interpret as how many times smaller the end-expiratory APT-ratio is relative to the end-inspiratory APT ratio. The APT-ratio proportion along the respiratory cycle is calculated for the malacia and non-malacia segment individually. An overall ratio is calculated which should be interpret as how many times smaller the APT-ratio change along the respiratory cycle is for the malacia segment relative to the non-malacia segment. In both segments, the end-inspiratory APT-ratio was considered as the 'whole' value because it corresponds to the maximum tracheal lumen dimension at that point.

$$\text{TLC} = \frac{\frac{\text{APT M}_{exp}}{\text{APT M}_{insp}}}{\frac{\text{APT N}_{exp}}{\text{APT N}_{insp}}} \tag{5}$$

The fourth model corresponds to *Formula 6*. This formula is similar to *Formula 4*, but now it considers end-inspiratory versus end-expiratory instead of malacia versus non-malacia segment. The APT-ratio difference between the malacia and non-malacia segment is calculated for end-inspiratory and -expiratory individually. A ratio between these two actual value changes is calculated which should be interpret as how many times smaller the end-expiratory change between the two tracheal segments is relative to the end-inspiratory change.

$$\text{TLC} = \frac{\text{APT N}_{exp} \text{ - APT M}_{exp}}{\text{APT N}_{insp} \text{ - APT M}_{insp}} \tag{6}$$

The fifth model corresponds to *Formula 7*. This formula is similar to *Formula 5*, but now it considers end-inspiratory versus end-expiratory instead of malacia versus non-malacia segment. The APT-ratio proportion between the malacia and non-malacia is calculated for the end-inspiratory and -expiratory individually. An overall ratio is calculated which should be interpret as how many times smaller the end-expiratory change between both segments is relative to the end-inspiratory change between both segments. In both respiratory cycles phases, the non-malacia APT-ratio was considered as the 'whole' value because it corresponds to the maximum tracheal lumen dimension for that specific respiratory cycle phase.

$$\text{TLC} = \frac{\frac{\text{APT M}_{exp}}{\text{APT N}_{exp}}}{\frac{\text{APT M}_{insp}}{\text{APT N}_{insp}}} \tag{7}$$

4.3 Results

In *Figure 19*, the scatterplots including a linear regression curve (red line) are plotted. Corresponding equation is also visualised in each scatterplot. Each blue dot corresponds to one patient. The x-axis represents the calculated outcome, based on each model formula. The y-axis represents the clinically determined TLC. In *Table 2*, corresponding statistics can be found.

Model 1 A positive relationship between the calculated value and the amount of tracheal lumen collapse is observed in model 1. The calculated value has an overall range between -1.0 and 1.9, with an outlier at x=3.7 It is not possible to cluster two groups, considering the cut-off value (33%) for determining the necessity of treatment. This observation holds for other percentages as well. Following descriptive statistics are observed: R^2=0.046, RMSE=22.12, and p=0.365, indicating statistically insignificance. The outlier had a high APT-ratio change along the respiratory cycle of the malacia segment, and not for the non-malacia segment. However, the impact of this outlier is limited, as its absences does not substantially improve the overall results.

Model 2 A positive relationship between the calculated value and the amount of tracheal lumen collapse is observed in model 2. The calculated value has an overall range between -1.0 and 1.9, with an outlier at x=3.3. It is not possible to cluster two groups, considering the cut-off value (33%) for determining the necessity of treatment. This observation holds for other percentages as well. Following descriptive statistics are observed: R^2=0.064, RMSE=21.91, and p=0.283, indicating statistically insignificance. The outlier is the same as in model 1, having a similar reason and impact.

Model 3 A positive relationship between the calculated value and the amount of tracheal lumen collapse is observed in model 3. The calculated value has an overall range between 0.004 and 0.065, with an outlier at x=0.1. It is not possible to cluster two groups, considering the cut-off value (33%) for determining the necessity of treatment. This observation holds for other percentages as well. Following descriptive statistics are observed: R^2=0.003, RMSE=22.61, and p=0.819, indicating statistically insignificance. The outlier had a small APT-ratio change along the non-malacia segment, which has more influence in this formula than for models 1 and 2. However, the impact of this outlier is limited, as its absences does not substantially improve the overall results.

Model 4 A negative relationship between the calculated value and the amount of tracheal lumen collapse is observed in model 4. The calculated value has an overall range between -2.9 and 5.9, with an outlier at x=72.4. It is not possible to cluster two groups, considering the cut-off value (33%) for determining the necessity of treatment. This observation holds for other percentages as well. Following descriptive statistics are observed: R^2=0.007, RMSE=22.56, and p=0.718, indicating statistically insignificance. The outlier is the same patient as in model 3. This patient had a large difference between the end-expiratory APT-ratio of the non-malacia and malacia segments, which was not observed in the end-inspiratory phase. This discrepancy results in this high value.

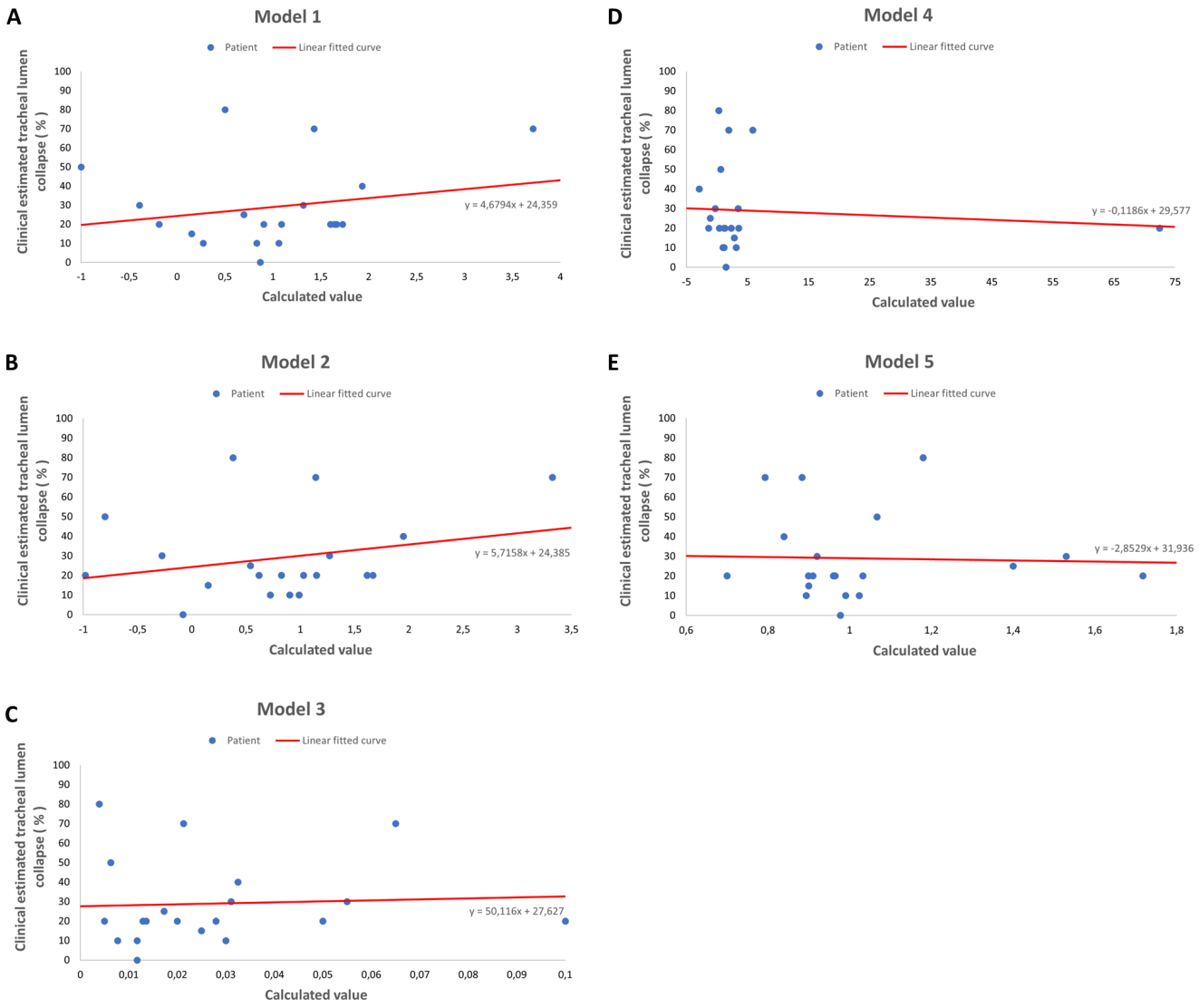

Figure 19: Scatterplots and linear regression curve (red line) for each explored model. Each blue dot corresponds to a patient. Each x-axis represents the calculated outcome, based on each model formula. Each y-axis represents the tracheal lumen collapse during the intervention.

A: Model 1 | TLC $= \frac{\frac{\text{APT M}_{insp} - \text{APT M}_{exp}}{\text{APT M}_{insp}}}{\frac{\text{APT N}_{insp} - \text{APT N}_{exp}}{\text{APT N}_{insp}}}$

B: Model 2 | TLC $= \frac{\text{APT M}_{insp} - \text{APT M}_{exp}}{\text{APT N}_{insp} - \text{APT N}_{exp}}$

C: Model 3 | TLC $= \frac{\frac{\text{APT M}_{exp}}{\text{APT M}_{insp}}}{\frac{\text{APT N}_{exp}}{\text{APT N}_{insp}}}$

D: Model 4 | TLC $= \frac{\text{APT N}_{exp} - \text{APT M}_{exp}}{\text{APT N}_{insp} - \text{APT M}_{insp}}$

E: Model 5 | TLC $= \frac{\frac{\text{APT M}_{exp}}{\text{APT N}_{exp}}}{\frac{\text{APT M}_{insp}}{\text{APT N}_{insp}}}$

Model 5 A negative relationship between the calculated value and the amount of tracheal lumen collapse is observed in model 5. The calculated value has an overall range between 0.7 and 1.7, without an outlier. It is not possible to cluster two groups, considering the cut-off value (33%) for determining the necessity of treatment. This observation holds for other percentages as well. Following descriptive statistics are observed: R^2=0.002, RMSE=22.62, and p=0.840, indicating statistically insignificance.

Table 2: Descriptive statistics corresponding to the linear regression in the scatterplots (n = 19).

Descriptive statistic	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Coefficient of determination (R^2)	0.046	0.064	0.003	0.007	0.002
Root mean squared error (RMSE)	22.12	21.91	22.61	22.56	22.62
***p*-value**	0.365	0.283	0.819	0.718	0.840

4.4 Discussion

The goal of this chapter was to translate a measurement outcome into a clinically usable scoring system for TLC assessment. The APT-ratio seems not to be the right measurement because none of the developed models is usable, considering the p-value. General differentiation between the requirement for a tracheopexy or not (considering the cut-off value of 33%) is also not possible based on the results. This supports also the conclusion that the APT-ratio is not useable.

It is not possible to correlate the results with other studies because similar studies were not found. One relevant paper was found, published by *Hysinger et al.*[1]. They calculated the CSA change for the malacia and normal segment separately without correlating them with each other.

There are eight limitation in this study. First, the end-inspiratory and -expiratory frames were manually selected with guidance of the changing point in posterior wall movement direction. However, it was sometimes difficult to check which frame was really end-inspiratory and -expiratory. It would be recommended to calculate the measurement along a video, instead of just one frame. Based on the biggest and largest value, it should be possible to extract which frame is the end-inspiratory and -expiratory frame. Second, the clinically estimated TLC was subjective. Some patients had a dubious estimated TLC value. For example, the trachea was almost fully collapsed but it was graded as 20% collapse. It is doubted if this value was correct. It would be recommended to let multiple people do the TLC estimation, also multiple times. This minimizes the inter- and intra-observer variability. Third, it was assumed that just below the vocal cords would be a normal collapse. However, there could be patients with a malacia in this trachea segment. Fourth, a small sample size was used making the results not generalizable. More patients would be recommended for the next time. Fifth, not all TLC percentages were represented. Data of the patients treated in 2022 and 2023 was used to get a diverse population. However, this was not successful. Sixth, the manual perspective correction varied each time which makes it unreliable. It was explored if manual correction would be better than not using any perspective correction. The APT-ratio was determined for multiple (following) frames of the same cartilage ring. It was observed that manual correction resulted in an

APT-ratio which better matched the ratio calculated in a perpendicular view at the cartilage ring. However, repeatedly performing this for the same segmentation resulted in different values. Seventh, it is expected that the inaccuracy is increased relative to the original segmentation due to all manual segmentation adaptions like horizontal alignment and the perspective correction. Last, the lens distortion was ignored in this study. It is recommended to consider if this aspect would be relevant or not since this distortion could be the same in all frames.

For further research, it is recommended to focus on how the area could be used as measurement. A monocular depth map could be considered, which is already discussed in *Paragraph 3.4*. It is hypothesized that the area is a better measurement. It might be possible to correlate multiple segmentation area's when a correction factor for the segmentation depth could be established. This could also help overcome the limitation of the subjective clinically determined TLC used for linear regression.

It was hypothesized that the transverse diameter would not change along the trachea. However, this is not true. This variation along the trachea supports also the conclusion that the APT-ratio is not a usable measurement. Furthermore, it was hypothesized that the anteroposterior diameter would be smaller than the transverse diameter. This is also not true. Some patients had a larger anteroposterior than transverse diameter which affects the APT-ratio because it is a fraction. A wide range of calculated values was observed, depending on whether the numerator or denominator was the largest value.

It can be doubted if linear regression was the best choice. However, it is hypothesized that using another type of regression would not result in a generalized model due to the small sample size. Outliers in the y-direction were dismissed, as it is likely the result of using an inappropriate measurement for tracheal lumen collapse.

4.5 Conclusion

The anteroposterior : transverse diameter ratio seems not to be a useful measurement for determining the percentage of tracheal lumen collapse. Further research should focus on how the area could be used as measurement. The subquestion addressed in this chapter is not answered.

5

Discussion

In *Chapter 2*, color-space analysis was used to identify the ROI suitable for segmentation from bronchoscopy frames. In *Chapter 3*, various image segmentation techniques were explored and evaluated. In *Chapter 4*, the feasibility of developing a scoring system was explored. This chapter aims to answer the main research question of this thesis: *How can the percentage of tracheal lumen collapse in neonates born with esophageal atresia objectively be assessed, using bronchoscopy videos?* Additionally, the relevance of this thesis will be discussed.

It is possible to segment the tracheal lumen with straightforward segmentation techniques. However, this thesis showed limited evidence that using Multi-Otsu thresholding in the value channel could best be used. The APT-ratio seems not to be a useful measurement as none of the used formula's resulted in a statistically significant linear regression analysis in *Chaper 4*.

There were six main limitations in this thesis. First, multiple manual steps were involved. The ROI's for the pixel value histograms and frames were manual selected, and manual segmentation adaptations were performed which could result in an increase the inaccuracy. Second, limited data was used, resulting in limited evidence. The population of neonates born with EA is small, and an online database containing bronchoscopy video's was not found. Third, only the subjective estimated TLC percentage was available. For further research, it is important to let multiple people do the TLC estimation, also a multiple times. However, this will still be subjective estimations. Fourth, the linear regression was not assessed using the ground truths, which would result in more evidence that the APT-ratio is not usable. Fifth, some neonates were excluded due to insufficient information. The population is small so it is crucial to obtain the missing information to ensure inclusion of these patients. Last, the algorithm was not applicable to all data. This

must be overcome to analyse the whole trachea.

The mortality and patient outcomes of neonates born with EA are improved during the years.[39] But still, there is no generally accepted grading system for tracheomalacia in these patients. More profit can be made to further improve the patients outcome. It is necessary that a patient-specific decision can be made. This requires the understanding of the crucial factors in clinical decision making. Nowadays, crucial decisions on the necessity of a posterior tracheopexy still rely on subjective estimations. This is not a problem in case of severe or hardly any TLC. However, there are also neonates with moderate collapse. Specifically these will benefit from an accurately quantified percentage. TLC quantification is crucial since the decision is based on a cut-off value, where 1% difference could be decisive in deciding the necessity of a tracheopexy. The WKZ aims to do a primary tracheopexy during the initial EA surgery. However, the effect of this primary intervention is unknown. To address this, the department will conduct a randomised trial. Maybe, 33% is not the right cut-off value, or other factors should also be considered to identify the neonates benefiting from a tracheopexy. Addressing such questions requires enhanced diagnostics accuracy regarding the percentage. This thesis contributes to that specific part because it presents the initial steps towards an objective quantification. When understanding deepens regarding all contributing factors, the potential emerges to develop a scoring system which supports the medical professionals in making patient-specific treatment decisions. This is advantageous for the neonate as it prevents a second surgery which could be challenging due to adhesions owing to the previous surgery. Besides, it could prevent the worse consequences of 'brief resolved unexplained events', like feeding difficulties and an extended hospital stay. For some, it will prevent an unnecessary tracheopexy.

As discussed in *Paragraph 1.9*, existing literature has tried to quantify the TM severity, primarily focusing on adults rather than neonates. These studies[1][13][20][21] used imaging techniques which are not preferred for neonates for various reasons. CT-scans pose risks to neonates due to radiation exposure, and it lacks the ability to visualise the dynamic posterior wall movement. MRI is challenging in neonates. For example sedation is required, they are connected to monitors, and they have a replogle tube in situ (a tube for continuous secretion drainage from the upper esophageal part, to prevent aspiration into the respiratory tract). Furthermore, it is also not possible to visualise the dynamic posterior wall movement. Rigid bronchoscopy is the gold standard, so more research is necessary on how this could be used for objective assessment instead of other diagnostics.

This thesis relied on trial-and-error due to limited literature, but it resulted in new insights. The simpler segmentation techniques used in this thesis performed well for tracheal lumen segmentation, so Deep Learning seems to be overkill for the segmentation itself. This is advantageous since it requires a lot of data, takes a long time to train a network, and dedicated hardware like a powerful graphics processing units are necessary. The translation from segmentation towards a clinical scoring system is still unresolved in this thesis. Additionally, the relationship between end-inspiratory and -expiratory frames in the malacia and non-malacia tracheal segments is still unresolved.

From a clinical point of view, several requirements are established regarding the objective assessment. It is essential that additional elements are not necessary in the trachea during the procedure, as bronchoscopy is a risky intervention. Real-time analysis of the bronchoscopy video is not required. However, the analysis

should be completed within 10 minutes after the procedure because this determines the surgery approach. Ideally, the whole trachea should be analysed, eventually per tracheal segment. This helps in comparing the percentages during subsequent bronchoscopies, or determine the appropriate suture placement for a tracheopexy. A fully automatic algorithm is not required, as long as an interactive approach enhances objectivity. For now, it is not known whether under- or over-quantifying the TLC would be worser. The outcomes of the proposed trial could contribute to establishing the requirement regarding this. In the end, a clinical study must be performed to prove the reliability of the algorithm. When this is successful, the objective assessment tool could be integrated into the clinical workflow.

6

Conclusion

In this thesis, color-space analysis was used to investigate which ROI can be segmented from bronchoscopy frames. Various image segmentation techniques were examined and evaluated. Finally, an attempt was made to explore the feasibility of scoring system development. All aimed at addressing the following research question: *How can the percentage of tracheal lumen collapse in neonates born with esophageal atresia objectively be assessed, using bronchoscopy videos?*.

The tracheal lumen is the ROI that can be segmented from a bronchoscopy video. The highest DSC value (DSC = 0.9681) was obtained for three-threshold Multi-Otsu thresholding in the value channel. However, the examined techniques had all quite similar DSC values. It remains uncertain whether this particular combination is really the best, given the constraints of the limited data. The APT-ratio seems not to be a useful measurement because none of the proposed formula's had a statistically significant linear regression analysis. Further research should focus on how the area could be used as measurement.

Bibliography

[1] E. B. Hysinger and H. B. Panitch, "Paediatric Tracheomalacia," *Paediatric Respiratory Reviews*, vol. 17, pp. 9–15, 1 2016. [Online]. Available: https://doi.org/10.1016/J.PRRV.2015.03.002

[2] A. Kamran and R. W. Jennings, "Tracheomalacia and Tracheobronchomalacia in Pediatrics: An Overview of Evaluation, Medical Management, and Surgical Treatment," *Frontiers in Pediatrics*, vol. 7, p. 512, 12 2019. [Online]. Available: https://doi.org/10.3389/FPED.2019.00512

[3] H. J. M. Albalawi, R. D. G. Alzahrani, and R. S. D. Alshwamin et al, "Epidemiology and Management of Pediatric Tracheomalacia," *Journal of Pharmaceutical Research International*, pp. 229–235, 8 2021. [Online]. Available: https://doi.org/10.9734/JPRI/2021/V33I42A32400

[4] D. P. Drake and A. Durward, "Tracheomalacia," *Pediatric Surgery*, pp. 485–493, 2020. [Online]. Available: https://doi.org/10.1007/978-3-030-41724-6_43

[5] E. S. van Tuyll van Serooskerken, S. H. Tytgat, and J. W. Verweij et al, "Primary Posterior Tracheopexy in Esophageal Atresia Decreases Respiratory Tract Infections," *Frontiers in Pediatrics*, vol. 9, p. 932, 9 2021. [Online]. Available: https://doi.org/10.3389/fped.2021.720618

[6] N. Durkin and P. De Coppi, "Management of neonates with oesophageal atresia and tracheoesophageal fistula," *Early human development*, vol. 174, 11 2022. [Online]. Available: https://doi.org/10.1016/j.earlhumdev.2022.105681

[7] "Esophageal Atresia Symptoms & Surgery — Children's Minnesota." [Online]. Available: https://www.childrensmn.org/services/care-specialties-departments/fetal-medicine/conditions-and-services/esophageal-atresia/

[8] P. W. Furlow and D. J. Mathisen, "Surgical anatomy of the trachea," *Annals of Cardiothoracic Surgery*, vol. 7, no. 2, pp. 25 560–25 260, 3 2018. [Online]. Available: https://doi.org/10.21037/ACS.2018.03.01

[9] S. E. Wert, "Normal and Abnormal Structural Development of the Lung," *Fetal and Neonatal Physiology, 2-Volume Set*, pp. 627–641, 1 2017. [Online]. Available: https://doi.org/10.1016/b978-0-323-35214-7.00061-5

[10] P. Varela, M. Torre, and C. Schweiger et al, "Congenital tracheal malformations," *Pediatric surgery international*, vol. 34, no. 7, pp. 701–713, 7 2018. [Online]. Available: https://doi.org/10.1007/s00383-018-4291-8

[11] K. Hammond, U. K. Ghori, and A. I. Musani, "Tracheobronchomalacia and Excessive Dynamic Airway Collapse," *Clinics in Chest Medicine*, vol. 39, no. 1, pp. 223–228, 3 2018. [Online]. Available: https://doi.org/10.1016/J.CCM.2017.11.015

[12] S. H. Tytgat, M. Y. van Herwaarden-Lindeboom, and E. S. van Tuyll van Serooskerken et al, "Thoracoscopic posterior tracheopexy during primary esophageal atresia repair: a new approach to prevent tracheomalacia complications," *Journal of Pediatric Surgery*, vol. 53, no. 7, pp. 1420–1423, 7 2018. [Online]. Available: https://doi.org/10.1016/J.JPEDSURG.2018.04.024

[13] S. Ebrahimian, S. R. Digumarthy, and B. C. Bizzo et al, "Automatic segmentation and measurement of tracheal collapsibility in tracheomalacia," *Clinical Imaging*, vol. 95, pp. 47–51, 3 2023. [Online]. Available: https://doi.org/10.1016/J.CLINIMAG.2022.11.020

[14] "Tracheomalacia / Bronchomalacia information - Airway Unit, Service of Otorhino-laryngology - CHUV." [Online]. Available: https://www.chuv.ch/en/voies-aeriennes/orva-home/patients-and-families/conditions-we-treat/tracheomalacia-bronchomalacia-information

[15] C. Wallis, E. Alexopoulou, and J. L. Antón-Pacheco et al, "ERS statement on tracheomalacia and bronchomalacia in children," *European Respiratory Journal*, vol. 54, no. 3, 9 2019. [Online]. Available: https://doi.org/10.1183/13993003.00382-2019

[16] C. Pelaia, A. Bruni, and E. Garofalo et al, "Oxygenation strategies during flexible bronchoscopy: a review of the literature," *Respiratory Research*, vol. 22, no. 1, p. 253, 12 2021. [Online]. Available: https://doi.org/10.1186%2Fs12931-021-01846-1

[17] S. F. Polites, M. Kotagal, and L. J. Wilcox et al, "Thoracoscopic posterior tracheopexy for tracheomalacia: A minimally invasive technique," *Journal of Pediatric Surgery*, vol. 53, no. 11, pp. 2357–2360, 11 2018. [Online]. Available: https://doi.org/10.1016/J.JPEDSURG.2018.08.004

[18] S. H. Tytgat, E. S. van Tuyll van Serooskerken, and J. W. Verweij et al, "Thoracoscopic posterior tracheopexy during primary esophageal atresia repair." [Online]. Available: https://doi.org/10.1016/j.jpedsurg.2018.04.024

[19] H. Dreyer, E. van Tuyll van Serooskerken, and L. Rodenburg et al., "Airway Epithelial Cultures of Children with Esophageal Atresia as a Model to Study Respiratory Tract Disorders," *Children (Basel, Switzerland)*, vol. 10, no. 6, 6 2023. [Online]. Available: https://doi.org/10.3390/children10061020

[20] K. Douros, G. Kremmydas, and V. Grammeniatis et al, "Helical multi-detector CT scan as a tool for diagnosing tracheomalacia in children," *Pediatric Pulmonology*, vol. 54, no. 1, pp. 47–52, 1 2019. [Online]. Available: https://doi.org/10.1002/ppul.24188

[21] P. Ciet, P. Wielopolski, and R. Manniesing et al, "Spirometer-controlled cine magnetic resonance imaging used to diagnose tracheobronchomalacia in paediatric patients," *The European respiratory journal*, vol. 43, no. 1, pp. 115–124, 1 2014. [Online]. Available: https://doi.org/10.1183/09031936.00104512

[22] I. B. Masters, M. M. Eastburn, and R. Wootton et al, "A new method for objective identification and measurement of airway lumen in paediatric flexible videobronchoscopy," *Thorax*, vol. 60, pp. 652–658, 2005. [Online]. Available: https://doi.org/10.1136/thx.2004.034421

[23] D. Hema and S. Kannan, "Interactive Color Image Segmentation using HSV Color Space," *Science and Technology Journal*, vol. 7, p. 1. [Online]. Available: http://doi.org/10.22232/stj.2019.07.01.05

[24] "OpenCV: Color conversions." [Online]. Available: https://docs.opencv.org/3.4/de/d25/imgproc_color_conversions.html#color_convert_rgb_hsv

[25] Z. N. Khudhair, A. N. Khdiar, and N. K. El Abbadi et al., "Color to Grayscale Image Conversion Based on Singular Value Decomposition," *IEEE Access*, vol. 11, pp. 54 629–54 638, 2023. [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3279734

[26] F. Garcia-Lamont, J. Cervantes, and A. López et al., "Segmentation of images by color features: A survey," *Neurocomputing*, vol. 292, pp. 1–27, 5 2018. [Online]. Available: https://doi.org/10.1016/j.neucom.2018.01.091

[27] J. Chai, H. Zeng, and A. Li et al, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Machine Learning with Applications*, vol. 6, p. 100134, 12 2021. [Online]. Available: https://doi.org/10.1016/J.MLWA.2021.100134

[28] N. O. Mahony, S. Campbell, and A. Carvalho et al, "Deep Learning vs. Traditional Computer Vision," vol. 943, 10 2019. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-17795-9

[29] J. Olveres, G. González, and F. Torres et al, "What is new in computer vision and artificial intelligence in medical image analysis applications," *Quantitative Imaging in Medicine and Surgery*, vol. 11, no. 8, p. 3830, 8 2021. [Online]. Available: https://doi.org/10.21037%2Fqims-20-1151

[30] Y. Yu, C. Wang, and Q. Fu et al, "Techniques and Challenges of Image Segmentation: A Review," *Electronics 2023, Vol. 12, Page 1199*, vol. 12, no. 5, p. 1199, 3 2023. [Online]. Available: https://doi.org/10.3390/electronics12051199

[31] J. Qi and H. Yang, "Research on image segmentation and edge detection technology based on computer vision," *Journal of Physics: Conference Series*, vol. 1994, no. 1, p. 012035, 8 2021. [Online]. Available: https://doi.org/10.1088/1742-6596/1994/1/012035

[32] S. Sahu, H. Sarma, and D. Jyoti Bora, "Image Segmentation and its Different Techniques: An In-Depth Analysis," *Proceedings of the 2018 3rd IEEE International Conference on Research in Intelligent and Computing in Engineering, RICE 2018*, 10 2018. [Online]. Available: https://doi.org/10.1109/RICE.2018.8509038

[33] S. Yuheng and Y. Hao, "Image Segmentation Algorithms Overview," 7 2017. [Online]. Available: https://doi.org/10.48550/arXiv.1707.02051

[34] A. Reinke, M. D. Tizabi, and M. Baumgartner et al, "Understanding metric-related pitfalls in image analysis validation," 2 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2302.01790

[35] R. Keuth, M. Heinrich, and M. Eichenlaub et al., "Weakly Supervised Airway Orifice Segmentation in Video Bronchoscopy," p. 9, 8 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2208.11468

[36] M. R. Hassan, R. R. Ema, and T. Islam, "Color Image Segmentation using Automated K-Means clustering with RGB and HSV Color Spaces," *Global Journal of Computer Science and Technology*, vol. 17, no. F2, pp. 33–41, 5 2017. [Online]. Available: https://computerresearch.org/index.php/computer/article/view/1587

[37] S. Bzovsky, M. R. Phillips, and R. H. Guymer et al, "The clinician's guide to interpreting a regression analysis," *Eye 2022 36:9*, vol. 36, no. 9, pp. 1715–1717, 1 2022. [Online]. Available: https://doi.org/10.1038/s41433-022-01949-z

[38] S. W. Lee, "Regression analysis for continuous independent variables in medical research: statistical standard and guideline of Life Cycle Committee," *Life Cycle*, vol. 2, 2 2022. [Online]. Available: https://doi.org/10.54724/lc.2022.e3

[39] G. Keefe, K. Culbreath, and E. M. Edwards et al., "Current outcomes of infants with esophageal atresia and tracheoesophageal fistula: A multicenter analysis," *Journal of Pediatric Surgery*, vol. 57, no. 6, pp. 970–974, 6 2022. [Online]. Available: https://doi.org/10.1016/j.jpedsurg.2022.01.060

A

Segmentations

In this appendix, the segmentation results without the mask post-processing are visualized. Additionally, the effect of using three-threshold Multi-Otsu thresholding, and K-means clustering with K=3 and K=5 are visualized.

A.1 Otsu thresholding

In *Figure 20*, some examples using Otsu thresholding are visualized. Each column belongs to one input frame. The images in row A belong to using the red channel as input, row B to the value channel, and row C to the gray channel. In general, the tracheal lumen is present in the segmentations. It was observed that the outline of the field of interest (FOI) was still present in the segmentation, also some other (irrelevant) parts. Therefore, it was decided to use mask post-processing to remove these parts.

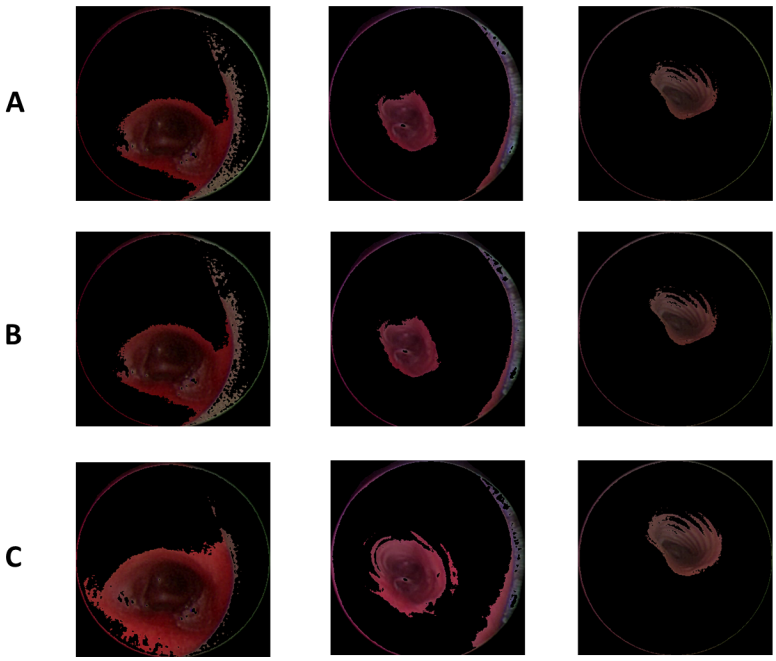

Figure 20: Some examples of using Otsu thresholding without mask post-processing. Each column belongs to one input frame.
A: Red channel;
B: Value channel;
C: Gray channel.

A.2 Multi-Otsu thresholding

A.2.1 Threshold value effect

In *Figure 21*, some examples of the segmentations using different thresholds determined with Multi-Otsu thresholding are visualized. In this case, three thresholds were used. The results are without mask post-processing and belong all to the same input frame. The images in row A belong to using the red channel, row B to the value channel, and row C to the gray channel. The first column is the segmentation using the lower threshold, the second column using the middle threshold and the third using the upper threshold. The fourth column are the segmentations using the range between the lower and middle threshold, and the last column is the range between the middle and upper threshold.

Using the low threshold resulted in too limited segmentations for all image representations. In most frames, it resulted in just a few spots deep into the tracheal lumen. The middle threshold resulted in segmentation of the lumen itself. Therefore, this method was considered as potential. However, for the gray channel it was in general too broad. The high threshold resulted in segmentations outside the tracheal lumen, making this threshold inappropriate at all. The lower range is the difference between the lower and middle threshold.

The outline tracheal lumen is segmented. However, the outline is the same as for the middle threshold. Therefore, there seems not to be an added value for this range because the lower threshold was also useless. The upper range is the difference between the high and middle threshold. This range is also not useful because the outline was far too broad for the segmentations. Considering all, it was decided to focus only on the middle threshold value.

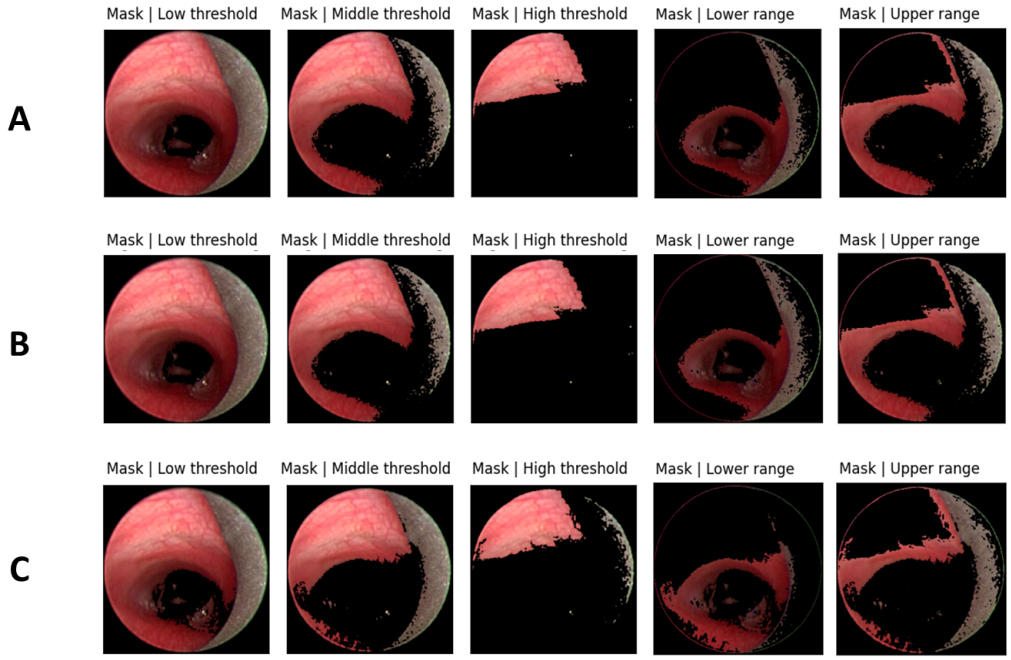

Figure 21: The effect of using each of the three thresholds and corresponding ranges, based on Multi-Otsu thresholding. All results belong to the same input frame.
A: Red channel;
B: Value channel;
C: Gray channel.

A.2.2 Middle threshold value

In *Figure 22*, some examples of using the middle threshold determined with Multi-Otsu thresholding are visualized. It has to be noticed that the inverse of the segmentation mask is used, otherwise the desired area would be removed instead of segmented. Each column belongs to one input frame. The images in row A belong to using the red channel as input, row B to the value channel, and row C to the gray channel.

The tracheal lumen is present in the segmentations. It was observed that the outline of the FOI was still present in the segmentation, also some other (irrelevant) parts. Therefore, it was decided to use mask post-processing to remove these parts.

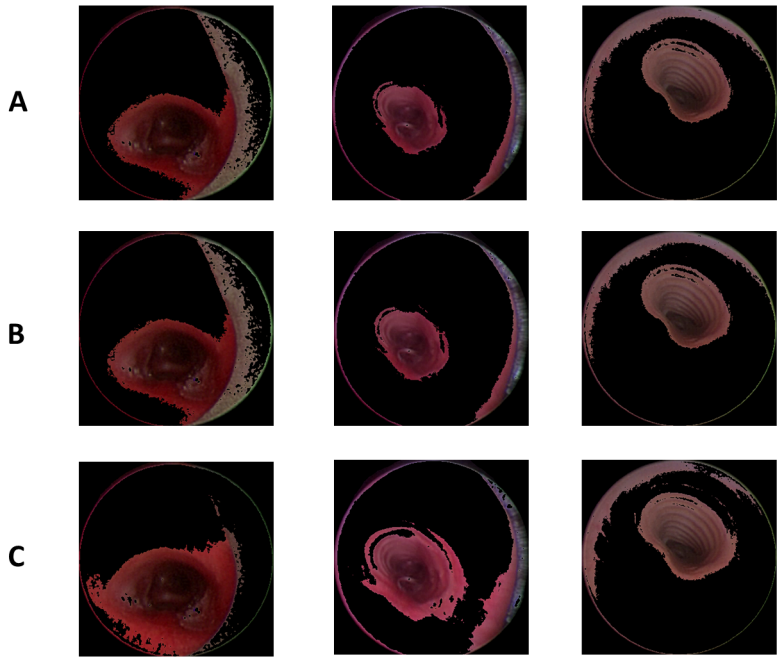

Figure 22: Some examples of using the middle value threshold, based on Multi-Otsu thresholding. These results do not include mask post-processing. Each column belongs to one input frame.
A: Red channel;
B: Value channel;
C: Gray channel.

A.3 K-means clustering

A.3.1 Number of clusters

In *Figure 23*, some examples of K-means clustering using K=3 are visualized. In *Figure 24*, some examples of K-means clustering using K=5 are visualized. In both figures, each column belongs to one input frame. The images in row A belong to using the red channel as input, row B to the value channel, and row C to the gray channel. Using five clusters seemed to be better than three clusters because the amount of irrelevant tissue was limited in the cluster for the tracheal lumen.

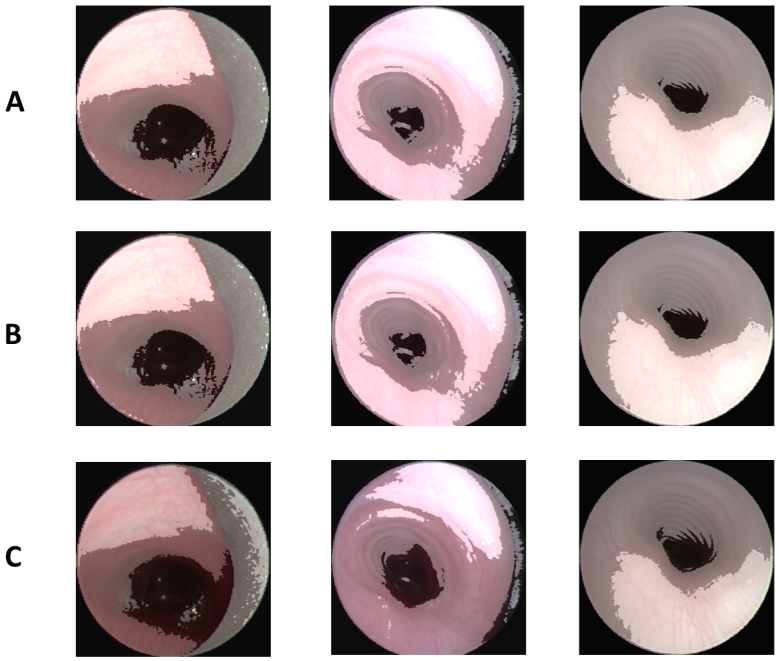

Figure 23: Some examples of using K-means clustering with K=3. Each column belongs to one input frame.
A: Red channel;
B: Value channel;
C: Gray channel.

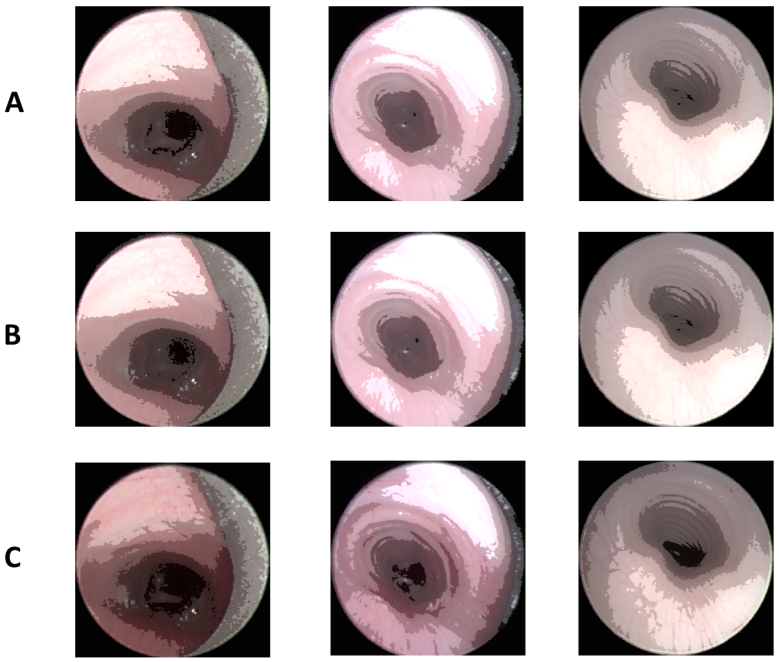

Figure 24: Some examples of using K-means clustering with K=5. Each column belongs to one input frame.
A: Red channel;
B: Value channel;
C: Gray channel.