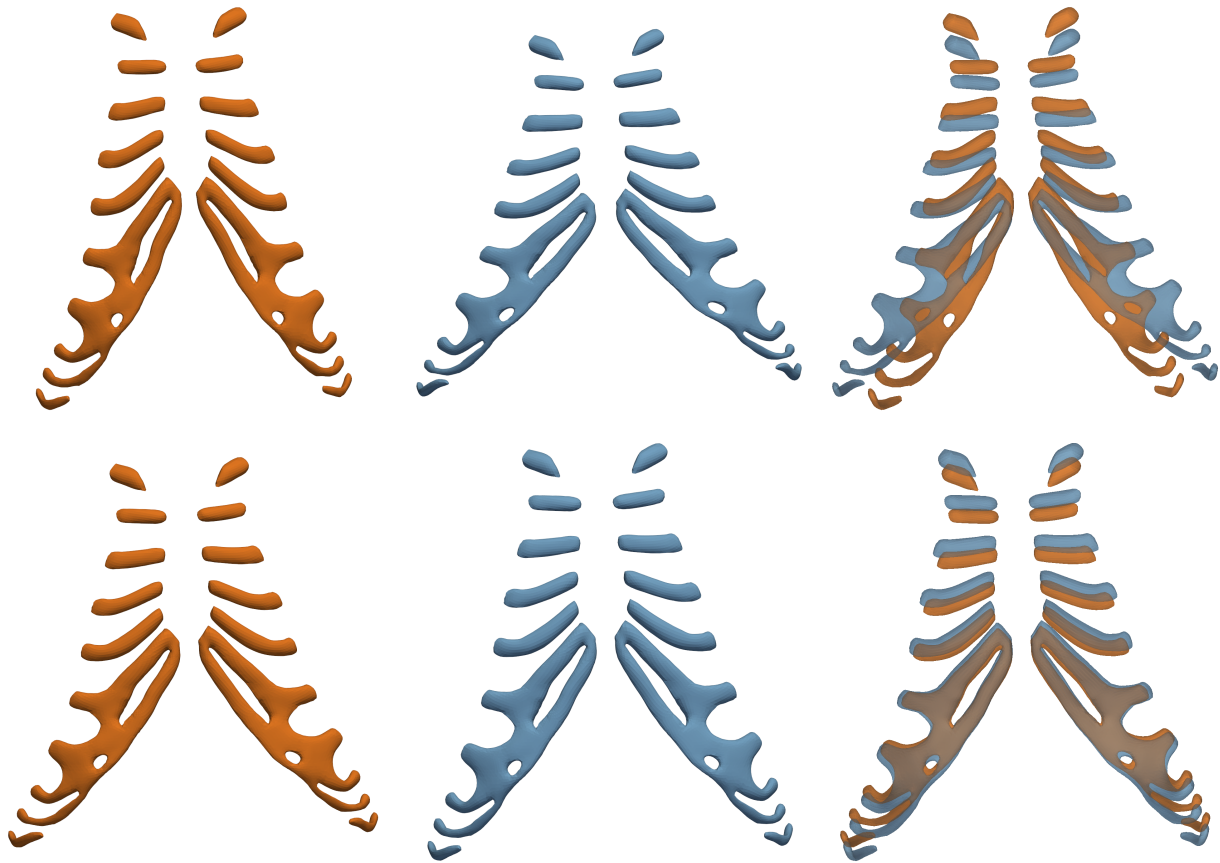




CHALMERS
UNIVERSITY OF TECHNOLOGY



Statistical Shape Model of Costal Cartilage in the Human Body

Using Skeletonization, PCA and Regression to explore population variability

Master's thesis in Complex Adaptive Systems

Emmi Olsson

DEPARTMENT OF MECHANICS AND MARITIME SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2026

www.chalmers.se

MASTER'S THESIS 2026

Statistical Shape Model of Costal Cartilage in the Human Body

Using Skeletonization, PCA and Regression to explore population variability

EMMI OLSSON



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mechanics and Maritime Sciences

Division of Vehicle Safety

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2026

Statistical Shape Model of Costal Cartilage in the Human Body
Using Skeletonization, PCA and Regression to explore population variability
EMMI OLSSON

© EMMI OLSSON, 2026.

Supervisor: Oscar Hallberg, Department of Mechanics and Maritime Sciences
Examiner: Johan Iraeus, Department of Mechanics and Maritime Sciences

Master's Thesis 2026
Department of Mechanics and Maritime Sciences
Division of Vehicle Safety
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: A visualization of variations in the costal cartilage for a female subject with varying height and BMI. The visualization was obtained using python.

Typeset in L^AT_EX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Statistical Shape Model of Costal Cartilage in the Human Body
Using Skeletonization, PCA and Regression to explore population variability
EMMI OLSSON
Department of Mechanics and Maritime Sciences
Chalmers University of Technology

Abstract

Human Body Models are digital representations of the human anatomy used for crash test injury prediction. They provide an alternative to physical crash test dummies by providing more exact injury predictions while being cheaper to employ. Vehicle crashes are the most common cause of blunt thoracic injuries. The costal cartilage connects the ribs to the sternum and injuries to the costal cartilage is believed to cause ribcage instability, pain and prolonged disability. The geometry of the costal cartilage is sparsely studied. Not only does the shape and size of the costal cartilage vary, the topology differs between individuals. In this thesis, the topology of the costal cartilage is studied using Skeletonization and graph search algorithms. Using the most common topology, a template mesh is cast to the subject geometries using Coherent Point Drift and Iterative Closest Point. The populational variance is studied using Principal Component Analysis, resulting in a Statistical Shape Model of the costal cartilage. Finally, Linear Regression is used to correlate the variations to the demographic variables Age, Sex, Height and BMI. The resulting regression model obtains an R^2 of 0.23, which is similar to the values obtained when studying equally complex geometries. The results indicate that the vertical spacing between cartilages increase with height, and that the width of the costal cartilage increases with BMI.

Keywords: Human Body Model, Statistical Shape Modeling, Skeletonization, Principal Component Analysis, Regression Analysis

Acknowledgements

I would like to thank my supervisor Oscar Hallberg for his never ending support, and my examiner Johan Iraeus for providing his input throughout the project. I want thank Oscar Denton, for sharing his expertise and answering my numerous questions. A big thank you to Autoliv, for giving me the opportunity to write this thesis. I would also like to thank my opponents, Elsa Arte and Alva Lindström, not only for their constructive feedback in regards to this project, but for being the best company a girl could ask for. Finally, I would like to thank everyone who have filled my years at Chalmers with laughs, who have shared my misery and supported me throughout. I have many dear friends and memories from these years, and for that I am forever grateful. An especially warm thank you to my good friend Lovisa Årebäck, and my mom. I could not have done this without you.

Emmi Olsson, Gothenburg, June 2026

Contents

List of Figures	xi
List of Tables	xv
1 Introduction	1
1.1 Purpose	2
1.2 Aim	2
1.3 Limitations	2
1.4 Ethical Considerations	2
2 Theory	3
2.1 Anatomy	3
2.1.1 The Costal Cartilage	3
2.2 Topological Skeletons and Skeletonization	4
2.3 Human Body Modeling	5
2.3.1 Statistical Shape Modeling	6
2.3.1.1 Landmarking	6
2.3.1.2 Morphing	6
2.3.1.3 Generalized Procrustes Alignment	6
2.3.1.4 Landmarkless Morphing	7
2.3.1.4.1 Coherent Point Drift	7
2.3.1.4.2 Iterative Closest Point	7
2.3.1.5 Principal Component Analysis	8
2.4 Regression	8
2.4.1 Feature Selection and Akaike Information Criterion	9
2.4.2 Coefficient of Determination, R -squared	9
2.4.3 Significance p of regression models	9
3 Methods	11
3.1 Data	11
3.2 Topological study	12
3.2.1 Preprocessing	13
3.2.2 Skeletonization and post-processing	14
3.2.3 Identification of Costochondral Joints	14
3.2.4 Labeling Costochondral Joints	17
3.2.5 Labeling branches	17
3.2.6 Identifying Interchondral Joints and Hollows	19

3.2.7	Identifying sternocostal joints	20
3.3	Statistical Analysis	20
3.3.1	Data and preprocessing	20
3.3.2	Landmarking and pre-morphing using RBF-TPS	20
3.3.3	Landmarkless Morphing	21
3.3.4	PCA	22
3.4	Regression Analysis	22
3.4.1	Feature Selection	22
3.4.2	Dataset Prediction and R -squared	23
4	Results	25
4.1	Topological Study	25
4.1.1	Overall algorithm performance	25
4.1.2	Occurrences of Hollows, Interchondral- and Sternochondral joints	26
4.1.3	The most common topology	28
4.2	Statistical Analysis	29
4.2.1	Distribution of Subjects	29
4.2.2	Principal Component Analysis	30
4.3	Regression	33
4.3.1	Predicting geometries using PC scores	35
5	Discussion	39
5.1	Data inconsistencies	39
5.2	Topological Study	39
5.2.1	Correspondence between results and anatomical features	40
5.2.2	Algorithm construction and potential data leakage	40
5.3	Statistical Analysis	41
5.3.1	ICP and the choice of template	41
5.3.2	Data quality and sample size	41
5.4	Regression	43
5.4.1	Choice of regression model, correlation and possible interaction effects	43
5.4.2	Specifics of the dataset and the subsequent effects on the results	43
5.4.3	Prediction of population trends, outliers and consistency with related studies	44
5.5	Generalizability and use of results	45
5.6	Future Work	45
6	Conclusion	47
A	Python & Matlab Packages	I
B	Shape variations for PC5-PC11	III

List of Figures

2.1	The figure shows a schematic view of the costal cartilage and its features. The cartilages are numbered, and marked in the image are sternocostal joints, interchondral joints, costochondral joints and hollows.	4
2.2	The figure shows an example binary image (left) and the skeletonizaion of the same (right). The image was generated by using data and code available from Scikit-learn (van der Walt et al., 2014).	5
3.1	The figure shows a schematic view of the steps taken to achieve the results for the topological study. The workflow is from left to right. .	12
3.2	The figure shows a side by side comparison of segmentation, before (3.2a) and after (3.2b) it has undergone preprocessing.	13
3.3	The figure shows the same skeletonized volume before and after short junction-to-endpoint branches has been removed.	14
3.4	The figure illustrates step two of the algorithm developed to identify costochondral joints. From left to right: the specific point is chosen. All points within a vertical interval of that point are extracted. The lateral-most point is accepted.	15
3.5	The figure illustrates step three of the algorithm that was developed to label branches in accordance with the costal cartilage that they pertain to. The steps are ordered from left to right.	16
3.6	The figure shows the effect of each step of the endpoint extraction-algorithm. From left to right: 1) all potential endpoints are extracted, 2.1-2.2) all candidate points are required to be the the most lateral points in their vertical interval, and 3) add points back if they align with the already accepted points.	16
3.7	The figure illustrates how the labeling of branches is conducted. Starting at a costochondral joint, we move to the adjacent branch. All branches that either are at a specific angle with that branch, or align with the previous branch, are given the same label as that costochondral joint.	18
3.8	The figure shows a comparison of before and after the branches of the skeleton have been subjected to merging.	19
3.9	The figure shows a side by side comparison of the template mesh, a subject segmentation and the template mesh once it has been pre-morphed to adapt to the subject geometry.	21

4.1	figures 4.1a and 4.1b show the number of times each distinct costal cartilage were detected, versus the number of costal cartilages that were detected per segmentation.	26
4.2	The figure shows how frequently each costal cartilage attaches to either the sternum body or the manubrium via a sternocostal joint, across 135 subjects.	26
4.3	The figure shows the prevalence of interchondral joints for each consecutive pair of cartilages across 135 subjects. The fraction is obtained by dividing the number of occurrences of each interchondral joint by the number of occurrences of the inferior-most cartilage in the pair. .	27
4.4	The figures show the prevalence of hollows in the subject segmentations. The fraction is obtained by dividing the number of occurrences of a hollow by the number of occurrences of the inferior-most cartilage.	27
4.5	The figure shows the most common topology that was found by thresholding each feature at $> 50\%$. The sternocostal joints (CC1-CC6 on left and right side) are marked by yellow circles. There are interchondral joints between CC6-CC7, CC7-CC8 and CC8-CC9 on the left and right side, with an additional interchondral joint between CC5-CC6 on the right side. There is a hollow between CC7-CC8 on the left.	28
4.6	The figure shows the distribution of ages among all of the remaining subjects, as well as the remaining male and female subjects, respectively. Each figure also shows the corresponding distribution among the 155 subjects for which the segmentations were made.	29
4.7	The figure shows the distribution of height among all of the remaining subjects, as well as the remaining male and female subjects, respectively. Each figure also shows the corresponding distribution among the 155 subjects for which the segmentations were made.	30
4.8	The figure shows the distribution of BMI among all of the remaining subjects, as well as the remaining male and female subjects, respectively. Each figure also shows the corresponding distribution among the 155 subjects for which the segmentations were made.	30
4.9	The figures show the explained variance ratio and cumulative explained variance ratio per principal component.	31
4.10	The figure shows the mean shape that was obtained by taking the mean value for each feature across all subjects.	31
4.11	The figure shows the Pearson Correlation for the independent variables Age, Sex, Height and BMI. Apart from Sex and Height, there are no strong correlations.	33
4.12	The figure shows p -value per significant PC (left) and R^2 -value per significant PC (right). The colors of the bar indicate which subset of independent variables that yielded the highest average AIC across five folds.	33
4.13	The figure shows how values of PC1 correlate with Sex, Age, Height and BMI. Each of the subplots display a positive correlation.	34

4.14	The figure shows how values of PC2 correlate with Sex, Age, Height and BMI. The subplots display a positive correlation for Age and BMI, but a negative correlation for Sex and Height.	34
4.15	The figure shows how values of PC5 correlate with Sex, Age, Height and BMI. The subplots display a positive correlation for Age, Sex and Height, but a negative correlation for BMI.	35
4.16	The figure shows how the shape and size of the costal cartilage varies between male and female subjects, given that there is no difference in Age, Height or BMI.	36
5.1	The figure shows a comparison between the template mesh that was originally used (figure 5.1a) and the template mesh that was used for the final analysis (figure 5.1b). Note how the final template is smoother than the original, and has CC10 detached from CC9.	42

List of Tables

3.1	The table displays the specific parameter values that were employed for the CPD-algorithm, along with a short description of what they control.	21
3.2	The table displays the specific parameter values that were employed for the ICP-algorithm, along with a short description of what they control.	22
4.1	The table shows the most common features of the costal cartilage. These features were detected in at least 50% of the subject segmentations.	28
4.2	The table shows the mean values of Age, Height and BMI for all of the remaining subjects, as well as male and female subsets separately.	30
4.3	Principal component shape variations for $\pm 2SD$ of PC1 to PC4.	32
4.4	The table shows shape variations for female subjects when all independent variables but one are set to the female dataset mean, seen in table 4.2. The varying independent variable is set to $\pm 1SD$ for either Age, Height or BMI, based on the female dataset. Each row corresponds to changes in the listed variable.	37
4.5	The table shows shape variations for male subjects when all independent variables but one are set to the male dataset mean, seen in table 4.2. The varying independent variable is set to $\pm 1SD$ for either Age, Height or BMI, based on the male dataset. Each row corresponds to changes in the listed variable.	38
5.1	Comparison between reference values reported by Fryar et al., 2021 and the dataset statistics for male and female subjects.	42
A.1	The table shows the Python libraries utilized during the project.	I
A.2	The table shows the Matlab Add-Ons utilized during the project.	I
B.1	Principal component shape variations for $\pm 2SD$ of PC5 to PC8.	IV
B.2	Principal component shape variations for $\pm 2SD$ of PC9 to PC11.	IV

1

Introduction

Swedish authority Trafikverket has since 1995 adopted *Vision Zero*, which states that “no one should die or be seriously injured in traffic” (Trafikverket, 2020). Human Body Models (HBMs) are digital representations of the human anatomy used for crash test injury prediction. They constitute an alternative to physical crash test dummies by providing more exact injury predictions while being cheaper to employ (Fahse et al., 2023). Available passive HBMs include, among others, the Total Human Model for Safety (THUMS) and the Global Human Body Models Consortium (GHBMC), (Mokhtar & Hyncik, 2025). These models are available in three sizes: small female, mid-size male and large male. Neither of the models take age or obesity level into account (Hwang et al., 2016). In a review article from 2018, Xu et al. stress the importance of crash tests being performed on groups that are more vulnerable to injury, for example elderly or obese people (Xu et al., 2018). SAFER HBM is a human body model whose developers aim to create an HBM that is both tunable and scalable. In contrast to physical crash test dummies, the SAFER HBM is not only specified for frontal and side impacts, but any direction of impact. The model is developed to predict injury risk and kinematics for both high-g and low-g events (Iraeus et al., 2024). While THUMS modeled a body based on the 50th percentile male as described by Schneider et al., 1983, the GHBMC model is built based on CT and MRI data of one living male subject (Xu et al., 2018). In contrast to this, both SAFER HBM and the VIVA+ models are based on statistical shape models (Iraeus et al., 2024)(John et al., 2022). This means the latter models are based on anatomical variation in the population, rather than the anatomy of a single person that is the median of the population in regards to e.g. height and weight.

Autoliv and Chalmers, along with Sahlgrenska University Hospital and Volvo Cars, have an ongoing research project aimed at improving traffic safety by improving the thorax of the SAFER HBM. The costal cartilage (CC) is situated in the thorax and connects the ribs to the sternum. Blunt thoracic injuries are most commonly caused by automotive crashes, and can cause both broken ribs and fractures in the costal cartilages. The costal cartilage injuries are sparsely studied, but are known to cause ribcage instability, pain and prolonged disability (Nummela et al., 2022). Similarly, the exact geometry of the costal cartilage is sparsely studied, partly because the density of cartilage is similar to that of adjacent tissue, making segmentation of them difficult (Pazokifard et al., 2015). Not only does the shape and size of the costal cartilage vary, but the topology differs between individuals. To date, and to the best knowledge of the author, no previous study on costal cartilage topology and variance has been conducted.

This thesis will constitute part of the work towards improving the thorax of SAFER HBM, by investigating how the topology, shape and size of the costal cartilage varies in the population, contributing to a deeper understanding about its effect on injuries sustained in traffic.

1.1 Purpose

The purpose of this thesis is to contribute to improved road safety by investigating how the costal cartilage varies in the population. The goal is to determine whether some of this variation can be linked to specific demographic attributes, and if so, describe this correlation. The results can then be used to model the costal cartilages of different subpopulations, and possibly correlate their anatomy to their injury pattern. Understanding how the costal cartilage varies in the population could contribute to a better model of the human thorax, which in turn could lead to an improved HBM for crash testing.

1.2 Aim

The aim of this thesis is threefold and is summarized in the list below:

1. To investigate if there is a common topology for the costal cartilage, and determine what that topology is.
2. To investigate how the shape and size of the costal cartilage varies in the population.
3. To investigate how much of the variance in costal cartilage shape and size that can be explained by the demographic variables Sex, Age, Height and BMI.

1.3 Limitations

There are limitations set in place for this project. Primarily, the project will be limited to segmentations that are readily available from a previous thesis work carried out in 2025 (Alexandersson & Moln , 2025). Secondly, the demographic in the data may not be representable of every anatomical variation of interest, and moreover may not equally represent all demographics. Any model based on a skewed or incomplete data set will inherit those traits.

1.4 Ethical Considerations

The data set used for this project is based on cadavers of people who have died suddenly and whose consent to be treated as an object of scientific interest has not been collected. The subjects should be treated as respectfully as possible, meaning no unnecessary information about the subjects should be utilized. Even if GDPR is not applicable to cadavers, the subjects should be treated with dignity.

2

Theory

This chapter will cover human anatomy, the costal cartilage and topological skeletons. Both Human Body Modeling and Statistical Shape Modeling will be introduced, along with the related techniques. Finally regression, including feature selection and evaluation, will be covered.

2.1 Anatomy

The human body is described using specific directions and planes. The superior direction is towards the head, while the inferior is away from the head. The anterior side of the body is the front side, while the posterior is the back side. The medial direction is towards the midline of the body, and the lateral direction is away from the midline of the body. The mid-sagittal plane is vertical and divides the body into left and right sides. The left- and right-hand side are defined by the body, and not the observer (Marieb & Hoehn, 2019).

2.1.1 The Costal Cartilage

The sternum is situated in the thorax and consists of three parts: the superior-most manubrium, the body, and the inferior-most xiphoid process. The Costal Cartilage (CC) is a body part that connects the ribs to the sternum, i.e. the breastbone, protecting the lungs while allowing the chest to expand and contract as we breathe. The costal cartilage connects to the ribs via costochondral joints, and to the manubrium and sternum body via sternocostal joints. The costal cartilage also connects to itself via interchondral joints. The cartilages are numbered from superior-most to the inferior-most (Haase & Shaikh, 2024). A schematic view of the costal cartilage can be seen in figure 2.1. The features of the costal cartilages vary between individuals. In some individuals, there are hollows in the cartilages. Both the number of sternocostal and interchondral joints vary between individuals, while the number of costochondral joints is fixed at 10 for both sides of the body. The existence and placement of hollows also varies. The properties of the cartilages are not necessarily symmetrical around the mid-sagittal plane (Haase & Shaikh, 2024).

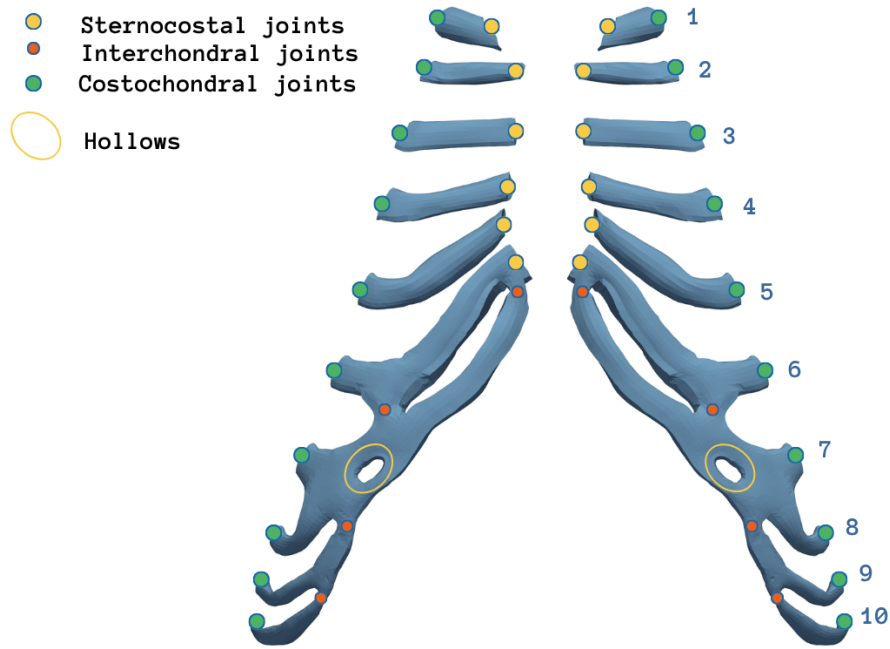


Figure 2.1: The figure shows a schematic view of the costal cartilage and its features. The cartilages are numbered, and marked in the image are sternocostal joints, interchondral joints, costochondral joints and hollows.

2.2 Topological Skeletons and Skeletonization

A topological skeleton is a thinned version of a binary image that represents its connectivity (Maragos & Schafer, 1984). Skeletonization is a means of reducing a binary input to a pixel representation that is one unit wide. The input is binary in the sense that an image consists of pixels that belong to an object, and pixels that do not. To obtain a skeleton of the object, the image is subjected to operations of erosion, whereby a pixel value is replaced with the minimum value of its neighborhood. This iteratively removes the boundary pixels of the object until only the skeleton, the 1-pixel-wide representation, remains (Zhang & Suen, 1984). An illustration of a skeletonized binary image can be seen in figure 2.2.

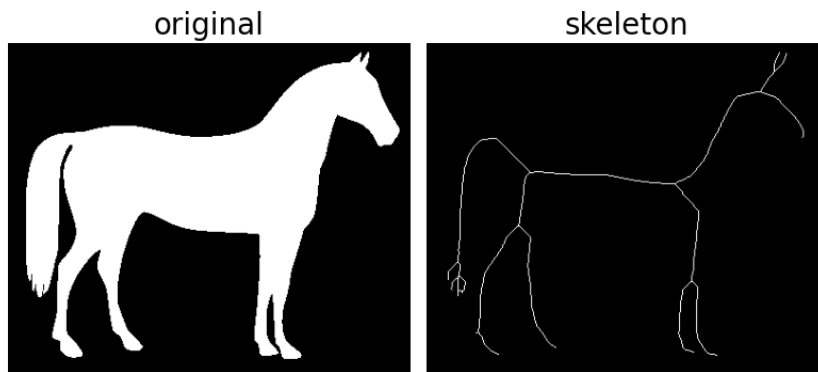


Figure 2.2: The figure shows an example binary image (left) and the skeletonization of the same (right). The image was generated by using data and code available from Scikit-learn (van der Walt et al., 2014).

In Scikit-image, skeletonization is performed using the algorithm proposed by Zhang and Suen, 1984, and results in a skeleton living in pixel or voxel space, depending on the dimension of the input (van der Walt et al., 2014). The skeleton can be further analyzed using Skan, a library constructed for skeleton analysis (Nunez-Iglesias et al., 2018).

In Skan, a skeletal branch is defined as any set of voxels that are connected to one another, with a euclidean distance of 1 or $\sqrt{2}$ voxels for adjacent and diagonally adjacent voxels, respectively. Using Skan, the skeleton can be summarized into a Pandas dataframe where each row constitutes a branch. For each branch, the summarization dataframe contains voxel coordinates for its source and destination nodes, collectively referred to as “endpoints”. The dataframe also contains the length of each branch (measured in euclidean distance) and the type of the branch. The type is dependent on the connectivity of the branch. An endpoint-to-endpoint branch is an isolated branch that connects to no other branch. A junction-to-endpoint branch connects to another branch only in one end, whereas a junction-to-junction branch has connected branches in each end. Finally, there are isolated cycles, which are branches that connect only to themselves and have the same source and destination coordinates (Nunez-Iglesias et al., 2018).

2.3 Human Body Modeling

Human Body Models (HBMs) are digital representations of the human anatomy used for crash test injury prediction (Fahse et al., 2023). Anthropomorphic Test Devices (ATDs), commonly known as crash test dummies, have historically been employed to predict injury response in a human subjected to a crash. ATDs are implemented both physically and virtually, but can only indicate injuries. Human Body Models are developed to capture the mass distribution, geometry, joints and passive joint stiffness of the human body. They allow a higher degree of biofidelity compared to the ATDs, thereby providing more exact predictions of injuries. In

addition, they are not as costly to employ as the physical ATDs (Fahse et al., 2023).

2.3.1 Statistical Shape Modeling

Statistical Shape Modeling is a means of obtaining Statistical Shape Models (SSMs). These are statistical models that describe the geometrical variation in a population, and can be applied to specific body parts to obtain information about variation in human anatomy. This has previously been done for, among other, the pelvis (Brynskog et al., 2021) and the ribs (S. A. Holcombe et al., 2024). The results obtained via statistical shape modeling can serve as a basis for an HBM, or be used for morphing HBMs to other sub populations, and this section will describe the general procedure.

2.3.1.1 Landmarking

Landmarks are anatomically significant points, and landmarking is the process of specifying these points in the plane or in space. Semi-landmarks are intermediate landmarks that are placed at an even distance between existing landmarks (Slice, 2005). While they are not anatomically significant on their own, semi-landmarks allow a more accurate description of the subject surface when used along with landmarks. By landmarking each sample, the object surfaces are described in such a way that all anatomically significant points are distinctly marked.

2.3.1.2 Morphing

Morphing is the process of deforming an object into the shape of another. This is done using a template mesh, i.e. a mesh of the correct topology, with valid connectivity, that is an approximation of the target geometries (Yeh et al., 2011). The template mesh is morphed to the geometry of each subject geometry using e.g. a Radial Basis Function with Thin Plate Spline (RBF-TPS) interpolation (Hwang et al., 2016). Fitting the template mesh to all subject geometries will ensure that every geometry is described using the same nodes and elements. This ensures comparability, meaning node-by-node comparisons are meaningful. Any differences discovered upon further analysis will be differences in shape and size, and not in geometrical representation.

2.3.1.3 Generalized Procrustes Alignment

Generalized Procrustes Alignment (GPA) is a method utilizing translation and rotation to align objects in space. Meshes, defined strictly by their landmarks, are aligned and rotated so that the sum of mean differences between their shapes is minimized, while keeping their respective geometries intact (Gower, 1975). The resulting rotation matrix is applied to the meshes. Performing GPA ensures that the orientation of the objects are the same which ensures that any differences later discovered will be differences in shape and size and not a rigid body motion.

2.3.1.4 Landmarkless Morphing

While manual landmarking of each subject mesh ensures that the points are comparable, it is labor intensive and sensitive to human error. An alternative approach is landmarkless morphing, which can be accomplished using the following methods. They treat both template and subject segmentation as point clouds, meaning they disregard the connectivity and surfaces of the meshes, taking only vertices into account. The landmarkless morphing results in each subject segmentation being represented using the same topology, and furthermore, that each subject segmentation is represented by a morphed template with $3n$ features, which are the cartesian coordinates of the n nodes that make up the template. This consistency in description is a prerequisite for the subsequent statistical analysis.

2.3.1.4.1 Coherent Point Drift

Coherent Point Drift (CPD) is a probabilistic point set registration algorithm, which is a type of algorithm that maps one set to another by assigning correspondences between points. The first step of CPD is to use a Gaussian Mixture Model (GMM) to describe the template. This in itself is a probability-maximizing problem, where the goal is to find the specific Gaussian distributions that, together, best describe the vertices of the template. Once these distributions are found, the GMM has been fit to the template. This model can be used to approximate the subject segmentation. The goal is to find the transformation of the GMM centroids that maximizes the likelihood of vertices in the subject segmentation being in the GMM centroids. To ensure the topology is preserved, the GMM centroids are forced to move coherently. In the non-rigid case of CPD, the coherency of different regions' movement is enforced by regularizing the displacement field, i.e. restricting how large movements are allowed. This is done by applying Tikhonov Regularization, which penalizes spatially non-smooth movement (Myronenko & Song, 2010).

2.3.1.4.2 Iterative Closest Point

Iterative Closest Point (ICP) is a rigid algorithm used to minimize the distance between two point clouds using translation and rotation. The target source cloud is kept fixed, and every point in the template point cloud is paired with its closest neighbor in the target point cloud. The template point cloud is then rotated and translated in the way that minimizes the distance between paired points. These steps are performed iteratively until convergence (Besl & McKay, 1992).

Optimal Step Non-Rigid ICP is an alternative implementation presented by Amberg et al., 2007. By applying local transformations to a specified deformation field, points in different regions of the point cloud are allowed to move differently, constituting non-rigid transformation. The local transformation is obtained by Minimizing Equation 2.1, where $\mathbf{X}$ is a transformation matrix, $E_d(\mathbf{X})$ penalizes the distance between template and target being large, $E_s(\mathbf{X})$ penalizes nearby vertices moving differently and $E_l(\mathbf{X})$ maps template landmarks to target vertices by penalizing their distance being large. The parameters α and β are used to weight the flexibility and impact of landmarks (Amberg et al., 2007).

$$E(\mathbf{X}) = E_d(\mathbf{X}) + \alpha E_s(\mathbf{X}) + \beta E_l(\mathbf{X}) \quad (2.1)$$

2.3.1.5 Principal Component Analysis

Principal Component Analysis (PCA) is a method of dimensionality reduction, data exploration and statistical analysis. Using the covariance matrix computed from the data, PCA identifies the eigenvector corresponding to the direction of the greatest spread in the features of the data (measured by their variance); this is the first principal component (PC). The eigenvector that corresponds to the direction of the second greatest variance, while also being orthogonal to the first PC, becomes the second PC, and so on. The method projects the data into a new feature space spanned by the orthogonal principal components (Gewers et al., 2021).

In statistical shape modeling, PCA can be used to understand how nodes move together across subject geometries. Each PC captures a specific movement of nodes, which describes a specific anatomical variation. A PC can for example describe a variation in thorax width. Then, an individuals' score for that PC becomes the measure of how wide their thorax is.

2.4 Regression

Regression analysis is a means to investigate whether observed variations can be correlated to predictor variables, i.e. to predict an outcome Y based on an input X . The regression task involves one or more independent variables, and one or more dependent variables. Multiple regression models the relationship between two or more independent variables (predictors) and one dependent variables (target) (Montgomery et al., 2012b). In relation to SSM, regression can be used to analyze the relation between population variation and demographic variables, such as Sex, Age, Stature and BMI.

The linear regression model is described in Equation 2.2, where x is the independent variable, y the dependent variable, β_0 is the intercept, β_1 is the slope and ϵ is an error term. This model attempts to describe linear trends in data by minimizing the error, i.e. the difference between the model prediction and the ground truth value (Montgomery et al., 2012c).

$$y = \beta_0 + \beta_1 x + \epsilon \quad (2.2)$$

A regression model can be evaluated over several Cross-Validation (CV) folds, whereby the dataset is randomly split into n subsets of training and test-data. The model is fit to $n - 1$ of the subsets and evaluated on the n :th. This is repeated until all of the subsets have been used for evaluation once. Testing model performance over different splits can be used to ensure the model performs consistently, and has not overfit to the dataset (Montgomery et al., 2012d).

2.4.1 Feature Selection and Akaike Information Criterion

It can be useful to reduce the number of independent variables, for example to reduce computational cost or increase interpretability (Jović et al., 2015). Reducing the number of variables can also be desirable in cases where independent variables are collinear, as such a relation between independent variables can yield unstable regression models, low significance or high standard deviations (Montgomery et al., 2012a). The selection of independent variables should be stable, i.e. the same subset should ideally be chosen independent of data split.

Akaike Information Criterion (AIC) is a means of evaluating the performance of a statistical model which rewards high model fit while penalizing the number of model parameters (Akaike, 1974). This is described by Equation 2.3, where $\hat{\ell}_{n,m}$ is the maximized log likelihood of model M_m given observations n , and d_m is the dimension of the model.

$$AIC_m = -2\hat{\ell}_{n,m} + 2d_m \quad (2.3)$$

2.4.2 Coefficient of Determination, R -squared

The Coefficient of Determination, also known as R -squared (R^2), measures how much of the variance observed in the data that is explained by a regression model (Emmert-Streib & Dehmer, 2019). A model that repeatedly predicts the average y regardless of input features would obtain an R -squared of 0.0, but an arbitrarily worse model can obtain a negative score. The best possible score is 1.0 (Pedregosa et al., 2011). Regression models can be evaluated using R^2 -scores, as defined in Equation 2.4, where SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (2.4)$$

Cross validated R -squared can be utilized to compare the performance of regression models. While R -squared shows how well each regression model describes the current data set, it says nothing about the stability of the model across data sets. Using cross validated R -squared solves this problem, as the R^2 -scores will be evaluated on unseen data, constituting a measure of generalization (Emmert-Streib & Dehmer, 2019).

2.4.3 Significance p of regression models

The significance, otherwise called the p -value, measures how likely it would be to obtain results at least as extreme as those observed, given that the null hypothesis is true (Devore et al., 2021b). In regression analysis, the null hypothesis is typically that there is no relationship between the dependent and independent variables (Montgomery et al., 2012b). A smaller p -value indicates that the observed results would be unlikely under the null hypothesis (Devore et al., 2021b). Deciding whether a regression model is significant can be done by computing the p -value of

the F-statistic, which is the ratio of the variance explained by the model and the variance left unexplained by the model. If the null hypothesis is true, the ratio would be near 1.0. Conversely, the larger the ratio, the greater the probability that the null hypothesis can be rejected (Devore et al., 2021a).

3

Methods

The project was divided into three parts, one for each aim. First, a study of topology was carried out on the entire data set. The purpose of this was to investigate whether some topologies are more common than others. The topological study was used as a basis for choosing this specific topology to represent all subjects, since later stages of the study required all subjects to be modeled using the same topology. Second, a statistical analysis of variance was carried out. The purpose of this was to determine what anatomical variations are most common. Third, a linear regression model was fitted to the data and evaluated. The purpose of this was to investigate whether the variance can be attributed to differences in Sex, Age, Stature or BMI. The project was carried out in Python 3.12.12. If other software was used, this is explicitly stated in the related section. A list of relevant packages utilized for the respective parts of the method can be found in appendix A. In this section, these three parts are explained chronologically.

3.1 Data

The New Mexico Decedent Image Database (NMDID) contains medical imagery of decedents along with their metadata. The medical imagery consists of CT-images and the metadata contains census records, health data, demographic data and circumstances of their death (Edgar et al., 2020). Using this data, Alexandersson and Moln , 2025 built an automatic segmentation model that generated 155 segmentations of the costal cartilage, manubrium and sternum body. Their automated segmentation of the costal cartilage was made for CC1-CC10, and out of the 155 there are 82 male subjects and 71 female. The resulting segmentations contain inconsistencies. The number of cartilages that have been captured vary, both between subjects and around the mid-sagittal plane for single subjects. The segmentations sometimes have missing faces, sometimes contain spikes due to incorrect prediction of one singular vertex and sometimes contain floating islands of cartilage. When comparing the segmentations to the ground truth, the reported DICE-score for the costal cartilage was 0.87, where a DICE-score of 1.0 would imply perfect overlap between target geometry and the segmentation. These segmentations constitute the dataset used for this thesis.

3.2 Topological study

The aim of the topological study is to investigate whether some topology or other is more common in the population. The study was conducted on both the left and right side of the thorax simultaneously, to capture any asymmetries around the mid-sagittal plane. This was done utilizing the segmentations of the costal cartilage, manubrium and sternum body.

Primarily three anatomical variations are of interest. Firstly, identifying cartilage that attaches to another cartilage, i.e. identifying interchondral joints. Secondly, identifying the number of cartilage that attaches to the sternum, i.e. identifying sternocostal joints. Lastly, identifying cartilage that either attaches to both the sternum and another cartilage, where there is a hollow between the sternochondral and interchondral joints, or analogously a cartilage that attaches only to another cartilage, but has several distinct interchondral joints with hollows in between.

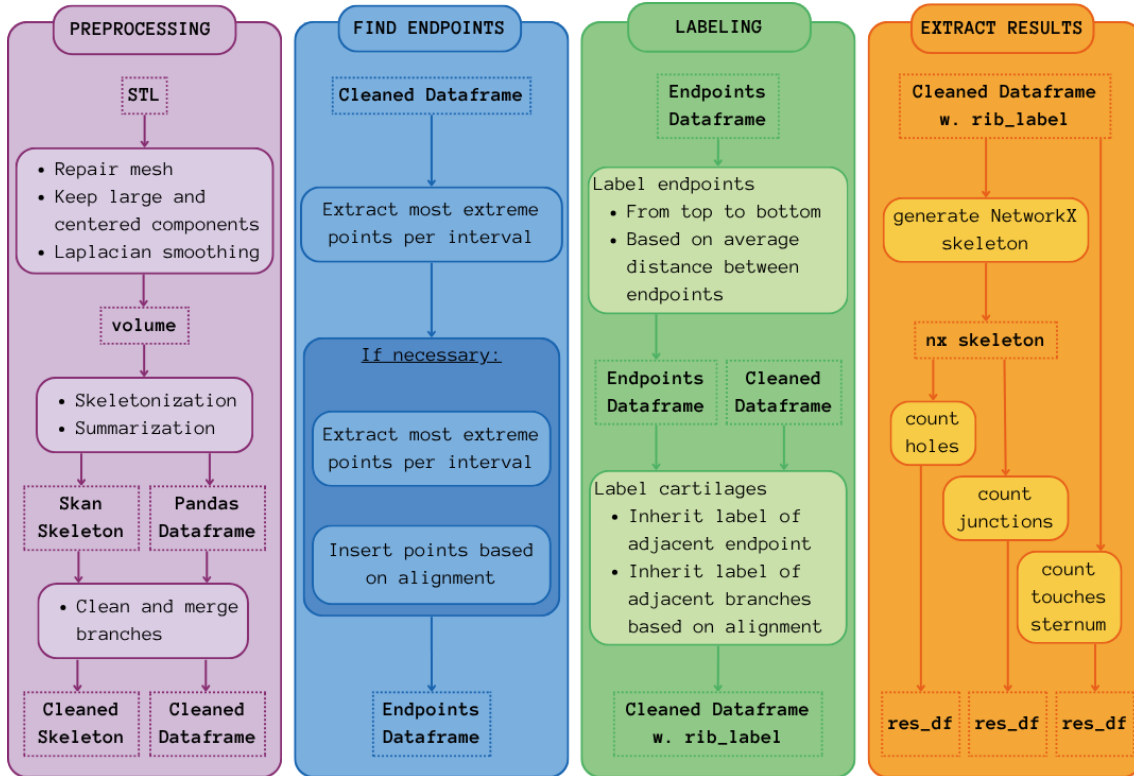


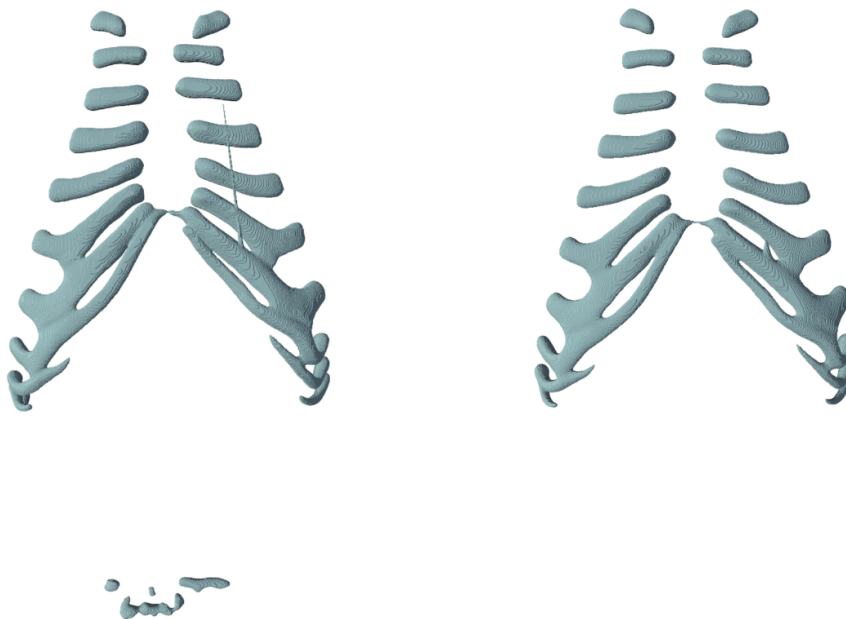
Figure 3.1: The figure shows a schematic view of the steps taken to achieve the results for the topological study. The workflow is from left to right.

The topology was studied by skeletonizing each segmentation and analyzing the source and destination coordinates of each branch. This allowed the costochondral joints to be identified and labeled. The joints were subsequently used for labeling branches. Once every branch was labeled, each branch either belonged to a cartilage and was given a cartilage label, or did not belong to a cartilage and was given label NaN. The labeled branches were used to identify interchondral joints, and to identify

the location of hollows and sternocostal joints. The study was conducted by implementing an automated Python-script, and a schematic overview of the algorithms used can be seen in figure 3.1.

3.2.1 Preprocessing

First the input data, which consists of segmentations of the costal cartilage, was preprocessed. This entailed repairing any missing faces, discarding any components that were too small (less than 0.5% of the total volume of the segmentation) or had their center of gravity further than 10 Mean Absolute Deviation (MAD) than the median distance from the other components. This removed any floating islands while leaving cartilages present. The remaining components were then smoothed by Laplacian filtering from `trimesh` (Dawson-Haggerty & Trimesh Contributors, 2019). The effects of the preprocessing can be seen in figure 3.2. The components were voxelized using $pitch = 0.5$ and each volume was filled. The specific translation from cartesian space to voxel space was stored for further use later on.

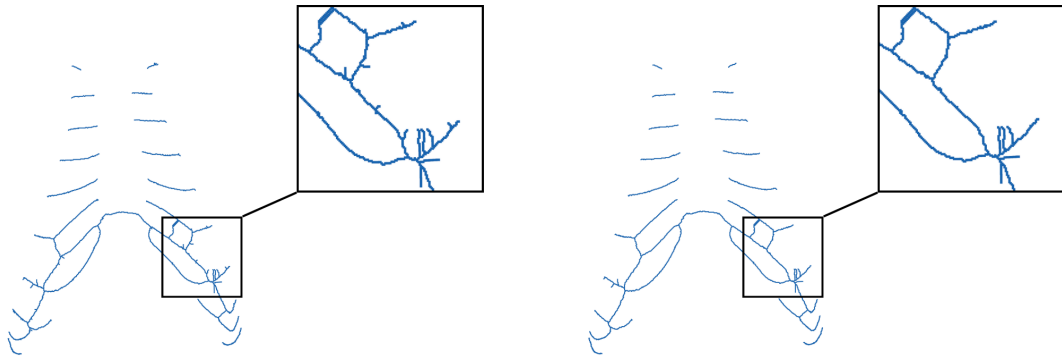


(a) The figure shows the segmentation of a specific subject anatomy before undergoing any preprocessing. Note the existence of floating islands and spikes. (b) The figure shows the segmentation of the subject anatomy after it has undergone the preprocessing steps, including the removal of floating islands and laplacian smoothing.

Figure 3.2: The figure shows a side by side comparison of segmentation, before (3.2a) and after (3.2b) it has undergone preprocessing.

3.2.2 Skeletonization and post-processing

Once the volumes are obtained, these are skeletonized using **Scikit-image**, and the skeleton is summarized using **Skan**. Further analysis is done on the resulting **Pandas** dataframe. Only branches that pertain to a cartilage are of interest. Therefore, any junction-to-endpoint branches that were shorter than 0.5 of the mean junction-to-endpoint branch length were removed. The effects of this cleaning can be seen in see figure 3.3. Additionally, all branches were assigned a *branch number*. Pairs of branches whose endpoints were within $0.2SD$ of the distance between all endpoints were tested for cosine similarity. If the directions of the two branches had a cosine similarity greater than 0.65, they were assigned the same branch number. Once this preprocessing was completed, the data was ready for further analysis.



(a) The figure shows the point cloud that is the result of skeletonizing the volume, before performing any further data cleaning. (b) The figure shows the same point cloud, after the junction-to-endpoint branches shorter than 0.5 of the mean junction-to-endpoint branch length have been cleaned from the data.

Figure 3.3: The figure shows the same skeletonized volume before and after short junction-to-endpoint branches has been removed.

3.2.3 Identification of Costochondral Joints

The costochondral joints are used as a point of departure for identifying each cartilage. Using the costochondral joints as ground truth, the costal cartilages can then be labeled based on what rib they attach to. To determine this, the algorithm utilizes the source- and destination coordinates in the summarization dataframe, gathered under the name “endpoints”.

The first step one is to extract the n lateral-most endpoints. Step two of the algorithm consists of sorting and discarding endpoints and is illustrated in figure 3.4. The sorting and discarding is done in steps, for left and right side of the body separately, according to the following logic:

- a. Sort the points in descending order, from superior-most to inferior-most. Then process the points in descending order.

- b. Let each point in turn be at the center of a z-interval. Discard the point if it is not the lateral-most point in its own interval.
- c. Finally, all that remains is a dataframe of accepted endpoints.

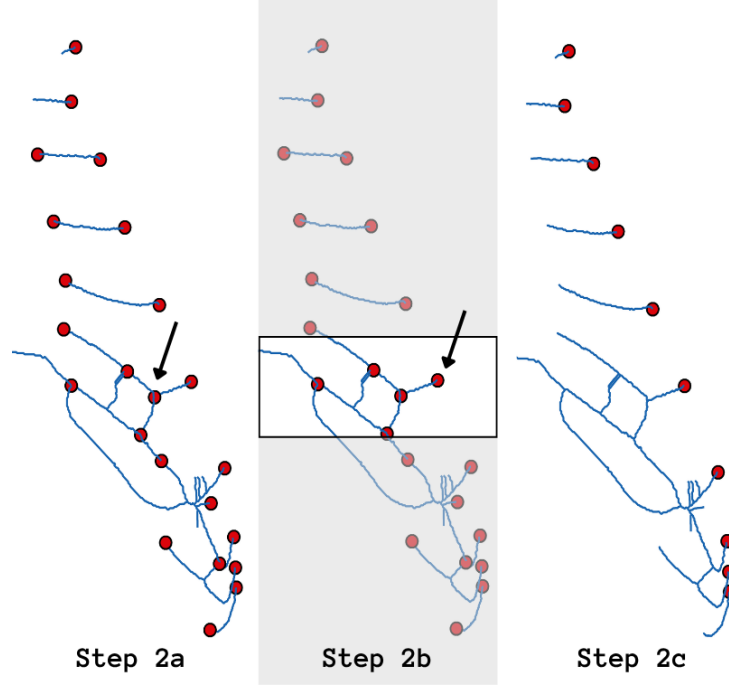


Figure 3.4: The figure illustrates step two of the algorithm developed to identify costochondral joints. From left to right: the specific point is chosen. All points within a vertical interval of that point are extracted. The lateral-most point is accepted.

The segmentations were made for 10 costal cartilages, but there were sometimes more (or fewer) cartilages present. This meant there were no ground truth regarding how many endpoints should be found in each segmentation. This means that simply restricting the algorithm to finding a certain number of points will not necessarily produce the desired result, and extra precaution needs to be taken to guarantee that the accepted points actually correspond to costochondral joints.

The third step is an additional fail safe, to avoid including points that should be omitted, or analogously, omitting endpoints that should be included. It consists of adding or discarding endpoints conditionally: If either side has more than 10 remaining points, the validity of the accepted points will be checked by performing step two once more. If either side has less than 12 points remaining, points will be added according to the following algorithm, which is illustrated in figure 3.5:

- a. For every consecutive pair of accepted points (a, b) , extract all endpoints that lie in the z-interval between the accepted points, i.e. that satisfy $z_b \leq z \leq z_a$. Process the endpoints in descending order.
- b. For each candidate point c among the extracted points, compare the vector between the two accepted points, $\vec{ab}$ to the vectors from extracted point to

either accepted point, $\vec{ca}$ and $\vec{cb}$. Any point whose corresponding vectors have an absolute cosine similarity greater than 0.9 are stored.

- c. Add the point to the dataframe of accepted points. If there are several stored points, only the lateral-most will be added to the accepted points.

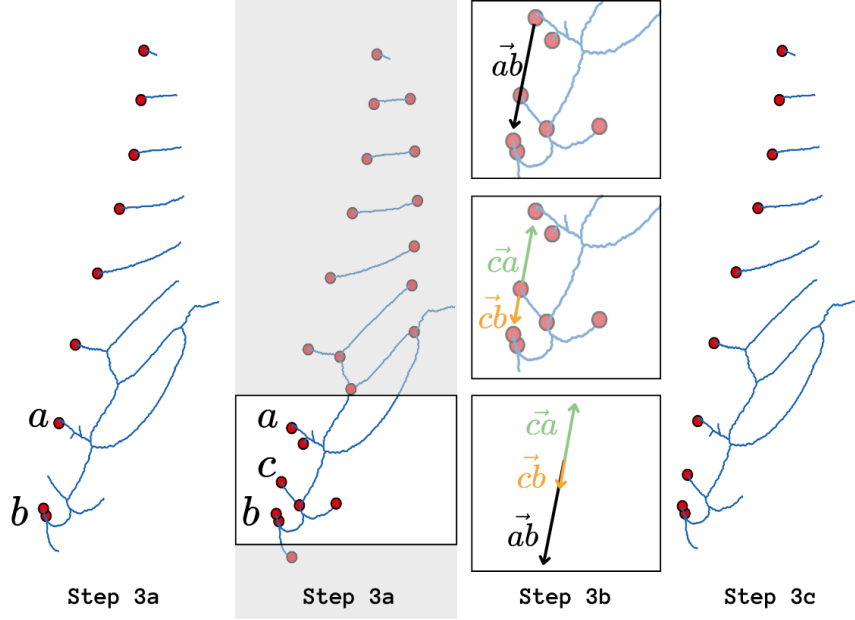


Figure 3.5: The figure illustrates step three of the algorithm that was developed to label branches in accordance with the costal cartilage that they pertain to. The steps are ordered from left to right.

The final output consists of a dataframe containing costochondral joints, the summarization dataframe with added columns for costal cartilage label and branch number, and a skeleton object. To see the effect of the endpoint extraction, see figure 3.6.

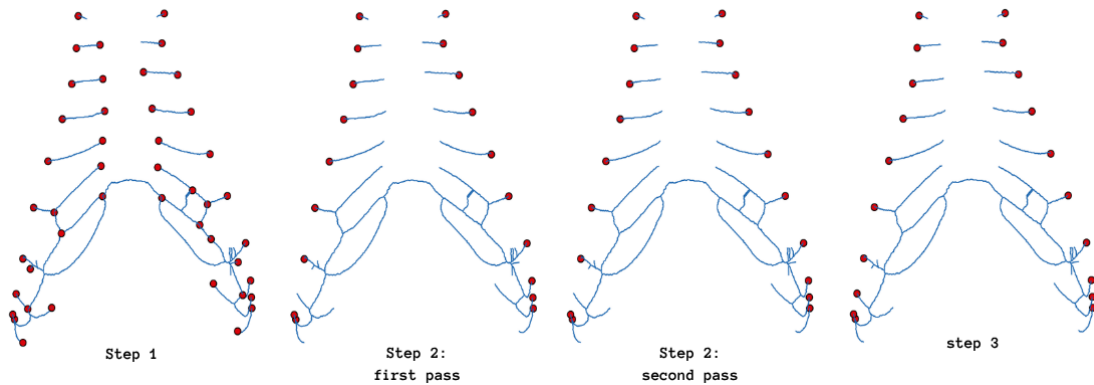


Figure 3.6: The figure shows the effect of each step of the endpoint extraction algorithm. From left to right: 1) all potential endpoints are extracted, 2.1-2.2) all candidate points are required to be the the most lateral points in their vertical interval, and 3) add points back if they align with the already accepted points.

3.2.4 Labeling Costochondral Joints

Once the costochondral joints are identified, they are labeled in ascending order, starting at the superior-most costal cartilage. The average vertical distance between joints is used to encourage correct labeling regardless of how many joints were present in, or detected from, the segmentations. If the current joint is approximately at an average distance away the previous detected joint, it will be assigned the next CC number. If not, it will be assigned a CC number according to Equation (3.1).

$$\text{CC number} = \text{round} \left(\text{previous CC number} \cdot \frac{\text{distance}}{\text{avg_distance}} \right) \quad (3.1)$$

If the CC1 joint were missing, it would cause an offset in the labeling. To avoid this, the z-values of the superior-most left and right endpoints are compared. If these are approximately the same, CC1 is assumed to be present for both sides. Otherwise, the superior-most is assumed to be CC1, and the other endpoint will be assigned a label according to Equation (3.1).

3.2.5 Labeling branches

Once the costochondral joints were labeled, these were used to label the cartilages, by assigning the same label to the branches that make up each cartilage. As Skan's requirements for points constituting one branch are quite strict, one costal cartilage is made up of several branches. The following algorithm searches among branches adjacent of each costochondral joint to find all branches that make up that cartilage.

All branches were initially given CC label NaN. Then, any junction-to-endpoint branch that shared source or destination coordinates with a costochondral joint inherited its label. In the case of junction-to-junction branches, they inherited the label of an adjacent branch if certain criteria were met. This allowed for one cartilage to be made up of several branches, which was obtained by the following algorithm. An illustration of the same is found in figure 3.7.

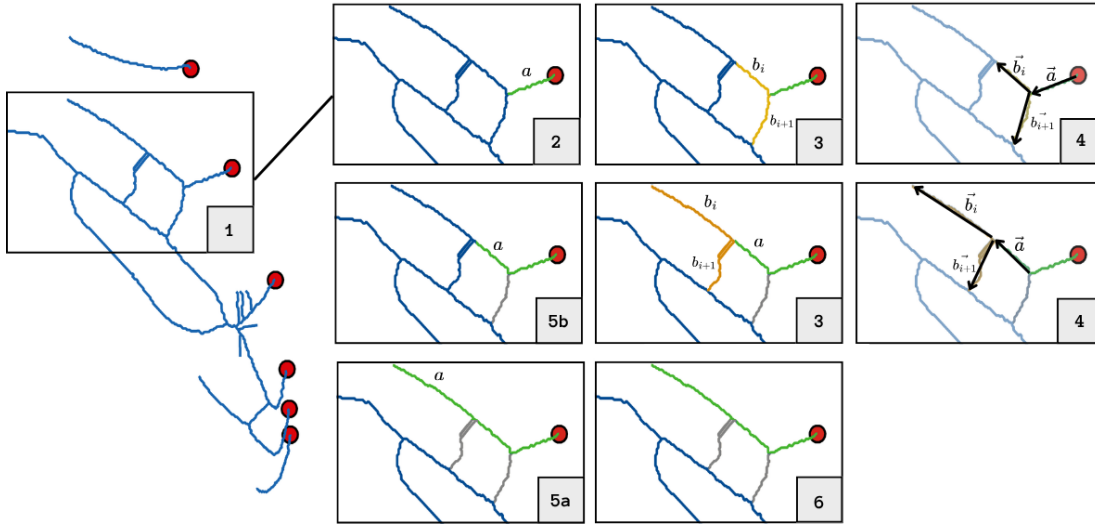
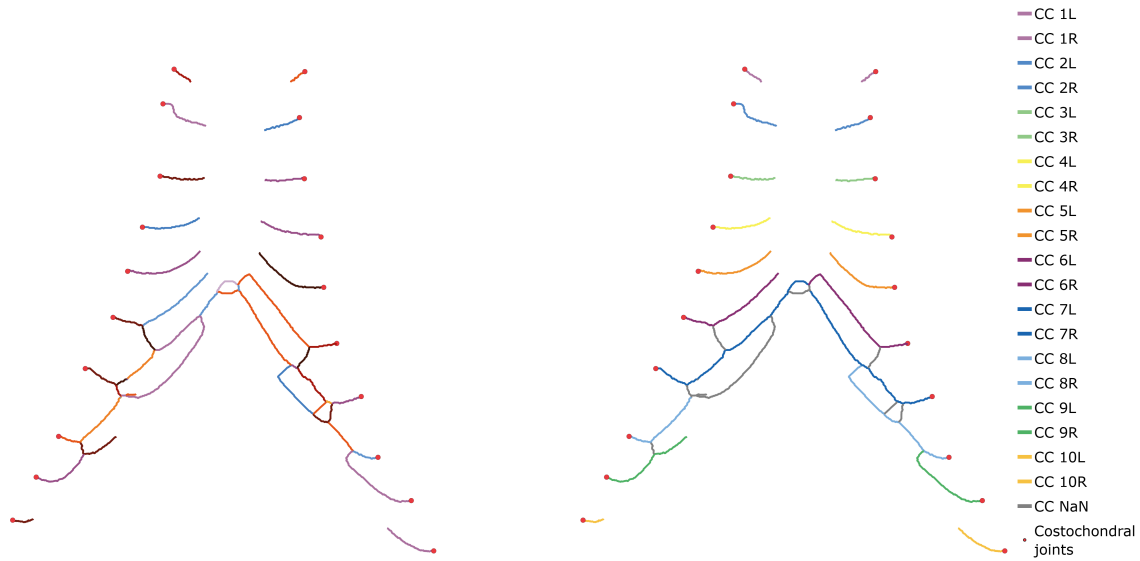


Figure 3.7: The figure illustrates how the labeling of branches is conducted. Starting at a costochondral joint, we move to the adjacent branch. All branches that either are at a specific angle with that branch, or align with the previous branch, are given the same label as that costochondral joint.

- Step 1** Start at a costochondral joint.
- Step 2** Move to the branch a that shares source or destination coordinates with that costochondral joint.
- Step 3** For that branch, identify all adjacent branches b_i that are unlabeled. Adjacent branches are branches that share source or destination coordinates with the current branch.
- Step 4** Define a vector using the source and destination coordinates for branch a . For every adjacent branch b_i : define a vector using the source and destination coordinates for that branch.
- Step 5** Compare the vectors and make actions using one of the following cases:
 - 5a)** The vectors have very good cosine similarity. Let adjacent branch b_i inherit the label from current branch a . Move to adjacent branch b_i , making it the new current branch.
 - 5b)** The branch b_i grows upwards, and its angle relative to the first branch of the cartilage is the smallest angle greater than 90° among all adjacent branches. Let adjacent branch b_i inherit the label from current branch a . Move to adjacent branch b_i , making it the new current branch.
- Step 6** Repeat steps 3 to 5 until there are no more branches to visit or no adjacent branch satisfies the conditions.
- Step 7** Move to next costochondral joint and repeat the process.

An example showing the effects of the labeling can be seen in figure 3.8a.



(a) The figure illustrates the extracted endpoints and all branches colored by endpoints, before any labeling of cartilage. Branches that pertain to no specific cartilage has occurred.

(b) The figure illustrates the extracted endpoints and the branches colored by branch number, before any labeling of cartilage. Branches that pertain to no specific cartilage are assigned label NaN and depicted with gray color.

Figure 3.8: The figure shows a comparison of before and after the branches of the skeleton have been subjected to merging.

3.2.6 Identifying Interchondral Joints and Hollows

Once the costochondral joints and branches had been assigned rib labels, these structures were used to identify interchondral joints and hollows. This was obtained by using the python library **NetworkX**, which is built for building and analyzing complex network structures (Hagberg et al., 2008).

A graph structure was generated from the summary dataframe. The nodes were constructed from the source and destination coordinates and the edges were constructed to correspond to a branch. The CC label, along with additional data from the skeleton, was stored as a property of the edges.

To identify the interchondral joints, a Breadth-First Search (BFS) was conducted, conceptually answering the question: “If we are at a branch with label n , could we traverse edges with label NaN and reach a branch with label $n + 1$?” The hollows were detected using the minimum cycles algorithm presented by Kavitha et al., 2008, which is implemented in NetworkX. The detected cycles were thresholded on cycle length. The existing interchondral joints and hollows were stored in dataframes which were exported to csv-format and saved for later use.

3.2.7 Identifying sternocostal joints

The sternocostal joints connect the costal cartilages to the sternum. To determine their existence, the segmentations of the sternum and manubrium were utilized in addition to the costal cartilage. The vertex coordinates of the meshes were extracted and transformed into voxel space by applying the same transformation that was applied during voxelization of the cartilages. Once transformed, the distances between vertices and branches were calculated. If the distance was small, and if the branch had a CC label, then that cartilage was assumed to attach to the sternum via a sternocostal joint. The existing sternocostal joints were stored in a dataframe which was exported to csv-format and saved for later use.

3.3 Statistical Analysis

The second part of the project consists of a statistical analysis of the subject geometries. The steps necessary to conduct this analysis will be explained in this section.

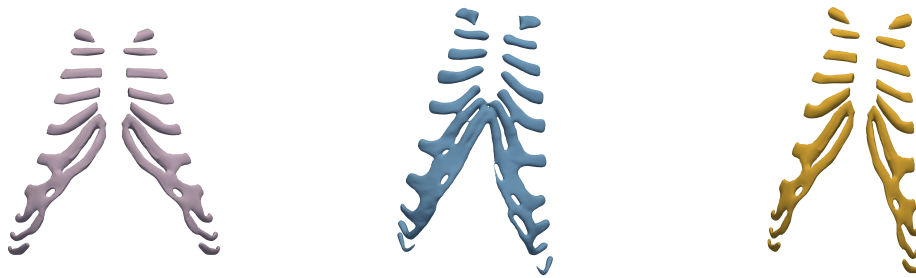
3.3.1 Data and preprocessing

For continuity, the smoothed meshes from the topology study were used as target geometries. Only subjects for which 20 or more costochondral joints were detected were included, reducing the dataset to 57 subjects. These meshes were additionally preprocessed, reducing the number of nodes for manageability. In addition, the coordinates of the costochondral joints that were obtained in the topology study were utilized, along with the template mesh and a small set of landmarks.

All smoothed subject volumes were translated, and the same translation was applied to the rib endpoints. The same goes for the template mesh and its landmarks. The segmentation volumes and the rib endpoints were rotated to ensure the orientation between the segmentations and template was consistent.

3.3.2 Landmarking and pre-morphing using RBF-TPS

The templates were pre-morphed using a small set of landmarks. The pre-morphing was done by landmarking the 20 costochondral joints in the template mesh. The landmarking was done manually using ANSA 25.1.0 (BETA CAE Systems), a pre-processing software developed for Computer Aided Engineering (CAE) model building. The template was then morphed by geometrically matching the landmarks to the corresponding costochondral joint. This was done in Python using a Radial Basis Function with Thin Plate Spline (RBF-TPS) interpolation (Hwang et al., 2016). Once this was completed, there were one preprocessed template mesh per subject. The effect of the pre-morphing can be seen in figure 3.9.



(a) The figure shows the template mesh, before it has undergone any pre-morphing.

(b) The figure shows a subject segmentation.

(c) The figure shows the template mesh after it has been pre-morphed to better represent the subject segmentation.

Figure 3.9: The figure shows a side by side comparison of the template mesh, a subject segmentation and the template mesh once it has been pre-morphed to adapt to the subject geometry.

3.3.3 Landmarkless Morphing

The pre-morphed templates were transformed to match their corresponding subject segmentation using a Matlab R2025b pipeline consecutively employing CPD- and ICP algorithms. The first pass was a non-rigid CPD implementation in that utilized a harsher downsampling and required movement of nodes to be substantial. This roughly aligned the point clouds of target and template. For exact parameter values, see table 3.1. For the second pass, an ICP implementation was utilized, for which a bigger portion of the vertices were included, and movement was required to be smaller. For exact parameter values, see table 3.2. After these passes, the exact surface fitting was tuned using ICP. Once all targets and templates had been processed, the resulting geometries were further aligned using GPA, implemented in Python.

Table 3.1: The table displays the specific parameter values that were employed for the CPD-algorithm, along with a short description of what they control.

Algorithm	Parameter	Value	Description
CPD	Downsample grid size	6	Size of the grid that will be represented by one averaged point.
	Beta	6	Gaussian Kernel Width. Determines to what degree points can move independently.
	Lambda	10	Regularization Parameter. Determines to what degree incoherent movement is penalized.

Table 3.2: The table displays the specific parameter values that were employed for the ICP-algorithm, along with a short description of what they control.

Algorithm	Parameter	Value	Description
ICP	Downsampling	4000	Number of vertices to keep during shape matching.
	Modes	1000	Number of deformation modes.
	Lb Range	20	Lower bound of deformation range.
	Ub Range	40	Upper bound of deformation range.

3.3.4 PCA

In the context of this project, each shape i was represented by $3n$ features, which are the cartesian coordinates of the n nodes that make up the template mesh. The average value of each feature across samples, resulting in an average shape. PCA was conducted on the vertex coordinates of the morphed meshes using Scikit-learn, a python library constructed by Pedregosa et al., 2011. Each principal component describes one type of shape variation, corresponding to an anatomical variation in the population. In addition to the PC-scores, the explained variance per PC was computed.

3.4 Regression Analysis

Regression analysis was utilized to predict variations, specifically PC-scores, given a set of demographic variables. The features Sex, Age, Height and BMI, that were available in the metadata from NMDID, were used as independent variables. The PC scores obtained in the previous step were the dependent variables. To investigate the potential collinearity between independent variables, a correlation matrix was computed for the features.

3.4.1 Feature Selection

The data was split into five training and test sets using Sklearns KFold (Pedregosa et al., 2011). Cross validation and AIC was used for feature selection. For each subset of features, the model was fit to each of the folds, and both R^2 - and AIC-scores were computed and stored. Any subset of features that yielded a lower mean AIC-score across all inner folds than the previous subsets was stored as the best subset. For that subset, the R^2 -score was averaged across all folds and stored. The best subset was then used to fit the model on all data, and the p -values were registered for each regression model. This process was repeated for each of the PCs that individually explained more than 1% of the variance. The linear regression model was implemented using the python package Statsmodels by Seabold and Perktold, 2010.

3.4.2 Dataset Prediction and R -squared

Once a best subset was chosen, the regression model was used to predict PC scores for every significant PC. This was done for all subjects in the dataset, using the best subset of metadata for each PC as independent variables. The R^2 -value for the entire regression model was computed.

4

Results

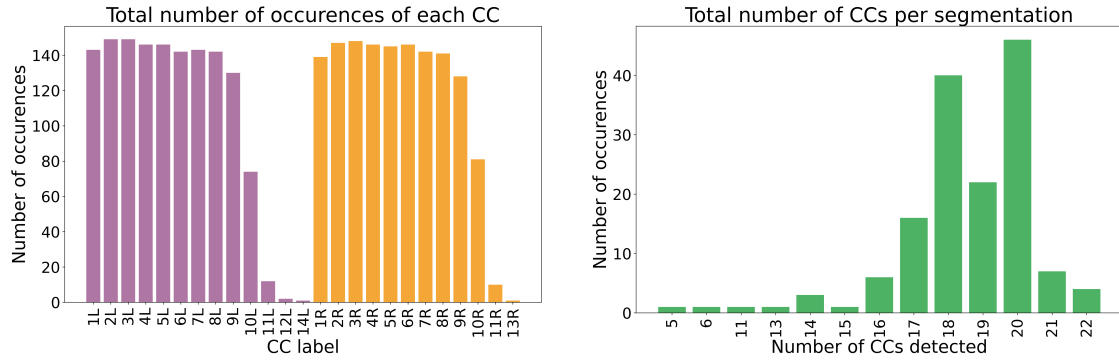
In the following section, the results from the three parts of this thesis will be presented in turn.

4.1 Topological Study

The analysis is done in three parts. First, the overall performance of the algorithm is assessed. Second, the number of occurrences of junctions, hollows and cartilages that touch the sternum are computed. Third and last, the most common combination of junctions, hollows and ribs touching sternum is presented.

4.1.1 Overall algorithm performance

Overall performance is assessed by documenting which costal cartilages are detected for each segmentation and how many costal cartilages are detected per segmentation. In total, 149 out of the 155 segmentations were processed. The results can be seen in figures 4.1a and 4.1b. Based on figure 4.1b it was decided to carry out the topological analysis only on subjects where 17 or more costal cartilages had been detected. This set of 135 segmentations will hence be called the reasonable segmentations.



(a) The figure shows the number of times each distinct costal cartilage was detected, accumulated over 149 subject segmentations. (b) The figure shows the number of costal cartilages that were detected per segmentation. Most commonly, 20 costal cartilages were detected.

Figure 4.1: figures 4.1a and 4.1b show the number of times each distinct costal cartilage were detected, versus the number of costal cartilages that were detected per segmentation.

4.1.2 Occurrences of Hollows, Interchondral and Sternochondral joints

In the following section, the results for each individual feature are presented. The results for the sternocostal joints are found in figure 4.2. Here, the number of sternochondral joints is scaled by the number of occurrences of that costal cartilage among the reasonable segmentations, resulting in a frequency on the y-axes. In general, CC1-CC6 have sternocostal joints, but they are less common for CC6.

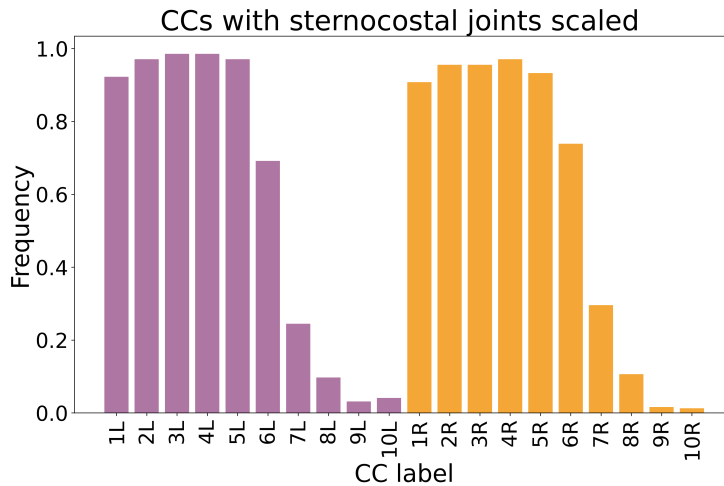


Figure 4.2: The figure shows how frequently each costal cartilage attaches to either the sternum body or the manubrium via a sternocostal joint, across 135 subjects.

Compiling the results for the interchondral joints gives figure 4.3. The number of occurrences are scaled by the number of occurrences of the inferior-most costal car-

tilage of the two that make up the junction, resulting in a fraction.

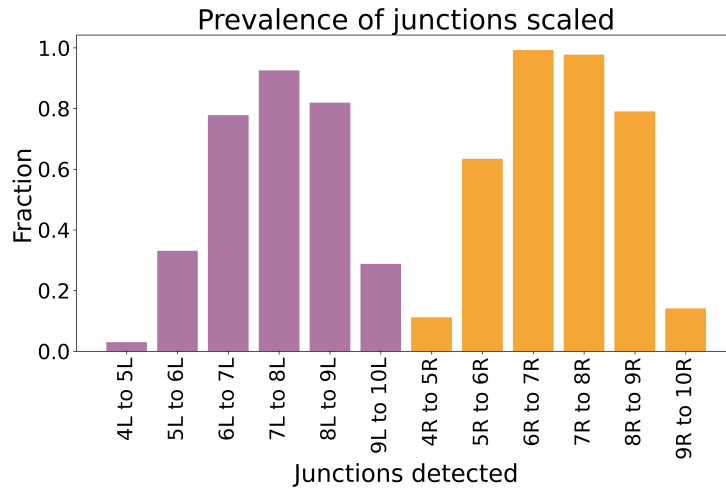
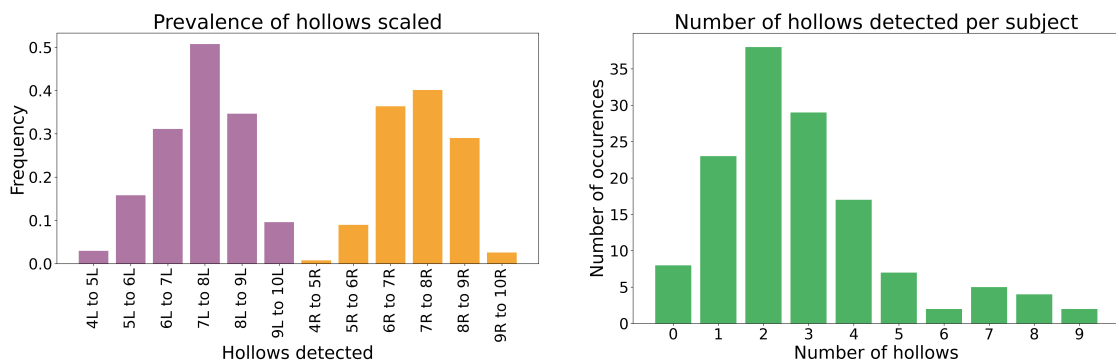


Figure 4.3: The figure shows the prevalence of interchondral joints for each consecutive pair of cartilages across 135 subjects. The fraction is obtained by dividing the number of occurrences of each interchondral joint by the number of occurrences of the inferior-most cartilage in the pair.

The results for the hollows are presented in figure 4.4. Figure 4.4a shows that there is a lateral difference, where a hollow is more common on the left side of the body than on the right. The one exception is hollows between CC6-CC7, which are more common on the right. Figure 4.4b shows it is most common to have two hollows.



(a) The figure shows the frequency of hollows between pairs of cartilages. (b) The figure shows the number of hollows detected per subject segmentation.

Figure 4.4: The figures show the prevalence of hollows in the subject segmentations. The fraction is obtained by dividing the number of occurrences of a hollow by the number of occurrences of the inferior-most cartilage.

4.1.3 The most common topology

The most common topology is determined separately for each feature, and later the combination of features is taken into account. Thresholding the frequency of each feature at 50%, only the most common ones remain, and the most common topology is obtained. The results can be found in table 4.1 and illustrated in figure 4.5.

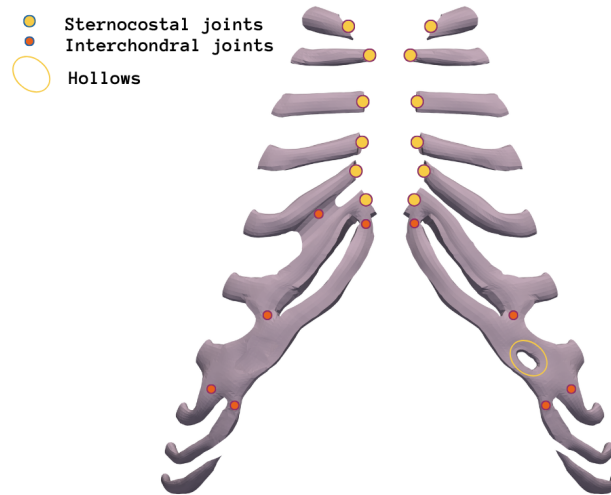


Figure 4.5: The figure shows the most common topology that was found by thresholding each feature at $> 50\%$. The sternocostal joints (CC1-CC6 on left and right side) are marked by yellow circles. There are interchondral joints between CC6-CC7, CC7-CC8 and CC8-CC9 on the left and right side, with an additional interchondral joint between CC5-CC6 on the right side. There is a hollow between CC7-CC8 on the left.

Table 4.1: The table shows the most common features of the costal cartilage. These features were detected in at least 50% of the subject segmentations.

Feature	Left	Right
Sternocostal joints	CC1-CC6	CC1-CC6
Interchondral joints	CC6-CC7, CC7-CC8 and CC8-CC9	CC5-CC6, CC6-CC7, CC7-CC8 and CC8-CC9
Hollows	CC7-CC8	None

The most common combination of junctions, hollows and sternocostal joints was computed. The results show a large variation, as there were three most common combinations of interchondral joints, sternocostal joints and hollows. Each of the three combinations were shared by only two out of the 135 samples that were deemed reasonable.

4.2 Statistical Analysis

In the following section the results from the statistical analysis are presented. The analysis was conducted on the subjects for which 20 or more costochondral joints were detected, which reduced the dataset to 57 subjects. Four additional subjects were removed, leaving 53 subjects. The results include a visualization of the distribution of sex, age, height and BMI in the remaining dataset, a mean shape across these samples, as well as visualizations of what anatomical variations that correspond to each principal component.

4.2.1 Distribution of Subjects

After visual inspection and removing subjects for which the morphing had failed, 53 subjects remained, out of which 29 were male and 24 female. Figure 4.6 shows the distribution of ages among the subjects. Figure 4.7 shows the distribution of height among the subjects. Finally, figure 4.8 shows the distribution of BMI among the subjects. All figures show the distribution among the 53 remaining subjects, and then in the male and female subpopulations, respectively. In addition, each figure also shows the corresponding distribution among the 155 subjects for which the segmentations were made. Table 4.2 shows the mean value for each sex and demographic variable.

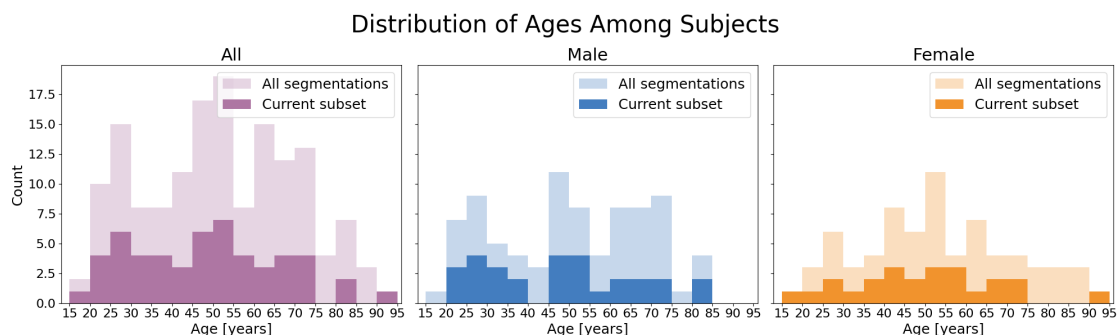


Figure 4.6: The figure shows the distribution of ages among all of the remaining subjects, as well as the remaining male and female subjects, respectively. Each figure also shows the corresponding distribution among the 155 subjects for which the segmentations were made.

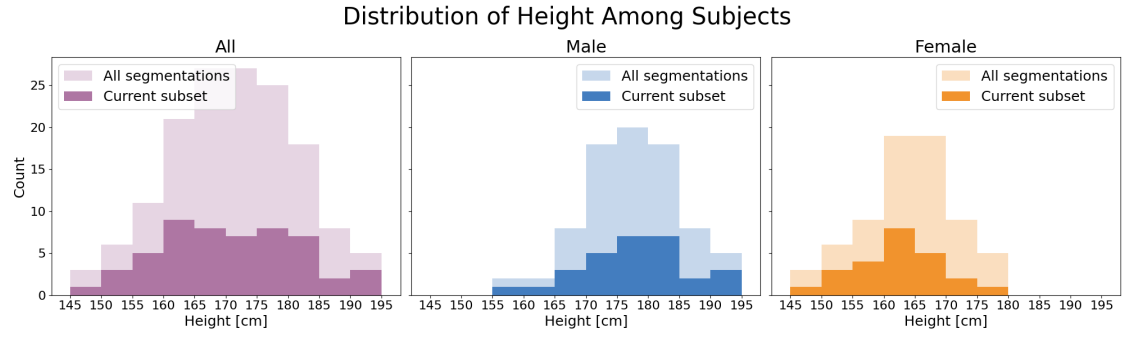


Figure 4.7: The figure shows the distribution of height among all of the remaining subjects, as well as the remaining male and female subjects, respectively. Each figure also shows the corresponding distribution among the 155 subjects for which the segmentations were made.

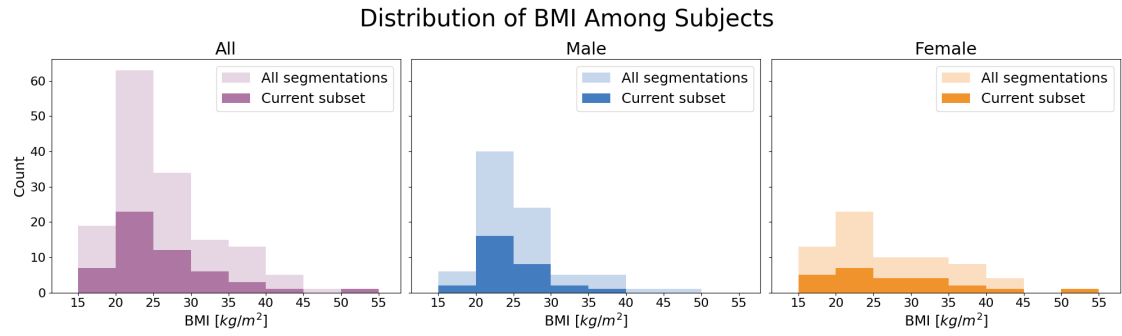


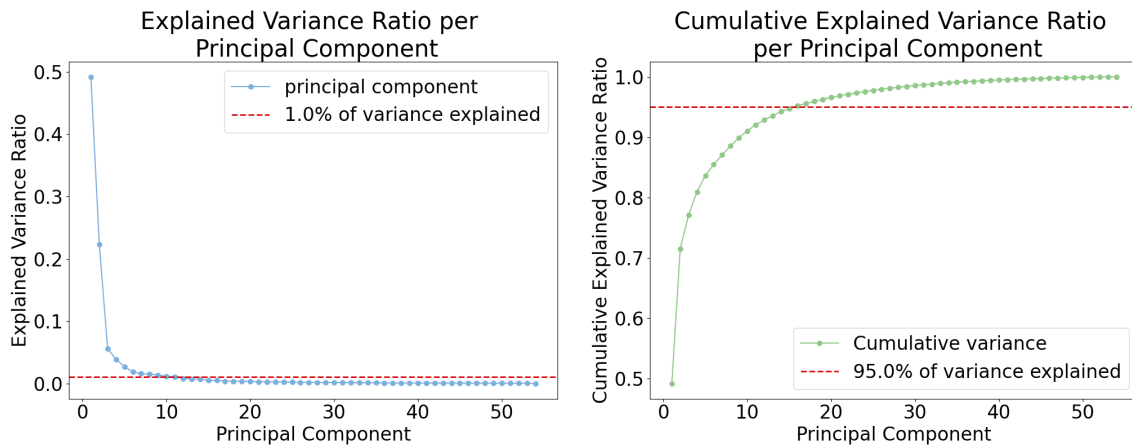
Figure 4.8: The figure shows the distribution of BMI among all of the remaining subjects, as well as the remaining male and female subjects, respectively. Each figure also shows the corresponding distribution among the 155 subjects for which the segmentations were made.

Table 4.2: The table shows the mean values of Age, Height and BMI for all of the remaining subjects, as well as male and female subsets separately.

Sex	Age [years]	Height [cm]	BMI [kg/m^2]
Both	48	170.21	25.97
Male	47	177.56	24.62
Female	49	161.33	27.59

4.2.2 Principal Component Analysis

After conducting PCA, it was found that the first 11 principal components explain at least 1% of the total variance in the dataset, which is illustrated in figure 4.9a. Additionally, figure 4.9b shows that the first 16 PCs are needed to explain in total 95% of the variance in the dataset.



(a) The figure shows the explained variance ratio per principal component. It shows that the first eleven PCs describe at least 1% of the variance each. (b) The figure shows the cumulative explained variance ratio per principal component. It shows that the first 16 PCs are needed to describe at least 95% of the total variance.

Figure 4.9: The figures show the explained variance ratio and cumulative explained variance ratio per principal component.

The mean value of all PCs, corresponding to the mean shape of the subjects in the dataset, is illustrated in figure 4.10.

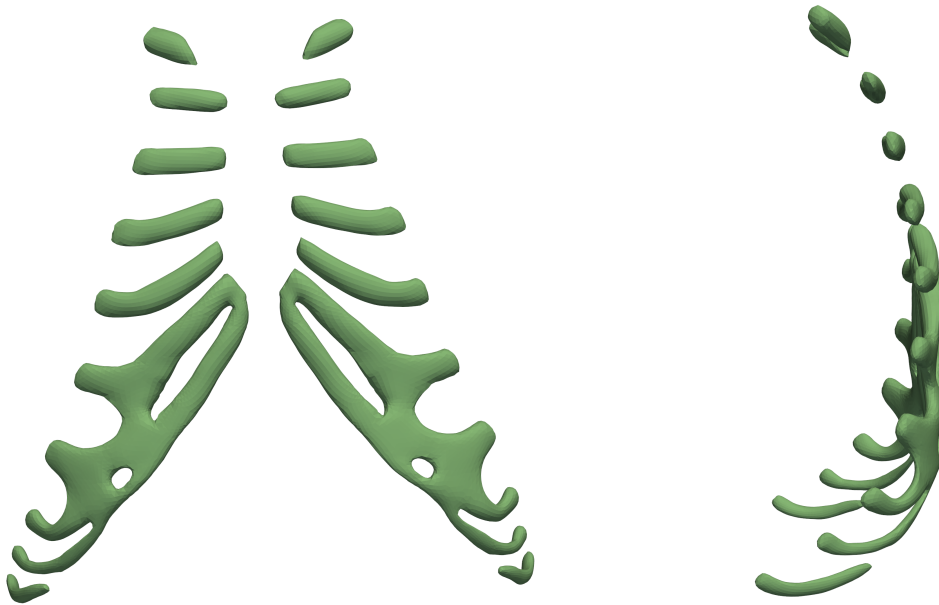
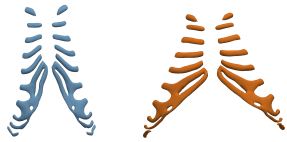
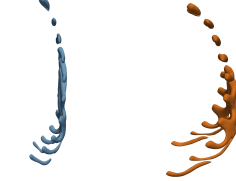
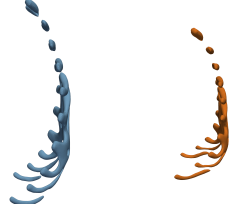
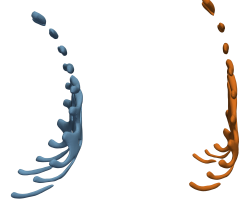
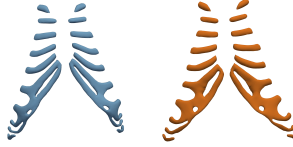
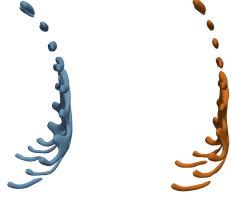


Figure 4.10: The figure shows the mean shape that was obtained by taking the mean value for each feature across all subjects.

Table 4.3: Principal component shape variations for $\pm 2SD$ of PC1 to PC4.

PC	Description	Front view [$-2SD, 2SD$]		Side view [$-2SD, 2SD$]	
PC1 49.20%	Width and curvature.				
PC2 22.33%	Length.				
PC3 5.56%	Smaller or larger spacing between inferior CCs.				
PC4 3.88%	Curvature of superior CC profile. Width and lateral distance for CC6-CC10.				

In table 4.3 the variations per principal component are presented for PC1-PC4. For each PC, it is listed how much of the variance is explained, an interpretation of what the changes are, and a side by side comparison of the mean shape $\pm 2SD$. Corresponding tables for PC5-PC11 can be found in Appendix B.

4.3 Regression

The correlation between the independent variables can be observed in figure 4.11. Only Sex and Height show strong correlation. The model selection yielded the results in figure 4.12, where for each significant principal component the best set of predictor variables are shown along with the average p -value for that set. The regression models were significant only for PC1, PC2 and PC5, with p -values below 0.05. The figure also shows the average R^2 -value over five folds for each of the significant models. The standard deviations are relatively small, and all models gets an average R^2 larger than zero. Note how the best subset included either sex or height, but not both.

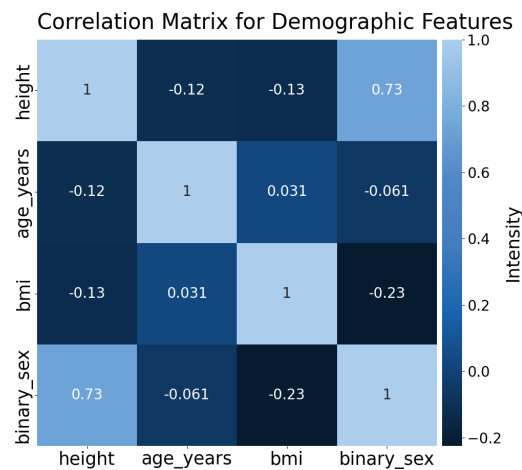


Figure 4.11: The figure shows the Pearson Correlation for the independent variables Age, Sex, Height and BMI. Apart from Sex and Height, there are no strong correlations.

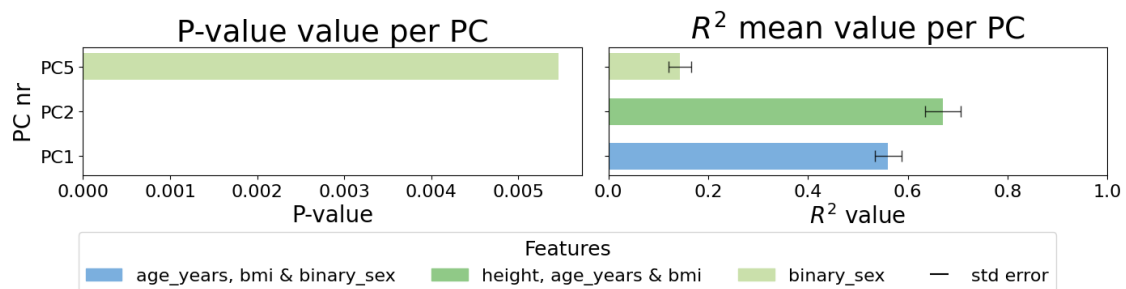


Figure 4.12: The figure shows p -value per significant PC (left) and R^2 -value per significant PC (right). The colors of the bar indicate which subset of independent variables that yielded the highest average AIC across five folds.

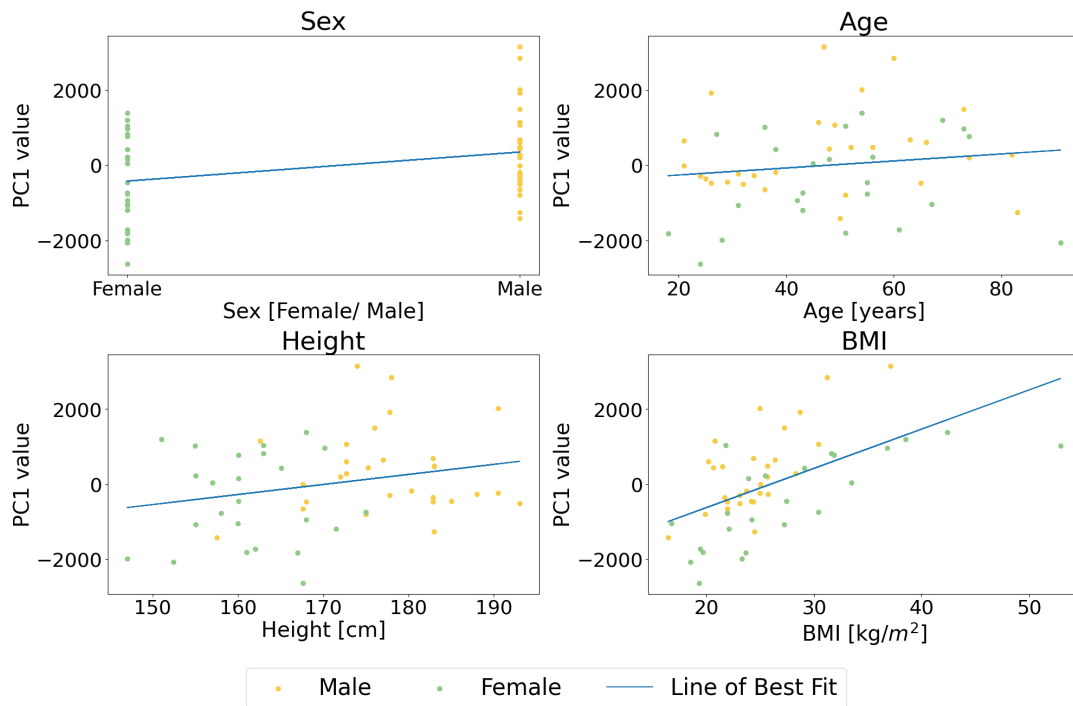


Figure 4.13: The figure shows how values of PC1 correlate with Sex, Age, Height and BMI. Each of the subplots display a positive correlation.

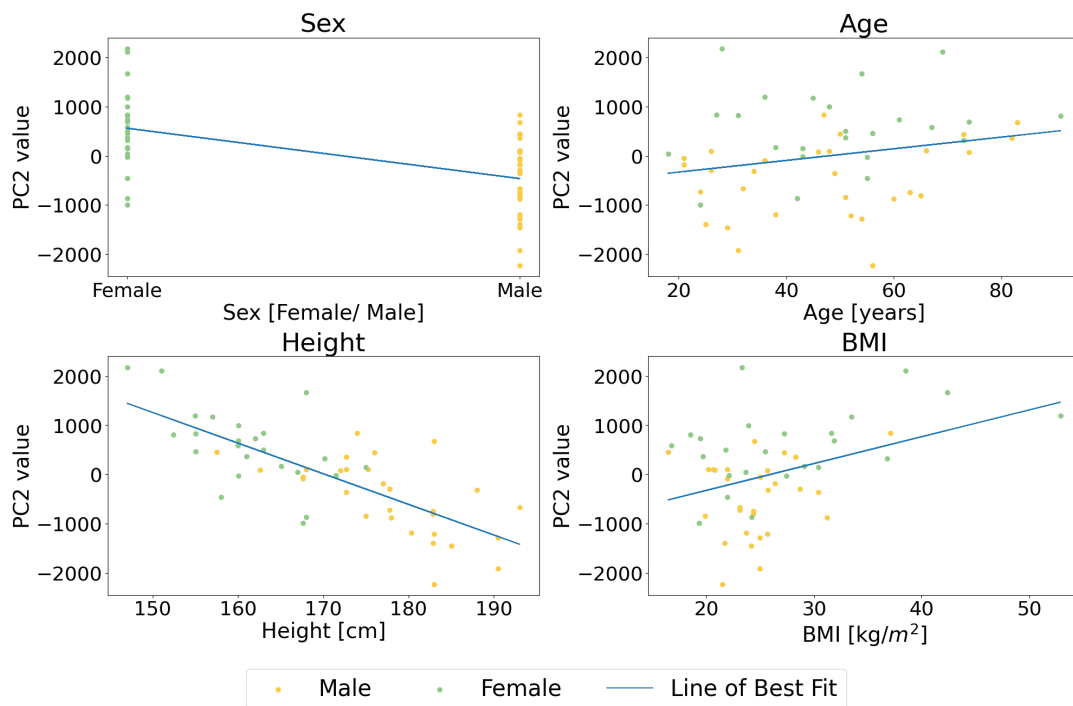


Figure 4.14: The figure shows how values of PC2 correlate with Sex, Age, Height and BMI. The subplots display a positive correlation for Age and BMI, but a negative correlation for Sex and Height.

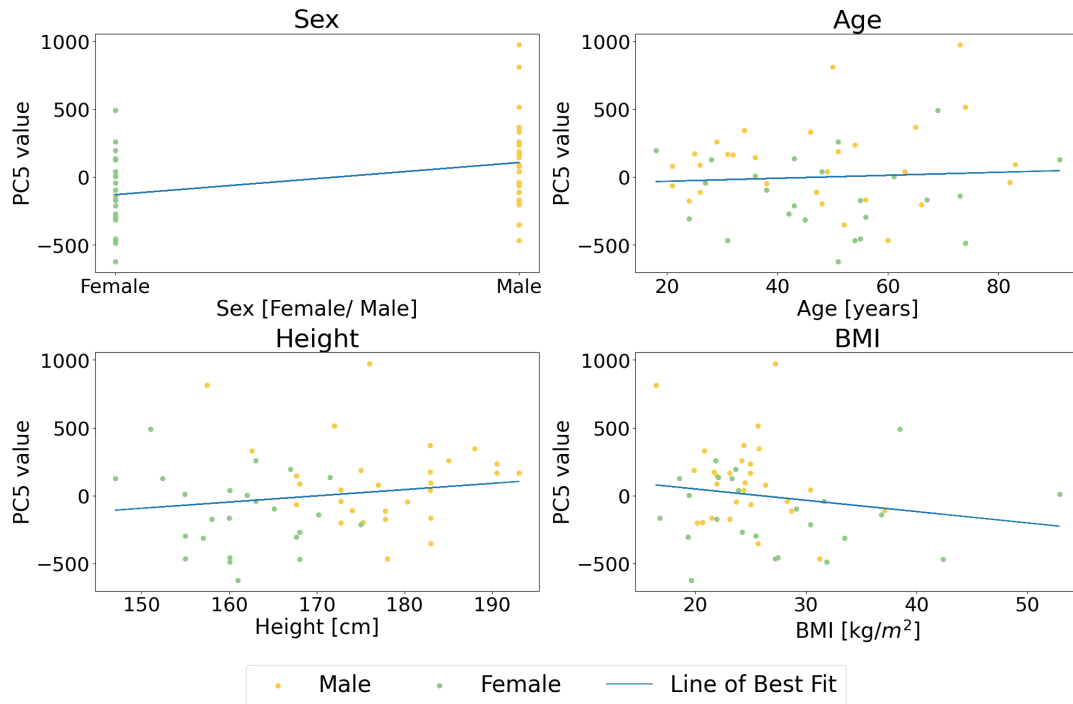


Figure 4.15: The figure shows how values of PC5 correlate with Sex, Age, Height and BMI. The subplots display a positive correlation for Age, Sex and Height, but a negative correlation for BMI.

Figures 4.13, 4.14 and 4.15 show the relationship between each of the independent variables, and PC1, PC2, and PC5 respectively. For PC1, the best predictor was sex, and bmi. Figure 4.13 shows that the scores for PC1 increases with age, and that males tend to have higher scores than females. The figure also shows a positive correlation between PC1-scores and BMI. For PC2, the best predictors were age, height and bmi. Figure 4.14 show a positive correlation between PC-scores and both sex and age. The figure also shows a negative correlation between height and PC-scores. For PC5, the best subset of predictors was sex. Figure 4.15 shows a positive correlation between PC5-values and sex, age and height, but a negative correlation between PC5-values and BMI. Together, PC1, PC2 and PC5 explain 77.09% of the total variance.

4.3.1 Predicting geometries using PC scores

Using the regression models to reconstruct each subject geometry based on pc scores, the model got an R^2 -value of 0.23 for the entire dataset. To understand trends in the population, the variance in shape was studied for the following cases. Figure 4.16 shows the changes that occur when all variables apart from sex is set to the mean in the subject metadata, showing the predicted difference between male and female subjects that share age, height and BMI. Tables 4.4 and 4.5 illustrate how the subject geometry changes when only one variable is changed, and the mean values used are displayed in table 4.2. Table 4.4 shows how female subjects are predicted to be affected by changes in the listed variable. All but the listed variable are set

to the mean of all female subject metadata. In table 4.5 the sex is instead fixed to 'male' and all but the listed variable is set to the mean of all male metadata.

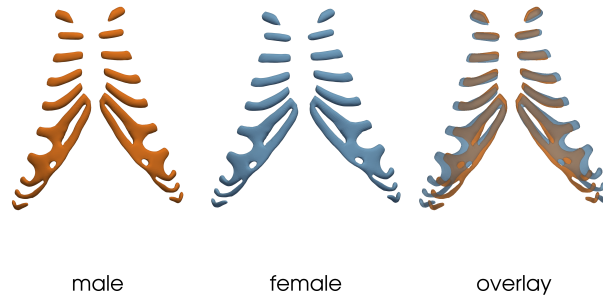


Figure 4.16: The figure shows how the shape and size of the costal cartilage varies between male and female subjects, given that there is no difference in Age, Height or BMI.

Table 4.4: The table shows shape variations for female subjects when all independent variables but one are set to the female dataset mean, seen in table 4.2. The varying independent variable is set to $\pm 1SD$ for either Age, Height or BMI, based on the female dataset. Each row corresponds to changes in the listed variable.

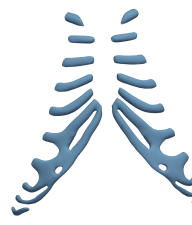
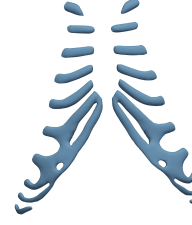
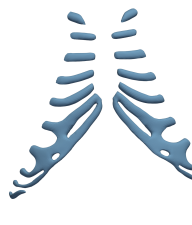
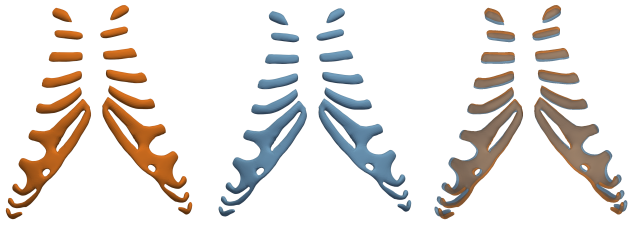
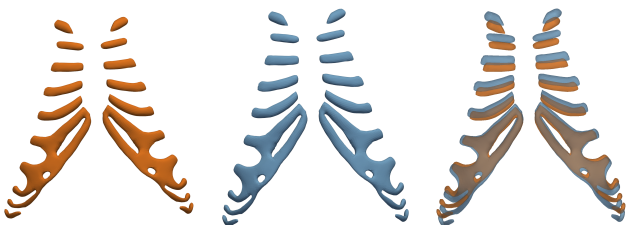
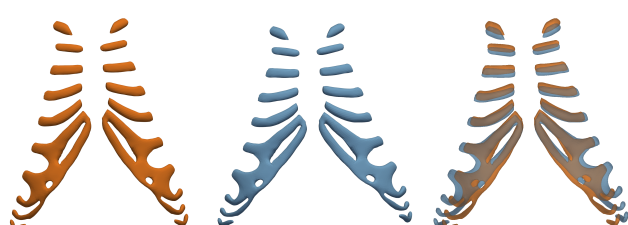
Female dataset mean $\pm 1SD$ for one independent variable		
Variable altered	Description of change	Front view of mean shape $\pm 1SD$
Age	Minimal difference.	
		
		overlay
Height	Vertical spacing increases with height.	
		
		overlay
BMI	Width increases with BMI. Height decreases.	
		
		overlay

Table 4.5: The table shows shape variations for male subjects when all independent variables but one are set to the male dataset mean, seen in table 4.2. The varying independent variable is set to $\pm 1SD$ for either Age, Height or BMI, based on the male dataset. Each row corresponds to changes in the listed variable.

Male dataset mean $\pm 1SD$ for one independent variable		
Variable altered	Description of change	Front view of mean shape $\pm 1SD$
Age	Minimal difference.	 -1SD +1SD overlay
Height	Vertical spacing increases with height.	 -1SD +1SD overlay
BMI	Width increases with BMI. Height decreases.	 -1SD +1SD overlay

5

Discussion

In order to contextualize and analyze the results presented in the previous chapter, a discussion will be carried out. The topology study, statistical analysis and linear regression will be discussed separately and in chronological order.

5.1 Data inconsistencies

As previously mentioned there are inconsistencies in the data, which makes it hard to ensure that the results correspond to the actual data, and not to what the data is supposed, or assumed, to look like. The inconsistencies were of different kinds. There were spikes on the surfaces, due to an inaccurate prediction of certain vertices. An example of this was shown in figure 3.2a. The spikes were removed by applying Laplacian smoothing to the subject meshes, which shrinks the volume inwards. A side effect of this is that the surface of each cartilage is altered, and that the cross-sectional area of the cartilages is shrunk slightly. Preserving these traits is not crucial in order to determine the topology. However, as the same volumes were used for the statistical analysis, the smaller cross-sectional area is likely carried into those results.

The segmentations were made for CC1-CC10 and had a DICE-score of ≈ 0.87 . This could entail both missing and superfluous cartilage in the segmentations. It is possible that some cartilages, and thereby joints, were not present in the segmentations, and subsequently could not be found by the algorithms. To mitigate floating islands of cartilage, the volumes were screened for size and distance to neighboring volumes. It is possible that this screening did not capture all floating islands, or conversely, that it did remove some CC10.

5.2 Topological Study

The topology study was conducted without any ground truth, making it important to discuss the reliability of the results. To this end, the results will be discussed in relation to the anatomy of the costal cartilage as described in literature, and the algorithm development.

5.2.1 Correspondence between results and anatomical features

In the results section it was shown that the algorithm generally detected costochondral joints for CC1-CC9. For approximately half of the subjects did the algorithm also find CC10. One possible reason for the detection of CC10 being low is that CC10 were missing in several subject segmentations. This could either be a consequence of lacking data quality or data cleaning. Generally, CC1-CC7 are expected to have sternocostal joints (Haase & Shaikh, 2024). Sternocostal joints were found for $> 90\%$ of CC1-CC5, for approximately 70% of CC6, and 20% of CC7. The low rate of detection of sternocostal joints in CC7 can possibly be explained by the fact CC6 and CC7 attach to the sternum very closely to one another. This could cause the algorithm to interpret the topology as CC6 attaching to the sternum, and CC7 only attaching to CC6.

In literature, CC8-CC10 are described having only interchondral joints. For CC6-CC7 sternocostal joints are expected, but interchondral joints are not excluded (Haase & Shaikh, 2024). The detected junctions correspond well to the junctions described in literature, apart from the interchondral joint between CC9-CC10, which was found to be less common. It is worth noting that the results are shown as fractions, where the total number of detections of an interchondral joint is scaled by the prevalence of the inferior-most cartilage. This means that the results are not an effect of potentially having few CC10 present in the segmentations, but show that the interchondral joint is rare among the detected CC10. Overall, despite the uncertainties stemming from the data, the results seem to indicate that the algorithm works well and that no systemic errors should be expected.

5.2.2 Algorithm construction and potential data leakage

The algorithm was designed to find and name costochondral joints, label cartilages, and subsequently identify and count hollows, junctions and sternocostal joints. During the development, the algorithm was tested on the data at hand. This means that the algorithm is built to cater to the specifics of this dataset, and should the dataset not be representative of the population at whole the algorithm could fail when encountering new data. Nevertheless, far from every subject was inspected individually during development, and the risk of having overfit the algorithm to this specific dataset is expected to be small.

Not knowing how many sternocostal- or costochondral joints that exist in each subject segmentation, it is tricky to figure out how many should be found. Have we found 18 because there are 18, or have we found 18 because we've missed two? Conversely, if we've found 20, does that mean those are the correct points? Questions like these necessitated stricter geometric requirements for the joints, rather than restricting the number of joints found. Algorithm development was on the large performed using visual inspection of the results, rather than mathematical criteria.

5.3 Statistical Analysis

In this section, the results from the PCA will be discussed. It will cover how the choice of template and morphing algorithm affects the results, as well as a discussion on how well the dataset corresponds to the general population.

5.3.1 ICP and the choice of template

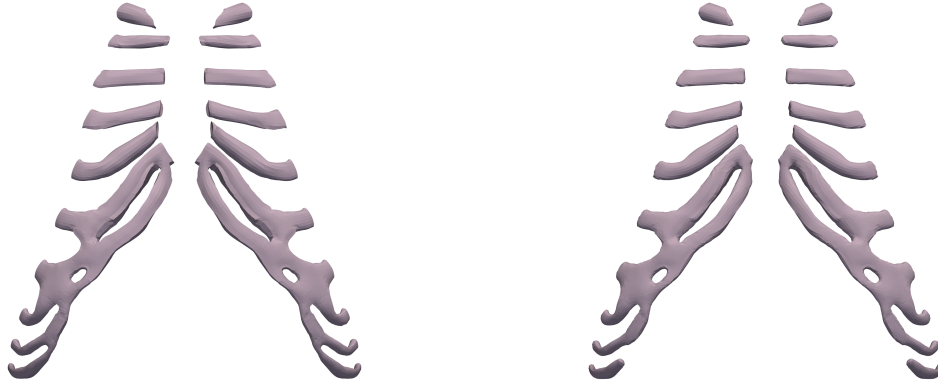
The properties of the template are inherently transferred to the morphed template when using ICP. While the algorithm iteratively finds pairs of closest points in target and template and subsequently deforms the template to minimize the distance, it will not alter the properties of the template. That is, if there is a hollow in the template, this hollow will not disappear simply because of an absence of hollow in the target. This is mainly due to both ICP and CPD treating the meshes as point clouds. In a point cloud, the connectivity is not considered, and a topological entity such as a hollow loses meaning.

Once a fairly accurate surface matching has been accomplished, the surface overlap can be tuned using nonrigid ICP. However, for a complicated and highly varied geometry such as the coastal cartilage, the surface matching is generally not good enough for the tuning to not break the connectivity of the mesh. In this project, tuning was applied cautiously. While the overlap was improved slightly, it is by no means perfect. To improve surface matching possibilities, it was decided to smooth the template mesh and detach CC10 from CC9, in accordance with the results from the topology study. A side by side comparison between the original template and the final version can be seen in figure 5.1.

The consequence of imperfect surface matching is that all morphed templates will carry more similarity to the template than the subject geometries do. As previously mentioned, the meshes were smoothed before being used as targets. Therefore, the statistical analysis is less reliable in regards to surface details and cross-sectional area, while still capturing the general variations in size and shape as well as path from costochondral joint to sternum.

5.3.2 Data quality and sample size

Due to the varying topology between the subjects, pre-morphing the templates aided the performance of the ICP and CPD algorithms, as it provided an initial rigid transformation of the template. Since the algorithms process point clouds, not the surface of the volume, they risk collapsing two cartilages into one, which proved true for segmentations with fewer CCs than the template. The subject segmentations that were utilized for the statistical analysis were therefore only the segmentations for which 20 or more costochondral joints were detected in the topology study. However, there were only 57 subjects that fulfilled these criteria, and three additional outliers were removed after visual inspection of the morphed templates. One subject was removed since their weight and height data was unavailable. Out of the remaining



(a) The figure shows the initial template mesh. (b) The figure shows the smoothed template mesh that was used for the final analysis. Note the separation between the inferior-most cartilages, CC9 and CC10.

Figure 5.1: The figure shows a comparison between the template mesh that was originally used (figure 5.1a) and the template mesh that was used for the final analysis (figure 5.1b). Note how the final template is smoother than the original, and has CC10 detached from CC9.

subjects, 29 were male and 24 were female, which is an imbalance that was seen already among the 155 segmentations, where 82 were male and 71 female.

Despite removing a substantial amount of data, the distribution of subject meta-data seems to approximate the distribution in the American population (Fryar et al., 2021). This is true for all variables but BMI, which is noticeably lower in the dataset used in this project. In table 5.1 the dataset means and medians are compared to the median values presented by Fryar et al., 2021. However, figure 4.8 shows that the BMI distribution of the 53 subjects is very similar to that of the 155 segmentations. The same goes for height, which was showed in figure 4.7. Inspecting 4.6, we can see a lack of elderly subjects that is carried over into the in the smaller subset. However, it seems reasonable to conclude that no groups have been disproportionately excluded from the study.

Table 5.1: Comparison between reference values reported by Fryar et al., 2021 and the dataset statistics for male and female subjects.

	Male			Female		
	Mean	Median	Reference	Mean	Median	Reference
BMI [kg/m ²]	24.6	24.4	28.5	27.6	24.9	28.4
Height [cm]	177.8	177.8	175.4	161.3	160.5	161.3

The number of subjects are few, and with the costal cartilage being a complex geometry with high degrees of individual variation, there are drawbacks to small sample sizes. The analysis becomes more sensitive to outliers. It is also possible that there exists trends in shape variation that could not be captured across so few samples. Lastly, since the distribution of sex is 29 males and 24 females, the mean shape will be slightly more representative of the male population than the female.

5.4 Regression

Predicting the PC values of each PC using the demographic variables yielded three significant PCs, each with an average $0.1 \leq R^2 \leq 0.7$, as was seen in figure 4.12. Although the average R^2 was greater than 50% for PC1 and PC2, meaning they explain a majority of the variance in PC-scores, using them for predicting individual subject geometries was not as successful. The R^2 for the entire regression model was as previously mentioned 0.23, meaning the reconstructions are better representations of the subject geometries than the mean shape is. This can be compared to the R^2 of 0.29 obtained by Brynskog et al., 2021 for the pelvis, which is another complex geometry.

5.4.1 Choice of regression model, correlation and possible interaction effects

For this project, Statsmodels' linear regression model was fit to the data using Ordinary Least Squares (Seabold & Perktold, 2010). Linear regression is not the only type of regression that could have been employed to investigate the relation between the independent and dependent variables. When visually inspecting figures 4.13, 4.14 and 4.15, it does not seem that an exponential, polynomial or sinusoidal model would have described the correlation more accurately. Figure 4.11, displays a strong correlation between sex and height. In this case, the feature selection for the significant PCs reduced the best subset to contain either sex or height, mitigating any potential effect the correlation could have had on the results.

Interaction occurs when the relationship between an independent variable x_1 and a dependent variable y is affected by a third variable x_2 (Montgomery et al., 2012b). Interaction was not included in the regression model obtained in this thesis, nor were the possible interaction effects explored. In other words the relationships between dependent and independent variables were modeled as if interaction did not exist.

5.4.2 Specifics of the dataset and the subsequent effects on the results

As previously mentioned, the subject segmentations were smoothed and the surface matching has room for improvement. This means that the morphed meshes will inherit traits of the template mesh, or analogously, that some features in the target

geometries are not captured by the morphing algorithms. This is likely to inflate the R^2 values, as similarity has basically been enforced in the morphing step. Despite the reconstructions being a better representation of subject geometry than the mean shape, over 75% of the variance remains unexplained.

The anatomical variations seem to be correlated to sex, height and BMI, but not age. It has been shown that calcification of a cartilage increases with age (S. Holcombe et al., 2017). Subjects of ages 75 and up are underrepresented in the segmentations, and increased calcification is listed as one possible reason (Alexandersson & Moln , 2025). The fact that elderly are underrepresented in the available data could decrease a possible age dependency, rendering it insignificant.

5.4.3 Prediction of population trends, outliers and consistency with related studies

Regarding the variations at population level, there are trends that can be observed. Regardless of sex, the vertical spacing between cartilages increases with height, and decreases with BMI, while the width of the costal cartilage increases with BMI. Between male and female subjects of same age, height and BMI, the main difference consists of females having wider costal cartilages. There is one outlier among the female subjects, with a BMI of 55. Since BMI seems to positively correlate with a wider thorax, and there are fewer female than male subjects, the initial suspicion was that the presence of this outlier could be partially responsible for the difference in costal cartilage width between the sexes. However, this hypothesis was tested by re-fitting the regression model to a subset of the dataset where all subjects had a BMI < 30, and using the newly fitted model to predict the subset. Having a BMI of 30 and above is classified as obesity (Anekwe et al., 2020). This setup showed the same trends as the previous model, which indicates that the outlier is not the sole reason we see a correlation between BMI and costal cartilage width.

These results can be compared to the ones presented for the ribs by Shi et al., 2014, who showed that the vertical spacing between ribs increases with height, and that the width of the ribcage increases with BMI. In the same study, it was also found that for males, the ribs are longer (i.e. extend further towards the sternum) than for females. Lynch et al., 2023 found, among other things, that there were significant positive correlations between subject height and thoracic height, and subject weight and thoracic width. They also found that the male ribcage is both taller, deeper and wider than the female. All of the above results are coherent with the findings of this thesis. While no previous studies on the topology and variance of the costal cartilage were found in the literature, the fact that these results align with similar studies certainly adds to their credibility.

5.5 Generalizability and use of results

The distribution of Sex, Age, Height and BMI in data that was utilized for this project was found to be close to the median reported by Fryar et al., 2021. However, both the dataset and the reference data originate from the U.S, which means they are not necessarily representative of people of different origin. This is not to say the results are not valuable, but that there are potential limitations to their accuracy. As has been noted previously, there are some limitations to the ages that are included in the regression model. The model obtained during this thesis is for example not representative of kids, as there were no such subjects present in the data.

Given that the distribution of subjects is representative of a general American population, the mean shape, along with the regression model that were obtained in this project, can be used to predict population trends for that demographic. Apart from the results already showed, it would for example be possible to predict the shape of costal cartilage for a person of average height and BMI for each age category, which could be useful as BMI tends to vary between age groups (Fryar et al., 2021). Further, predicting the shape of the costal cartilage implies predicting exactly where in the thorax the sternum is situated. In broad terms, knowing the most common topology and being able to predict the shape and size of the costal cartilage given demographic variables can contribute to a more accurate mathematical description of the thorax, which in turn can contribute to a more accurate prediction of injuries.

5.6 Future Work

As previously mentioned, to date and to the best knowledge of the author, this study is the only one of its kind. This entails that there are lots of opportunities for future work!

The topology study was conducted without ground truth and algorithm performance was verified visually. Although not every sample was inspected during algorithm development, no test set was set aside for verification, risking data leakage. A future work could be to verify algorithm behavior on unseen data. Another data-related future task stems from the fact that there were few elderly present in the dataset. Similarly, there were no children present in the dataset. It would be interesting to see if and how an increased portion of elderly in the dataset, or the inclusion of children, would affect the regression model and the subsequent predictions.

Although 155 segmentations were available, in the end only 53 segmentations were utilized for PCA and regression. The reason was that the algorithm did not detect 20 costochondral joints for all segmentations, either due to them not being present in the segmentation or due to some flaw in the algorithm. Looking further into this issue could lead to a larger dataset being made available, which could add to the results.

While the mean shape obtained during this project has both a surface and a cross sectional area, the accuracy of these measurements were not the main focus of this thesis. A better surface matching could possibly be obtained by using fewer samples, and either use traditional landmarks, or focus on tuning the point cloud algorithms.

No interaction effects were included, nor investigated, when building the regression model that this thesis work resulted in. Investigating whether interaction effects exist, and how they potentially affect the regression model, would be a beneficial next step for understanding how the costal cartilage varies in the population.

It was not within scope for this thesis to look into how the topology, shape and size of the costal cartilage affect injury risk. However, it would certainly be an exciting next step to investigate if there is any one group that is more prone or susceptible to injury, and investigate whether that can be linked to costal cartilage topology, shape or size.

6

Conclusion

The aims of this thesis were threefold. The first aim was to study the topology of the costal cartilage and determine whether there was any most common topology and what features that topology displayed. The second aim was to study the variance of the costal cartilage in the population. The final aim was to investigate whether there were any correlations between this variance, and the Sex, Age, Height or BMI of the subjects. The topology study was conducted using visual inspection to verify the results, which showed both substantial individual variation and common trends in costal cartilage topology. The statistical analysis showed that variations in costal cartilage length and width account for the absolute majority of the total variance, and the regression analysis showed that these variations can be partly attributed to the demographic variables Sex, Age, Height and BMI. The results are consistent with previous findings on thorax and rib variation, although any direct comparison is limited by the apparent absence of prior studies specifically addressing costal cartilage variation. This author hopes that the results presented can contribute to a better understanding of the costal cartilage, and add to the improvement of the thorax of the SAFER HBM.

Bibliography

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Alexandersson, A., & Molnő, A. (2025). *Deep Learning for Segmentation and Landmarking of the Human Anatomy* [Master’s thesis, Chalmers University of Technology].
- Amberg, B., Romdhani, S., & Vetter, T. (2007). Optimal Step Nonrigid ICP Algorithms for Surface Registration. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8. <https://doi.org/10.1109/CVPR.2007.383165>
- Anekwe, C. V., Jarrell, A. R., Townsend, M. J., Gaudier, G. I. G., Hiserodt, J. M. H., & Stanford, F. C. (2020). Socioeconomics of Obesity. *Current Obesity Reports*, (9), 272–279. <https://doi.org/10.1007/s13679-020-00398-7>
- Besl, P. J., & McKay, N. D. (1992). Method for registration of 3-D shapes. *Sensor Fusion IV: Control Paradigms and Data Structures*, 1611, 586–606. <https://doi.org/10.1117/12.57955>
- Brynskog, E., Iraeus, J., Reed, M. P., & Davidsson, J. (2021). Predicting pelvis geometry using a morphometric model with overall anthropometric variables. *Journal of Biomechanics*, 126. <https://doi.org/10.1016/j.jbiomech.2021.110633>
- Dawson-Haggerty, M., & Trimesh Contributors. (2019). *Trimesh 4.11.1 (software)* [Retrieved: 11 Feb 2026]. <https://trimesh.org/index.html>
- Devore, J. L., Berk, K. N., & Carlton, M. A. (2021a). Regression and Correlation. In *Modern Mathematical Statistics with Applications* (pp. 703–822). Springer.
- Devore, J. L., Berk, K. N., & Carlton, M. A. (2021b). Tests of Hypotheses Based on a Single Sample. In *Modern Mathematical Statistics with Applications* (pp. 501–563). Springer.
- Edgar, H., Daneshvari Berry, S., Moes, E., Adolphi, N., Bridges, P., & Nolte, K. (2020). New Mexico Decedent Image Database. <https://doi.org/10.25827/5s8c-n515>
- Emmert-Streib, F., & Dehmer, M. (2019). Evaluation of Regression Models: Model Assessment, Model Selection and Generalization Error. *Machine Learning and Knowledge Extraction*, 1(1), 521–551. <https://doi.org/10.3390/make1010032>
- Fahse, N., Millard, M., Kempter, F., Maier, S., Roller, M., & Fehr, J. (2023). Dynamic human body models in vehicle safety: An overview. *GAMM-Mitteilungen*, 46(2). <https://doi.org/10.1002/gamm.202300007>

- Fryar, C. D., Carroll, M. D., Gu, Q., Afful, J., & Ogden, C. L. (2021). Anthropometric Reference Data for Children and Adults: United States, 2015–2018. *National Center for Health Statistics. Vital Health Stat*, 3(46).
- Gewers, F., Ferreira, G., De Arruda, H., Silva, F., Comin, C., Amancio, D., & Costa, L. D. (2021). Principal Component Analysis: A Natural Approach to Data Exploration. *ACM Computing Surveys*, 54(4). <https://doi.org/10.1145/3447755>
- Gower, C. (1975). Generalized Procrustes Analysis. *Psychometrika*, 40(1), 33–51.
- Haase, D. R., & Shaikh, H. S. (2024). Anatomy of the Ribs, Sternum, and Costal Margin. *Journal of Orthopaedic Trauma*, 38(12S), 1–6. <https://doi.org/10.1097/BOT.0000000000002919>
- Hagberg, A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th Python in Science Conference (SciPy2008)* (pp. 11–15).
- Holcombe, S., Ejima, S., & Wang, S. (2017). Calcification of Costal Cartilage in the Adult Rib Cage. *Proceedings of the 2017 International IRCOBI Conference on the Biomechanics of Injury*, 737–744.
- Holcombe, S. A., Huang, Y., & Derstine, B. A. (2024). Population trends in human rib cross-sectional shapes. *Journal of Anatomy*, 244, 792–802. <https://doi.org/10.1111/joa.13999>
- Hwang, E., Hallman, J., Klein, K., Rupp, J., Reed, M., & Hu, J. (2016). *Rapid Development of Diverse Human Body Models for Crash Simulations through Mesh Morphing* (tech. rep.). SAE International. <https://doi.org/10.4271/2016-01-1491>
- Iraeus, J., Brynskog, E., John, J., Östh, J., Pipkorn, B., & Davidsson, J. (2024). *Development of SAFER HBM v11 - Report 2024:06, Version 1.1* (tech. rep.). Chalmers University of Technology - Department of Mechanics and Maritime Sciences.
- John, J., Klug, C., Kranjec, M., Svenning, E., & Iraeus, J. (2022). Hello, world! VIVA+: A human body model lineup to evaluate sex-differences in crash protection. *Frontiers in Bioengineering and Biotechnology*, 10. <https://doi.org/10.3389/fbioe.2022.918904>
- Jović, A., Brkić, K., & Bogunović, N. (2015). A review of feature selection methods with applications. *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1200–1205. <https://doi.org/10.1109/MIPRO.2015.7160458>
- Kavitha, T., Mehlhorn, K., Michail, D., & Paluch, K. E. (2008). An $\tilde{O}(m^2n)$ Algorithm for Minimum Cycle Basis of Graphs. *Algorithmica*, 52, 333–349. <https://doi.org/10.1007/s00453-007-9064-z>
- Lynch, S. D., Taylor, S. L., Greene, K. A., Devane, K. S., & Weaver, A. A. (2023). Characterizing thoracic morphology variation to develop representative 3D models for applications in chest trauma. *Computers in Biology and Medicine*, 163. <https://doi.org/10.1016/j.combiomed.2023.107211>

- Maragos, P. A., & Schafer, R. W. (1984). Morphological skeleton representation and coding of binary images. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5), 1228–1244.
- Marieb, E. N., & Hoehn, K. (2019). The Human Body: An Orientation. In *Human Anatomy & Physiology*. Pearson education.
- Mokhtar, A. A., & Hyncik, L. (2025). A comprehensive review of human body model in different crash scenarios: active and passive models. *International Journal of Crashworthiness*, 30(1), 1–13. <https://doi.org/10.1080/13588265.2024.2352242>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012a). Multicollinearity. In *Introduction to Linear Regression Analysis* (pp. 190–216). John Wiley & Sons.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012b). Multiple Linear Regression. In *Introduction to Linear Regression Analysis* (pp. 56–91). John Wiley & Sons.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012c). Simple Linear Regression. In *Introduction to Linear Regression Analysis* (pp. 24–55). John Wiley & Sons.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012d). Validation of Regression Models. In *Introduction to Linear Regression Analysis* (pp. 243–254). John Wiley & Sons.
- Myronenko, A., & Song, X. (2010). Point Set Registration: Coherent Point Drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12), 2262–2275. <https://doi.org/10.1109/TPAMI.2010.46>
- Nummela, M. T., Pyhälä, T. T., Bensch, F. V., Heinänen, M. T., & Koskinen, S. K. (2022). Costal cartilage fractures in blunt polytrauma patients — a prospective clinical and radiological follow-up study. *Emergency Radiology*, 29, 845–854. <https://doi.org/10.1007/s10140-022-02066-w>
- Nunez-Iglesias, J., Blanch, A. J., Looker, O., Dixon, M. W., & Tilley, L. (2018). A new Python library to analyse skeleton images confirms malaria parasite remodelling of the red blood cell membrane skeleton. *PeerJ*, 6. <https://doi.org/10.7717/peerj.4312>
- Pazokifard, B., Sowmya, A., & Moses, D. (2015). Automatic patient-customised 3D reconstruction of human costal cartilage from lung MDCT dataset. *International Journal of Computer Assisted Radiology and Surgery*, 10, 465–472. <https://doi.org/10.1007/s11548-014-1086-9>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Schneider, W., Robbins, D. H., Pflug, M. A., & Snyder, R. (1983). *Anthropometry of Motor Vehicle Occupants, Vol. 2*. National Highway Traffic Safety Administration.
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. *9th Python in Science Conference (SciPy 2010)*, 92–96.

- Shi, X., Cao, L., Reed, M. P., Rupp, J. D., Hoff, C. N., & Hu, J. (2014). A statistical human rib cage geometry model accounting for variations by age, sex, stature and body mass index. *Journal of Biomechanics*, 47(10), 2277–2285. <https://doi.org/10.1016/j.jbiomech.2014.04.045>
- Slice, D. E. (2005). Modern Morphometrics. In *Modern Morphometrics in Physical Anthropology*. Kluwer Academic / Plenum Publishers.
- Trafikverket. (2020). *This is Vision Zero* [Accessed: 2026-01-27]. <https://bransch.trafikverket.se/en/startpage/operations/Operations-road/vision-zero-academy/This-is-Vision-Zero/>
- van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T., & scikit-image contributors. (2014). scikit-image: Image processing in Python. *PeerJ*. <https://doi.org/10.7717/peerj.453>
- Xu, T., Sheng, X., Zhang, T., Liu, H., Liang, X., & Ding, A. (2018). Development and Validation of Dummies and Human Models Used in Crash Test. *Applied Bionics and Biomechanics*, 2018. <https://doi.org/10.1155/2018/3832850>
- Yeh, I.-C., Lin, C.-H., Sorkine, O., & Lee, T.-Y. (2011). Template-Based 3D Model Fitting Using Dual-Domain Relaxation. *IEEE Transactions on Visualization and Computer Graphics*, 17(8), 1178–1190. <https://doi.org/10.1109/TVCG.2010.124>
- Zhang, T. Y., & Suen, C. Y. (1984). A Fast Parallel Algorithm for Thinning Digital Patterns. *Communications of the ACM*, 27(3).

A

Python & Matlab Packages

Table A.1: The table shows the Python libraries utilized during the project.

Library	Version
Numpy	2.3.5
NetworkX	3.6.1
Pandas	3.0.0
Scikit-image	0.26.0
Scikit-learn	1.8.0
Scipy	1.17.0
Skan	0.13.1
Trimesh	4.11.1

Table A.2: The table shows the Matlab Add-Ons utilized during the project.

Add-on	Version	Creator
Anatomic Shape Modelling Toolbox	-	Oscar Denton, Chalmers
Computer Vision Toolbox	R2025b	MathWorks
Deep Learning Toolbox	R2025b	MathWorks
Image Processing Toolbox	R2025b	MathWorks
Iso2Mesh	1.9.0	Qianqian Fang
Lidar Toolbox	R2025b	MathWorks
Statistics and Machine Learning Toolbox	R2025b	MathWorks

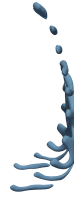
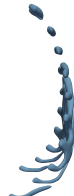
B

Shape variations for PC5-PC11

Table B.1: Principal component shape variations for $\pm 2SD$ of PC5 to PC8.

PC	Description	Front view $[-2SD, 2SD]$		Side view $[-2SD, 2SD]$	
PC5 2.69%	Width of sternum. Curvature of CC10.				
PC7 1.55%	Lateral imbalance, vertical tilt.				
PC8 1.5%	Lateral curvature, rotation along vertical axis.				

Table B.2: Principal component shape variations for $\pm 2SD$ of PC9 to PC11.

PC	Description	Front view $[-2SD, 2SD]$		Side view $[-2SD, 2SD]$	
PC9 1.36%	Outwards rotation of the inferior CCs.				
PC10 1.13%	Width of lower sternum body.				
PC11 1.03%	Lateral distance between pairs of CC6.				



CHALMERS
UNIVERSITY OF TECHNOLOGY