3D LUNG VOLUME CALCULATION FOR SCOLIOSIS

BARRIERS TO HEALTHCARE FOR THE MARSHALLESE IN NORTHWEST ARKANSAS

An Undergraduate Thesis Portfolio
Presented to the Faculty of the
School of Engineering and Applied Science
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Biomedical Engineering

By

Sam Schach

May 6, 2021

3D Lung Volume Calculation for Scoliosis Patients

A Research Paper submitted to the Department of Biomedical Engineering
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Biomedical Engineering

Tony Albini, Will Farley, and Sam Schach

By

Sam Schach

March 25, 2021

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISOR
Keith Bachmann Department of Orthopedic Surgery

Abstract

Early onset scoliosis is a three dimensional curvature of the spine that occurs in patients 10 years or younger. Physicians tend to pursue non-surgical options first. If these are unsuccessful, surgical intervention becomes necessary. In scoliosis cases, physicians use total lung capacity as the metric for determining the optimal time for surgery. Current standards of treatment do not offer an accurate way for physicians to measure total lung capacity for early onset scoliosis patients. This project offers a potential solution to this problem. The total lung capacity can be calculated by subtracting the mediastinum volume from the rib cage volume. A multivariate linear regression was used to create a predictive equation with patient demographics as input variables to predict the mediastinum volume. This equation predicted mediastinum volume at a higher accuracy compared to previous work and had a multiple R^2 of 0.87. To calculate the rib cage volume, a convolutional neural network was built. Using X-ray images, a computer could train itself to identify the rib cage. Although the rib cage volume was not calculated, the convolutional neural network was able to identify the rib cage to a limited degree. The model built here can be further improved upon to calculate a volume. If successfully calculated, the mediastinum and rib cage volumes could be combined to calculate the total lung capacity.

Keywords: Early Onset Scoliosis, Lung Volume, Predictive Equation, Convolutional Neural Network

Introduction

Early onset scoliosis (EOS) is a spinal deformity that presents itself before 8-10 years of age[1]. The four types of EOS are idiopathic, congenital, neuromuscular, and syndromic[2]. Idiopathic scoliosis has no known cause. Congenital scoliosis is present at birth. Neuromuscular scoliosis is caused by a systemic condition such as cerebral palsy or muscular dystrophy. When scoliosis is part of a larger condition, such as Marfan's Syndrome, it is called syndromic scoliosis. While the overall prevalence of EOS is unknown, scoliosis has an incidence of two to three percent of the US population[3,4]. While 80% of scoliosis cases are idiopathic, idiopathic EOS cases make up less than one percent of all scoliosis cases[4,5]. Although EOS is a rare condition, a treatment plan needs to be created for each patient. If left untreated, EOS can bring about an early death due to pulmonary complications[1].

Non-surgical options are often initially used to treat EOS to slow progression of the disease[6]. A common surgical option is spinal fusion, which corrects the spinal deformity and curvature seen in scoliosis patients. An incision is made either through the back or the side of the patient, and two metal rods are then placed alongside the vertebrae to force it to straighten[7]. If this procedure is performed too early, the patient can develop restrictive pulmonary disease[8]. For this reason, spinal fusion is not a routine procedure at a young age. In order to proceed with this surgery, it is crucial to determine the total lung capacity (TLC) of the lungs. TLC allows the physician to determine if the patient will have proper pulmonary function after spinal fusion surgery. Two common methods for measuring TLC are spirometry or pulmonary function testing (PFT), and computed tomography (CT)[9]. However, there are drawbacks to these methods. It is difficult for young children or those with disabilities to perform the PFT properly. CT scans also have multiple drawbacks. They are relatively expensive, subject patients to roughly ten times the amount of radiation of an X-ray[10]. The current options for determining TLC offer no clear set of rules or indications to perform surgery. Instead, the best option is to rely on the doctor's expertise and experience[6].

This project investigates the hypothesis that sagittal and coronal X-ray images can be used in combination with patient demographics to calculate TLC. While X-rays may not be able to identify the lungs, they do display the rib cage. By using a convolutional neural network, a computer can learn to detect the rib cage from images. From this the volume of the chest cavity can be calculated. Patient demographics are used to predict mediastinum volume. If the volume of the mediastinum, the central portion of the thoracic cavity that is not the lungs, is subtracted from the volume of the chest cavity, TLC will be calculated.

Results

Creating a Predictive Equation for Mediastinum Volume Using Patient Demographics as Inputs

Calculation of Mediastinum Volume from CT Scans

Mediastinal volume was computed using the MATLAB software package Pulmonary Toolkit. Our work was built upon the code of the previous capstone group and needed to

be debugged before it was able to run[11]. In order to make handoffs to future groups easier, comments were added to the code to make it easier to understand and modify.

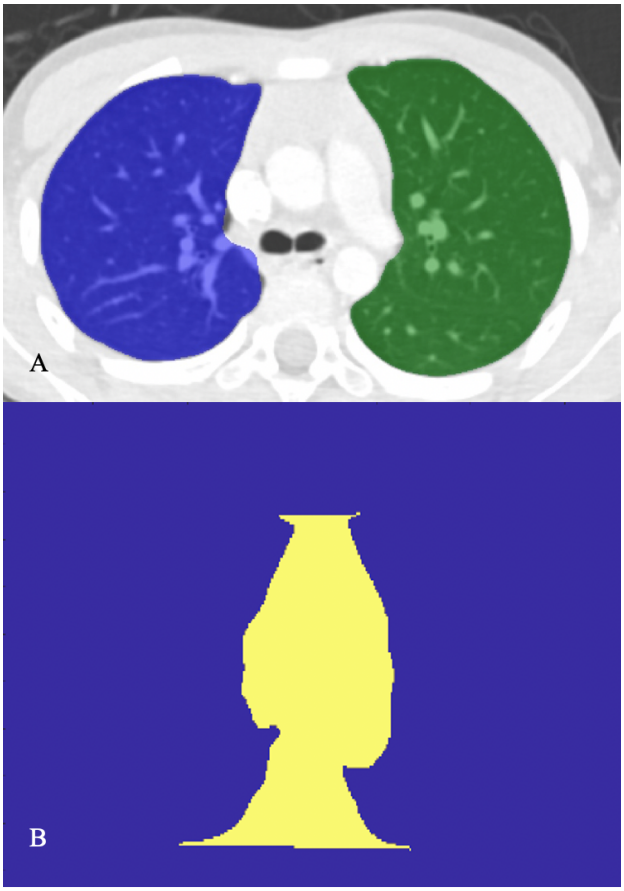


Figure 1. Lung Segmentation and Mediastinum Area. Both images are the result of lung segmentation using the Pulmonary Toolkit. Panel A is the segmentation of the lungs directly from the Pulmonary Toolkit and panel B is the area of the mediastinum (in yellow) from one of the CT slices.

The Pulmonary Toolkit uses DICOM files from CT scans as its input. It finds the boundaries of both the right and left lung and then segments them (Figure 1). Each slice of the CT scan is examined and the area in between the lungs is calculated (Figure1). The mediastinum area of each slice is used to compute the volume.

Our group was provided with some initial calculated volumes but they lacked information on patient height. Our team hypothesized that height may be an important predictor of mediastinum volume, and in order to pair it with the appropriate volume, new volumes needed to be calculated. A total of 80 CT scans were used to create the predictive algorithm. Some scans were not used because they either lacked important patient demographics or were unable to be segmented.

Validating Mediastinum Volumes

CT scans can be used to estimate mediastinum volume, but they cannot provide the true value. Our team did not have access to the true mediastinum volumes, and without them it is difficult to validate the calculated volumes. The heart is the largest component of the mediastinum, and therefore, heart volumes from literature were used to make sure our calculated values were reasonable[12]. The volumes from the literature are reported in averages and are organized by weight and gender. Since cardiac volume is an underestimate of mediastinum volume, it is expected that our calculated volumes should be greater than the cardiac volumes. Our calculated data set did not have enough data points to compare exact weights so ten pound ranges centered on the target weight were used. As seen in Supplemental Table 1, most of the calculated mediastinum volumes were found to be significantly greater than the cardiac volumes. The calculated mediastinal volumes being larger than their component part does not completely verify them, but it shows that our data is not a gross underestimate. An underestimate of mediastinum volume would lead to an overestimate of TLC. This could have negative repercussions for the patient if their operation resulted in restrictive pulmonary disease due to their actual TLC being too low for the operation. The larger p-values in Supplemental Table 1 are likely due to small sample sizes and ten pound weight ranges. The large amount of N/As for

females in the upper weight ranges can be attributed to both a small sample size and lower average female weights.

Since the heart is the largest component of the mediastinum, mediastinum volume trends should mirror those of cardiac volume. As seen in Figure 2, both cardiac and calculated mediastinal volume increase as weight increases with males having larger volumes in both cases. The similarity in trends suggests that our calculated data may be representative of mediastinum volume.

Linear Regressions Using Individual Variables

Individual variables were plotted against mediastinum volume so that potential inputs to a predictive equation could be identified and so that comparisons can be made against previous results[11]. Specifically, we analyzed how height, age, weight, and gender related to mediastinum volume. We found that age, weight, and height were positively correlated and that on average males had larger mediastinum volumes than females. The regression curves shown in Supplemental Figure 1 show that weight had the highest correlation coefficient, 0.85, and height had the lowest at 0.74. As seen in Supplemental Table 2, our linear regressions had much higher R squared values than the previous year for all measured variables. This is likely due to a narrowed patient age range.

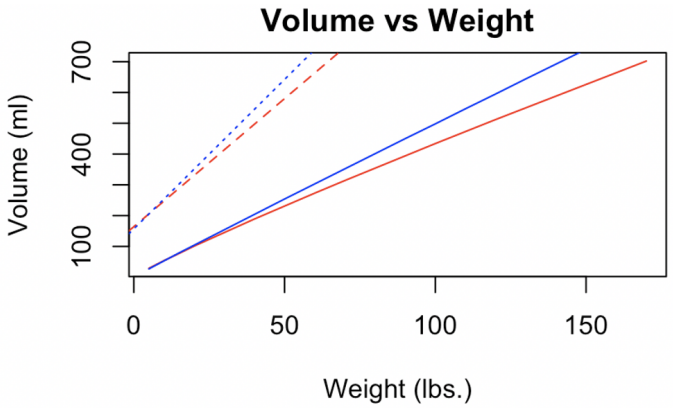


Figure 2. Mediastinal and Cardiac Volume vs Weight. Both calculated mediastinal volumes (ml) and cardiac volumes (ml) from literature are plotted against weight (lbs.). Red lines denote the the female subset while blue denotes males. Solid lines are the cardiac volumes from literature and the dashed lines are lines of best fit for calculated mediastinal volume.

Predictive Model

All four available patient demographics correlated with mediastinum volume and were included in the initial multivariate predictive models. The random forest (RF), multivariate adaptive regression splines (MARS), and multivariate linear regression (MVLR) models were tested to find the model that best fit the data. As seen in Table 1, the multivariate linear regression model is the highest performing model across all metrics. The correlation coefficients, intercept, and their respective p-values for the multivariate linear regression model can be found in Supplemental Table 3. All coefficients in the equation except the intercept were statistically significant at a significance level of $\alpha = 0.01$. Since this equation will not be used when all variables are at or near zero, a non-significant intercept is not problematic. Each of the coefficients from patient demographics are positive indicating that an increase in them will result in an increase in mediastinum volume. This was expected due to the positive linear correlations described earlier. The multivariate linear regression model takes weight, age, and gender as its inputs. Age and height were found to be redundant variables and age provided for a slightly better predictive model. For this reason, height was excluded from the final predictive equation.

Table 1. Comparison of Predictive Models. For each model R^2, mean absolute percent error, and root mean square error (RMSE) are displayed. Models are listed by descending R^2.

Model	R^2	Mean Absolute Percent Error	RMSE
Multivariate Linear Regression	0.87	15.16%	175.26
Random Forest	0.83	16.13%	247.23
Multivariate Adaptive Spline Regression	0.71	18.66%	319.45

Rib Cage Identification

Generating Masks

To train a U-Net model, masks of input images are required, which in this case are the biplanar X-rays. Masks are the specified boundaries the U-Net model uses to learn the desired region of interest. For example, a boundary was drawn along the rib cage. From the original image and the mask, the model can learn where the desired object is located. Our dataset did not contain masks; therefore, they were made manually. The VGG Image Annotator developed by the Visual Geometry Group was used to create the masks. An example of the masks can be seen in Figure 3.

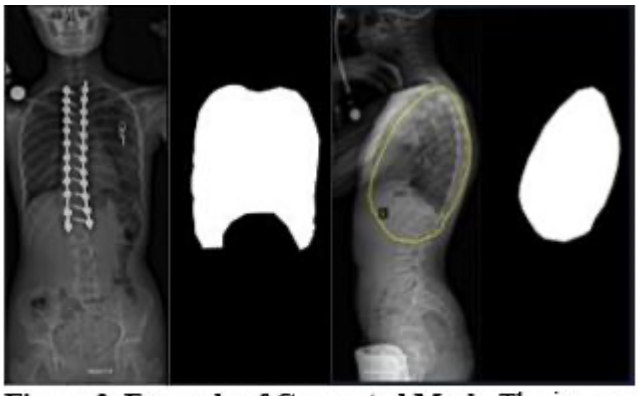


Figure 3. Example of Generated Mask. The images display examples of generated masks, along with the original image, for both sagittal and coronal plane X-rays. Additionally, the coronal image shows a yellow line, which is the annotation used to generate the mask.

Final Model Specifications

The Adam compiler, which uses the default Adam algorithm, in the keras library was used to construct the optimal model. The final model consisted of an input layer, 4 downsampling layers, 4 max pooling layers, 5 bottleneck layers, 4 upsampling layers, 4 concatenating layers, and one output layer. The model trained itself on the same images over multiple epochs. In one epoch, the model trains itself once on the training images. Therefore, with multiple epochs. the model learns from each run and retrains itself. The model attempts to improve its predictions after each run. A total of 9 epochs were used for the sagittal model. A varying learning rate was also introduced to the model. The learning rate alters how quickly the model changes its prediction based on accuracy. If the accuracy remained constant, the learning rate would decrease. Additionally, the model would cease training when the same accuracy was obtained after successive epochs. The model stopped training after 9 epochs because of the constant output accuracy.

Jaccard Index and Dice Coefficient

The model output two accuracy measurements: the jaccard index and Dice coefficient. The Jaccard index is found by dividing the area of overlap between the prediction and actual values divided by its union. Therefore, the Jaccard index indicates how well the predicted region covers the actual region of interest. The Dice coefficient is found by dividing the area of overlap by the total number of pixels then multiplying this value by 2. The Dice coefficient indicates how much of the total area is covered by the area of overlap. The Jaccard index obtained from the sagittal model was 0.691364 and 0.66833 from the coronal model. The Dice coefficients obtained were 0.5 and 0.5485 from the sagittal and coronal models respectively.

Discussion

The correlation coefficients for the regressions we made this year were much higher than the previous year's[11]. This is likely due to the exclusion of patients outside of the pediatric age range. The goal of predicting TLC is to determine if the lungs have sufficiently developed, and this is not a large concern for older patients. Furthermore, our focus is on the pediatric population. These older patients were also outliers for the previous group, and this explains why our correlation coefficients were higher.

A study using pulmonary function tests for two patient groups whose average ages were 14.3 and 15.4 years old found that the worst PFTs had TLCs of 3L and the best were 3.9L. Our RMSE was 175.26 ml, and this is 0.058% of the worst TLC[13]. On its own, the error from the mediastinum volume prediction makes it difficult to rely on this method as the sole measure of TLC for surgery, but it may be accurate enough to provide a physician a general idea of their patient's TLC. This may allow a physician to not subject the patient to unnecessary CT scans if their predicted volume is much lower than the threshold. After chest volume is able to be calculated, the error for TLC will likely be higher and the equation for mediastinum volume prediction may need to be further optimized.

The obtained Dice coefficient value was not as high as we had aimed. In the aforementioned study, Wessel et al. was able to get a dice coefficient equivalent to 0.73 for identification from sagittal X-rays[14]. The two U-Net models produced Dice coefficients near 0.5. The larger Jaccard index values may indicate that the model was working effectively. Future tuning of the U-Net model could improve the model's accuracy.

Limitations

For both the mediastinum and rib cage volume calculations, there were no actual measurements of volume. Therefore, the obtained measurements could not be compared to actual values. Using cardiac volumes for mediastinum volume validation made it difficult to judge how accurate the predicted volumes were. It could only be found that the predicted volumes were within reason. This raises the possibility that the predictive model was created on inaccurate data. For the rib cage identification algorithm, there was no true mask for the rib cage. Instead, masks were manually generated. This manual creation introduces

human error. The model would then be trained on incorrect data.

The metadata from the CT scans provided a limited amount of patient demographics and some had incomplete metadata that caused them to be excluded from analysis. Additional patient demographics and complete CT metadata would allow for more potential input variables and data points respectively.

Additionally, the U-Net model was unable to output image predictions on the testing data. This lack of information does not allow a visualization of the rib cage prediction. Therefore, it can not be stated with certainty that the model is actually predicting the region of interest with 69% or 66% accuracy in the sagittal or coronal planes respectively.

Future Work

In future work, an emphasis should be placed on improving the accuracy of the rib cage identification algorithm. One possible solution would be acquiring a large dataset that contained both CT scans or spirometry data along with the patient's biplanar X-ray images. First, this would allow a comparison between the predicted volume measurements and the actual measurements. Second, a larger dataset could improve the accuracy of the rib cage identification model. For instance, the coronal model was only trained on 57 images. More data could improve the model's prediction accuracy. Additionally, further model tuning could improve its accuracy. One method would be to further augment the data. There are many data augmenting techniques that were not pursued. These methods have the potential to improve the model.

Finally, creating a unified algorithm and simple software for both rib cage identification and mediastinum volume calculation would provide physicians an easy tool to use in clinical practice. The rib cage identification and mediastinum calculations were written separately.

Materials and Methods

Predicting Mediastinum Volume

Dataset

Dr. Keith Bachmann provided a data set of deidentified CT scans of patients primarily in the pediatric age range . The CT scans were DICOM files and patient information could be found in the files' metadata.

Mediastinum Volume Calculation

The Pulmonary Toolkit MATLAB package was used to calculate mediastinum volume. The package found the boundaries of the lungs and segmented them. The right and left lung boundaries were identified. The code starts at the top of the lungs (closer to the head) and moves downwards until it finds the start of the both of the lungs. These indexes are used to create the upper boundary of the central mediastinum. If one lung is tilted higher than the other, the interior upper edges of both are found and connected using Bresenham's line package. Given endpoints, Bresenhams's line can calculate intermediate points which will form part of the upper boundary.

The lower boundary of the mediastinum is found by finding the slice in the xz-plane that has the greatest lung area. This slice is used as it will likely be from a central portion of the lung and representative of the lower boundary. The code finds the z position at which the x component of the xz slice is discontinuous, and this marks the beginning of the diaphragm and the lower boundary of the mediastinum. The lungs curve around the diaphragm so a straight line along the x axis would be discontinuous.

After the boundaries have been identified, the slices are stepped through and the area between the lungs is calculated. Any gaps in the area are filled in and the area of all the slices is used to calculate mediastinal volume.

Predictive Model

The predictive model was created using mediastinal volume as the response variable and weight, age, gender, and height as the initial input variables. The data was cleaned in excel and then imported into R for analysis. Initial linear regressions were performed to determine the linearity of the relationship between input variables and mediastinum volume and to serve as a benchmark to previous work. The patient population was trimmed to a pediatric age range as this range is relevant to EOS.

The dataset was randomly divided into a training and testing dataset. Three predictive models were tested on the testing dataset RF, MARS, and MVLR. Each individual variable had a close fit to a linear regression when compared against mediastinum volume. This is the reason for the testing of the MVLR. The MARS model builds upon the MVLR model to make it more flexible and better fit non linear trends. This was used in case our initial impressions of the data were incorrect. The RF model is a machine learning model that can be used for regression and is not prone to

overfitting. Three different models were so that any unseen trends in the data could be caught.

Rib Cage Identification

Dataset

For the rib cage identification, a dataset was given by Dr. Keith Bachmann. The dataset consisted of a set of biplanar X-ray images taken of scoliosis patients. Biplanar means for each patient there was both a sagittal and coronal image. The images were in DICOM format.

Convolutional Neural Network

The first step in completing the rib cage identification was to identify the optimal convolutional neural network (CNN) for image segmentation. A study illustrated that Mask R-CNN had the potential to identify the rib cage. Although this study was successful, it was only implemented on posterior-anterior X-ray images[14]. There was trouble implementing this network due to its dependency on older libraries and packages. Therefore, other CNN architectures were examined. U-Net is a CNN that was initially trained on medical images. Therefore, the network should train well on this dataset. The U-Net architecture consists of two parts. The first is typical of a traditional CNN. It involves downsampling and pooling the most important features of the training dataset. The second part involves upsampling and concatenating the images. At each layer, the feature map obtained from the downsampling is cut in half [15].

The code for rib cage identification was implemented in Python. It was written on Google Colab Notebooks to allow collaborative coding. Sci-kit learn and keras, which are Python libraries and software, were used to develop the model. The pydicom library was used to convert the images from DICOM to PNG files.

Since the dataset consisted of biplanar X-ray images, two separate models were created. The first model was trained on the sagittal dataset, while the second was trained on the coronal images. Additionally, this involved splitting the original images into their respective image group. In order to train the model, some images had to be removed due to their poor quality. In some cases, parts of the rib cage were not in the image or had objects obstructing its view as seen in Supplemental Figure 2. Eleven images were removed from the posterior-anterior dataset and 28 images from the lateral dataset.

Finally, to train the model each image needed to be the same size. Therefore, the images were all resized to 224 by 224 dimensions. The resizing also helps the model run more efficiently. The model will run faster with a smaller dataset versus larger images. The dataset needed to be split into training, testing, and validation images. The model uses the training and validation images to teach itself. The testing images are used to observe how accurate the model is after creation. The dataset was randomly split.

End Matter

Author Contributions and Notes

T.A., S.S., and W.F. designed research, T.A., S.S., and W.F. performed research, T.A., S.S., and W.F. wrote software, S.S. analyzed data; and T.A., S.S., and W.F. wrote the paper. The authors declare no conflict of interest.

Acknowledgments

We would like to thank our primary advisor Dr. Keith Bachmann for his guidance throughout this project.

References

1. Skaggs, D. *et al.* Early Onset Scoliosis Consensus Statement, SRS Growing Spine Committee, 2015. *Spine Deform.* **3**, 107 (2015).
2. Blanco, J. Early Onset Scoliosis: Overview. *Hospital for Special Surgery* https://www.hss.edu/conditions_early-onset-scoliosis-overview.asp (2018).
3. Riseborough, E. J. & Wynne-Davies, R. A Genetic Survey of Idiopathic Scoliosis in Boston, Massachusetts. *JBJS* **55**, 974–982 (1973).
4. Scoliosis – Symptoms, Diagnosis and Treatment. https://www.aans.org/.
5. Beauchamp, E. C., Anderson, R. C. E. & Vitale, M. G. Modern Surgical Management of Early Onset and

Adolescent Idiopathic Scoliosis. *Neurosurgery* **84**, 291–304 (2019).

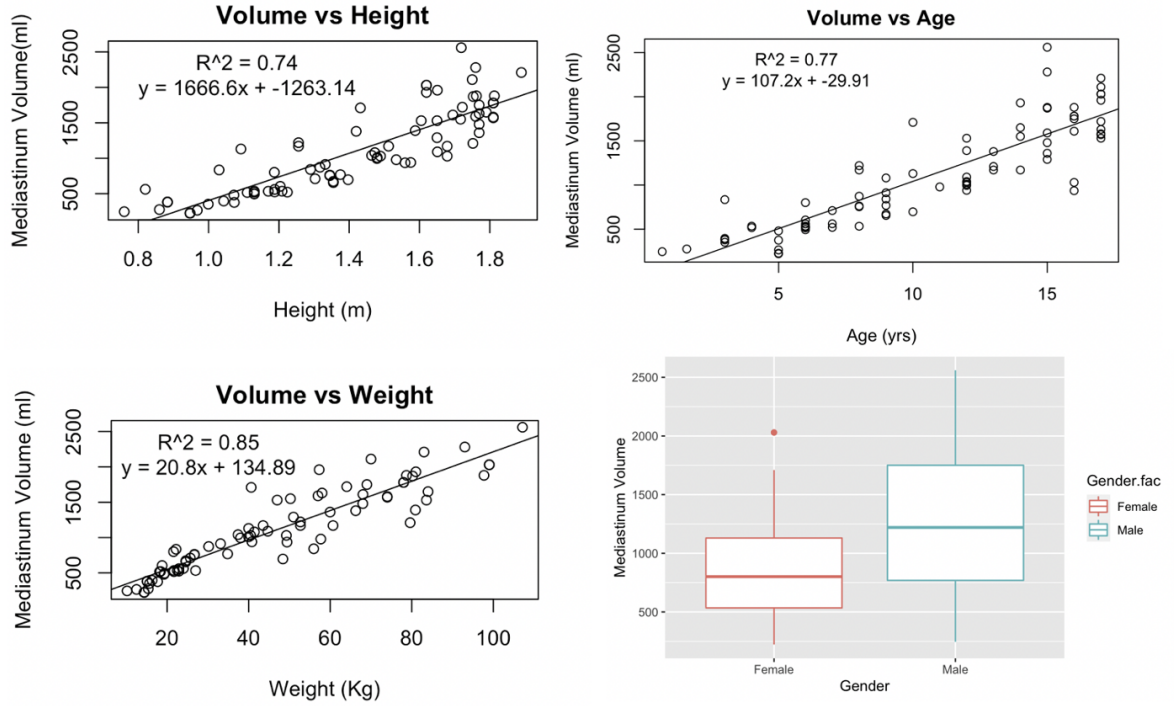
6. Cunin, V. Early-onset scoliosis – Current treatment. *Orthop. Traumatol. Surg. Res.* **101**, S109–S118 (2015).
7. Spinal Fusion for the Treatment of Idiopathic Scoliosis in Children. https://orthop.washington.edu/patient-care/articles/childrens/spinal-fusion-for-the-treatment-of-idiopathic-scoliosis-in-children (2009).
8. Karol, L. A. Early definitive spinal fusion in young children: what we have learned. *Clin. Orthop.* **469**, 1323–1329 (2011).
9. Delgado, B. J. & Bajaj, T. Physiology, Lung Capacity. in *StatPearls* (StatPearls Publishing, 2020).
10. Mettler, F. A., Huda, W., Yoshizumi, T. T. & Mahesh, M. Effective doses in radiology and diagnostic nuclear medicine: a catalog. *Radiology* **248**, 254–263 (2008).
11. Wang, D. *et al.* Three-Dimensional Modeling of Lung Volumes in Application to Scoliosis. (2020).
12. Nghiem Quang X., Schreiber Melvyn H., & Harris Leonard C. Cardiac Volume in Normal Children and Adolescents. *Circulation* **35**, 509–522 (1967).
13. Ledonio, C. G. T. *et al.* Pulmonary function tests correlated with thoracic volumes in adolescent idiopathic scoliosis. *J. Orthop. Res. Off. Publ. Orthop. Res. Soc.* **35**, 175–182 (2017).
14. Wessel, J., Heinrich, M. P., von Berg, J., Franz, A. & Saalbach, A. Sequential Rib Labeling and Segmentation in Chest X-Ray using Mask R-CNN. (2019).
15. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds. Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) vol. 9351 234–241 (Springer International Publishing, 2015).

Supplemental Material

Supplemental Table 1. Hypothesis Testing Results.
This table shows the p-values from the hypothesis test where H_0 = Mediastinum volume equals mean cardiac volume and H_a = Mediastinum volume is greater than cardiac volume. N/A denotes weight ranges that did not have enough samples to perform a hypothesis test.

Weight (lbs.)	Female p-value	Male p-value
30	0.00233**	0.001***
40	0.01225*	0.003174**
50	0.00002631***	0.03384*
60	0.01588*	N/A
70	N/A	0.01176*
80	N/A	0.08368 .
90	0.03485*	0.001861**
100	0.07832 .	N/A
110	0.01129*	N/A
120	N/A	0.01326*
130	0.05733.	0.002025**
140	N/A	N/A
150	N/A	0.005204**
160	N/A	0.002022**
170	N/A	0.0005806***

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



Supplemental Figure 1. Correlating Individual Variables with Mediastinum Volume. These graphs show trends between individual patient demographics and mediastinum volume.

Supplemental Table 2. Comparison of Linear Regressions from this Year and the Previous Year. Mediastinum volume was plotted against individual demographic variables and their R^2 values are displayed in the table. Old R^2 denotes last year's data and New R^2 denotes this year's data[11]. N/A is written if there is no data for that variable.

Y vs X	Old R^2	New R^2
Volume vs Weight (M)	0.54	0.88
Volume vs Weight (F)	0.23	0.79
Volume vs Age (M)	0.50	0.81
Volume vs Age (F)	0.18	0.72
Volume vs Height (M)	N/A	0.77
Volume vs Height (F)	N/A	0.67

Supplemental Table 3. Multivariate Linear Regression Coefficients. The values of the model's regression coefficients and their respective p-values are shown below.

Coefficient	Estimate	P-value
Intercept	-55.645	0.46
Weight	14.665	1.2*10^-7
Age	35.916	0.0035
Gender	180.169	0.0056

Supplemental Figure 2. Removed Images from Rib Cage Identification Data. The image on the left shows objects that impede the view of the rib cage. Further, the image on the right illustrates an example when part of the rib cage is not within the X-ray scan.