Deep Learning for Classifying Adolescent Idiopathic Scoliosis

Submitted By: Michael Ling

Dept. of Mechanical Engineering

Northwestern University

Date Submitted: 8/20/2022

Table of Contents

Deep Learning for Classifying Adolescent Idiopathic Scoliosis

Michael Ling

ABSTRACT: Scoliosis is a sideways curvature of the spine and is very common, with there being more than 3 million US cases per year. Adolescent idiopathic scoliosis is the most common type of scoliosis affecting as many as 4 in 100 adolescents, and as the name suggests, affects adolescents who are undergoing puberty and their growth spurt, but for unknown reasons. While most forms of scoliosis are but a minor encumbrance, scoliosis can grow more severe over the course of life until it becomes painful and disabling. In these cases or when requested, surgery is used to correct the spinal deformity with the most common method being spinal fusion in which vertebrae are fused together to eliminate the curvature. In order to standardize the surgical procedures, classification systems were created beginning in 1905 with Schulthess. But in 2001, with the introduction of the Lenke Classification System for classifying adolescent idiopathic scoliosis replacing the King's classification system, the classification system has seen widespread usage and has become the gold standard because of its ability to capture the complex nature of scoliosis. This is made possible by virtue of the 3 components of the classification system with 6 curve types, 3 lumbar spine modifiers, and 3 sagittal thoracic modifiers. But with such complexity lies room for error which can be minimized through the implementation of deep learning algorithms. There currently exist methods for using deep learning to measure Cobb angles but none make complete classifications within the Lenke Classification System. But through the use of a convolution neural network a complete classification can be made without any intermediate steps, reducing diagnosis time and room for error. In the end, the convolutional neural network was able to predict curve types with an accuracy of 99.167%. Despite the high accuracy, further improvements must be made such as regularizing to account for overfitting towards the training data and increasing the size of the dataset to be trained on. Once this is accomplished, while the accuracy will likely decrease, the model can eventually be used to cross check classifications or make them instead of a human.

KEYWORDS: *Convolutional Neural Network, Adolescent Idiopathic Scoliosis, Deep Learning, Lenke Classification System*

INTRODUCTION

Because of the complexity of the Lenke classification that is necessary to account for the complex nature and variation within scoliosis, there are many steps that are necessary to classify a spine. As a result, it is not only time consuming, but it also is subject to human error and inaccuracy. Most of the time, this won't change the overall classification but by passing the classifying to deep learning, more efficient and accurate classifications can be made. This accuracy is important because depending on the curve type, different surgical procedures will be implemented which will impact the health and recovery of the patient. Convolutional neural networks (CNNs) are equipped especially well to deal with this task because of how they specialize in working with images because of how they utilize convolution to analyze images and extract features. In this project, in order to construct the CNN, Pytorch was used and in order to train the model, images were provided courtesy of the Lurie's Children Hospital. From this project, I hoped to develop a model with a high accuracy but also general enough to be applicable in real life settings and to test how well deep learning, specifically CNN's, worked in medical applications.

The Lenke Classification System

In 2001, the Lenke Classification System replaced the existing King's classification system as the default classification system. This was because the King's classification system only had five possible classifications depending on the two dimensional curvature of the spine. However, the spine is a three dimensional object and scoliosis affects it not only laterally but also rotationally. The Lenke System accounts for this in how it utilizes a sagittal plane x-ray, capturing more information regarding the curvature of the spine. In addition, the Lenke classification system had greater inter and intraobserver reliability meaning that classifications were more precise. The Lenke system consists of three components, a curve type, a lumbar spine modifier, and a sagittal thoracic modifier allowing for much more complexity than King's system.

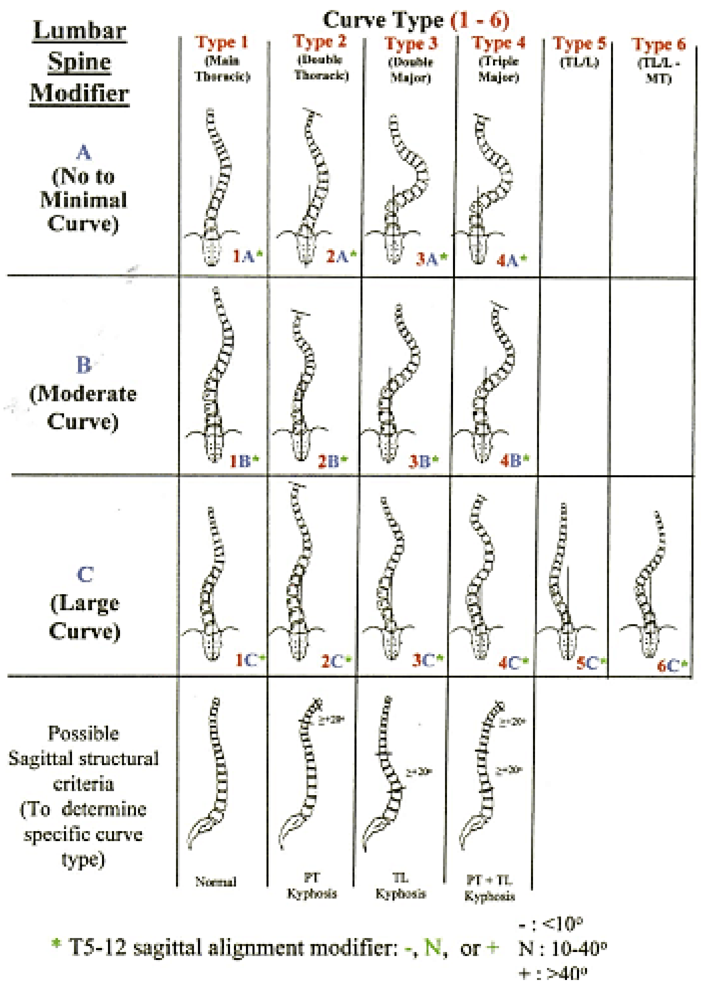

Fig 1. Lenke Classification System [1]

In order to classify a spine, one must measure Cobb angles. This is done by identifying the apex vertebra, finding the most tilted vertebrae above and below the apex vertebra, and drawing lines from the cephalad end of the upper vertebrae and caudal end of the lower vertebrae and measuring the resulting angle.

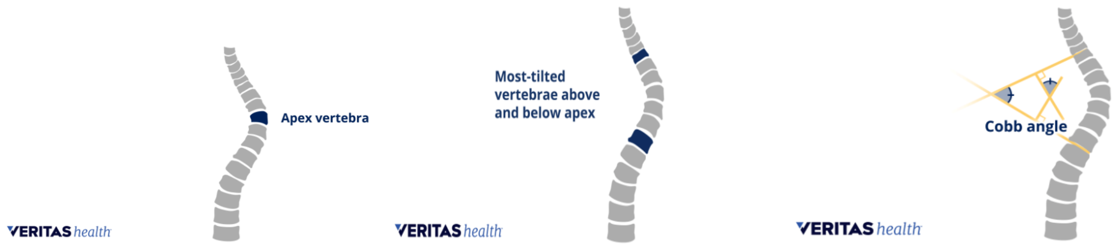

Fig 2. Cobb Angle Measurement [2]

Based on the location of the apex vertebrae, the curve is classified as either: Proximal Thoracic, Thoracic, Thoracolumbar, or Lumbar.

- Thoracic curves, the apex of which is located between the second thoracic vertebral body and the eleventh and twelfth thoracic intervertebral disc, include proximal thoracic curves with the apex at the third, fourth, or fifth thoracic level and main thoracic curves with the apex between the sixth thoracic body and the eleventh and twelfth thoracic disc.
- The apex of thoracolumbar curves is located between the cephalad border of the twelfth thoracic vertebra and the caudal border of the first lumbar vertebra.
- The apex of lumbar curves is located between the first and second lumbar disc and the caudal border of the fourth lumbar vertebra.

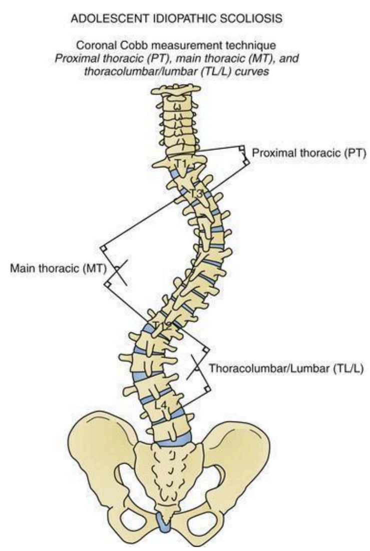

Fig 3. Proximal Thoracic, Main Thoracic, Thoracolumbar, Lumbar Curve Apex Locations [3]

In addition, each curve can be either structural or nonstructural. Structural curves, described by their location, lack normal flexibility and are termed as major (if they have the largest Cobb measurement) or minor. Minor curves can be structural or nonstructural.

Then based on the magnitude of the Cobb angle and which combination of the aforementioned curves are present in an image, a final classification can be determined

- Type 1—main thoracic: The main thoracic curve is the major curve, and the proximal thoracic and thoracolumbar/ lumbar curves are minor nonstructural curves.
- Type 2—double thoracic: The main thoracic curve is the major curve, while the proximal thoracic curve is minor and structural and the thoracolumbar/lumbar curve is minor and nonstructural.
- Type 3—double major: The main thoracic and thoracolumbar/lumbar curves are structural, while the proximal thoracic curve is nonstructural. The main thoracic curve is the major curve and is greater than, equal to, or no more than 5° less than the Cobb measurement of the thoracolumbar/lumbar curve.
- Type 4—triple major: The proximal thoracic, main thoracic, and thoracolumbar/lumbar curves are all structural; either of the two latter curves may be the major curve.
- Type 5—thoracolumbar/lumbar: The thoracolumbar/ lumbar curve is the major curve and is structural. The proximal thoracic and main thoracic curves are nonstructural.
- Type 6—thoracolumbar/lumbar-main thoracic: The thoracolumbar/lumbar curve is the major curve and measures at least 5° more than the main thoracic curve, which is structural. The proximal thoracic curve is nonstructural.

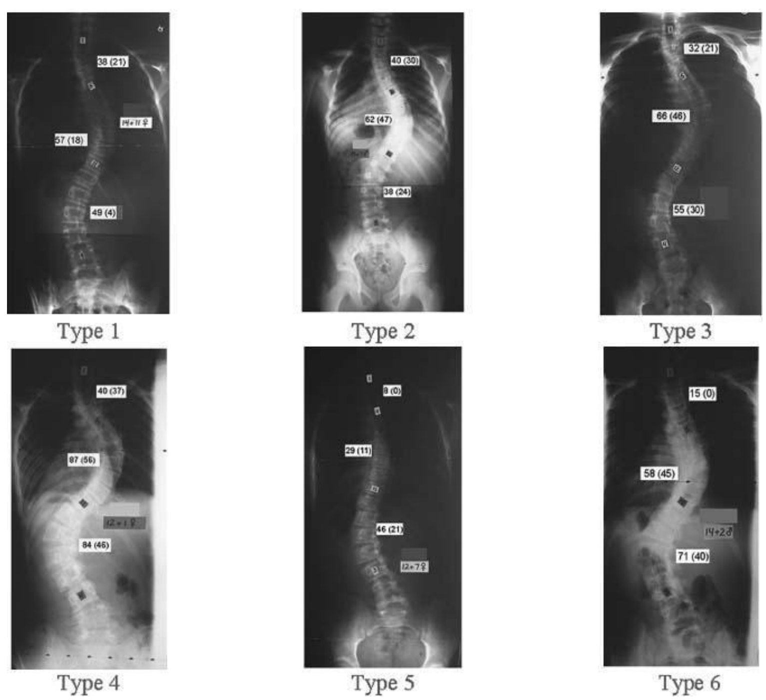

Fig 4. Types of Curves [4]

After a curve type is assigned, a lumbar spine modifier and sagittal thoracic modifier are added as well but the process for classifying these will be omitted for reasons presented later.

Dataset

The data used was provided by the Lurie's Children Hospital and consisted of coronal and sagittal x-rays. However for every coronal x-ray, the corresponding sagittal x-ray for the same patient was not present. Since none of the images were already labeled, it was impossible to label the images correctly since both a coronal and sagittal x-ray would be necessary. Thus, the CNN constructed was limited to only classifying curve types and lumbar spine modifiers. However, it was decided that the CNN should focus solely on the classification of curve types in order to maximize accuracy as including lumbar spine modifiers would increase the classes 3 fold. Data was labeled for training by the previously mentioned method and classified as a number from 0-6 inclusive with 0 corresponding to no detectable scoliosis or a straight spine and the classes 1-6 being the corresponding Lenke curve type. One issue that was encountered was the fact that there were only 60 images in the entire dataset as opposed to the hundreds to thousands of images typically used in deep learning. To resolve this issue, data augmentation was used to both increase the amount of images but also to reduce possible overfitting. Data augmentation operations used included: flipping the images horizontally, adding Gaussian blur, and adding noise. This effectively increased the data set by 8 fold.

Model Architecture

When dealing with data in the form of images, CNNs are especially useful because of how they implement convolution. The convolution operation is done by multiplying the corresponding elements of a filter/kernel and an area of the image and summing the resulting numbers onto a destination layer or feature map.

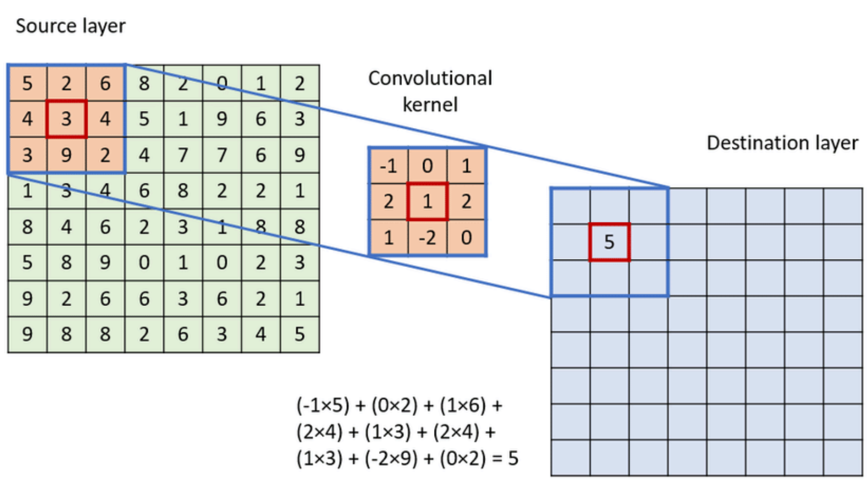

Fig 5. Convolution Operation [5]

The reason for CNN's effectiveness in dealing with images lies in how the convolution operation considers the neighboring pixels of a certain pixel. In addition, the elements or weights within a filter can be optimized through training in such a way that when the filter is applied, a certain feature from the image can be extracted. A simple example of this is how when the Sobel filter is applied, because of how the weights in the filter are organized, the resulting image is able to highlight edges, places where there is a significant change in pixel value. Through training, the CNN is able to develop filters of its own that look for features within the image such as curvature and edges, that are important in determining the classification of an image.

X – Direction Kernel

-1	0	1
-2	0	2
-1	0	1

Y – Direction Kernel

-1	-2	-1
0	0	0
1	2	1

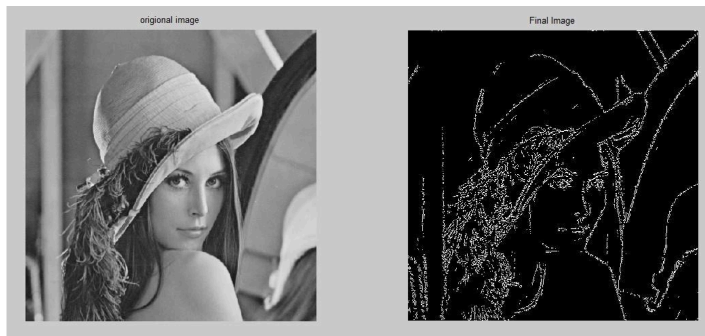

Fig 6. Sobel Filter and Resulting Image [6][7]

Within the CNN pooling layers are also frequently used after convolution operations for dimension reduction and to prevent overfitting. Typical pooling operations include max, min, and average. These poolings operations simply apply the selected operation to a select dimension of elements at a time. Unlike convolution, the same element in the original layer is never used twice and this is why the stride, or how much the filter moves along the image, is the same size as the filter.

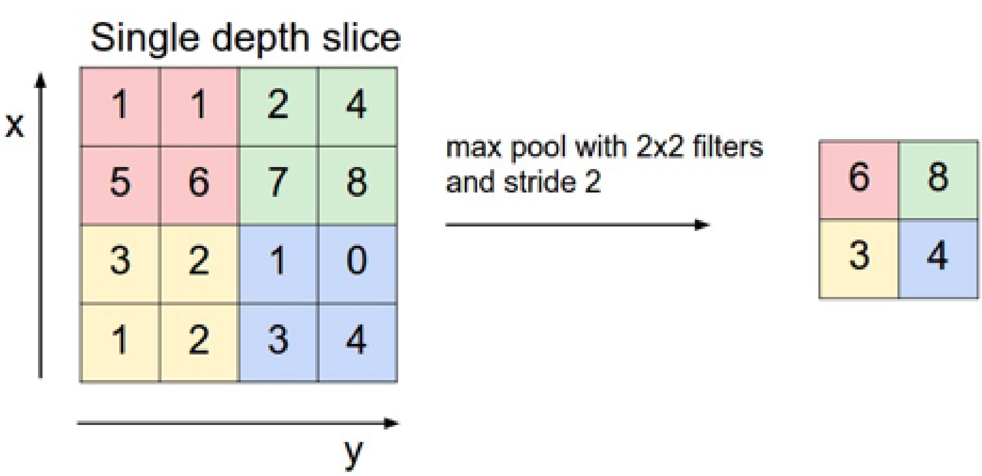

Fig 7. Max Pooling [8]

After an arbitrary amount of convolution and pooling layers, the final numbers are passed to a series of fully connected layers that are the same as the standard neural network. In the end a softmax is applied and the neuron with the greatest probability is the one that is chosen and will correspond to one of the 6 curve types or no detectable scoliosis. A cross entropy loss function is employed as this is a classification problem and is used as a measure of how far off the model is.

In the model, all images to 256x256 pixels and hyperparameters included: a learning rate of .001, a batch size of 4, and training over 4 epochs. The CNN architecture that was used consisted of 8 3x3 filters with stride 1 and with padding, followed by a ReLU activation function, and then a 2x2 max pool. Then 16 3x3 filters with stride 1 and with padding, followed by a ReLU activation function, another 2x2 max pool, and then into a fully connected layer with 131072 neurons that feeds into the final 7 output neurons.

RESULTS

After splitting the dataset into ¾ training data and ¼ test data and training over 4 epochs, the following accuracies were achieved:

Curve Type	**Accuracy (%)**
No Scoliosis	96.875
1	100
2	100
3	100
4	100
5	100
6	100
Overall	**99.167**

Figure 8. Accuracies after Training

CONCLUSION

While the resulting accuracies of the model seem promising, one needs to be careful. Because of the initial miniscule size of the dataset and how the data augmentation was implemented, the images that had the Gaussian blur and noise were almost virtually identical to the initial 60 images and the flipped versions. This means that the trained model likely had already seen some of the testing data and is why the accuracy is so high. This lack of variation in data means that the model is very prone to overfitting. This can be resolved by training over less epochs, increasing the size of the initial dataset before data augmentation, using stronger blur and noise, implementing L1/L2 regularization, utilizing dropout regularization, or even changing the overall architecture of the CNN. Once this is accomplished and the model becomes more general, it is possible for the model to be used in real life medical settings in classifying and diagnosing patients in a more timely and accurate way. While the resulting model was far from what I imagined it to be at the beginning, downsizing from a full classification to only curve types, it was interesting to undergo the trouble of not having the data to perform a task and adjusting for it. This project has also taught me the difficulty of creating an accurate model that also is general enough to be truly applicable and usable and I look forward to improving my project in the future to meet this criteria.

Acknowledgements and References:

❖ Thank you to my advisor Chanwook Park and the rest of the Northwestern University Mechanistic Data Science team as well as the Lurie's Children Hospital for making this project possible!

❖ [1] https://www.scoliosistools.com/
❖ [2] https://www.spine-health.com/conditions/scoliosis/cobb-angle-used-measure-scoliosis-curves
❖ [3] https://neupsykey.com/pediatric-and-adult-scoliosis/
❖ [4] https://neupsykey.com/the-lenke-classification-system-for-adolescent-idiopathic-scoliosis/
❖ [5] https://d2l.ai/chapter_convolutional-neural-networks/channels.html
❖ [6] https://www.projectrhea.org/rhea/index.php/An_Implementation_of_Sobel_Edge_Detection
❖ [7] https://medium.datadriveninvestor.com/understanding-edge-detection-sobel-operator-2aada303b900
❖ [8] http://cs231n.github.io/convolutional-networks/
❖ Lenke LG, Betz RR, Harms J, Bridwell KH, Clements DH, Lowe TG, Blanke K. Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis. J Bone Joint Surg Am. 2001 Aug;83(8):1169-81. PMID: 11507125.
❖ Lenke LG, Betz RR, Clements D, Merola A, Haher T, Lowe T, Newton P, Bridwell KH, Blanke K. Curve prevalence of a new classification of operative adolescent idiopathic scoliosis: does classification correlate with treatment? Spine (Phila Pa 1976). 2002 Mar 15;27(6):604-11. doi: 10.1097/00007632-200203150-00008. PMID: 11884908.

Code: https://github.com/mling23/Deep-Learning-for-Classifying-Adolescent-Idiopathic-Scoliosis