SPARK Program Student Research Project, Summer 2022

Deep Learning for Classifying Adolescent Idiopathic Scoliosis

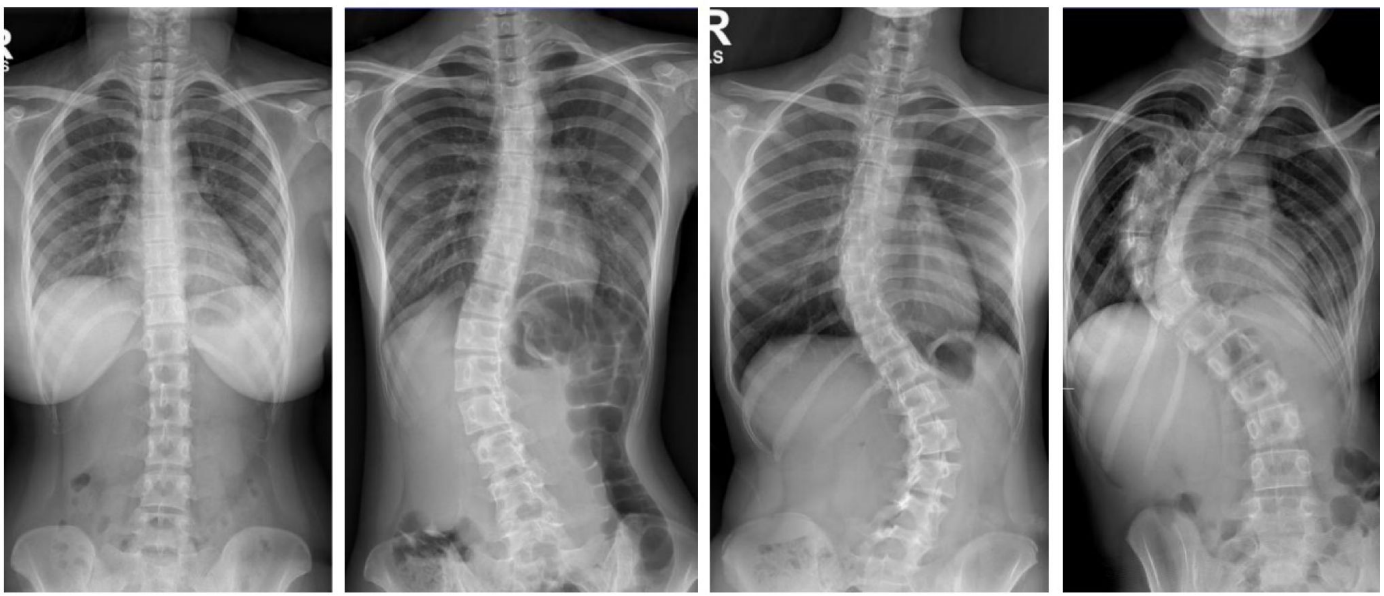
Student: Michael Ling
Advisor: Chanwook Park

Northwestern | McCORMICK SCHOOL OF ENGINEERING

Background

What is scoliosis?

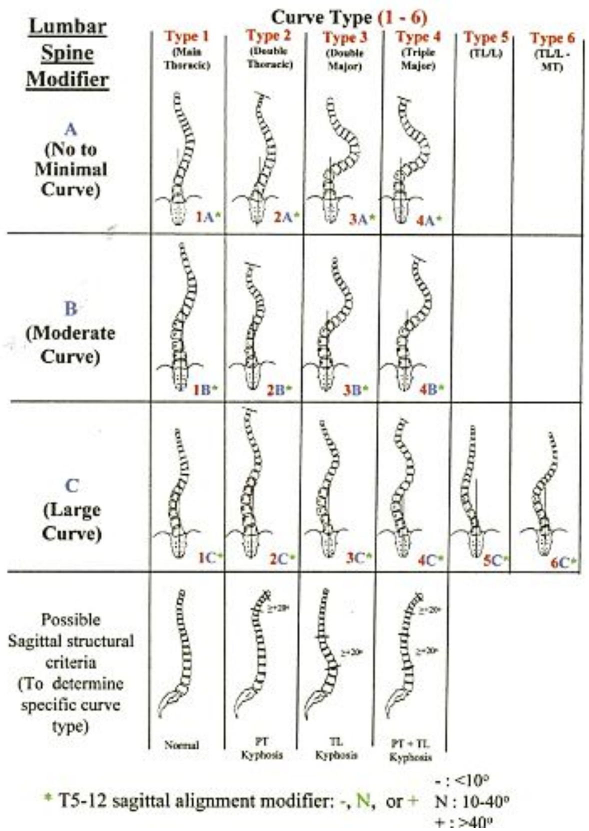
- A sideways curvature of the spine.
- Very common, more than 3 million US cases per year
- Scoliosis occurs most often during the growth spurt just before puberty.
- Most cases are mild with few symptoms. Some children develop spine deformities that get more severe as they grow.
- Severe scoliosis can be painful and disabling.



Motivation for Project

Lenke Classification System

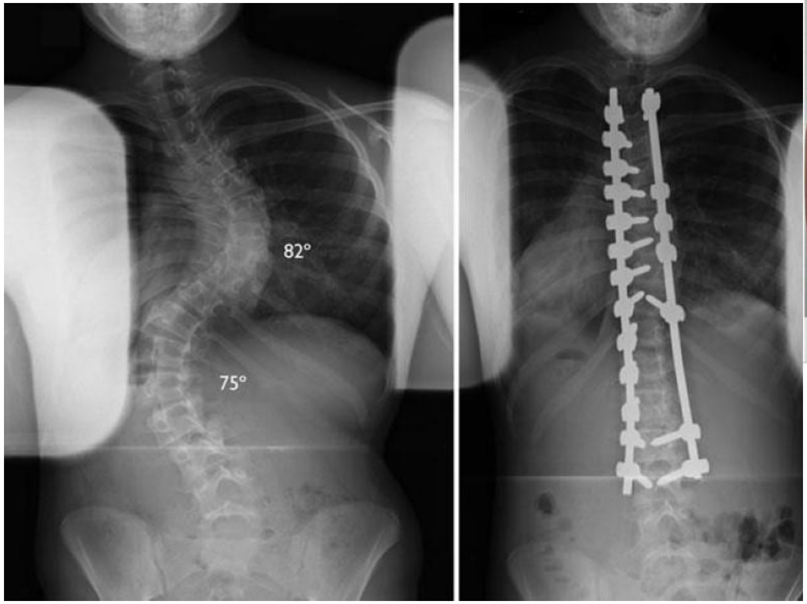
- ❖ A standardized way of diagnosing scoliosis
- ❖ Replaced the existing King Classification System in 2001
- ❖ This system is relatively complicated in order to account for the vast variation in types of scoliosis
- ❖ Many steps necessary to properly classify



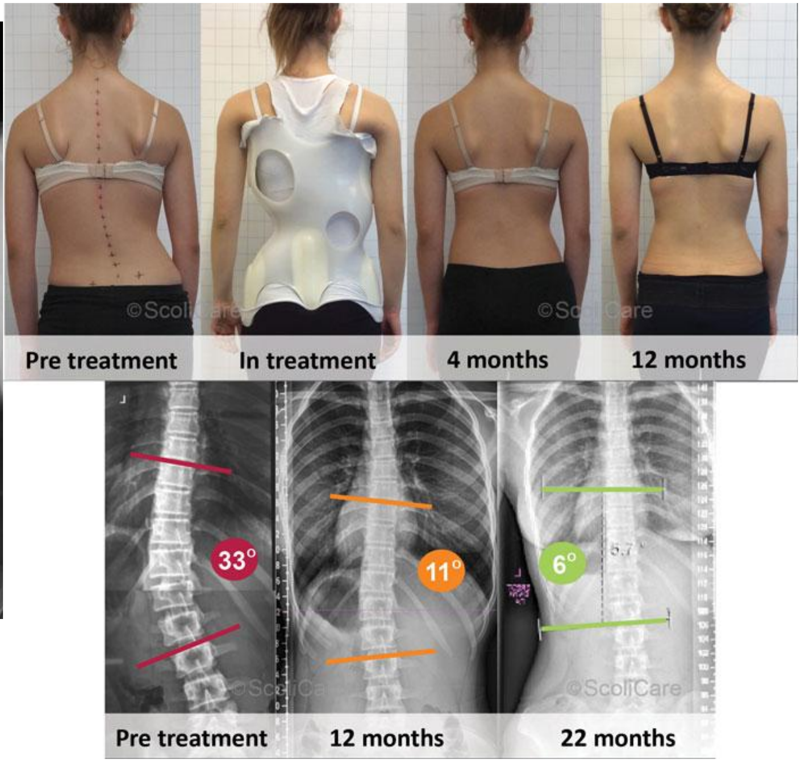
https://www.scoliosistools.com/

Motivation for Project

- In order to remove the human error and discrepancies between medical professionals, a classification done using deep learning can be utilized so that the most accurate diagnosis can be made
- Making the correct classification is important for treatments as different classes will require different procedures
- Braces are effective in patients who have not reached skeletal maturity. If the patient is still growing and their scoliosis isn't severe, a brace may be recommended to prevent the curve from progressing.
- In a spinal fusion the idea is to realign and fuse together the curved vertebrae so that they heal into a single, solid bone



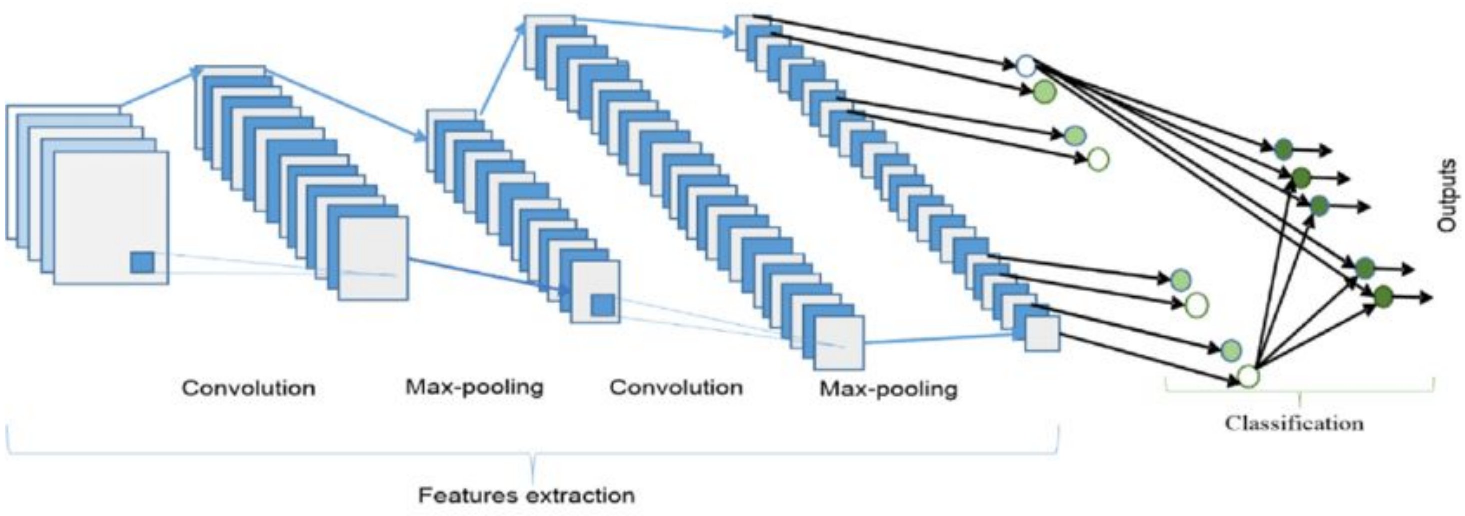
https://orthoinfo.aaos.org/en/treatment/surgical-treatment-for-scoliosis/



https://scolicare.com/scolibrace/

Project Overview

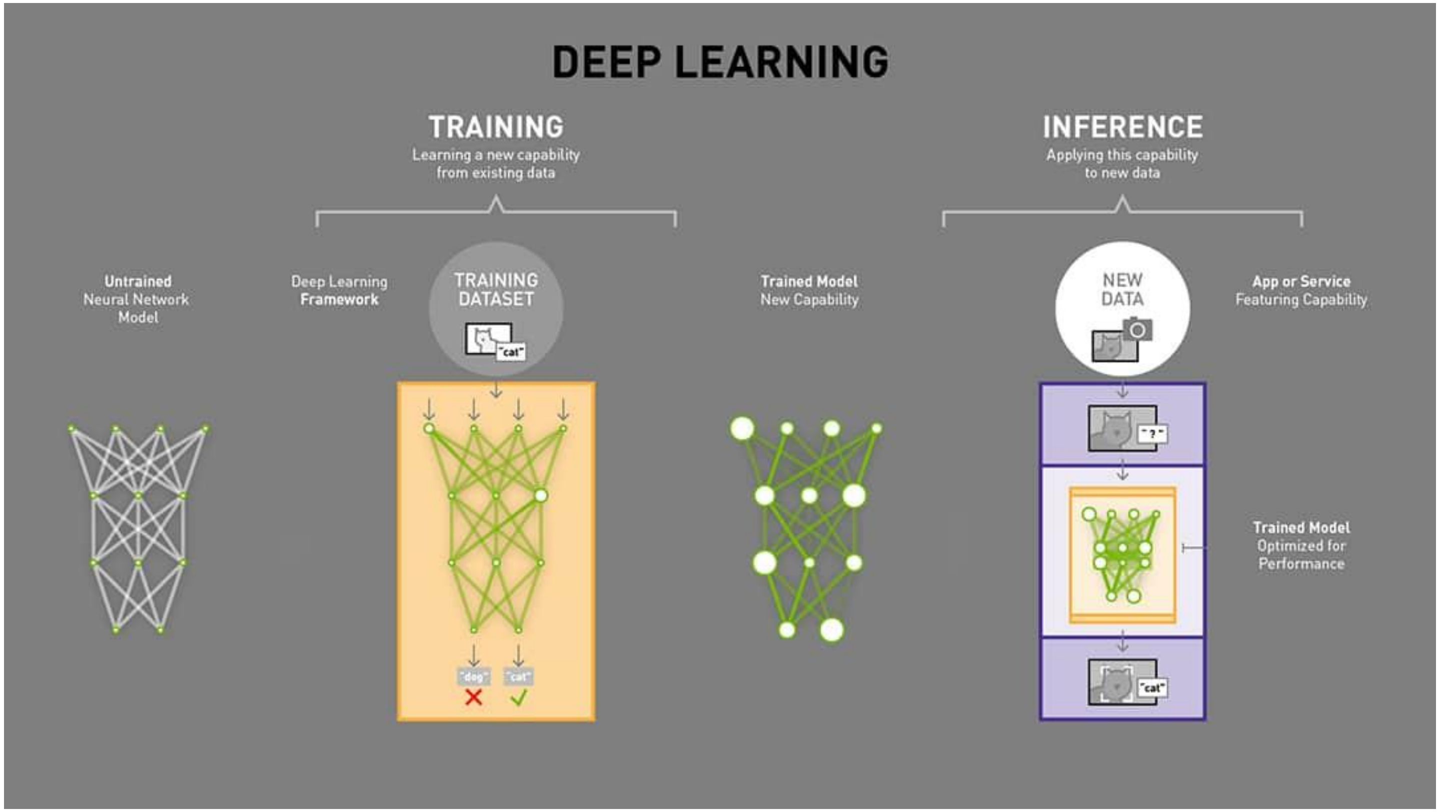
- ❖ To develop a program to classify a spinal x-ray image into one of the classes of the Lenke Classification System
- ❖ This will be done through the training of a convolutional neural network



https://www.researchgate.net/figure/The-overall-architecture-of-the-Convolutional-Neural-Network-CNN-includes-an-input_fig4_331540139

Data Labeling

- ❖ In order to train the neural network, labeled training data must be created
- ❖ All images were provided courtesy of the Lurie's Children Hospital
- ❖ This is done by measuring Cobb angles



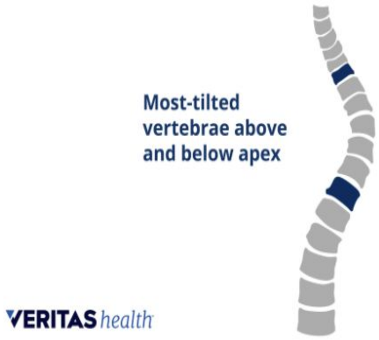
https://blogs.nvidia.com/blog/2016/08/22/difference-deep-learning-training-inference-ai/

Data Labeling

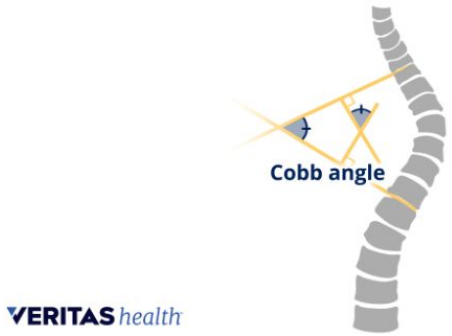
❖ To begin classifying images, Cobb angles must be measured this is done in the following way:

1. Identify the Apex Vertebra

2. Find the most tilted vertebrae above and below the apex vertebra



3. Draw lines from the cephalad end of the upper vertebrae and caudal end of the lower vertebrae and measure the cobb angle

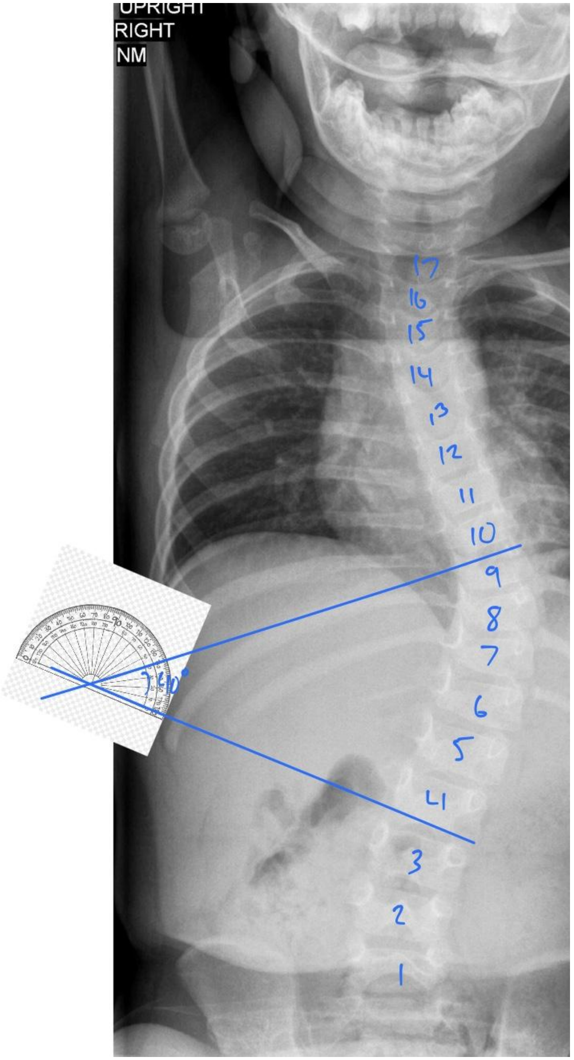


https://www.spine-health.com/conditions/scoliosis/cobb-angle-used-measure-scoliosis-curves

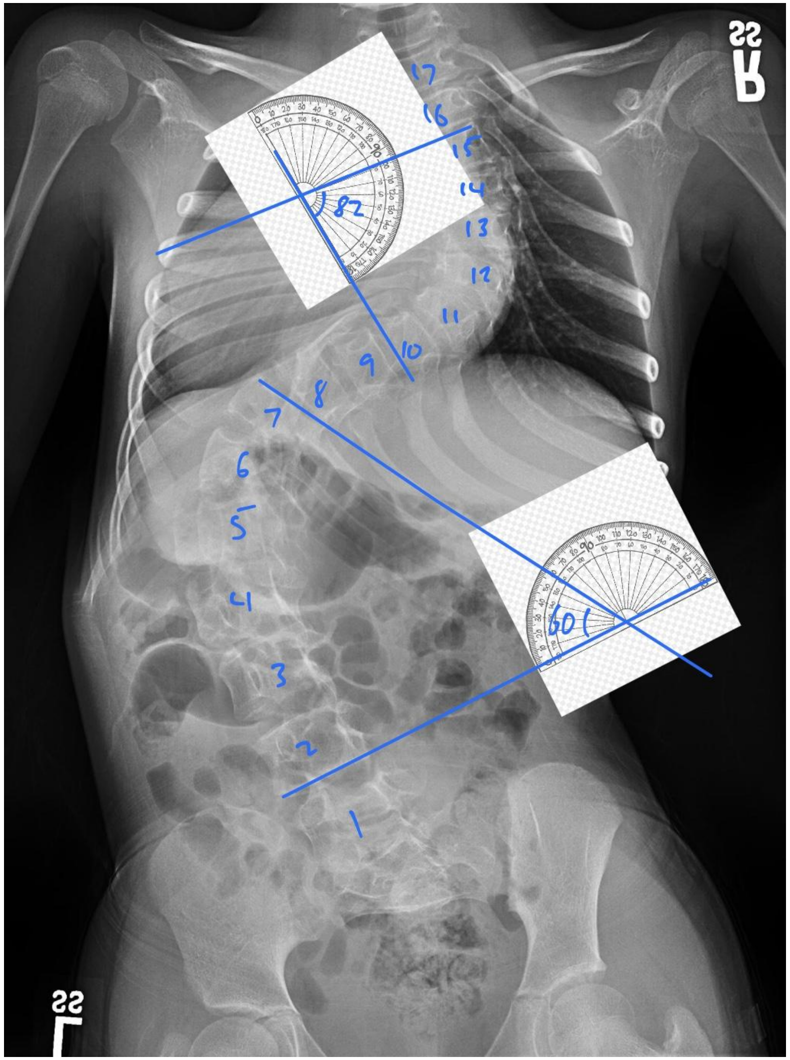
Data Labeling

- 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 caudad border of the first lumbar vertebra.
  - The apex of lumbar curves is located between the first and second lumbar disc and the caudad border of the fourth lumbar vertebra.
- In addition, each curve can be either structural or nonstrucutual. 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.

Examples



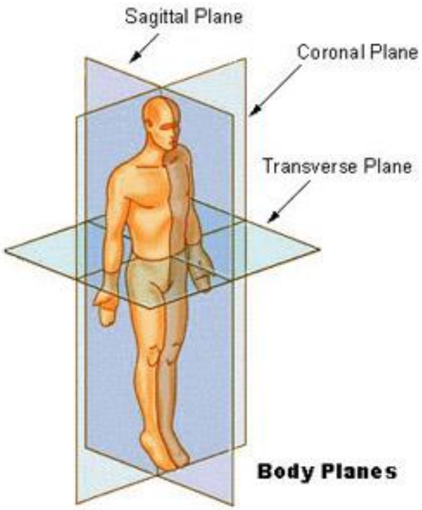
Type 1

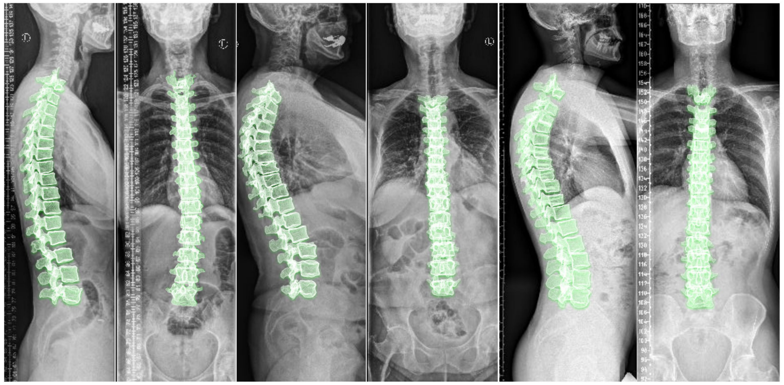


Type 3

Limitations

- In order to make a complete classification using the Lenke Classification System, both a coronal x ray as well as a sagittal x ray are needed
- Scoliosis is 3-dimensional
- However in the images provided, only 1 or the other were provided, and not both for a single patient
- The sagittal thoracic modifier could not be classified
- For simplicity's sake, the CNN that I decided to implement would only classify into 1 of the 6 curve types and would not include the lumbar spine modifier nor the sagittal thoracic modifier
- Modifiers are relatively easy to classify by hand and is less prone to error





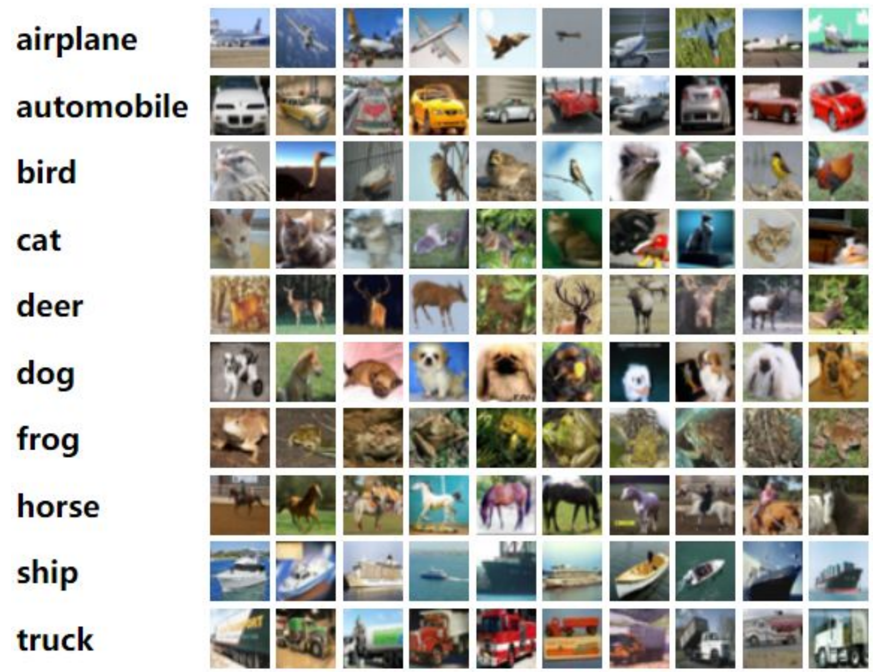
https://www.mdpi.com/2379-139X/8/1/39/htm

https://training.seer.cancer.gov/anatomy/body/terminology.html

Problem

- ❖ Usually with deep learning, datasets of at least hundreds of pictures are used for training
- ❖ The MNIST dataset which is used for digit recognition has 60,000 images
- ❖ The CIFAR-10 dataset which has 10 classes
- ❖ Meanwhile I only had access to 60 images

https://en.wikipedia.org/wiki/MNIST_database



https://paperswithcode.com/dataset/cifar-10

Data Augmentation

- In order to offset this, I implemented data augmentation, which is defined as "techniques used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data. It acts as a regularizer and helps reduce overfitting when training a machine learning model"
- The operations include flipping the image horizontally, adding gaussian blur, and adding noise
- Increased the amount of images by 8 fold
- Reduces overfitting

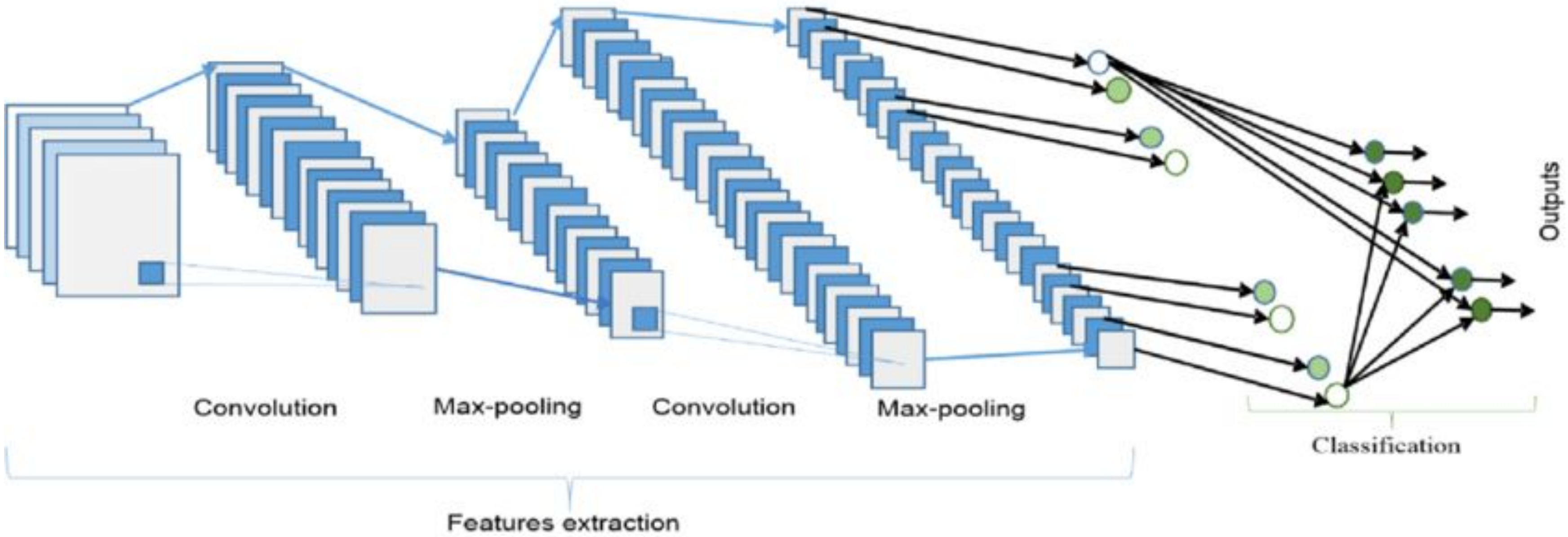
Original

Flipped

Blurred

Noised

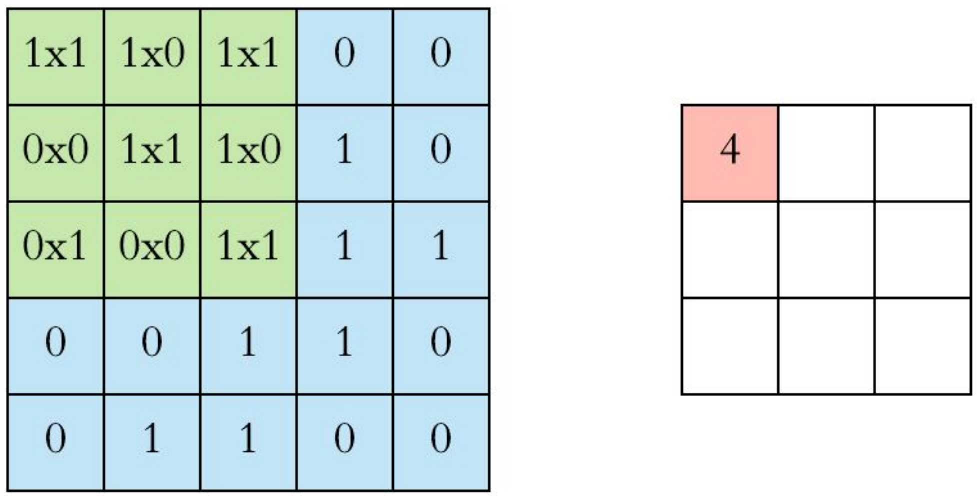
CNN Architecture



https://www.researchgate.net/figure/The-overall-architecture-of-the-Convolutional-Neural-Network-CNN-includes-an-input_fig4_331540139

Convolution Review

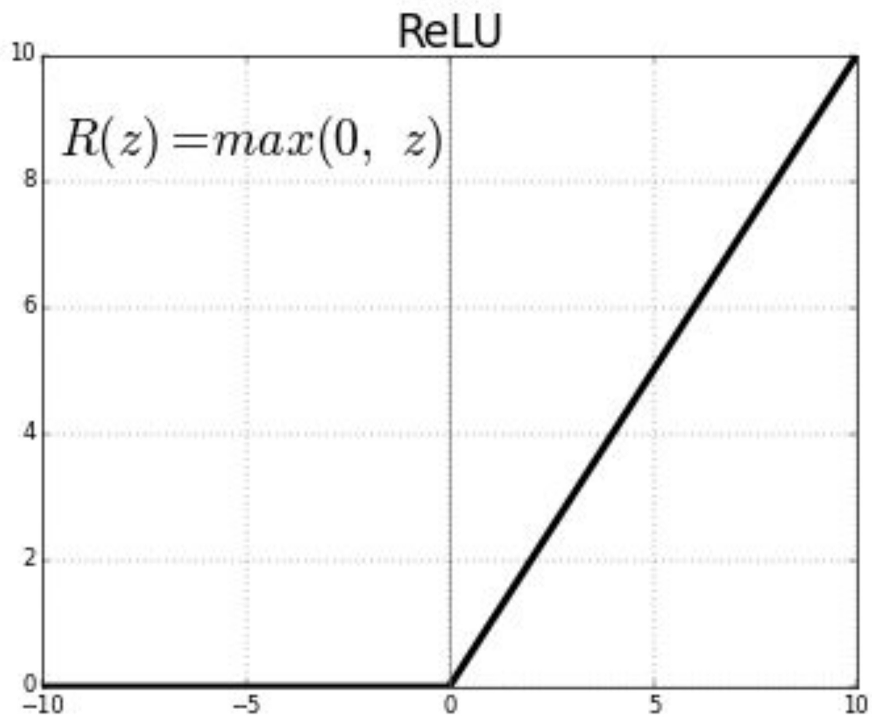
- In every CNN, the convolution operation is applied in convolution layers
- The convolution operation is done by multiplying an array of weights (kernel/filter) by a section of the original image and summing up the resulting numbers



https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Activation Function

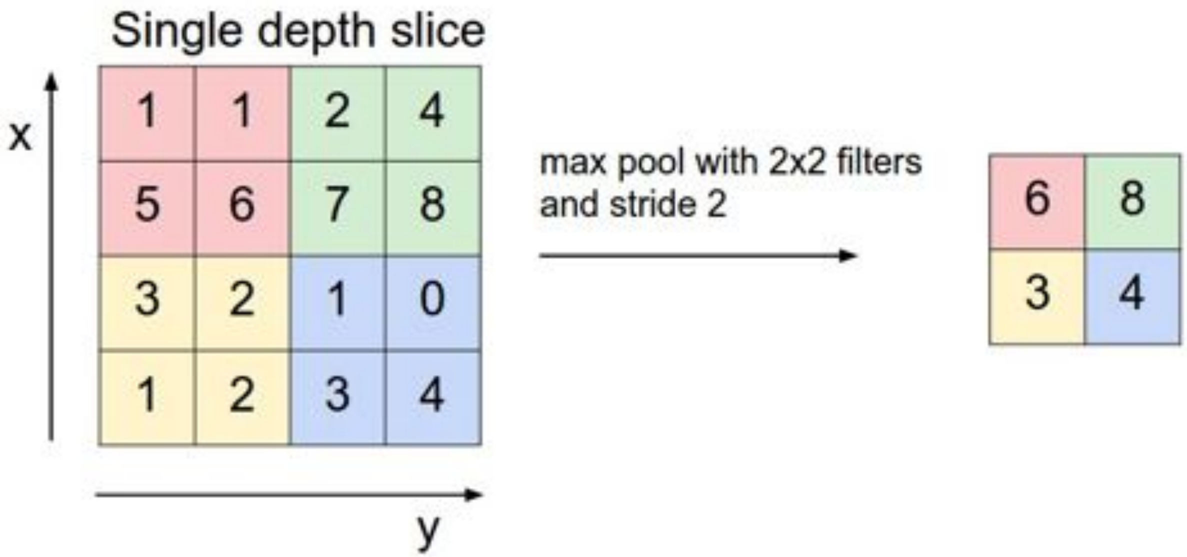
- ❖ In any good neural network, there needs to be non-linearity involved, and this can be done through the implementation of activation functions such as the RELU function
- ❖ This can be applied following the convolution layer and later in the fully connected layers



https://medium.com/@kanchansarkar/relu-not-a-differentiable-function-why-used-in-gradient-based-optimization-7fef3a4cecec

Pooling

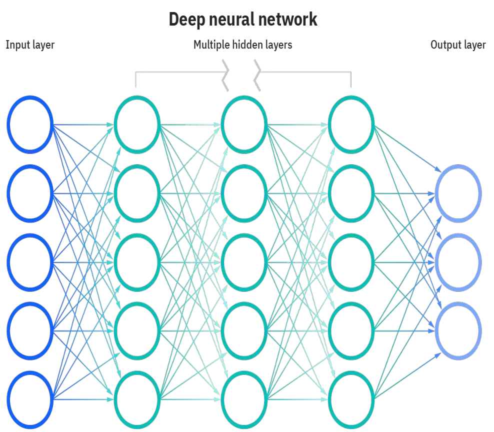
- Following each convolutional layer, a pooling layer is usually added to reduce the dimensionality which reduces the number of parameters and computation in the network
- This shortens the training time and controls overfitting
- Common types of pooling include: max, average, min



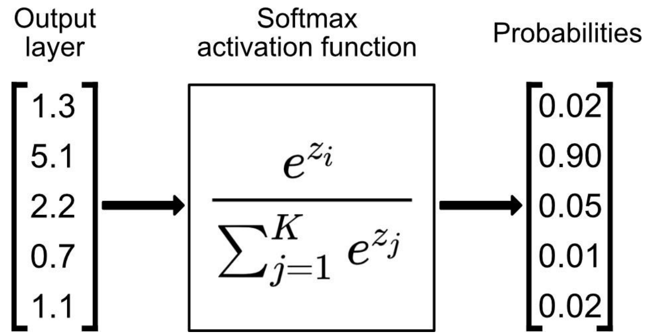
http://cs231n.github.io/convolutional-networks/

Fully Connected Layers

- After an arbitrary amount of convolution and pooling layer cycles, the outputs are put into a series of fully connected layers which simply follow the standard neural network
- Since this is a classification problem, in the output layer, which has 7 neurons, a softmax is applied so that the neuron with the highest probability is the final output of the network with each neuron corresponding to a specific Lenke curve type
- Cross entropy loss function



https://www.ibm.com/cloud/learn/neural-networks



https://towardsdatascience.com/softmax-activation-function-explained-a7e1bc3ad60

$$L = -\frac{1}{m}\sum_{i=1}^{m} y_i \cdot \log(\hat{y}_i)$$

https://levelup.gitconnected.com/grokking-the-cross-entropy-loss-cda6eb9ec307

How does a CNN learn?

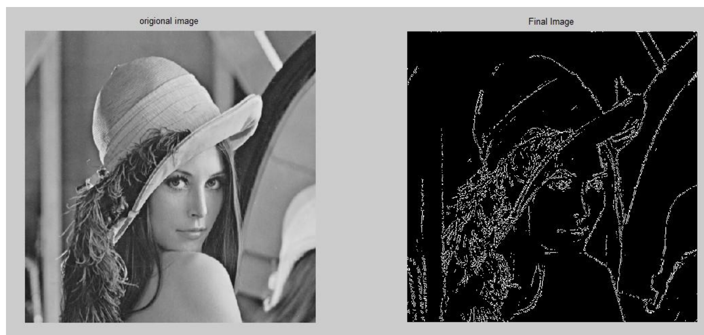
- Since each of the kernels/filters used in the convolutional layers are composed of weights, these can be optimized just like in a normal neural network to minimize loss
- Kernels/filters can be created to extract certain features from an image, e.g. Sobel Filter for edge detection
- During training, the network will learn what types of features to extract to minimize the loss for the specific task the network is being trained to solve by updating the weights in the filters in the convolutional layers

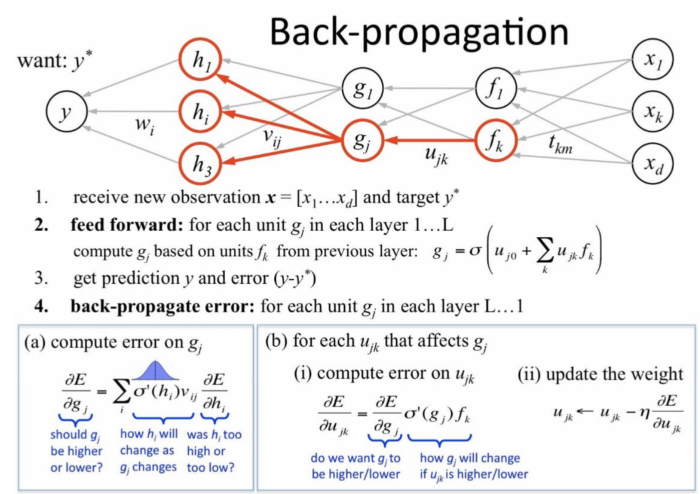
X – Direction Kernel

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

Y – Direction Kernel

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





https://www.projectrhea.org/rhea/index.php/An_Implementation_of_Sobel_Edge_Detection

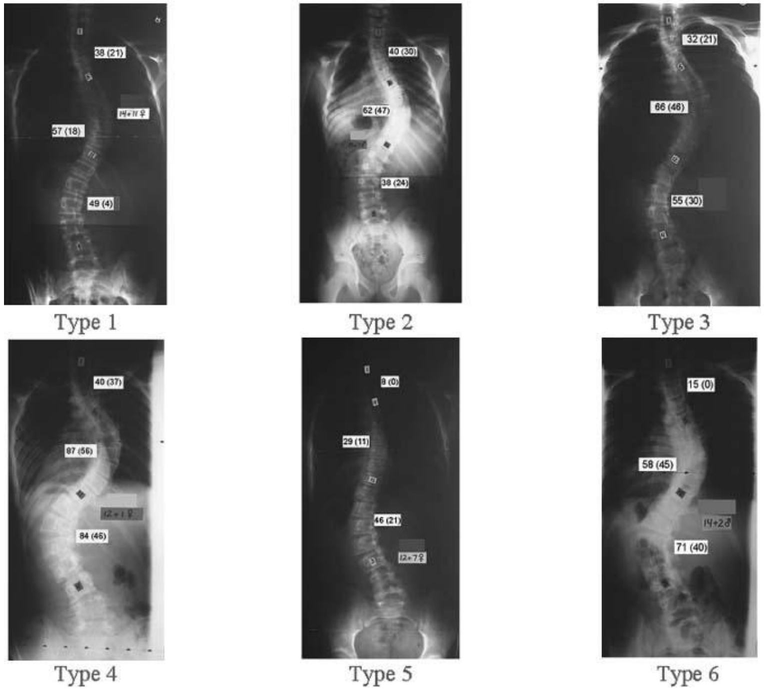
https://medium.datadriveninvestor.com/understanding-edge-detection-sobel-operator-2aada303b900

https://medium.com/@jorgesleonel/backpropagation-cc81e9c772fd

Results

- After training over 4 epochs this is what I came up with
- Each number from 0-6 corresponds to a curve type with 0 being no detectable scoliosis and 1-6 being the corresponding Lenke curves

```
Accuracy of the network: 99.1666666666667 %
[31, 27, 5, 10, 7, 7, 32]
[32, 27, 5, 10, 7, 7, 32]
Accuracy of 0: 96.875 %
Accuracy of 1: 100.0 %
Accuracy of 2: 100.0 %
Accuracy of 3: 100.0 %
Accuracy of 4: 100.0 %
Accuracy of 5: 100.0 %
Accuracy of 6: 100.0 %
```



https://neupsykey.com/the-lenke-classification-system-for-adolescent-idiopathic-scoliosis/

Improvements

- ❖ Going forward I will look to make sure that the model is not overfitting
  - ➢ More data augmentation
    - ■ Making the blurs and noise more drastic
  - ➢ Finding a larger base dataset
  - ➢ L1/L2 Regularization
  - ➢ Stopping training earlier
  - ➢ Changing the CNN architecture
  - ➢ Dropout Regularization
- ❖ Getting classifications cross checked by a professional

L1 Regularization

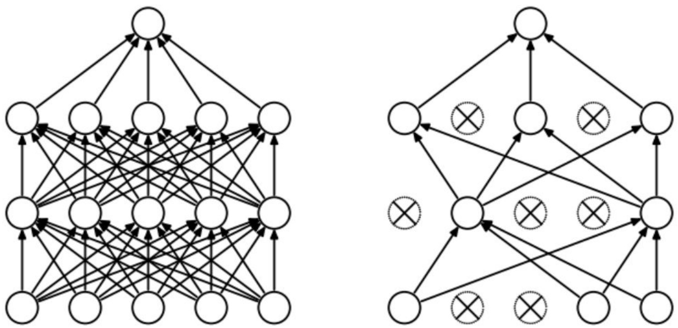
$$\text{Cost} = \sum_{i=0}^{N}\left(y_i - \sum_{j=0}^{M} x_{ij}W_j\right)^2 + \lambda \sum_{j=0}^{M} |W_j|$$

L2 Regularization

$$\text{Cost} = \sum_{i=0}^{N}\left(y_i - \sum_{j=0}^{M} x_{ij}W_j\right)^2 + \lambda \sum_{j=0}^{M} W_j^2$$

Loss function — Regularization Term

https://medium.com/analytics-vidhya/l1-vs-l2-regularization-which-is-better-d01068e6658c



(a) Standard Neural Net (b) After applying dropout.

https://www.tech-quantum.com/implementing-drop-out-regularization-in-neural-networks/

https://www.freepik.com/premium-photo/doctor-looking-chest-x-ray_5480795.htm

Thank You

to my advisor Chanwook Park and the rest of the Mechanistic Data Science Team as well as the Lurie's Children Hospital for making this project possible!

to my advisor Chanwook Park and the rest of the Mechanistic Data Science Team as well as the Lurie's Children Hospital for making this project possible!

References

- 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.
- Brownlee, Jason. “How Do Convolutional Layers Work in Deep Learning Neural Networks?” *Machine Learning Mastery*, 16 Apr. 2020, https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/.
- Chatterjee, Chandra Churh. “Basics of the Classic CNN.” *Medium*, Towards Data Science, 31 July 2019, https://towardsdatascience.com/basics-of-the-classic-cnn-a3dce1225add.