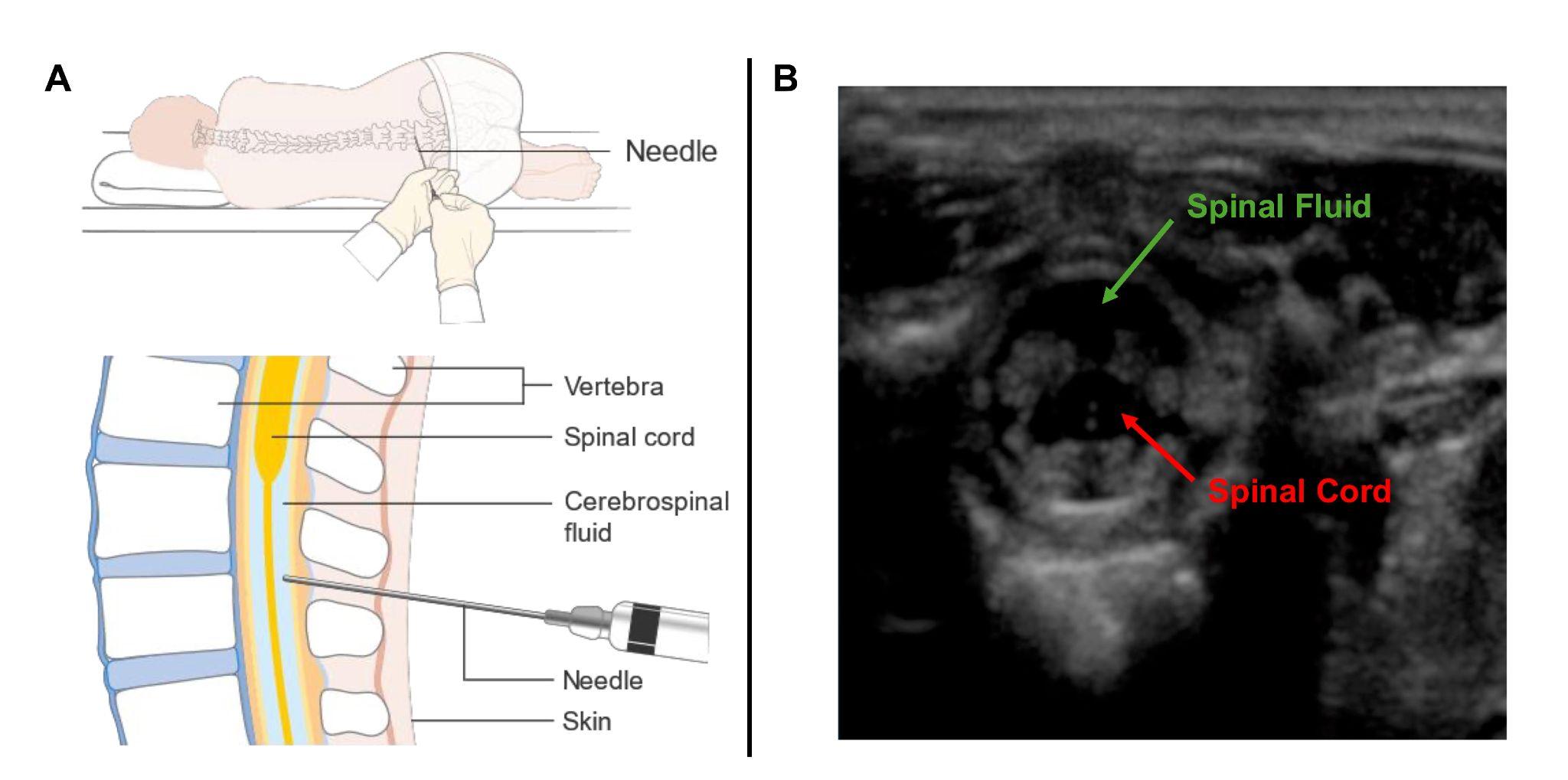
**Introduction:**

**Project Overview:**

This capstone project is a collaboration between the Columbia University Data Science Institute, the Lab of Dr. David Kessler, and Elsevier. The broad objective is to utilize computer vision techniques to improve and automate detection of spinal cord and spinal fluid within ultrasound images from infants. Eventually, these models will become the basis for clinical tools to help physicians read ultrasound images as they perform lumbar punctures in infants, enabling physicians to easily identify spinal cord and spinal fluid and perform the procedure with the highest level of accuracy.

**Background:**

Lumbar puncture (LP), also known as spinal tap, is a procedure used to diagnose or treat a variety of health conditions including meningitis, encephalitis, and Guillain-Barré syndrome.[[1]](#footnote-0) For this procedure, a hollow needle is inserted in the lower back and into the space surrounding the spinal column, in order to extract cerebrospinal fluid (CSF) or inject therapeutics. For highest procedure success, the needle is inserted in the L3-L4 or L4-L5 interspace, below the level of the spinal cord and yet where a sufficiently large amount of CSF is still present.[[2]](#footnote-1) A diagram of this procedure can be found in Figure 1A.



[IMAGE OF LUMBAR PUNCTURE PROCEDURE](https://upload.wikimedia.org/wikipedia/commons/f/f6/Diagram_showing_how_you_have_a_lumbar_puncture_CRUK_157.svg) + Sample US image (labeled with cord / fluid), something [like this](https://drive.google.com/file/d/1Jk7WmhPHtKFttxukJ8XchQUoIs-rK2RI/view?usp=drive_link)

* Caption: **Fig. 1: Overview of Lumbar Puncture and Spinal Ultrasound Images. A.** In a lumbar puncture procedure, a needle is inserted into the lower back to extract cerebrospinal fluid while avoiding the spinal cord. **B.** Cross section of the vertebral interspace region, with spinal fluid and spinal cord labeled in green and red, respectively.

This procedure can be performed on patients of all ages, including on infants, where it frequently is used to diagnose meningitis.[[3]](#footnote-2) Despite the high usage of this procedure in infants, the success rate is remarkably low (50-60%). Studies performed by Dr. David Kessler have shown that preprocedural ultrasound (US) can improve visualization of the desired interspace region by allowing physicians to directly view the spinal column cross-section (Figure 1B), but even this addition to the procedure is not perfect; reading US images requires expertise, something which not all physicians performing LP have.[[4]](#footnote-3) To aid physicians in reading US images, the Kessler lab aims to develop artificial intelligence models to automatically identify key features within the US images, such as spinal cord and spinal fluid. The development and optimization for such models is the focus for this capstone project.

**Concept Slide / Model Overview:**

The primary objective for this project is to take any ultrasound image as input, and answer the following three questions:

1. Is the image quality sufficiently high for proper analysis?
2. Is CSF detectable within the image?
3. Is the spinal cord present within the interspace? (if so, extraction of CSF in that region is dangerous)

To accomplish these three objectives, we propose the model concept outlined in Figure 2.

[DIAGRAM OF MODEL CONCEPT](https://drive.google.com/file/d/1phkSYCxGRhoVsKGUhRoNqnpPYkOC-dqm/view?usp=drive_link)

Caption: **Fig. 2: Proposed Workflow for Analyzing Ultrasound Images.** To analyze ultrasound images of the vertebral insterspace, the image quality is first assessed to ensure the image is usable. If the image quality is sufficiently high, the image can be separated by the presence of spinal cord and spinal fluid. The computational process for doing so involves first pre-processing the images, then using a multiclassification model to identify quality, spinal cord presence, and spinal fluid presence, and finally outputting the predicted state for each of the three classes.

**Dataset Overview:**

Vision Transformer (ViT) Models:  
  
Vision Transformer (ViT) models are a class of deep learning architectures that apply Transformer-based attention mechanisms, originally designed for natural language processing, to image data. Unlike traditional convolutional neural networks (CNNs), ViT treats an image as a sequence of flattened, non-overlapping patches, each patch of size \(p \times p \times 3\) being vectorized into a single token. Given an input image of dimensions \(H \times W \times 3\), the image is divided into \(N = \frac{H \times W}{p^2}\) patches. Each patch is linearly embedded into a fixed-dimension vector, and positional encodings are added to retain spatial information. ViT’s self-attention mechanism computes pairwise similarities between all patches through a matrix product, with the computational complexity scaling as \(O(N^2 \cdot d)\), where \(d\) is the embedding dimension. This global attention enables ViT to capture long-range dependencies across the entire image, offering superior performance on large-scale datasets. However, ViT requires extensive pre-training and larger datasets, as it lacks the inductive biases (like translation invariance and local connectivity) inherent in CNNs. Fine-tuning techniques like LoRA and quantization can help reduce the model's size and computational overhead while maintaining accuracy during inference.

Fine Tuning with LoRA:

**Shrinjay’s Part** [**Isaac Tucker Peabody**](mailto:itp2109@columbia.edu)**: Please take a look!**

I applied a pre-implemented ResNet-18 model to an older dataset of images to evaluate its performance, which yielded satisfactory results. However, when I implemented the same model on a new set of videos, I inferred that the model’s performance on the extracted images from the videos was not as strong. This performance drop could be attributed to differences in the characteristics of the new data, such as video-specific challenges like motion blur, lighting variations, or differences in object perspectives. Currently, I am focusing on meta-learning techniques for image classification to improve model generalization across different datasets, especially when there are shifts in data distribution. Meta-learning can enable the model to adapt more effectively to new tasks or datasets with minimal retraining, improving its robustness and performance.

1. https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/lumbar-puncture [↑](#footnote-ref-0)
2. https://www.merckmanuals.com/professional/neurologic-disorders/how-to-do-lumbar-puncture/how-to-do-lumbar-puncture [↑](#footnote-ref-1)
3. https://pmc.ncbi.nlm.nih.gov/articles/PMC11017152/ [↑](#footnote-ref-2)
4. https://pubmed.ncbi.nlm.nih.gov/29645365/ [↑](#footnote-ref-3)