CNN/Neural Network Route:

1. Predicting spinal curvature angles for scoliosis
   1. Input: X-ray images
   2. Labels: spinal curvature angles
   3. Ideas for models: Unet++, Centernet
   4. Dataset 16 on this website: <http://spineweb.digitalimaginggroup.ca/Index.php?n=Main.Datasets>
   5. Other - I’ve kinda already worked on this project so we can reuse some of our starter code

Other:

1. Something with spotify data (cause there's so much of it lol)
   1. Auto tagging system
      1. Problem identification:
         1. With many different ways to describe songs, it is often difficult to categorize songs in a single playlist. Furthermore, users can have different (and many) interpretations of a single song, making it overwhelming to categorize a large selection of songs and then finding them later.
      2. Task:
         1. The proposed application would automatically tag songs, and give the user the freedom to create personal tags for their profile. Multiple tags could be assigned to a song or an artist overall, giving the user the ability to search their playlists by tags. Furthermore, the tags would serve as smaller sub playlists that the user could choose to play and shuffle tags together. This would greatly increase the selection of playlists for users thus limiting the restriction of songs to one single playlist.
      3. Action:
         1. A naïve approach to this problem would be to create a dataset with songs, each with their labeled tags (tags can be based on mood, genre, location etc. the limitation of this approach is that only a few tags will be trained and thus could be used for auto-tagging). For each tag, a binary classifier (if the song belongs to a given tag or not) could be trained by using different audio features of the songs (**leveraging the Spotify Audio Features API**) and correlating them with their tag (the label). Logistic regression, SVM and random forest algorithms would be sufficient models to construct these classifiers. The resulting trained classifiers could then be used to predict the tag (from the list of tags that were trained) given a song and its extracted audio features. The Million Song Dataset could be used in this process. However, a drawback of this approach is that tags are user defined, so there is no standard for a tag. This approach could be a fast and useful way to analyze a user’s playlists and auto generate playlists based on their subjective understanding of the tags. For example, ‘happy songs’ might mean different things to different users. By training on data provided by the user from their playlists, the predicted tags and songs that belong to that tag would be personalized to them. The biggest limitation to this, however, is the availability of user defined tags. The initial data might need to be collected based on which songs a user has added in a playlist and the corresponding playlist name.
      4. Recent research has modeled audio tagging as a multi label classification problem. A deep learning approach to this problem would require a sufficient amount of data (i.e songs with their tags). Given that this data can be collected, recent research has explored deep architectures using waveforms as input with small convolutional filters, and log-mel spectograms with a convolutional neural network designed to learn temporal features. The main difference between these two designs is their use of musical domain knowledge or not. It would be worth while to examine models that rely on domain knowledge in the case where there is not sufficient data for a deep architecture. These two approaches would be useful to explore as they would provide a powerful model to auto generate tags for large music collections. This approach would be useful to define a semantic space of predicted tags that could then be used to categorize large music libraries.
   2. **map a genre to a song using k means clustering**
   3. Dataset:
      1. <https://machinelearningknowledge.ai/mini-ml-project-predicting-song-likeness-from-spotify-playlist/>
      2. <https://towardsdatascience.com/spotify-data-project-part-1-from-data-retrieval-to-first-insights-f5f819f8e1c3>
      3. Honestly, getting data won’t be hard