Task 2 – Research plan

The current structure of my research plan for quantifying scoliosis using neural networks on imperfect landmark data is:

1. **Main claim and sub-claims**
   - My main claim is that a machine learning approach using neural networks to estimate spinal curvature from imperfect can retrieve a curvature estimate, within clinically acceptable limits of error, for a range of imperfection severities.
   - Relating my main claim to the state-of-the art, I claim that ultrasound landmark data are likely to contain several types of imperfections not addressed by the literature.
     - I hope to back this claim up with references to authors who report excluding data because of missing points, or who talk about the difficulties of ultrasound. Otherwise, a separate study could be performed. Such a study would be difficult because it would require many scoliotic patients and ultrasonographers. After scanning the patients, the ultrasonographers would mark the transverse processes in whatever volumetric representation is used. The locations of these landmarks could be compared to ground-truth values located from CT to find typical error modes. Hopefully the literature support will be sufficient, because we certainly cannot find enough patients, much less expect them to undergo CTs.
2. **Factors, scope, and limitations**
   - Neural networks is a vast field offering many tools for many more problems. With limited experience in the field, and limited time for implementation, my investigation will have to be constrained to selected neural network approaches. This will mean experimenting with different architectures or learning algorithms, for example.
   - My hope is to differentiate myself from previous work by including cases of severe scoliosis. As mentioned in Task 1, Cheung et al. only considered patients with Cobb angles less than 30^{o}, whereas in severe cases, this can be more than doubled or tripled. Other authors' maximum angles are similar, some up to 45^{o}.
     - As much as I want to consider severe cases, it might be unrealistic for the amount of data I have. 124 patients are a lot when you're marking all of their landmarks, but not for machine learning. Hopefully it will be enough data to train the network for typical cases, but the fraction of 124 cases which are also severe might be too small.
   - I intend to demonstrate the robustness of a neural network with respect to imperfect data. I will perform this entirely with artificially degraded landmark data taken from ground-truth CT. This means my scope is limited to the computational method; I cannot make immediate statements about its clinical applicability with verifying that the imperfections imposed on the data are representative of those actually to be found in ultrasound.
3. **Methodology**
   - I will investigate network robustness with respect to 3 types of error, intended to simulate error likely to be found in clinical ultrasound data. These types of error are random noise, deleted points representing unlocatable landmarks, and landmarks accidentally placed on ribs.

- These 3 errors are applied to landmark location data in various amounts. The controlled parameters are the standard deviation of the noise, and the proportion of points to be deleted and misplaced.
- Preprocessing will likely be necessary to get a curvature estimate from the neural net.
    - Missing values will probably need to be estimated and placed in the set. This could be done by another neural network, or other statistical means.
    - Points misplaced on ribs should either be shown to have negligible impact on network performance, or they should be detected and corrected. They could be detected and corrected, again, by neural means or otherwise.
- The network will be trained on some portion of the data, and tested on the rest. Its ability to accurately estimate the Cobb angle corresponding to that data will be measured during training and testing, and used for evaluation.

4. **Evaluation approach**
    - The evaluation of the network approach will be carried out on the basis of curvature estimation accuracy, network insensitivity to the 3 errors, and network training performance.
5. **Performance metrics**
    - Curvature estimation accuracy is computed in terms of the difference between the network's output and ground-truth angle measurements taken from CT.
    - Network insensitivity to data imperfection can be demonstrated by finding the ranges of various combinations of the error types over which the network's output is within clinically acceptable error limits.
    - The speed in terms of required training epochs, and sizes of training data sets are examples of network training performance metric which could be investigated.
    - Combining accuracy and training performance, the relationship between curvature angle and network error could be computed.
6. **Benchmarking**
    - The proposed method is not intended to provide a means to quantify scoliosis from ultrasound landmarks where there isn't one. It could be used to make such landmark based procedures more robust. Therefore, the method should provide curvature estimates of comparable accuracy to those of the procedures.
    - The accuracy of the method can also be evaluated in terms of clinically acceptable limits of error. The ranges of error amounts present in the data over which the system performs within clinically acceptable limits should be investigated.
7. **Validation plan**
    - The accuracy of the network will be validated by examining the difference between its curvature estimations and ground-truth, i.e. its error. Its error will be compared to both recent ultrasound landmark-based scoliosis quantification methods' accuracies, and to clinically acceptable limits of error. The hope is that the network will demonstrate robustness over a range of error amounts.
    - The ability of the network to generalize, given our available training data, can be investigated as the relationship between network accuracy and scoliotic severity as the ground-truth angle. The network could be trained on all but one landmark set, then run on it, for all data sets, with their varying degrees of curvature, to validate generalizability.