Project Update

We think we have reached a milestone in the project having achieved the best **test MSE** to date. Our most notable changes to our approach in the last week are listed below:

Sorted Stratified k-Fold CV A cross validation technique for sparse label distributions in a regression task context.

1. Start with a set of k empty folds F.
2. Sort the training set $\mathbb{X}$ in ascending order by label. Split $\mathbb{X}_{\text{sorted}}$ into k tiers (largest N/k points, second largest N/k points, etc ...).
3. For each element in each tier, randomly assign that element to exactly one fold in F.

This ensures that each fold has approximately the same distribution of labels. A similar selection method is used to build the test set.

Frame skipping Given that little information is to be learned between to sequential frames of a video clip x_i, x_{i+1}, we elected to skip every other frame. We are experimenting further with sequence length, but the results from these experiments are from a sequence length of 60, meaning 120 frames are sampled from a video clip, those 120 frames are reduced to a sequence of 60 after the skip step.

Architecture changes Extended temporal stretch of low level convolutional filters, added more batch normalization, shifted order of each layer block to $Conv3D \rightarrow ReLU \rightarrow MaxPool3D \rightarrow BatchNormalization$. We are also experimenting with changes to batch size.

Augmentation Added back height/width shift, vertical/horizontal flip, rotation changes.

Dataset expansion Added more videos to dataset.

Query By Committee Take the top models from various runs and poll an average prediction on the test set.

Results

Below showcases our results so far.

Minimum Individual MSE	72.316
Minimum Average MSE across 5-folds	164.039

QBC Committee Size	MSE
10	67.157
5	63.415
4	67.358
2	60.970

Next steps

Our plans for next steps are to experiment wth increasing or decreasing the sequence length supplied to the network and reincorporating respiratory rate, either as a separate network or as training one system on both values. Any feedback is appreciated.