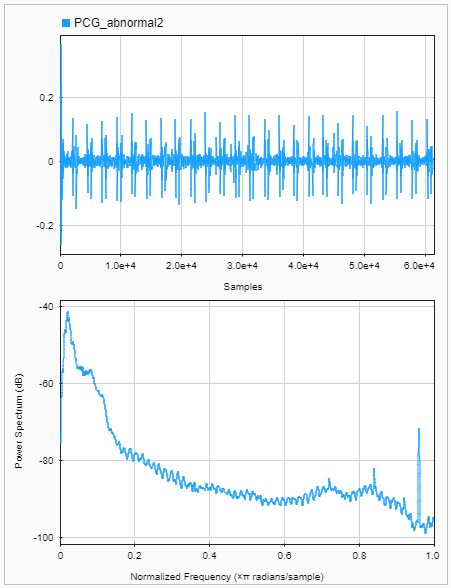
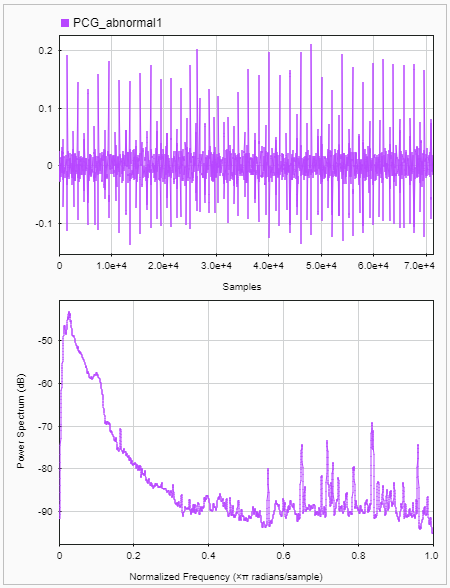
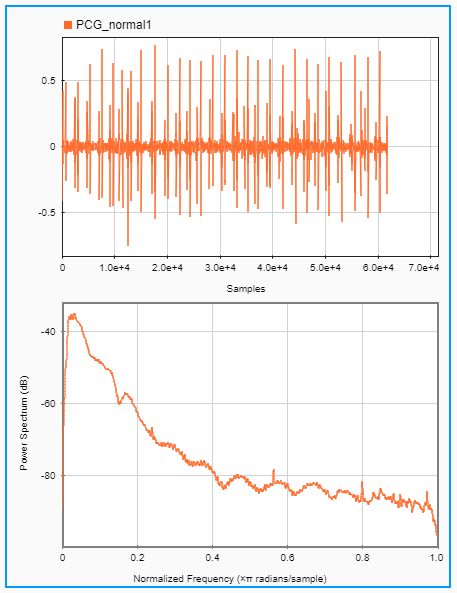
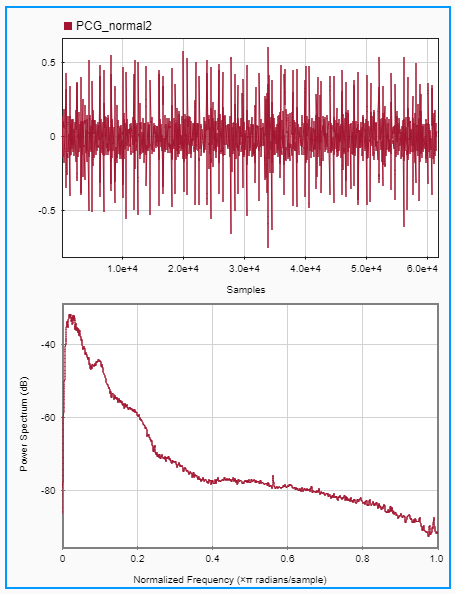
* Plot the power spectral densities of two additional normal and abnormal PCG’s (the document has an example for each). Do all the three abnormal PCG’s show a clear spike near 0.2 radian?



No, not all abnormal PCG shows a clear spike near 0.2 radian

* How many \*.wav files are there in the data/training directory (need to use Matlab to count)? How many training files have abnormal and normal PCG’s?

There are 3240 \*.wav files, 665 abnormal and 2575 normal PCG.

* How many \*.wav files are there in the data/validation directory (need to use Matlab to count)? How many \*.wav validation files have abnormal and normal PCG’s?

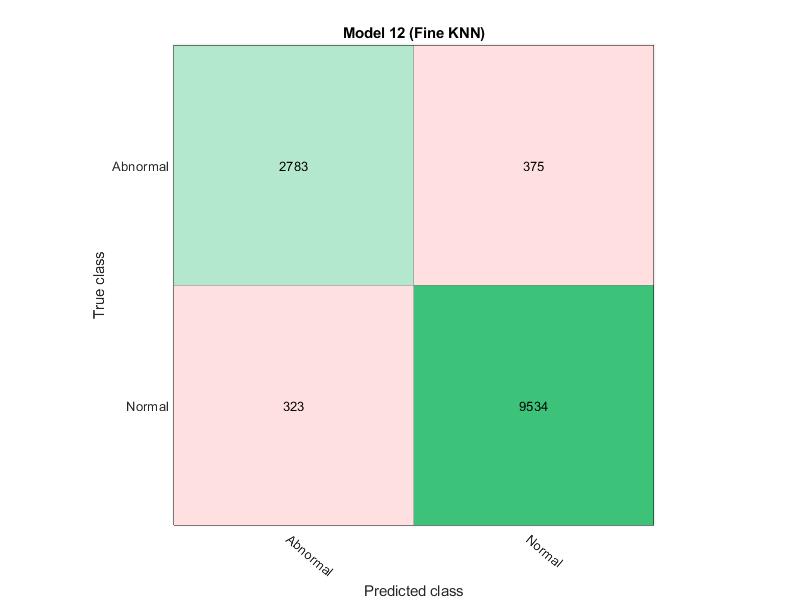
There are 301 \*.wav files, 151 abnormal and 150 normal PCG.

* What is the size of the variable feature\_table? Report the means and standard deviations of the 28 features for abnormal and normal PCG’s. By comparing the means, which features are likely most/least important for the classification.

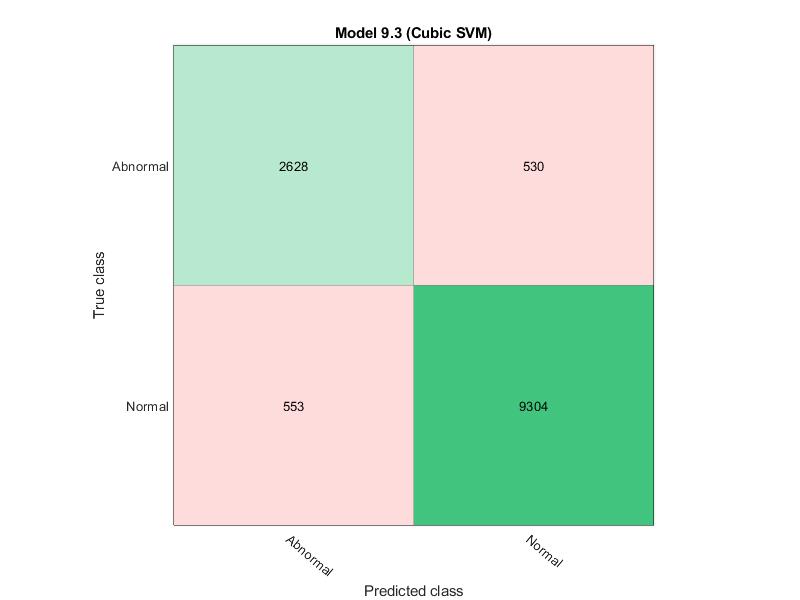
The feature\_table has a size of 13015x28. Sample Skewness seems the least important, and MFCC1 seems the most important.

* Run five classifiers using classificationLearner from at least three different groups (all models in Decision Trees are considered to be in the same group). You should have at least one classifier having an accuracy below 75% and one above 92%. Present the confusion matrices for each. Briefly explain the principle each classifier uses (in a few sentences).

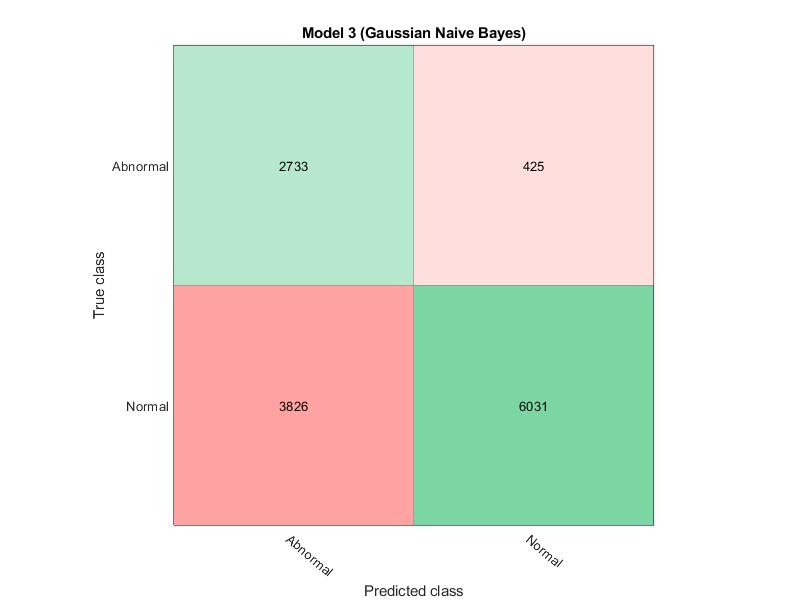
Fine KNN (Accuracy: 94.6%)

 The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.

Cubic SVM (Accuracy: 91.7%)

 The algorithm creates a line or a hyperplane which separates the data into classes.

Gaussian Naïve Bayes (Accuracy: 67.3%)

 Gaussian Naive Bayes is a variant of Naive Bayes that follows Gaussian normal distribution and supports continuous data.

* For your best model, reduce the false negative to less than 1% by modifying the cost function.
* In the Perform Feature Selection using Neighborhood Component Analysis section, what are the three most and least important features? Support your response with your own data

The three most important features are MFCC1, MFCC4 and MFCC3. For the least important features, all meanValue, medianValue, standardDeviation, meanAbosulteDeviations, quantile25, quantile75, signalQR, sampleSkewness, signalEntropy, spectralEntropy, dominantFrequencyMagnitude, dominantFrequencyRatio are not needed.

