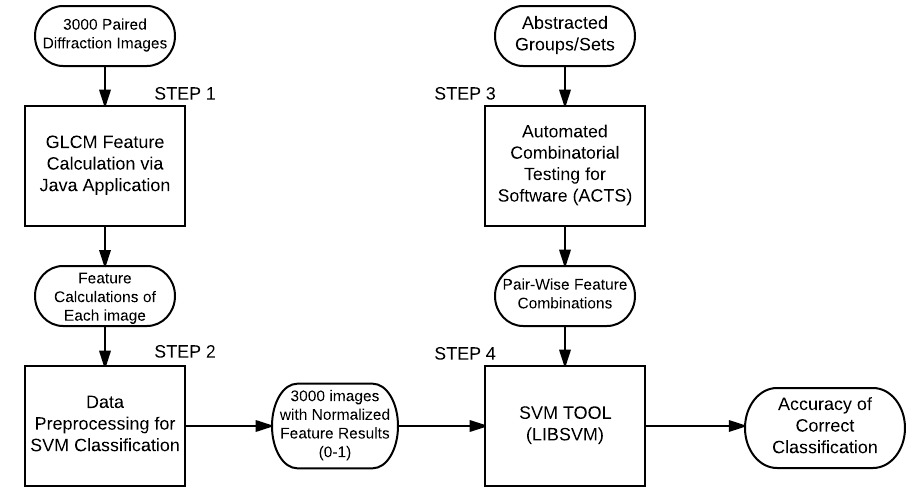
Pair-Wise Feature Selection Process



1. Data Flow Diagram of the Feature Selection/Image Classification Process

The experimentation included the following main steps:

Step 1. Calculating feature values from diffraction images based on Grey Level Co-occurrence Matrix (GLCM). 6000 images, specifically 3 sets of 1000 paired images, are processed using a Java-based GLCM application. This application derives the numerical value for 40 features, both p-polarized and s-polarized features. The output of this application is a .csv file which contains the image name in the first column, followed by 40 columns of numerical values which represent the efficacy of the feature assigned in that column. Each column represents a specific feature that was calculated.

Step 2. Preprocess feature results to allow classification via SVM Tool. To make the feature results appropriate for SVM classification, the data must be normalized to a scale of 0-1. The preprocessing outputs temporary data consisting of 3000 images with scaled features.

Step 3. Calculating Pair-Wise Feature Combinations. The 40 features, 32 viable features and 8 null features, are abstracted into groups, each group contains a set of features. Using the grouping abstraction as input, the ACTS tool creates an output of feature combinations. The number of groups determines the number of features in a combination. The degree of the t-wise (2-way, 3-way) will determine the number of combinations. The output of the ACTS tool is a csv file which contains an index, followed by feature numbers, these numbers correlate to the order of features recorded in step 1. These feature numbers indicate which features will be used to perform classification.

Step 4. Perform classification of diffraction images. The SVM classification tool requires 3 inputs, the training data, the testing data, and the Pair-Wise feature combinations. The 3000 scaled feature images are dynamically divided using 10 folder cross validation; this process splits the 3000 images into 10 equal sets, 9 of these sets will act as training data, while the remaining set will be used as testing data. The SVM tool then create a SVM classifier based on the training data set to classify the testing data into cell, strip and debris using a single combination of features provided by the Pair-Wise Feature Combinations. This process outputs the accuracy of classification for the given testing set and feature combination. Once this is complete, the result is stored temporarily and the cross validation is performed 9 more times in order for each respective set to act as the testing data, This allows for all 3000 images to be tested. Once the accuracy is calculated for each of the 10 folders, the average accuracy of all 10 folders calculated and recorded. This entire process is then repeated for each pair-wise combination.

Once step 4 has been complete, the process is repeated starting from step 3.