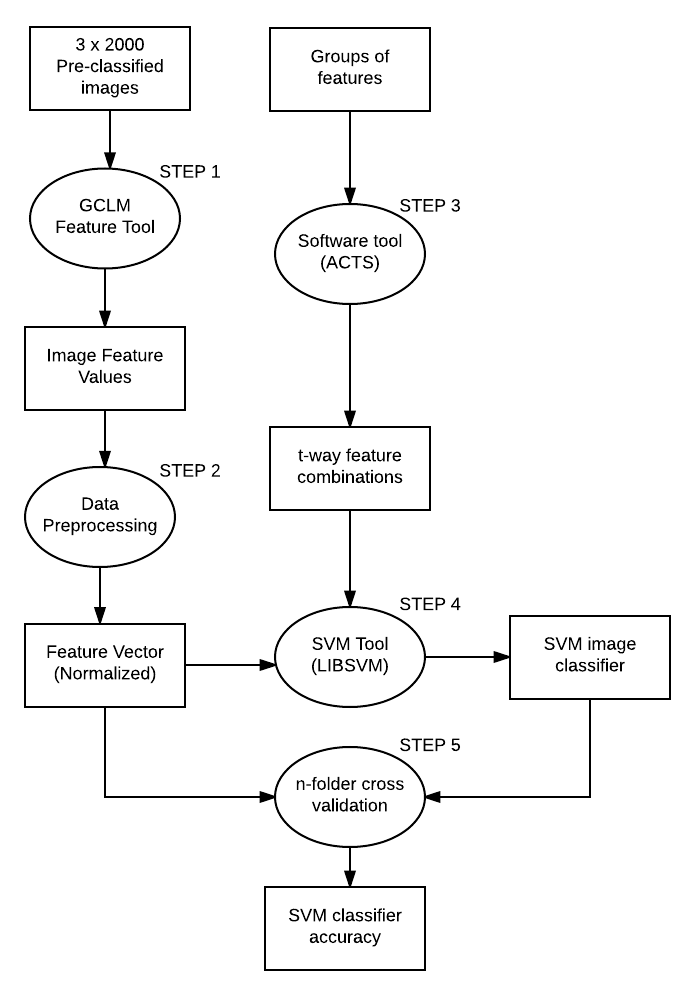
Pair-Wise Feature Selection Process

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 2000 paired images, are processed using a Java-based GLCM application. These images have been pre-classified as either cell, strip or debris. This application derives the numerical value for 40 image 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 feature values which represent the efficacy of the feature assigned in that column. Each column is assigned a specific feature that is 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 step then creates a feature vector, by combining the normalized feature values with the respective image labels. The preprocessing outputs feature vector consisting of image classification labels with normalized features.

Step 3. Calculating T-way feature combinations. The 40 features (32 viable features and 8 null features) are divided into relevant groups, each group contains a subset of features. Using the grouping 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 feature number of combinations. Using the ACTS tool, feature combinations were created for 2, 3, 4, 5, and 6-way 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. Create classifier for diffraction images. The SVM classification tool requires 2 inputs, the training-testing data, and the Pair-Wise feature combinations. The 3000 scaled feature images are dynamically divided into 10 equal parts, this process splits the 3000 images into 10 equal sets, and 9 of these sets will act as training data, while the remaining set will be used as testing data. Once the testing and training data is established, the SVM tool uses the combination of features produced in step 3 to create and train a classifier. This process is repeated 9 more times so that each subset of image labels is treated as the testing data.

Step 5. Calculate classifier accuracy with 10 folder cross validation. For each classifier created in step 4, the classifier is used to label one of the testing subsets also created in step 4. Using the label results of the SVM classifier, and the pre-classified labels in the feature vector, the accuracy of the classifier is determined by the number of correctly labeled images. This step is repeated 10 more times so that the SVM classifier will label each subset of the images. The average of these 10 accuracies is then recorded as the accuracy of the classifier.

Overall Process.

Steps 4 and 5 are repeated for every respective combination of t-way combinations created in step 2. Once all combinations of features have been used to determine classifier accuracy, the process begins from step 2 given a new grouping of features.