

An Assessment of the Impact of Dimensionality Reduction on the Speed and Accuracy of Hyperspectral Image Classification

Abhineet Singh¹, Savita Verma², Mangal Raj³, Anupam Agrawal⁴

Abstract

This paper investigates the extent to which the accuracy and speed of classifying Hyperspectral remote sensing images are affected by the application of varying degrees of dimensionality reduction. Three methods have been used for dimensionality reduction: PCA, ICA and random band selection; SVM has been used for classification. The results have been evaluated on both natural and man-made scenarios.

Keywords

Hyperspectral, dimensionality reduction, classification, support vector machine, sequential minimal optimization, PCA, ICA.

1. Introduction

Hyperspectral imagery is a relatively recent development in remote sensing technology that has the potential to allow the extraction of much finer details about a region on earth's surface than was possible with older technologies like multispectral imagery. A hyperspectral image consists of a large number of very narrow contiguous spectral bands, each around 10 to 20 nm wide. The number of bands can vary from few tens to several hundreds and usually cover visible through middle infrared spectrum ranges (usually between 0.4 to 2.4 micrometers). This property of hyperspectral images enables the construction of a continuous spectral signature for each pixel of the image which can be compared to the spectral signatures of known substances to identify the composition of the land surface at a very fine level of detail.

Abhineet Singh, Department of IT, IIT Allahabad, Allahabad, India.

Savita Verma, Department of IT, IIT Allahabad, Allahabad, India.

Mangal Raj, Department of IT, IIT Allahabad, Allahabad, India.

Anupam Agrawal, Department of IT, IIT Allahabad, Allahabad, India.

However, hyperspectral images suffer from the disadvantage of requiring very high computational and storage/transmission bandwidth because of the large quantity of data involved.

In addition, the presence of large number of narrow bands leads to the problem of redundancy since neighbouring bands often contain very similar information and are therefore highly correlated. Thus, some form of dimensionality reduction is needed to eliminate this redundancy while still retaining sufficient information to achieve acceptable classification accuracy.

The ICA based dimensionality reduction method used here is detailed in [1, 2] and uses band selection rather than feature extraction [3]. The FastICA algorithm [4] has been used to implement ICA. The algorithm used for implementing the SVM classifier is called Sequential Minimal Optimization and has been introduced in [5] with a more detailed description, including some ideas for implementation, being presented in [6].

2. Methodology

Datasets and Pre-Processing

Two data sets have been used for evaluating the results:

a) *Indian Pines Test Site (IP)*: This is a 224 band image of north-west Tippecanoe County, Indiana captured by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor. The image has a size of 145x145 pixels and is divided into 16 classes according to the ground truth map. Out of the originally captured 224 bands, 4 contained no information and had been discarded at source, so that the downloaded data had only 220 bands.

b) *Washington DC Mall (WDC)*: This is a 210 band image of a mall in Washington DC, captured as part of the Hyperspectral Digital Imagery Collection Experiment (HYDICE). The image is 307x1280 pixels in size and is divided into 6 classes according to the ground truth map.

Both datasets and their ground truth maps are available at [7]. Before carrying out dimensionality reduction and classification, several pre-processing steps have been applied to simplify the subsequent

tasks. These include atmospheric correction, elimination of water absorption bands (both using ENVI software package) and removal of bands containing too much noise by visual inspection. These methods resulted in the elimination of 35 bands from AVIRIS data set (20 were in water absorption region rest had too much visible noise) and 19 bands from the 210 band HYDICE image. Thus, 185 bands were retained for the former and 191 bands for the latter.

Algorithms

Results have been reported here for three different approaches to dimensionality reduction (DR). The first one uses the ICA separating matrix to evaluate the relative information content of each of the bands and selects the most informative ones [1, 2]. The second approach is based on Principal Component Analysis (PCA) and uses eigenvectors to select the bands corresponding to the direction of maximum covariance. The third method is called random band selection (RBS) and involves selecting a few unmodified bands at random from the original data. This last method has been used to evaluate the advantages of applying the aforementioned transformations (ICA/PCA) to the data over simply using a random subset of the original bands. Dimensionality reduction is followed by the use of a SVM based classifier [8] to detect different classes of the land cover from the images. SVM uses margin maximization to find a linear boundary between two classes of data. When the two classes are not linearly separable in the input space, a kernel function is first used to transform the data into a higher dimensional feature space where the classes are linearly separable. The optimal separating hyper plane is then estimated in this feature space and mapped back to the input space using the inverse kernel function to provide the complex decision surface required to separate the classes.

Since SVM is a binary classification method, it must be adapted for carrying out multi-class classification. There are several strategies for accomplishing this [8], three of which have been used in this work:

a) *Parallel one against all (OAA)*: This involves building one SVM for each class such that each SVM compares the corresponding class against all others. The test sample is then tested against each of these and is classified to the class, the SVM corresponding to which produces the maximum output. b) *Parallel one against one (OAO)*: This involves building one SVM for comparing each possible pair of input classes thus leading to the construction of $k(k-1)/2$ SVMs for k class problem. The test sample is

evaluated against each of these and is assigned to the class with the maximum number of victories. c) *Binary hierarchical tree with balanced branches (BHT)*: This involves building a hierarchy of SVMs such that each compares two groups of classes with roughly equal apriori probabilities. The SVM corresponding to the root of the tree divides the input classes into two roughly equal parts; each of its children further divide these into two equal subgroups and so on till each node corresponds to only one class. A test sample is evaluated by passing it down the tree from parent to child depending on which child wins at each step. The classification decision is made implicitly by the leaf node it ends up at.

3. Testing and Results

Setup

All tests have been conducted on an Intel Core-i7 950 based system clocked at 3.06GHz and having 6 GB of RAM. MATLAB has been used for implementing the algorithms. The total pixel counts for the IP and WDC datasets are 21025 and 392960 respectively (these reduce to 10366 and 136992 respectively after discarding the pixels marked as background in the ground truth map). Due to limitations of time and computational resources, only subsets of these pixels, selected randomly from the entire image, have been used for evaluating the results.

Since the distribution of pixels between different classes cannot be predicted in any random selection, it is possible for most pixels to belong to only one or two classes while leaving several other classes with too few pixels. Such an uneven distribution can result in the SVM classifier becoming biased towards the majority classes and giving incorrect classification for the others. To avoid such misclassifications, only those classes whose pixel counts are above a certain threshold are used for training (and subsequent testing) in any given run; remaining classes are discarded. Half the remaining samples are used for training and the other half for testing. The value used for this threshold here is 5% of the total samples (i.e. 50,100 and 500 respectively for 1000, 2000 and 10000 samples).

Experiments have been conducted with two configurations of the SVM classifier: one combines a small value of the learning parameter (parameter 'C' in [6]) with a simpler linear kernel while the other one uses a much larger value but with a more complex non-linear kernel based on the radial basis function (RBF). The former configuration (C=0.05/Linear

kernel) is much faster at the cost of lower accuracies while the latter (C=40/RBF kernel) is slower but gives better accuracies, especially in more difficult scenarios.

Classification accuracy is measured in terms of the percentage of total testing pixels classified correctly. Run time is the total time taken (in seconds) for performing both dimensionality reduction and classification.

Result Summary

- 1) **No dimensionality reduction:** To establish a baseline for performance comparisons, classification accuracy and run time were evaluated for both datasets without applying any dimensionality reduction. To establish a baseline for performance comparisons, classification accuracy and run time were evaluated for both datasets without applying any dimensionality reduction. These are summarized in Table 1.

Table 1: Classification Results with No Dimensionality Reduction

| Classification Accuracy (percent) | | | | | | |
|-----------------------------------|--------------------|-------|-------|------------------------|-------|-------|
| Number of samples | Washington DC Mall | | | Indian Pines Test Site | | |
| | OAA | OA0 | BHT | OAA | OA0 | BHT |
| 1000 | 92.96 | 91.75 | 91.95 | 76.83 | 79.91 | 77.78 |
| 2000 | 94.27 | 92.56 | 90.95 | 81.43 | 84.13 | 80.07 |
| 10000 | 94.09 | 93.89 | 93.49 | 88.32 | 92.47 | 87.80 |
| 25000 | 94.25 | 94.66 | 94.22 | NA | NA | NA |
| Classification Time (seconds) | | | | | | |
| 1000 | 432 | 103 | 218 | 610 | 239 | 467 |
| 2000 | 1678 | 327 | 441 | 2561 | 992 | 1662 |
| 10000 | 18782 | 4914 | 10623 | 27333 | 8701 | 13071 |
| 25000 | 84663 | 18473 | 52692 | NA | NA | NA |

- 2) **Varying SVM multi-class strategy:** It is evident from Table 1 that OAO multi class strategy not only gives better accuracy than the others (in most cases) but is also the fastest, often by a large margin too. To verify if these observations hold true with less bands too, tests were conducted using RBS-DR method. The results are presented in Fig. 1.
- 3) **Applying dimensionality reduction:** Plots were obtained for both datasets by varying the retained dimensions from 5 to 175 in steps of 10 and evaluating the accuracy and run time for each such run. All tests were conducted using OAO multi class

strategy with the slower and more accurate configuration of the SVM classifier. The results are summarized in Fig. 2 and 3.

- 4) **Varying SVM configuration:** As mentioned earlier, two different configurations of the SVM classifier were used for evaluating the results. These results are summarized in Fig. 4. All the reported results were obtained using OAO multi class strategy combined with PCA-DR method. The results of other DR methods followed a similar trend and are not reported here to avoid cluttering the graphs.

Result Analysis

Several points can be observed from the results. To begin with, it can be noted from Fig.1 that the trend observed earlier from Table 1 holds true for fewer dimensions too; OAO is usually the most accurate and also the fastest while OAA, though giving good accuracies, is slower by very large margins. BHT performs in between these two though its speed is much closer to that of OAO.

Moving on to Fig. 2 and 3, the first point to be noted is the significant difference in performance of the same DR method between the two datasets, both in terms of accuracy and processing time. The WDC dataset not only gave much higher accuracy figures than IP dataset but also required significantly less time to process. There are two probable reasons for this. Firstly, the WDC dataset features a man-made scenario with classes consisting of well-defined structures like buildings, roads, water bodies and gardens, that exhibit relatively less intra-class variations in their reflectance characteristics. The IP dataset, on the other hand, represents a natural undeveloped scenario where classes consist of individual types of plants like wheat, soybean, corn, oats and so on. These classes are bound to exhibit a lot more intra class variations than the comparatively well separated classes of the WDC dataset. Secondly, the IP dataset also features much larger number of classes than the WDC dataset (16 vs. 6) which obviously makes the task of correctly classifying it more difficult.

To summarize, it can be stated that the former involves a much finer level of classification and therefore requires more bands to extract sufficient information from the data. Hence its accuracy curves tend to reach saturation around 50-60 bands (Fig. 3)

as opposed to the 10 band saturation achieved by the WDC dataset (Fig. 2)..

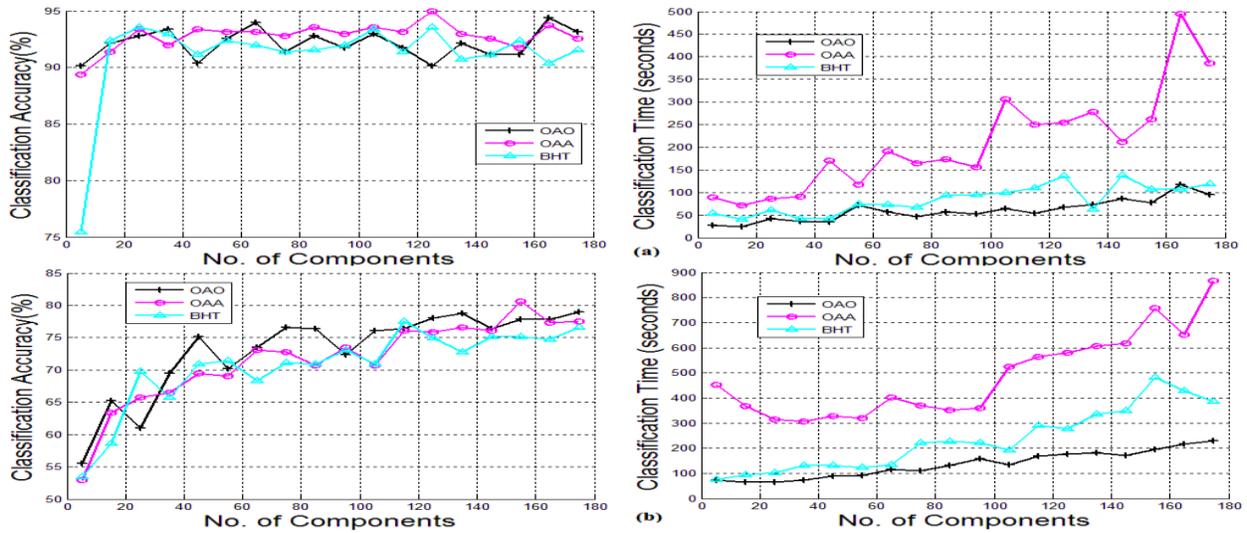


Fig. 1: Classification results for different SVM multi class strategies using 1000 samples with RBS-DR: (a) Washington DC dataset (b) Indian PINES dataset

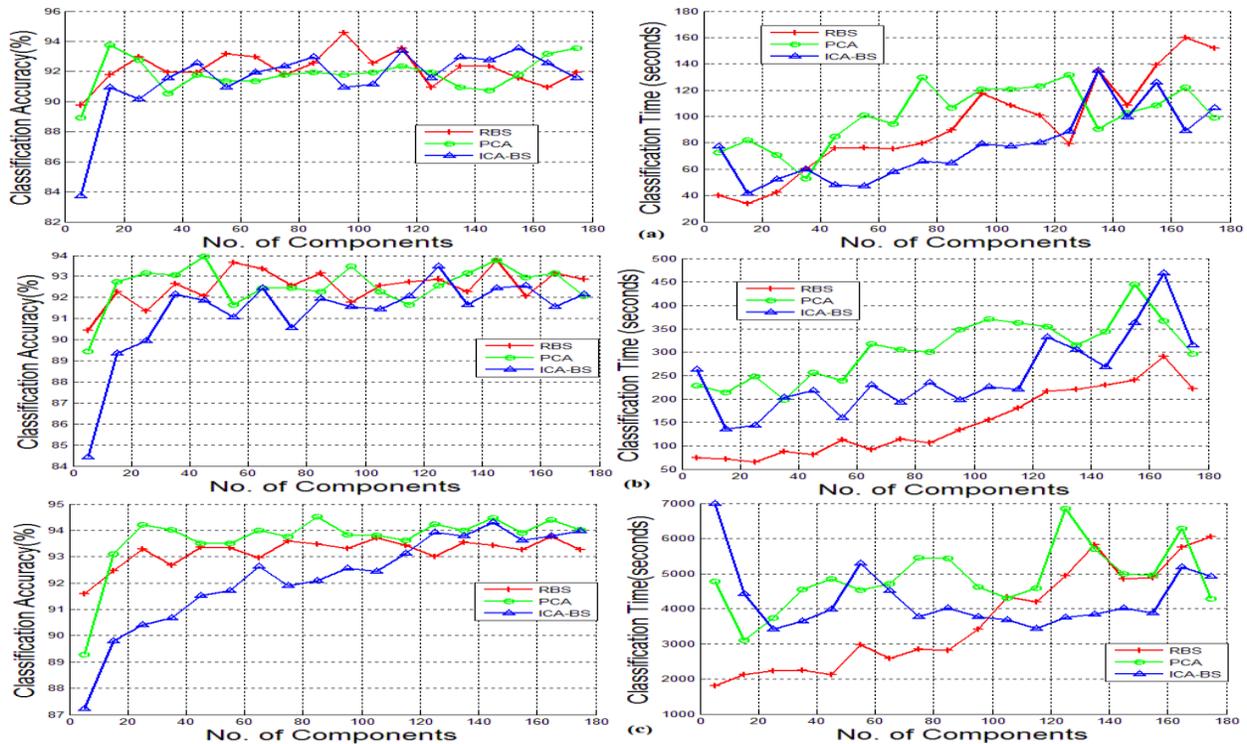


Fig. 2: Classification results for different DR techniques using Washington DC dataset and OAO strategy: (a) 1000 samples (b) 2000 samples (c) 10000 samples

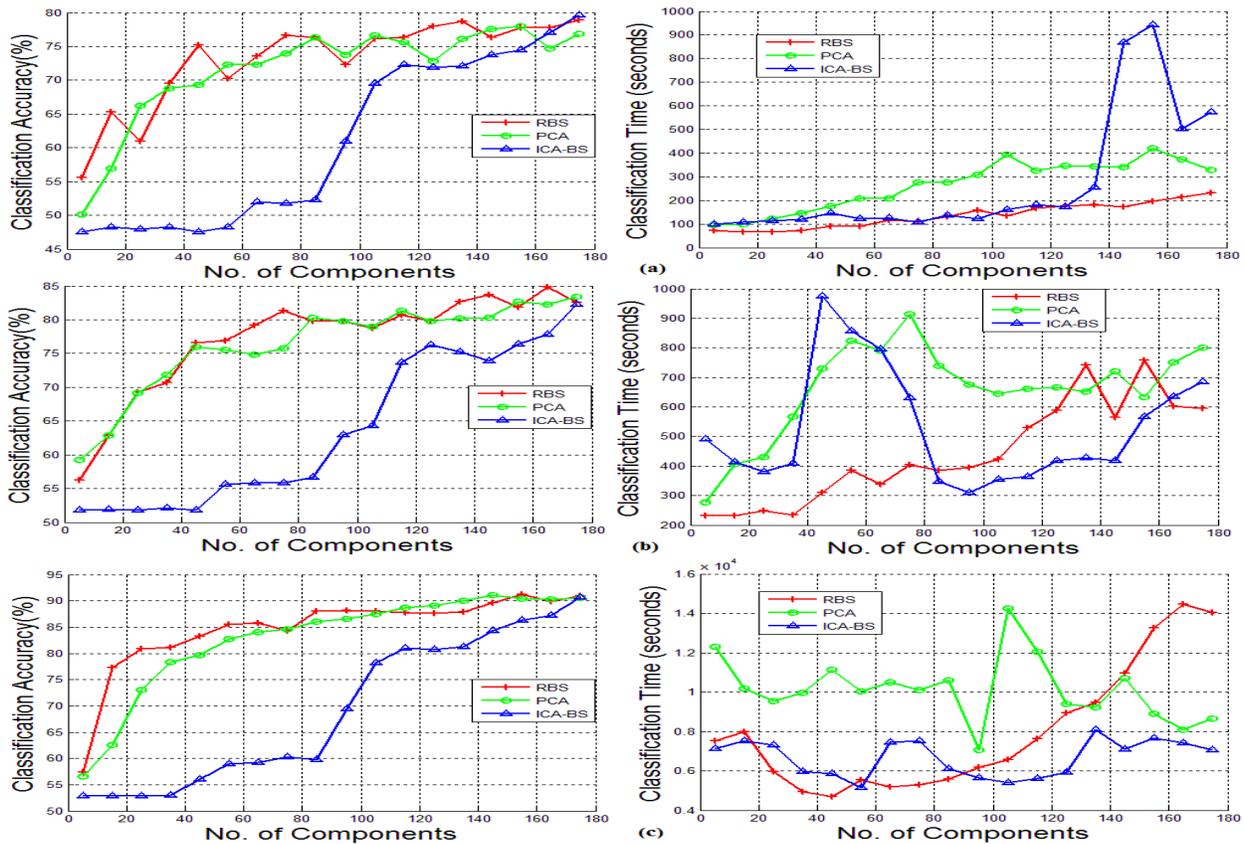


Fig. 3: Classification results for different DR techniques using Indian Pine dataset and OAO strategy: (a) 1000 samples (b) 2000 samples (c) 10000 samples

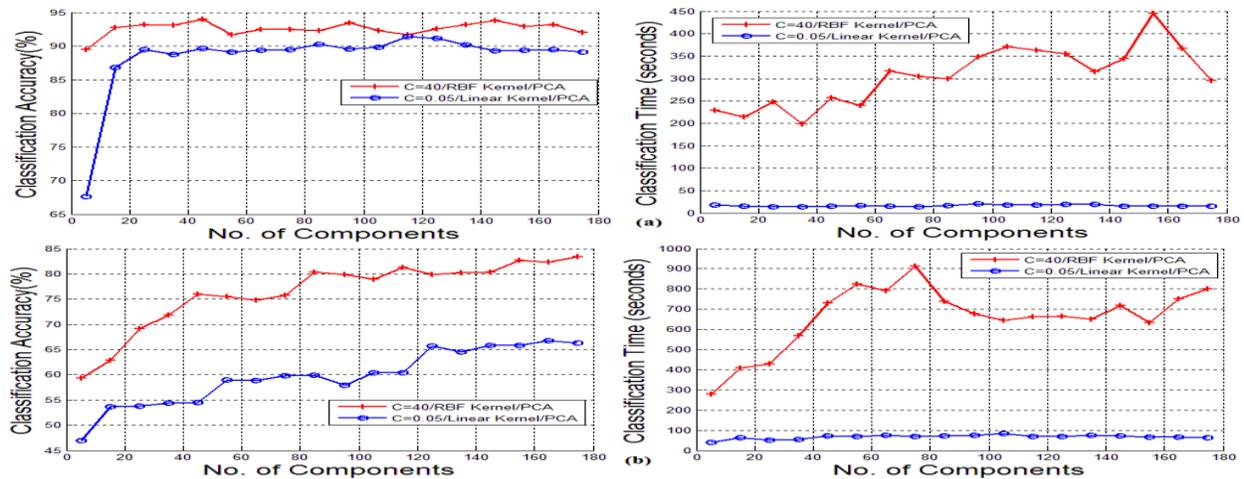


Fig. 4: Classification results for different SVM configurations using 2000 samples with PCA-DR: (a) Washington DC dataset (b) Indian Pine dataset

Another important observation is that even though ICA is the most mathematically robust of the three techniques tested here, it invariably gave the worst

results. In addition, the performance gap between ICA and the other methods was much wider in the more difficult case of the IP dataset where it takes

around 120 bands to provide classification accuracy even comparable to (but still worse than) the others. This is contrary to expectations since a more robust method would be expected to perform better in more challenging scenarios. In fact, with a few exceptions, the simplest (and trivial) DR technique of RBS gave the best performance out of the three, though PCA was usually too close, at least in terms of accuracy, to give it a clear advantage.

It can further be noted from the curves of PCA and ICA-BS in Fig. 2(c) and 3(c) that the classification time does not increase on using more bands once the sample size becomes large enough, a tendency that can be observed when less samples are involved (Fig. 2(a, b), 3(a, b)). This in turn indicates that applying sophisticated DR techniques provides no advantage with respect to processing time provided the number of samples involved is large enough. It can also be seen that the trivial RBS-DR technique shows an almost linear increase in processing time in all 3 test cases. This does not indicate any performance disadvantage for this technique, however, since its run time remains comparable to the other methods even when it reaches maximum; it merely indicates that RBS-DR gradually loses its speed advantage over other methods as the number of bands increases. Comparing the two configurations of the SVM classifier further reinforces the earlier observation regarding the discrepancy in the classification difficulty levels of the two datasets since the performance difference between the two configurations is much less for the WDC dataset (Fig. 4(a)) than for the IP dataset (Fig. 4(b)). Another noticeable point here is the flat nature of the time curves of the simpler configuration ($C=0.05$) as opposed to the (mostly) increasing time curve of the more intensive configuration ($C=40$), signifying that the use of more dimensions negatively impacts the classification speed only when a more rigorous classifier is used. Finally, the number of bands required for the accuracy curves to reach relative saturation was same for both configurations, thus indicating that intensiveness and complexity of the classifier have little influence on the number of bands required to extract (most of) the useful information from hyperspectral data; it depends only on the complexity of the dataset involved.

4. Conclusion and Future Work

Several tests were carried out using three different DR methods combined with three multi class strategies and two configurations of a robust SVM

based classifier. Results were evaluated on both natural and man-made scenarios and several interesting points, pertaining to the impact of dimensionality reduction relative to the complexity of the classifier and the dataset involved, were cleared by the analysis of these results. However, all the DR methods used here were comparatively simple ones and conducting additional tests with more mathematically rigorous DR techniques may throw additional light on the unexpected poor performance of the ICA DR method and the lack of any advantage of PCA over the trivial RBS technique.

References

- [1] H. Du, H. Qi, X. Wang, R. Ramanath and W. E. Snyder, "Band Selection Using Independent Component Analysis for Hyperspectral Image Processing" *Proceedings. 32nd Applied Imagery Pattern Recognition Workshop*, pp. 93-98, Oct. 2003.
- [2] H. Du "Dimensionality Reduction using Parallel Independent Component Analysis in Hyperspectral Image Analysis" Internet: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.133.4883&rep=rep1&type=pdf>, Feb 6, 2012.
- [3] J. Wang and C. Chang "Independent Component Analysis-Based Dimensionality Reduction With Applications in Hyperspectral Image Analysis." *IEEE Transactions on Geoscience and Remote Sensing*, pp.1586-1600, June 2006.
- [4] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis" *IEEE Transactions on Neural Networks*, Volume: 10 Issue: 3, pp.626 – 634, May 1999.
- [5] J. C. Platt "Sequential Minimal Optimization: A Fast Algorithm for training Support Vector Machines" Technical Report MSR-TR-98-14, Microsoft Research, 1998. Available at http://www.bradsblock.com/Sequential_Minimal_Optimization_A_Fast_Algorithm_for_Training_Support_Vector_Machine.pdf.
- [6] "Sequential Minimal Optimization for SVM" Internet: www.cs.iastate.edu/~honavar/smo-svm.pdf, June 2013.
- [7] "MultiSpec Hyperspectral Images," <https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>, June 2013.
- [8] F. Melgani and L. Bruzzone, "Classification of Hyperspectral Remote Sensing Images With Support Vector Machines" *IEEE Transactions on Geoscience and Remote Sensing*, Volume: 42 Issue: 8, pp. 1778 – 1790, Aug. 2004.