Project Module 2 : Features Extraction

Bahnisikha DUTTA, Mathieu LAPEYRE, Palash SHASTRI

Georgia Institute of Technology, ECE 6780, Group 9

Atlanta, GA

Keywords: Features extraction; Features reduction; Features reduction; PCA; t-SNE; Mutual Information; Statistical Dependency.

I. CLINICAL NEED

Labeling a histopathology image as having cancerous regions or not is a critical task in cancer diagnosis [1]. Existing image classification techniques require detailed manual annotations for the cancer pixels, which are time-consuming to obtain. With the dramatic improvements in computational power, computer-assisted diagnosis has been made easier. Today, pathologists need the support of a clinical decision system for diagnosis and prognosis of cancer.

II. PROBLEM STATEMENT

After achieving segmentation, the second thing to do is features extraction. The goal of feature extraction is to construct a set of informative data (the features) from the original input (the images) to help human having a better interpretation of the original data. This feature extraction can be based on stained images or not. For pathologists, diagnosis criteria are inevitably described using terms such as “nucleus” and “cell.” It is thus important to develop computer vision methods capable of such

object-level analysis. To achieve such feature extraction, we need to retrieve as many features as we can within 3 categories:

- Color features
- Shape features
- Texture features

Then, we want to reduce the number of features by selecting the most relevant ones, based on several criteria. Finally, we want to reduce the number of features again, to only 2 or 3 dimensions so that a human can easily distinguish the images between each other. This will be done through dimensionality reduction and the final result will be analyzed thanks to several performance metrics and comparison to ground truth labels.

III. LITERATURE CRITIQUE

Our features extraction approach is based on several papers that have been published for the last 20 years. We focused on papers which present different techniques about features extraction, features selection, features reduction and performance evaluation. We reviewed several papers to be able to choose the most efficient techniques for each of these domains. We wanted a varied set of features that could capture the inherent differences between images within different datasets. You can find the results of these reviews below:

Multifeature Prostate Cancer Diagnosis and Gleason Grading of Histological Images, *IEEE Transaction on Medical Imaging, Vol. 26, No. 10, October 2007[11]:* This paper presents an elaborate study of various image features for cancer diagnosis and Gleason grading of the histological images of prostate. We mainly referred this paper for our color extraction step. Most papers talk about color

extraction methods based on extracting the color histograms from the images with 'n' bins where each bin represents a feature. This paper had an unique color extraction method in the sense that it preprocessed the cancer images first by removing the white background pixels so as to remove any bias that may affect the histogram analysis.

Object- and Spatial-Level Quantitative Analysis of Multispectral Histopathology Images for Detection and Characterization of Cancer, *L. Boucheron et al, University of California Santa Barbara, 2008 [4]*: This paper is a very comprehensive review of techniques that have been and can be used for histopathology image analysis. For this module's purpose, feature extraction was the most important information that was referred to. This paper describes a lot of shape features that can be utilized for classification purpose. The major advantage of using this paper as reference was providing us with an overview of the commonly used shape features. The thesis doesn't describe how you would combine statistics for a tiled image and hence we estimate a distribution amongst the tiles using the statistics described in the feature extraction section.

Textural Features for Image Classification, *Robert M. HARALICK et al, IEEE Transaction on System, Man and Cybernetic, 1973 [2]*: Texture has been widely used to classify images between each others. Texture features can be model-based (fractals, autoregressive, markov random fields, …) or non model based (GLCM, Frequency domain, GLRL, Gabor filters, …). In his thesis [3], Michael J. Chantler stated that non-model based texture features yield better result than the model-based ones. Of course this highly depend on the input images, but overall co-occurrence matrix is reportedly better than other approaches on several general applications. This is why we chose GLCM and its easily computable 14 features for texture-based extraction.

	Co-occurrence	Sum and difference	Laws' masks	Run length	Fractal dimension	PSD wedges and rings	PSD peaks
Weska	1	1		2		2	
Conners	1	1		2		2	
duBuf	1		1		2		
Castrec	1	1	2				
He	1					2	3

Figure 1: Comparative study of texture features. 1 is best. [3]

Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits, *Pohjalainen et al, Computer Speech & Language, 2015[10]:* This paper had an elaborate list of feature ranking and selection algorithms namely Statistical Dependency(SD), Mutual Information(MI), Sequential Forward Selection(SFS), Sequential Backward Selection(SBS), Minimal-redundancy–maximal-relevance (MRMR) feature subset selection and Random Subset Feature Selection (RSFS). All the algorithms were discussed at great length along with proper explanation of their pros and cons. Various ways of combining multiple feature selection methods were also discussed. In addition, the paper also had an elaborate section on various performance evaluation methods for evaluating the feature selection algorithms.

Visualizing data using t-SNE, *Laurens van der Maaten et al, Journal of Machine Learning Research, 2008 [5]*: There are plenty of different approaches to reduce the dimension of a dataset so that it can be visualized in a 2D or 3D scatterplot. The t-Distributed Stochastic Neighbor Embedding is a recent non-linear technique that is particularly well suited for the visualization of high-dimensional datasets. Lee and Verleysen [6] reviewed many non-linear techniques, which keep the low-dimensional representations of very similar data points close together. However, we retained t-SNE as it preserves both local and global structure of the dataset.

PCA vs LDA, *Aleix M. Martonez et al, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001 [7]*: PCA and LDA are popular linear reduction techniques, especially for face recognition. One might think that LDA is better than PCA, but in some cases (especially with small training set) PCA performs better. PCA is a statistical procedure that orthogonally transforms the original n coordinates of a data set into a new set of n coordinates whose first principal component has the largest possible variance. Each succeeding component has the highest possible variance under the constraint that it is orthogonal to the preceding components. PCA and LDA are both linear techniques that can not keep local behavior of very similar datapoints, but we decided to use PCA as we can examine the performance between a nonlinear technique (t-SNE) and PCA.

IV. OUR APPROACH

Our approach to the problem at hand can be described by the following flow chart:

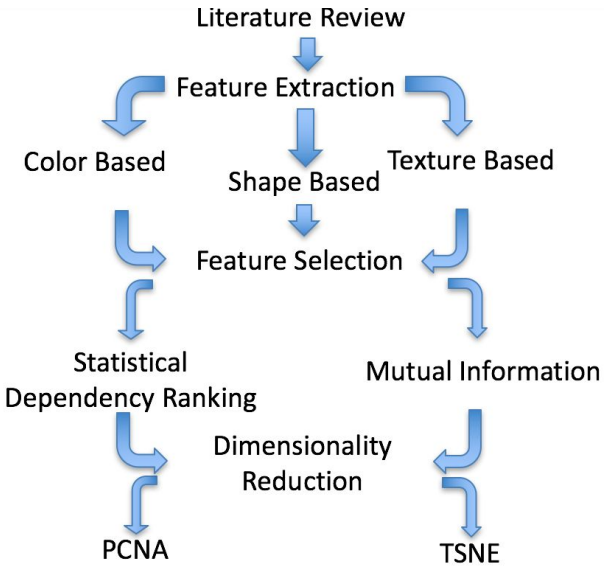


Figure 2: System design flowchart

Each of these steps are explained at length in the following sections:

A. FEATURES EXTRACTION

i. Color Features

Color based features are very important in histology applications since the biologists stain tissues to highlight special structures. These special structures in our case is the cell and nucleus and it is quite crucial that we have a full idea of their color characteristics in order to classify these images into various grades of tumor which is our ultimate goal. We followed the strategy attained by Tabesh et al for our color extraction step.

First, the color tissue image is transformed from the RGB space into the YCbCr space . The luminance (Y) component is then thresholded using an empirically determined, fixed, global threshold , 222 in our case. The color histograms were then obtained from each of the RGB channels for 16 bins. We additionally repeated the same process for other color spaces namely LAB and HSV. So, we had 16*9=144 color features in total. (Refer Appendix)

ii. Shape features

Shape features play a very important role in understanding object level characteristics. Shape features measure the underlying object level metrics that could prove very useful for classification purposes. The most commonly used shape features capture the area, completeness, orientation and structural information like perimeter information, convexity, eccentricity etc. The entire list of shape features used can be found in the features table [refer to table].

The underlying methodology for shape feature extraction can be summarized in the following steps:

1. Label Matrix is obtained using Unsupervised and Supervised Segmentation Algorithms

2. Each individual segmented image (3 in total) is converted to a binary image
3. A smoothing operator is applied to smooth out portions that are less than 40 pixels in area
4. Matlab operator **regionprops** is applied to obtain region characteristics for the binary image

After the above steps are undertaken, various region vectors are obtained as output. We measure eight important statistics that capture the underlying distribution of vector - **minimum, maximum, mean, median, standard deviation, inter quartile range, skewness and kurtosis**. These statistics have been used in literature for shape features [4][12].

iii. Texture features

The main tool we used to extract texture features is based on gray-level co-occurrence matrix (GLCM). The GLCM is applied to each of 3-stained images for one original image. The original image is masked with the stained image and feed the texture extraction function.

The underlying methodology for shape feature extraction can be summarized in the following steps:

1. Load stain-masked images (nucleus, cytoplasm, background) as grayscale images
2. Compute GLCM matrix in 5-pixel displacements and 4 orientations.
3. Compute 22 features for each matrix
4. Merge the 3 files (one per stain) into one

Below are some examples of the features that can be computed from a GLCM matrix:

Correlation: $$\frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

Cluster Prominence: $$\sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i, j)$$

Homogeneity: $$\sum \frac{p(i, j)}{1 + (\mathrm{i} - \mathrm{j})^2}$$

…

Where ux, uy, bx, by are respectively the mean and the standard deviation of the rows and columns of the element p(i,j), which is the (i,j)th entry of the GLCM.

Altogether, there are 440 texture features for each dataset (5dx*4theta*22features). The list of those features can be seen at the end of this document.

B. FEATURES SELECTION, RANKING

i. Statistical Dependency (SD)

The goal of the *statistical dependency* (SD) method is simply to measure whether the values of a feature are dependent on the associated class labels, or whether the two simply co-occur by chance. Each feature value is first quantized into one of the *Q*S levels, where the feature-specific quantization scale is adaptively determined such that each bin will contain roughly an equal amount of samples over the entire data set. The bins are chosen in this way, instead of a conventional uniform quantization scale, in order to lend some statistical validity to the occurrence of different quantization levels. The statistical dependence between the discretized feature values *y* and the class labels *z* is evaluated according to the formula

$$\mathrm{SD} = \sum_{y \in Y} \sum_{z \in Z} p(y, z) \frac{p(y, z)}{p(y) p(z)}$$

The larger the SD, the higher is the dependency between the feature values and the class labels.

ii. Mutual Information (MI)

Mutual information is a method that draws its importance from information theory. Information gain is the primary underlying principle. Mutual Information tries to find the maximum joint probabilistic distribution between two random variables. The basic equation is shown below-

$$I(x;y) = \iint p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy.$$

MI measures how much information the presence/absence of a term contributes to making the correct classification decision. As we have a lot of features for each image and clearly the feature space not being a linear one, it makes sense to compare between features randomly rather than using linear correlation based methods. Our MI based approach takes as input the original feature vector, alongwith ground truth labels with quantization levels as input. Quantization levels just discretize the input feature vectors into specified number of bins to calculate the probabilistic distribution amongst features and for faster performance as well.

[features_idx,weights] = MI(features,labels,Q)

Output of our MI operation is the feature indices ordered in their decreasing relevance. From here we evaluate the performance of different feature subsets which is described in the evaluation section.

iii. Top-5 list of selected features

TOP5	D1	D2	D3	D1	D2	D3
1	Orientation	Perimeter	Eccentricity	Convex Area	Perimeter (Median)	Eccentricity
2	Extent	Minor axis	Orientation	Major Axis	Perimeter	Value

					(Standard Deviation)	Channel
3	Area	Major axis	Contrast	Minor Axis	Minor Axis	Extent
4	Perimeter	Green channel	Dissimilarity	Perimeter	Saturation Channel	Major Axis
5	Eccentricity	Area	Area	Area	Hue Channel	Perimeter

Figure 3: The list of top features

C. FEATURES REDUCTION

i. Principal Component Analysis (PCA)

Principal Component Analysis is abundantly used in modern statistical analysis. It is a linear dimensionality reduction technique that is simple, non-parametric and can display relevant information from confusing data sets with minimal effort [8]. PCA finds the principal components of a dataset and rearrange the data in terms of its principal axes to display it in a 2D or 3D plot.

The underlying methodology for PCA reduction can be summarized in the following steps:

1. Calculate the covariance matrix
2. Find the eigenvectors and eigenvalues of the covariance matrix
3. Sort the variances in decreasing order
4. Project the original dataset on the new reference system

ii. t-Distributed Stochastic Neighbor Embedding (t-SNE)

The t-Distributed Stochastic Neighbor Embedding (t-SNE) technique is a new dimensionality reduction technique which performs really well on high dimensional datasets. It won the Kaggle Merck Viz Challenge in 2012 [9]. T-SNE represents each object by a point in a two-dimensional scatter plot,

and arranges the points in such a way that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points. When you construct such a map using t-SNE, you typically get much better results than when you construct the map using something like principal components analysis or classical multidimensional scaling, because (1) t-SNE mainly focuses on appropriately modeling small pairwise distances, i.e. local structure, in the map and (2) because t-SNE has a way to correct for the enormous difference in volume of a high-dimensional feature space and a two-dimensional map. As a result of these two characteristics, t-SNE generally produces maps that provide much clearer insight into the underlying (cluster) structure of the data than alternative techniques [5].

t-Distributed stochastic neighbor embedding minimizes the divergence between two distributions: a distribution that measures pairwise similarities of the input objects and a distribution that measures pairwise similarities of the corresponding low-dimensional points in the embedding. One defines joint probabilities pij that measure the pairwise similarity between objects xi and xj by symmetrizing two conditional probabilities as follows:

$$p_{j|i} = \frac{\exp(-d(\mathbf{x}_i, \mathbf{x}_j)^2/2\sigma_i^2)}{\sum_{k \neq i} \exp(-d(\mathbf{x}_i, \mathbf{x}_k)^2/2\sigma_i^2)}, \qquad p_{i|i} = 0$$

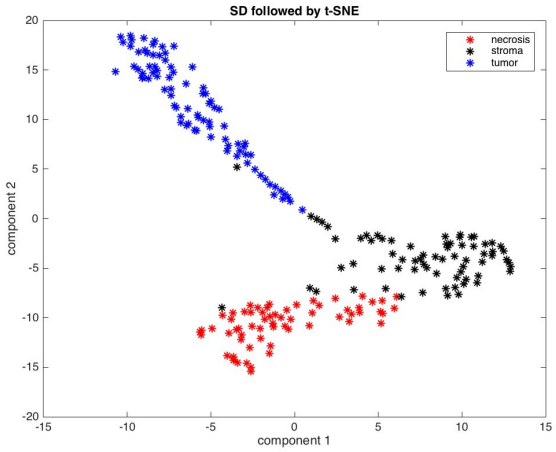
$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}.$$

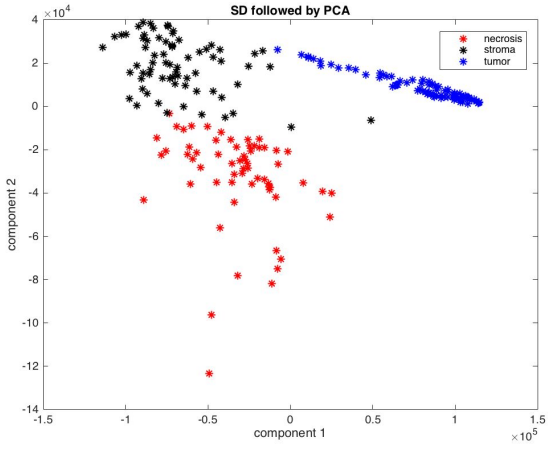
where σi is the variance of the Gaussian that is centered on datapoint xi.

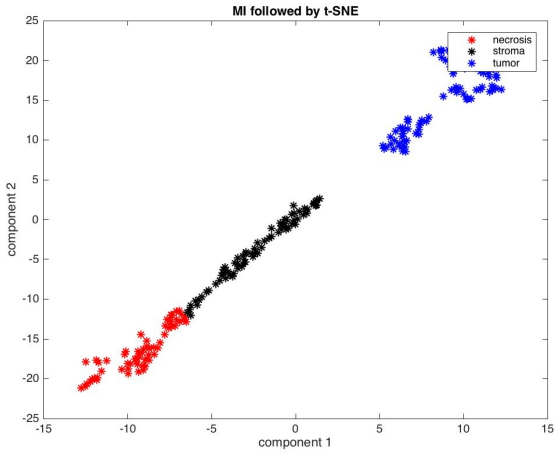
In contrast to, e.g., PCA, t-SNE has a non-convex objective function. The objective function is minimized using a gradient descent optimization that is initiated randomly. As a result, it is possible that different runs give you different solutions.

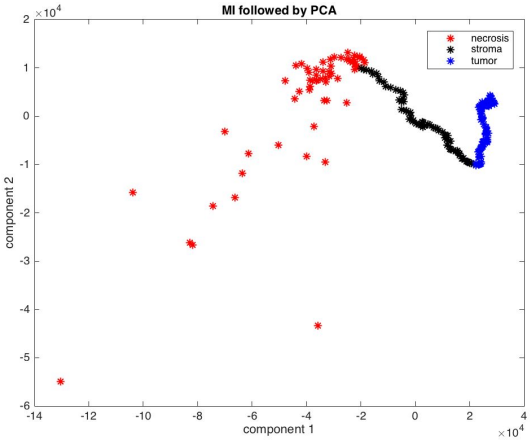
V. RESULTS

In this chapter we highlight some results obtained from a combination of feature selection and feature reduction techniques. All the plots can be found in the Appendix. Below are the plots for Dataset1, which is the dataset that gave the best results (>90% accuracy). Dataset2 are Dataset3 results can be found in the Appendix, the results are not fairly good as expected because of the skewness of the dataset and less perceivable differences amongst various grades, we may need to follow additional techniques such as the ones described in the Future Work section.









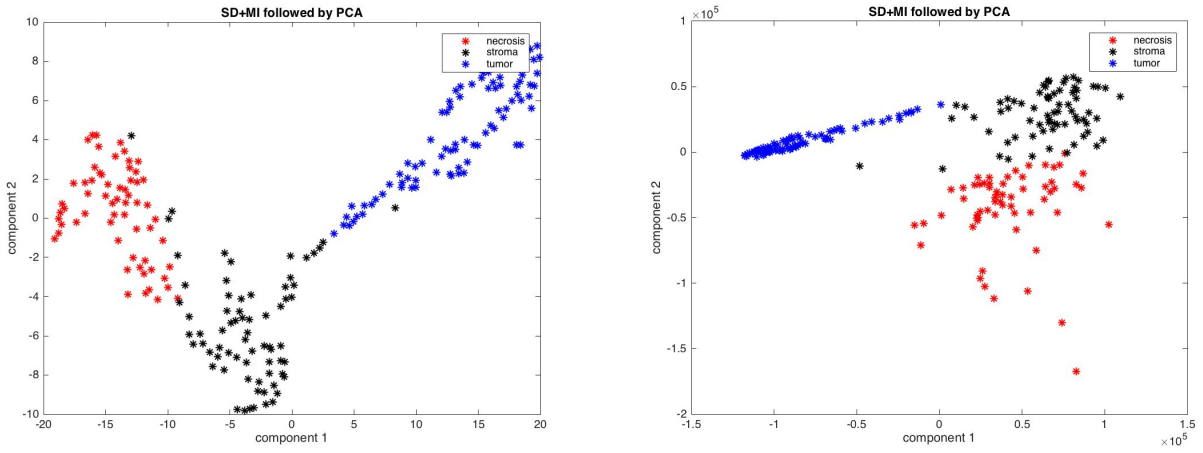


Figure 4: Results for Dataset1 using different techniques of features selection and features reduction

VI. EVALUATION

This section will report the performance of our feature extraction/selection/reduction algorithms on all datasets. Feature comparison tables for inter method and intra method for top 100 selected features are reported below.

Mutual Information and Statistical Dependency methods have been compared below for all the datasets individually as well as a comparison between them is reported for each dataset. Below feature comparison tables justify the intuitive information that we can comprehend while looking at the images. Images in Dataset1, though differ in color and texture, the structural information varies significantly, as quantified by the top features selected as well as the overlap of features for both the methods. For Dataset2 and Dataset3, structural information still plays an important role but you can also see the presence of color and texture amongst the top features as well as in the overlap tables below.

Method	Dataset1-Dataset2	Dataset2-Dataset3	Dataset1-Dataset3
MI	Extent, Perimeter, Convex Area, Major Axis, Solidity	Perimeter, Eccentricity, Extent, Area, Cluster Shade, Difference variance, Dissimilarity	Euler Number, Solidity,Extent, Perimeter
SD	Area, Convex Area, Perimeter, Orientation, Eccentricity	Area, Saturation Channel, Perimeter, Minor Axis, Euler Number	Convex Area, Euler Number, Perimeter, Major Axis, Extent, Solidity

Table 2: Comparison of common features for each method amongst datasets

Method	Dataset1	Dataset2	Dataset3
MI-SD	Area, Euler Number, Perimeter, Convex Area, Red Channel	Perimeter, Convex Area, Eccentricity, Extent, Saturation Channel	Blue channel, Orientation, Eccentricity, Value Channel, Perimeter, Information measures, Difference Variance, Dissimilarity

Table 3: Comparison of common features for each dataset amongst methods

Below plots showcase performance metrics of our feature selection methods. Accuracy and average recall plots are shown below. Below plots were obtained by iterating through 16 feature subsets of increasing length and metrics were evaluated using a KNN classifier for 5 nearest neighbors with a training/validation split of 60/40.

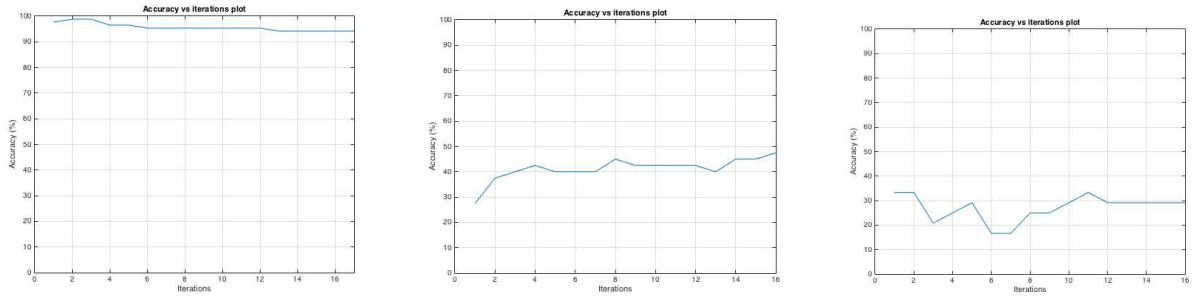


Figure 6: Accuracy vs iteration plot (out of 100%). Dataset1 yields great results while Dataset2 and Dataset3 give poor results

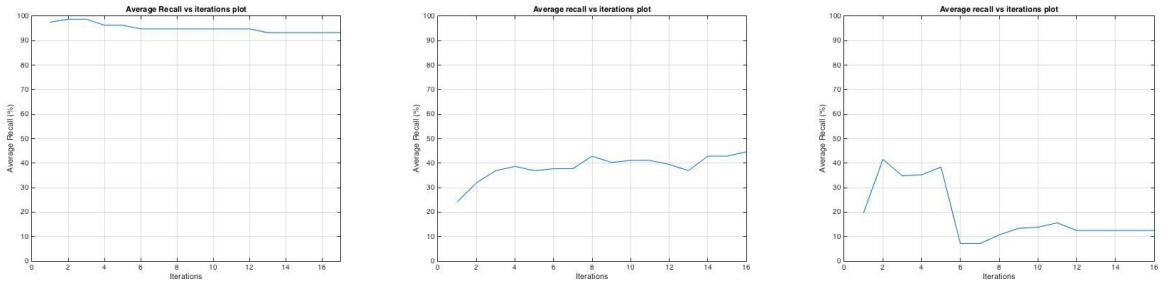


Figure 6: Average recall vs iteration plot (out of 100%). Dataset1 yields great results while Dataset2 and Dataset3 give poor results

VII. THE GUI

The targeted end-users of this project are the pathologists who want to get support when diagnosing cancer images. To this end, we need to design a GUI that can help the pathologist analyzing several images in an appealing way. The parameters described above for each segmentation technique must be accessible and the interface should be user friendly.

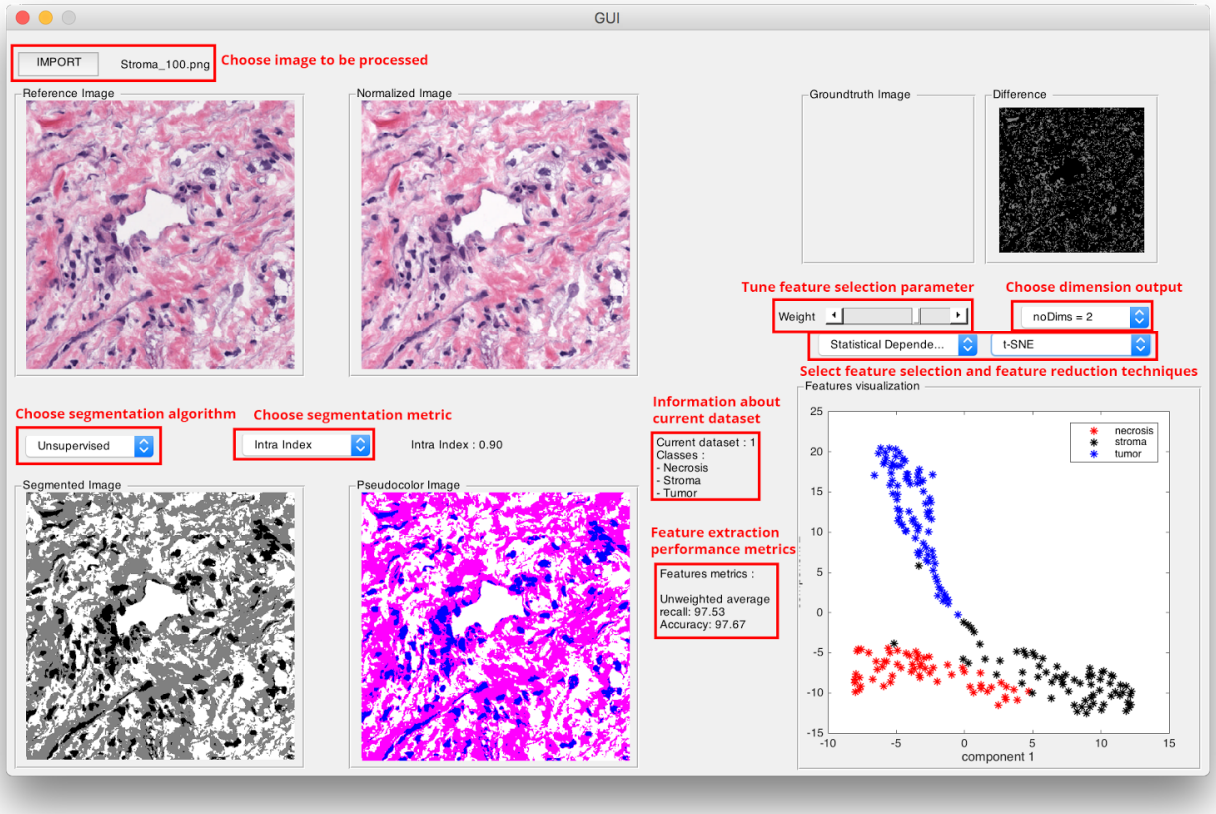


Figure 7: The interactive GUI

The interactive feature of the GUI let you choose the feature selection and feature reduction algorithms. We can also select the number of dimensions or tune some feature selection parameter. For each change, the data automatically updates and a new result is displayed.

VIII. FUTURE WORK

There are plenty of improvements that can be done on this part of the project. We would like to incorporate more bins in our color feature extraction to have more discreet results. Texture features can be computed from a lot of tools such as GLCM, but also Wavelet, Fractals or other object-based techniques. It would be really interesting to test them all and find the well suited one for histopathological images. For Feature Selection, we would like to try more combination of methods and see if we attain any improvement in results. After detailed analysis, we learnt that nuclei information plays a very important role for different grades. For example, Grade 2 images will have more circular and uniform distribution of nuclei, while Grade 4 will have unstructured pleomorphic tendency for nuclei and non-uniformly distributed as well. We would like to incorporate this information to further enhance our feature set and improve performance for Dataset2 and Dataset3. For feature reduction, there are many techniques that have been developed throughout the years and some of them could be really efficient on histopathological images. Thanks to the design of our GUI, we would be able to incorporate these new techniques easily.

References

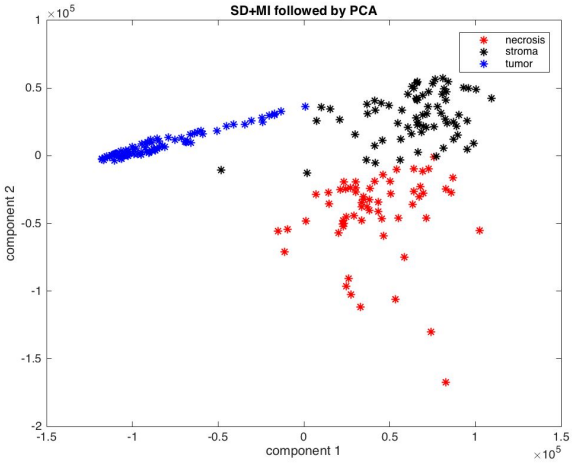
1. Yan Xu, Weakly supervised histopathology cancer image segmentation and classification, 2014
2. Textural Features for Image Classification, Robert M. HARALICK et al, IEEE Transaction on System, Man and Cybernetic, 1973
3. The effect of variation in illuminant direction on texture classification, Michael J Chantler, Department of Computing and Electrical Engineering Heriot-Watt University, 1994
4. L. Boucheron, "Object-and spatial-level quantitative analysis of multispectral histopathology images for detection and characterization of cancer," PhD thesis, University of California, Santa Barbara, 2008
5. Visualizing data using t-SNE, Laurens van der Maaten et al, Journal of Machine Learning Research, 2008
6. Nonlinear dimensionality reduction, JA Lee, M Verleysen, Information Science and Statistics, Springer, 2007
7. PCA vs LDA, Aleix M. Martonez et al, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001
8. A tutorial of Principal Component Analysis, John Shlens, 2003
9. http://blog.kaggle.com/2012/11/02/t-distributed-stochastic-neighbor-embedding-wins-merck-viz-challenge/
10. Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits, Computer Speech & Language ,Volume 29, Issue 1, January 2015, Pages 145–171 .
11. Multifeature Prostate Cancer Diagnosis and Gleason Grading of Histological Images , IEEE Transaction on Medical Imaging, Vol. 26, No. 10, October 2007
12. Histological Image Feature Mining Reveals Emergent Diagnostic Properties for Renal Cancer, IEEE International Conference on Bioinformatics and Biomedicine, 2011

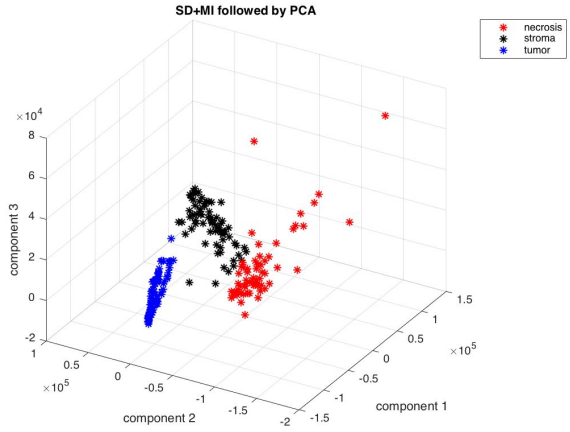
APPENDIX

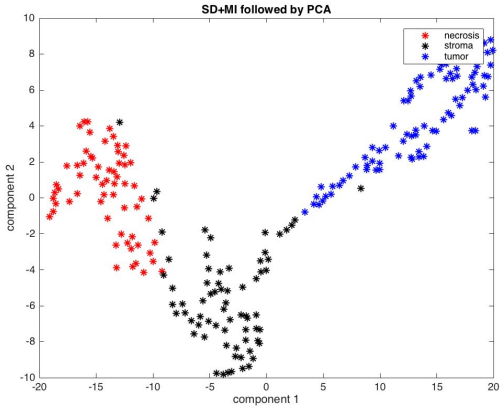
TYPE	FEATURE	COUNT
	Red Channel	16
	Green Channel	16
	Blue Channel	16
	L Channel	16
COLOR	A Channel	16
	B Channel	16
	Hue Channel	16
	Saturation Channel	16
	Value Channel	16
	Area	8
	Major Axis	8
	Minor Axis	8
	Eccentricity	8
	Orientation	8
SHAPE	Convex Area	8
	Solidity	8
	Filled Area	8
	Euler Number	8
	Extent	8
	Perimeter	8
	Autocorrelation	20
	Correlation	40

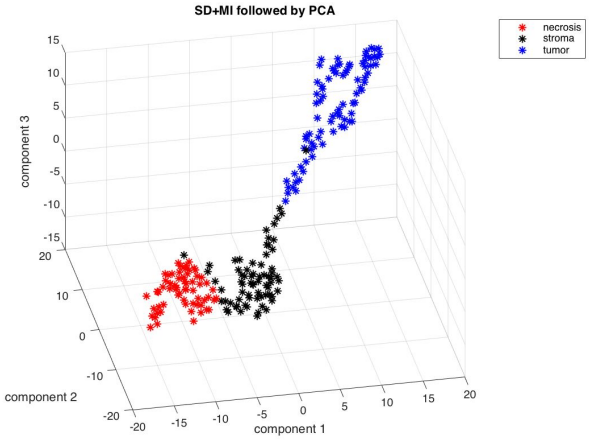
	Contrast	20
	Cluster Prominence	20
	Cluster Shade	20
	Difference Entropy	20
TEXTURE	Difference Variance	20
	Dissimilarity	20
	Energy	20
	Entropy	20
	Homogeneity	40
	Information Measure of Correlation	40
	Maximum Probability	20
	Sum Average	20
	Sum Entropy	20
	Sum Variance	20
	Inverse Difference Moment	60
TOTAL		(16*9)+(88*3)+(440*3) = **1728**

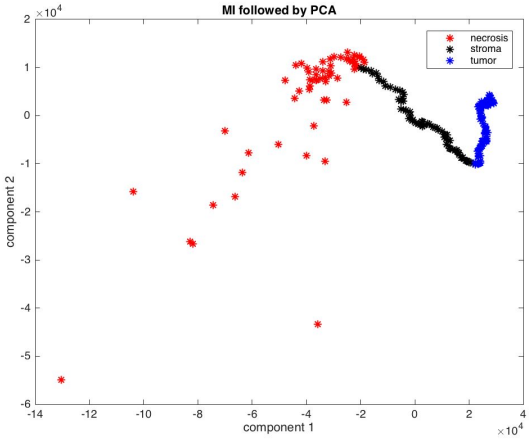
Appendix 1: List of all features used in the project

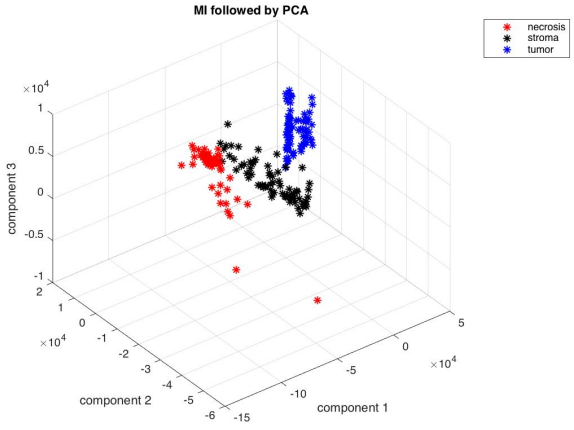

SD+MI followed by PCA
necrosis
stroma
tumor
component 1
component 2
×10^5

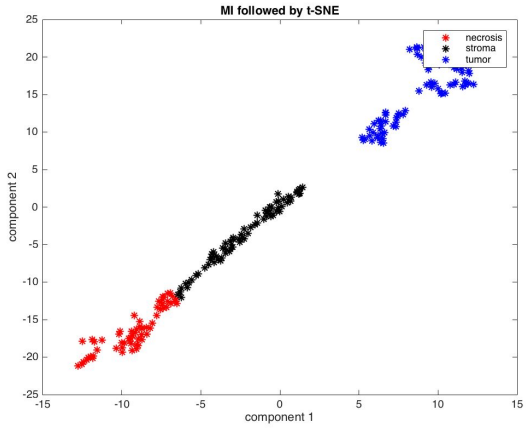

SD+MI followed by PCA
necrosis
stroma
tumor
component 1
component 2
component 3
×10^5
×10^4

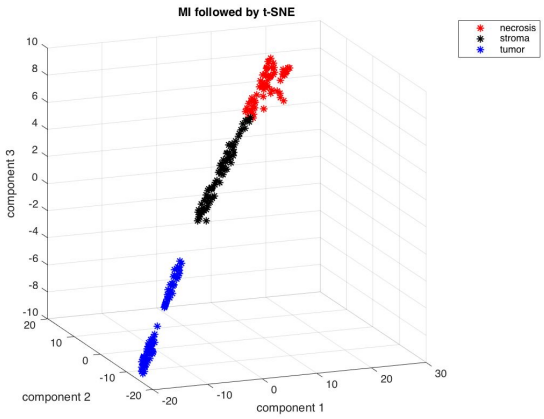

SD+MI followed by PCA
necrosis
stroma
tumor
component 1
component 2

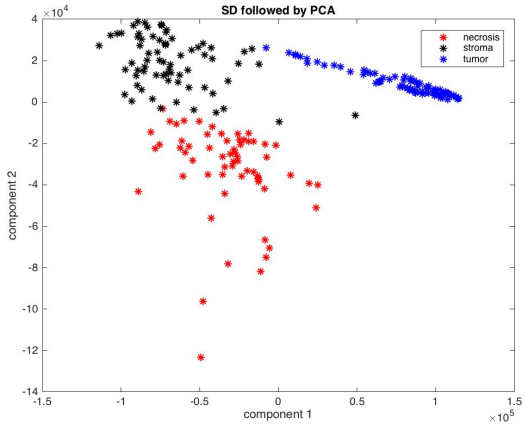

SD+MI followed by PCA
necrosis
stroma
tumor
component 1
component 2
component 3

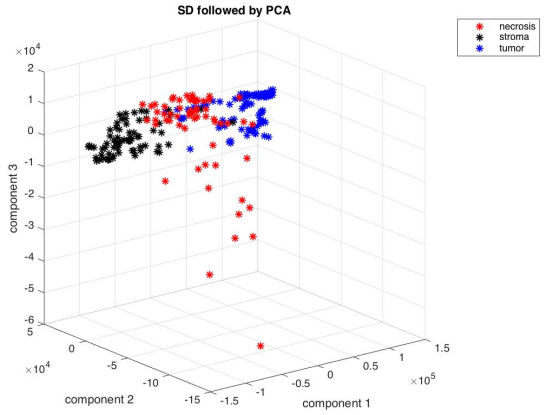

MI followed by PCA
necrosis
stroma
tumor
component 1
component 2
×10^4

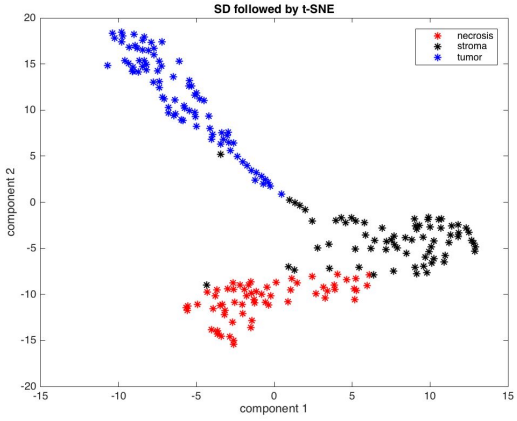

MI followed by PCA
necrosis
stroma
tumor
component 1
component 2
component 3
×10^4

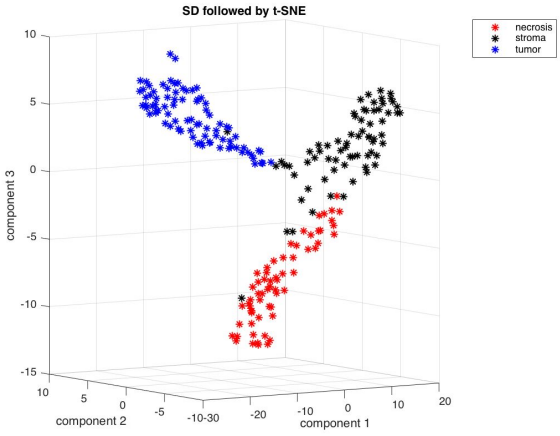

MI followed by t-SNE
necrosis
stroma
tumor
component 1
component 2

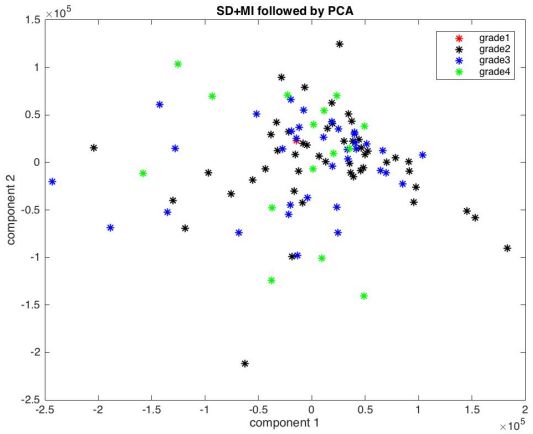

MI followed by t-SNE
necrosis
stroma
tumor
component 1
component 2
component 3

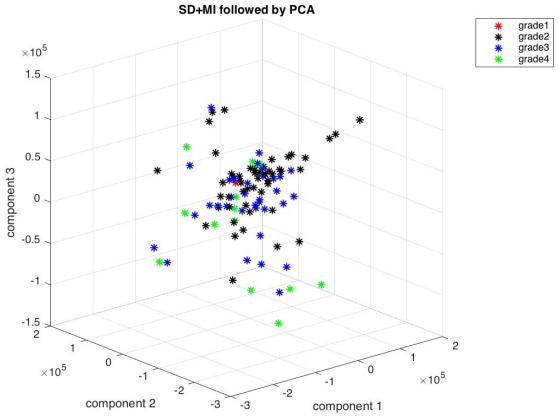

SD followed by PCA
necrosis
stroma
tumor
component 1
component 2

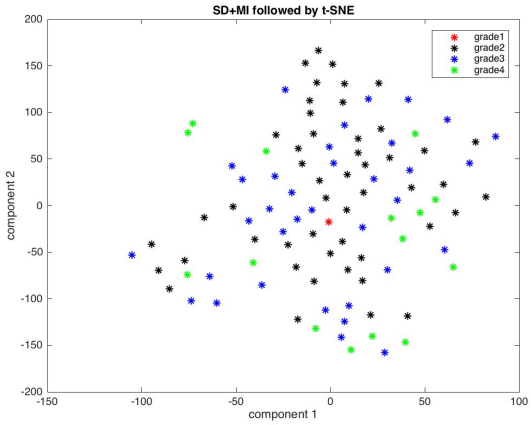

SD followed by PCA
necrosis
stroma
tumor
component 1
component 2
component 3

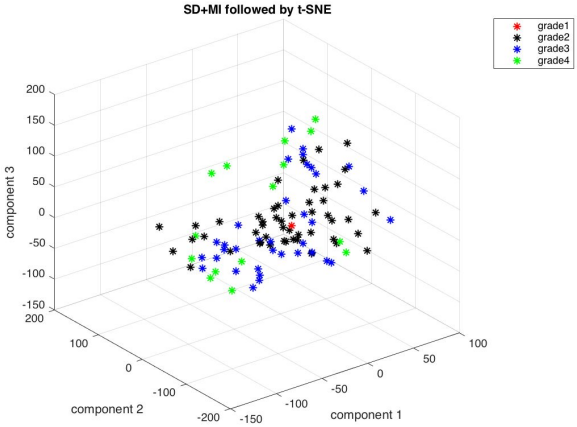

SD followed by t-SNE
necrosis
stroma
tumor
component 1
component 2

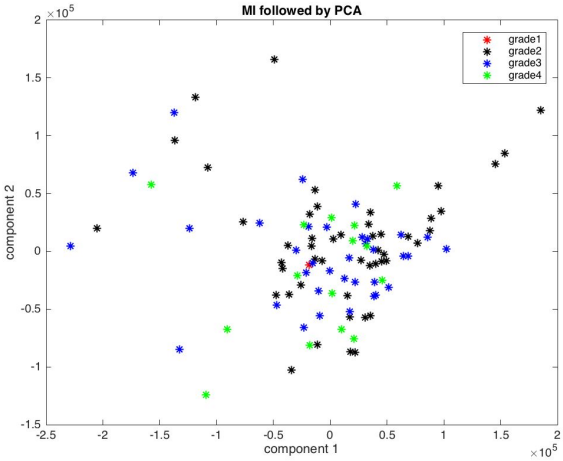

SD followed by t-SNE
necrosis
stroma
tumor
component 1
component 2
component 3

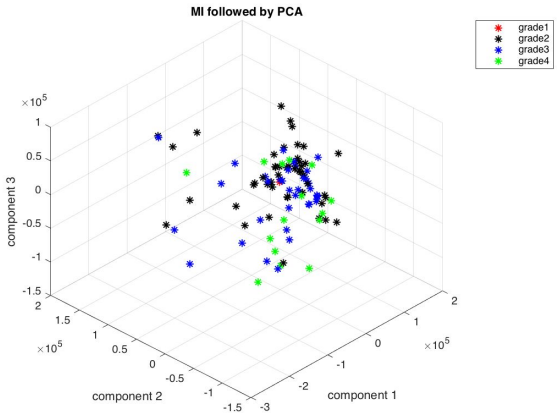

SD+MI followed by PCA
grade1
grade2
grade3
grade4
component 2
component 1
×10^5

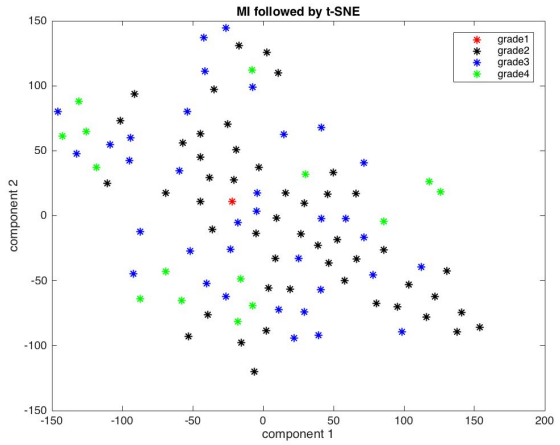

SD+MI followed by PCA
grade1
grade2
grade3
grade4
component 3
component 2
component 1
×10^5

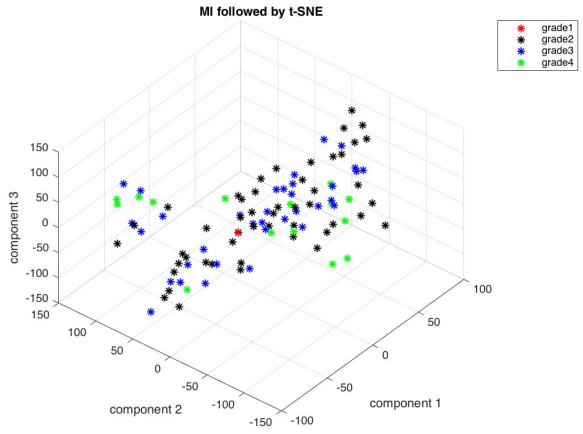

SD+MI followed by t-SNE
grade1
grade2
grade3
grade4
component 2
component 1

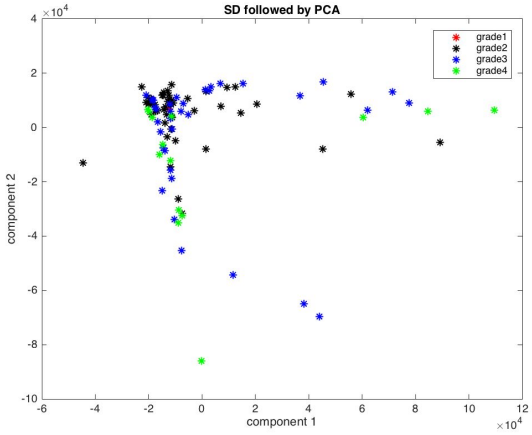

SD+MI followed by t-SNE
grade1
grade2
grade3
grade4
component 3
component 2
component 1

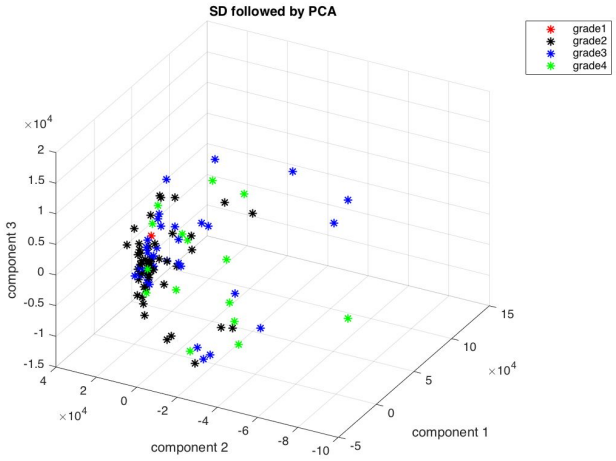

MI followed by PCA
grade1
grade2
grade3
grade4
component 2
component 1
×10^5

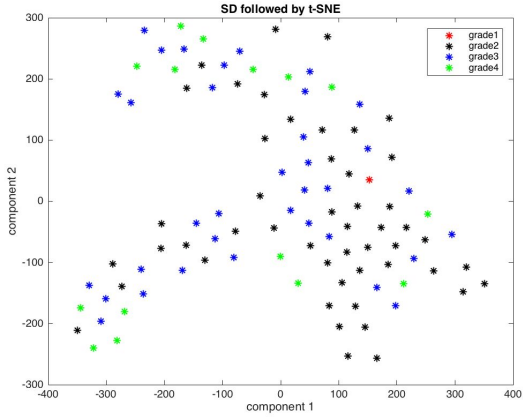

MI followed by PCA
grade1
grade2
grade3
grade4
component 3
component 2
component 1
×10^5

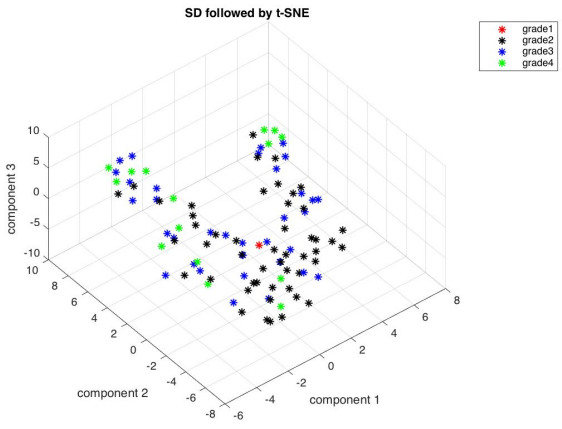

MI followed by t-SNE
grade1
grade2
grade3
grade4
component 1
component 2

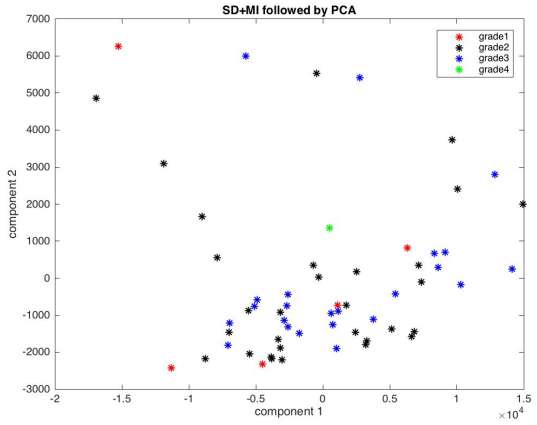

MI followed by t-SNE
grade1
grade2
grade3
grade4
component 1
component 2
component 3

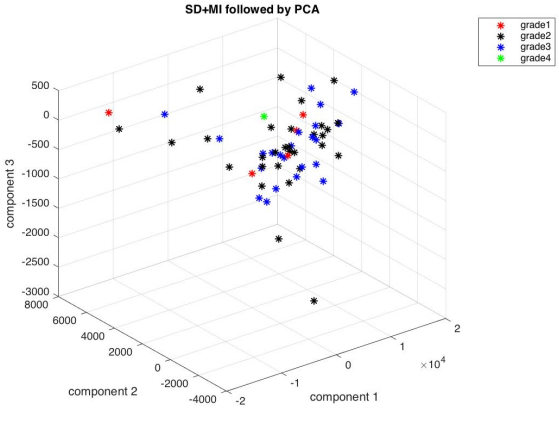

SD followed by PCA
grade1
grade2
grade3
grade4
component 1
component 2

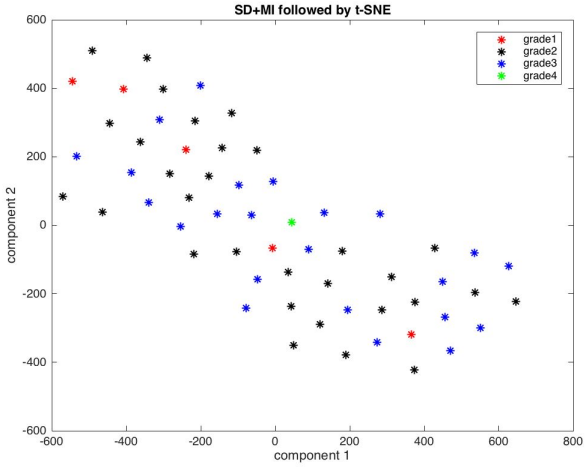

SD followed by PCA
grade1
grade2
grade3
grade4
component 1
component 2
component 3

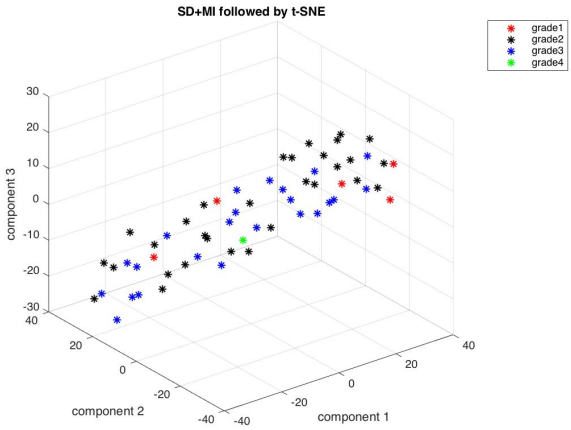

SD followed by t-SNE
grade1
grade2
grade3
grade4
component 1
component 2

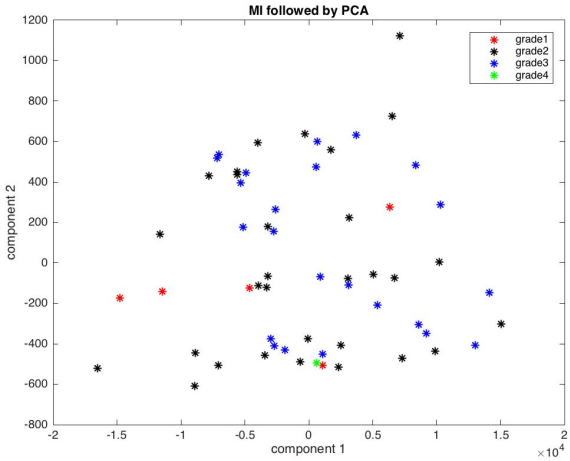

SD followed by t-SNE
grade1
grade2
grade3
grade4
component 1
component 2
component 3

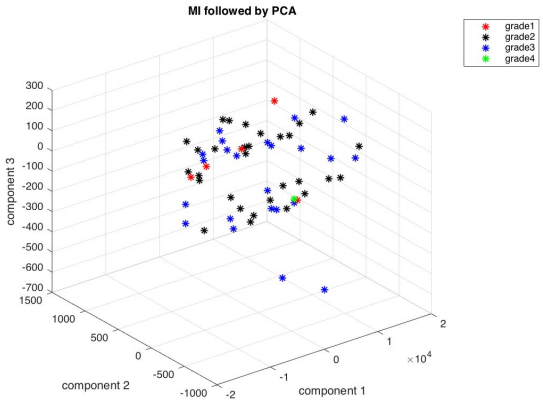

SD+MI followed by PCA
grade1
grade2
grade3
grade4
component 2
component 1
×10^4

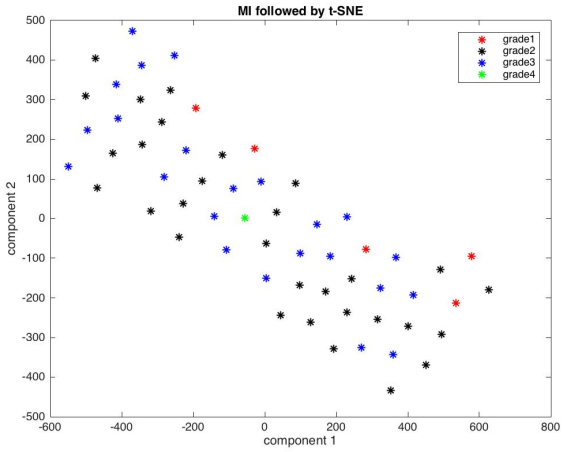

SD+MI followed by PCA
grade1
grade2
grade3
grade4
component 3
component 2
component 1
×10^4

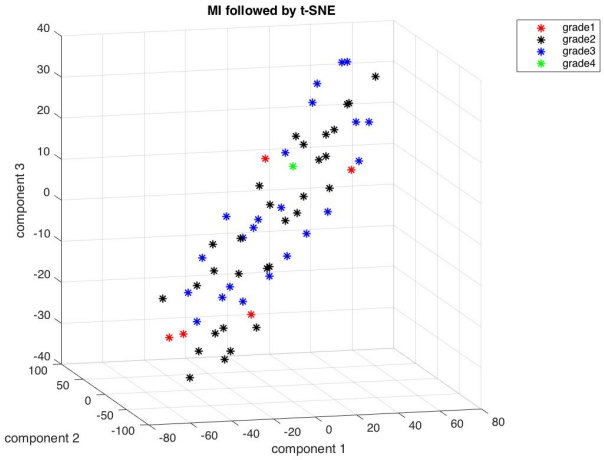

SD+MI followed by t-SNE
grade1
grade2
grade3
grade4
component 2
component 1

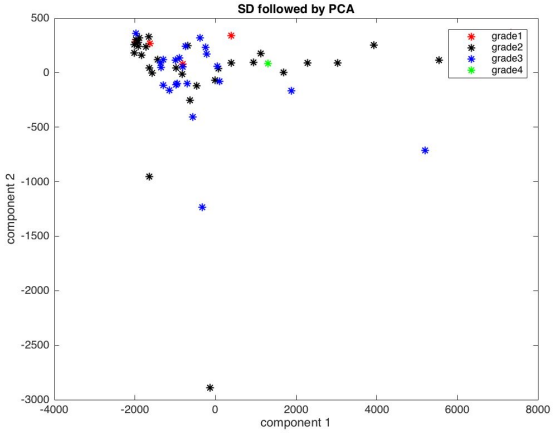

SD+MI followed by t-SNE
grade1
grade2
grade3
grade4
component 3
component 2
component 1

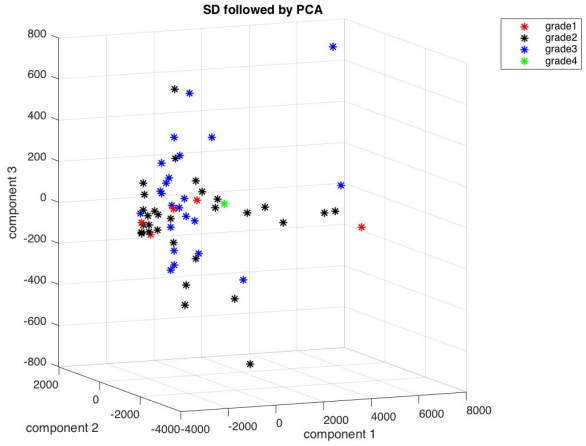

MI followed by PCA
grade1
grade2
grade3
grade4
component 2
component 1
×10^4

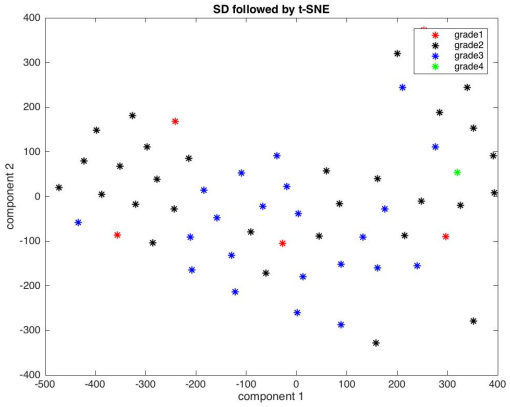

MI followed by PCA
grade1
grade2
grade3
grade4
component 3
component 2
component 1
×10^4

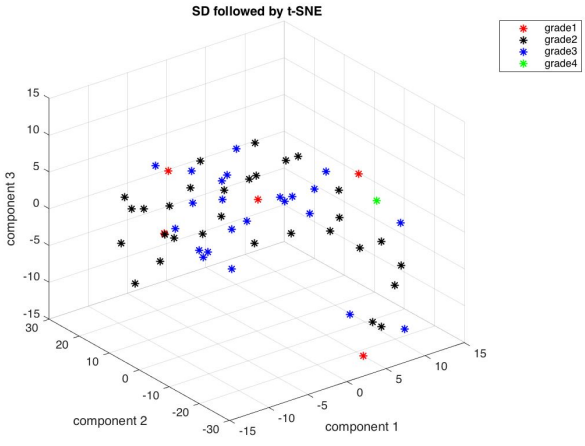

MI followed by t-SNE
grade1
grade2
grade3
grade4
component 2
component 1


MI followed by t-SNE
grade1
grade2
grade3
grade4
component 3
component 2
component 1


SD followed by PCA
grade1
grade2
grade3
grade4
component 2
component 1


SD followed by PCA
grade1
grade2
grade3
grade4
component 3
component 2
component 1


SD followed by t-SNE
grade1
grade2
grade3
grade4
component 2
component 1


SD followed by t-SNE
grade1
grade2
grade3
grade4
component 3
component 2
component 1

Appendix 2: Plots of results for Dataset 1, Dataset 2 and Dataset 3 according to different techniques of feature selection and feature reduction