Radboud University Nijmegen

Bachelor Thesis

Visualizing breast cancer data with t-SNE

Student:
Ka ka Tam
s4074157
k.tam@student.ru.nl

Supervisors:
Dr. Ida Sprinkhuizen-Kuyper
Dr. Elena Marchiori

October 25, 2013

Contents

Abstract

One in eight women will get breast cancer in her lifetime and in 2008 it has caused 458.503 deaths among the world [15]. Despite that technology has made considerable improvements in the last decades, there is still room for more advances. A technique that possibly can contribute to this field is t-SNE [24]. The aim of this thesis is to investigate whether t-SNE is able to present the breast cancer data in an interpretable way and possibly improves the classification performances. We employ two approaches to explore the applicability of t-SNE. In the first approach we compare the visualizations and in the second approach the classification performances are compared. We found that classification on the original data performed significantly better than on t-SNE data. This suggests that t-SNE is not applicable to the breast cancer data set.

1 Introduction

According to the World Health Organization(WHO), one out of eight women will be diagnosed with breast cancer [15]. It leads to breast cancer mortality to thousands of women each year. Many factors have an impact on the high mortality rate. Which in less developed countries are due to lack of diagnostic facilities, expertise and treatment. In contrast to western countries these factors are obesity and longer life expectancy. Despite the considerable progressions in technology there is still room for more improvement.

One of the most common breast cancer screening method is mammography. The mammograms are obtained by recording the transmission of X-ray through the breast. These X-ray images are capable of detecting small tumors. By using computer-aided diagnosis(CAD) [7], the radiologist can be assisted to find suspicious regions and therefore giving a diagnostic decision whether a region is malignant or benign. With this kind of assistance, it could prevent that malignant masses remains undetected by the radiologist. The occurrence that a malignant tissue will be overlooked is from 10 to 30 percent. And also, benign lesions would not be sent for biopsy when abnormalities are found after the examination of screening. This means that small tissue has to be removed for further investigation. Several reasons can be addressed to the radiologists misjudgment, such as subtle nature of the visual findings, poor image quality, fatigue, or oversight. By using efficient computer algorithms for a diagnostic purpose can be less time consuming and lead to more sensitive results.

In the field of Machine Learning has been done a lot of work to investigate dimensionality reduction methods and classification algorithms to address this problem. However, a lot of the dimensionality reduction methods are not able to provide insightful visualizations due to the high number of features. However, a "new" visualization technique has been presented by van der Maaten. The student-t Stochastic Neighbor Embedding(t-SNE) [24] has given promising results for the handwriting dataset, compared to other dimensionality reduction techniques such as the Sammon mapping [19], Isomap [21], LLE [18]. The intuitive idea of t-SNE is to map the high-dimensional points into a lower dimension such that it preserves the local structure, which means that the nearest neighbors of one point in the high-dimensional space should also be the same nearest neighbor of that point in the low-dimensional space.

1.1 Research question

The main question is to investigate whether useful information is extractable from the breast cancer data set by using t-SNE. This research question will be divided into sub-questions as following:

1. What are the characteristics of the resulting visualization? And will the resulting visualization be presented in distinctive clusters? Did the resulting visualization preserve the original structure?

2. If no useful information will be extracted, then the sub-research question will look for the answer why this technique is not applicable on the data set. Are there any properties of the t-SNE or data set that will lead to inapplicability? But did the resulting visualization preserve the original structure?

1.2 Outline

In section 2, I will discuss the relevant work that has been done in this field. Furthermore the methods that we used are discussed in section 3, such as the visualization algorithms, classifier and the experimental setup. In section 4, we will present the results. Eventually, the conclusion and discussion will be discussed in section 5.

2 Background

In this section, we will discuss the related work that has been done on visualizing data using t-SNE. Despite t-SNE is a "new" visualization technique, this technique has been used for research in several domains, such as in cancer research [12], computer security [6] and mostly in biology [3] [12].

2.1 Related work t-SNE

In the research of Jamieson et al. [13], they have investigated whether unlabeled data is able to enhance the breast CADx performance. This study was performed by comparing the dual-stage CADx schemes with a single-stage scheme. The dual-stage CADx schemes is constructed such that for the first stage the dimensionality of the breast cancer data is reduced which subsequently will be used in the second stage. The first stage employs both labeled and unlabeled data. In the second stage the labeled data from the reduced dimension embedding is used to train a classifier and eventually the performances will be evaluated. The single-stage scheme combines the two stages, and uses therefore a semi-supervised algorithm. They have find evidence that using unlabeled data in CADx methods may improve the classifier performance.

In another study for which t-SNE had been explored on Single Nucleotide Polymorphisms (SNPs). "SNPs are one the largest sources of new data in biology." [16] Usually these are visualized by using PCA, however, Platzer proposed to use t-SNE and compare the results with PCA in several ways. He has discovered that t-SNE performs better than PCA in all of the proposed criteria for evaluating the visualizations. Besides of judging

the visualizations visually which is rather a subjective measure, he also used measurements for the structuredness of the data. Such measurements are the Dunn's Validity Index [5] and the Silhouette Validation Method [17].

3 Methods

3.1 Breast cancer data set

The breast cancer that we used is provided by the Diagnostic Image Analysis Group from the UMC St Radboud. These are mammograms which are taken from different views at different stages of screening. At screenings for the very first time, the person will be screened from a medial lateral oblique (MLO) and cranio caudial(CC) view. The first screening view is taken from an angled view, this is preferred over a 90 degrees projection which is able to image more breast tissue [11]. The latter screening view takes the image above the breast. For next screening session, only one view MLO will be used for screening. But when there is an indication that taking an image from a CC view would be beneficial to the process for detecting malignant masses, then a image from this view will be obtained either.

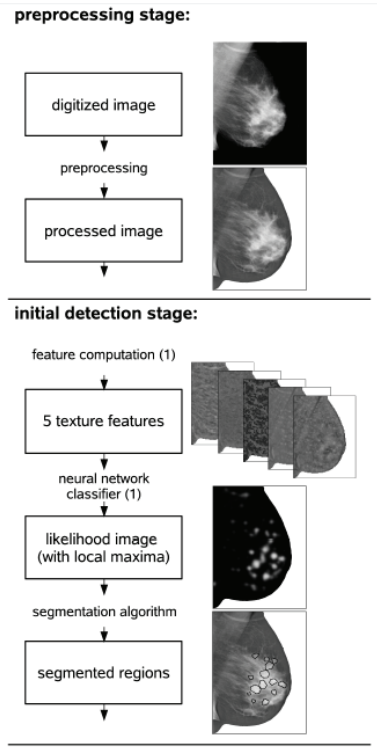


Figure 1: Schematic overview of preprocessing and initial detection stage [9]

The breast cancer data consist of 455.396 regions, and is achieved by completing two stages, respectively preprocessing and initial detection. The schematic overview is shown in figure 1.

3.1.1 Preprocessing

Subsequently, the mammograms are digitized at a pixel resolution of 50 μm and are sampled down to a resolution of 200 μm. After digitization of all mammograms, these will

be preprocessed by performing three steps, namely segmentation into three areas (breast region, background and pectoral muscle), pectoral equalization and peripheral enhancement. Then, in order to detect candidate mass regions, an initial detection described by Karssemeijer and te Brake has been performed [9] [14] [20]. This stage is based on image features which are extracted locally. These image features are including two gradient concentration measures, two spiculation measures and one measure indicating the scale at which most spiculation is present. Then, an ensemble of 5 neural networks are used to compute a likelihood score for each location on a regular grid, which all of these likelihood scores together makes up a likelihood image. Furthermore, the likelihood image is smoothed and each local maximum that exceeds a certain low threshold will be selected as a candidate mass region. Subsequently, for each of the local maximum a candidate mass region is segmented using a segmentation method based on dynamic programming. Hence, when candidate masses will be interpreted, the outcome will be classified into benign or malignant tissue. [9]

3.1.2 Features

Each segmented region consists of a set of 71 features for which can be divided into several categories. These categories can be found in figure 2. "One of the feature groups measure the suspiciousness of normal tissue. These context features are described in Hupse and Karssemeijer" [10]. Moreover, they have assigned 71 random values to each candidate region for which these random values are referred as noise features. These noise features were distributed uniformly between 0 and 1. The purpose to add these noise feature is to investigate whether the feature selection algorithms were able to distinguish between useful and non-useful features.

Category	Number	Description
Initial detection	11	Local spiculation and mass measures, suspiciousness level, spiculation and mass measures for region
Normal tissue	17	Suspiciousness measures for normal tissue context
Location	10	Relative location and distances to skin, pectoral muscle, chest nipple
Shape and size	7	Acutance, compactness and measures of region size
Linear texture	5	Presence of linear texture in the region and its surround
Dense tissue	7	Features that determine the amount of dense tissue located inside and outside the region
Contrast	8	Difference between gray level distribution in the region and its surround
Border	6	Features measuring the continuity of the segmentation border
Total	71	

Figure 2: Features computed for each segmented region by Hupse [10]

3.2 t-SNE

As mentioned in the introduction, the t-SNE will be used to visualizing the breast cancer data. This method is implemented in MATLAB by van der Maaten [23] for which different versions are available, but for this project I have used the simple implementation. Moreover, t-SNE is originated from Stochastic neighbor Embedding [8].

The algorithms works as following, first, for each pair of points in the high dimensional space $X = \{x_1, x_2, x_3, \ldots, x_n\}$, also called the H-space, the pairwise Euclidean distances between data points will be converted into conditional probabilities, which is the probability of data point x_i picks data point x_j as its nearest neighbor. As result that, if the distance between these points is small, then the conditional probability should be high. Once all pairwise conditional probabilities are computed, then these conditional probabilities will be averaged using the following equation, $p_{ij} = p_{i|j} + p_{j|i}/2n$. This is important due to the fact that $p_{i|j}$ and $p_{j|i}$ are not equal, which enables all data points to make a significant contribution to the cost function.

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2)/2\sigma_i^2}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2)/2\sigma_i^2} \tag{1}$$

The conditional probabilities are dependent on both the Euclidean distances between data points as well as the perplexity, which can be represented as the number of effective nearest neighbors. This parameter is the value which has to be specified by the user, such that the optimal number of effective neighbors can be found for all data points. This is important due to sparse and dense regions in the data. Sparse regions requires a higher variance while dense regions a lower variance. This is illustrated by figure 3. Thus, it is necessary that each data point can find the optimal number of nearest neighbor, such as outliers or distant points can be mapped near to one of its neighbors. Otherwise, the visualization will consist of many outliers or data points that are widely spreaden out.

The variance, for the data points in H-space, is distributed under a Gaussian distribution, and can be found using the second equation.

$$Perp(P_i) = 2^{H(P_i)} \tag{2}$$

$$\text{for which } H(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i} \tag{3}$$

After calculating all conditional probabilities the algorithm initializes a solution. This solution is randomly generated under a Gaussian distribution and updates the cost function at each iteration.

$$Y^{(0)} = \{y_1, y_2, y_3, \ldots, y_n\} \text{ for which } n \text{ is the size of the data set} \tag{4}$$

Subsequently, pairwise distances will be computed for data points in the lower dimension (V-space), for which the variance is distributed, in contrast to the data points in H-space, under a student t-distribution. By using a student t-distribution with one degree of freedom, which makes it heavier-tailed, as consequence that moderate distant points will also be mapped close to its neighbors.

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}} \tag{5}$$

For each iteration, the positions of the data points in the lower dimension will be be updated such that the gradient of the cost function is minimal. The update function is as follows:

$$Y^{(t)} = Y^{(t-1)} + \eta \frac{\delta C}{\delta Y} + \alpha(t)(Y^{(t-1)} - Y^{(t-2)}) \tag{6}$$

for which the gradient of the cost function is $\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \|y_i - y_j\|^2)^{-1}$

(7)

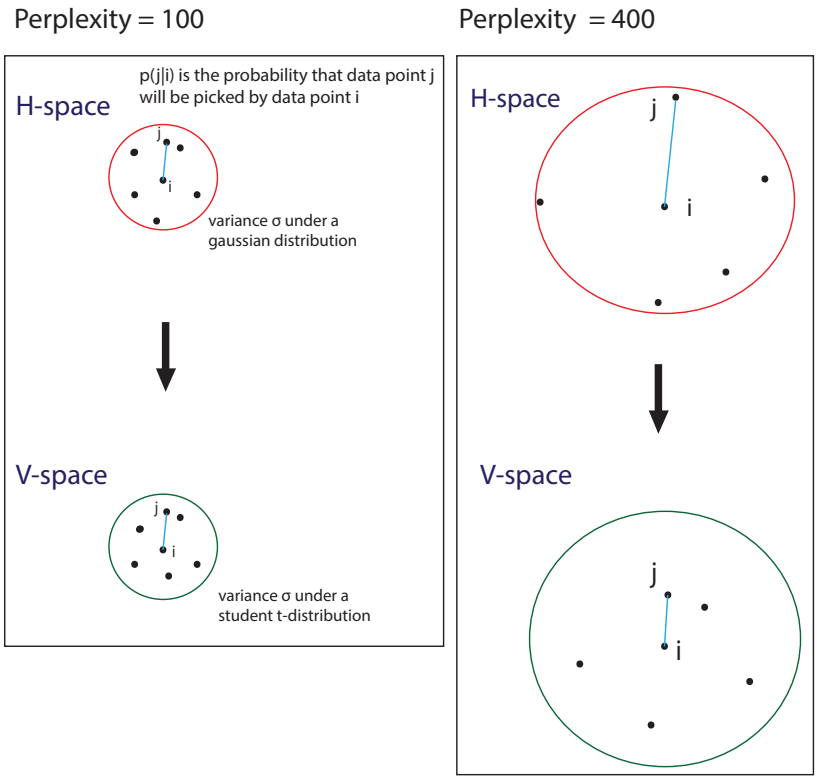


Figure 3: This figure illustrates the effect of a given perplexity value in data set for which the points are distant or close to each other. This shows that a higher perplexity value produces a higher variance, which enables that distant data points can also be mapped near to its nearest neighbor.

3.3 Laplacian Eigenmaps

Another visualization method that we have used in order to compare the results with t-SNE visualizations is Laplacian Eigenmaps [1]. This method has a similar purpose as t-SNE which means that the local structure will be preserved and is relatively insensitive to outliers and noise. The algorithm consist of a few steps, such as building a adjacency graph for a set of points in high-dimensional space, choose the weight for edges, eigen-decomposition of the graph and form the low-dimensional embedding.

Firstly, all data point in the high-dimensional space $X = \{x_1, x_2, x_3, \ldots, x_n\}$ will be represented as nodes. The adjacency graph is constructed by connecting the nodes for

which the nodes are close to each other. This can be done in two ways, either connecting two nodes only if $\|x_i - x_j\| < \epsilon$ for which $\epsilon \in \mathbb{R}$. Or using n nearest neighbors to determine which nodes has to be connected which means if node x_i is of the n nearest neighbors of node x_j or x_j is among the n nearest neighbor of node x_i, then these two nodes will be connected.

Secondly, weights have to be assigned to the connections. This can be achieved by two variations as well. For the simple-minded variant, if and only if node x_i and x_j are connected, then weight $w_{ij} = 1$, otherwise the weight $w_{ij} = 0$. Another variaton called heat kernel calculates the weight as following $w_{ij} = \exp^{-\frac{\|x_i - x_j\|^2}{t}}$ if the nodes are connected, otherwise w_{ij} is zero.

Thirdly, assume that graph G, which is constructed in the previous steps, is connected. Otherwise, proceed with this step for each connected component. The eigenvalues and eigenvectors are computed for the following eigenvector problem:

$$Ly = \lambda Dy \tag{8}$$

where D is the diagonal weight matrix, its entries are column sums of W, $D_{ii} = \sum_j W_{ji}$, $L = D - W$ for which L is the Laplacian matrix. "Laplacian is a symmetric, positive semidefinite matrix that be thought of an operator on functions defined on vertices of G." Let $y_0, \ldots, y_{k-1}$ be the solutions of equation 8, for which the solutions are ordered according to their eigenvalues:

$$\begin{gathered} Lf_0 = \lambda_0 D f_0 \\ Lf_1 = \lambda_1 D f_1 \\ \cdots \\ Lf_{k-1} = \lambda_{k-1} D f_{k-1} \\ 0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{k-1}. \end{gathered}$$

The eigenvector f_0 corresponding to eigenvalue 0 will be left out and the next m eigenvectors are used for embedding in m-dimensional Euclidean space:

$$x_i \to (f_1(i), \ldots, f_m(i)).$$

This algorithm is also implemented by van der Maaten in MATLAB. [22] As explained, the algorithm consist of multiple variations, therefore step 1 (constructing graph) is implemented by using n nearest neighbors for which the default value is 12. The advantages of using this method are: it simplifies choosing the to be connected nodes and it prevents that disconnected graphs will occur. However, the disadvantage of this method is that the geometrical intuitions are less obvious. Moreover, step 2 (choosing weights) can also be done in two ways. For this step, weights are based on heat kernel for which parameter t is set to 1. [2]

3.4 Experimental Setup

In order to investigate whether t-SNE is an applicable method for the breast cancer data set, we will employ two approaches. One approach will investigate the visualizations by t-SNE and therefore we use Laplacian Eigenmaps for comparison. The second approach that we use is to compare the classification performances of t-SNE with the performances

of Laplacian Eigenmaps and the original data. The second approach is illustrated by figure 4.

In table 1, all t-SNE parameters are listed, most of these parameter values are set to default, except for perplexity. Each experiment for which it involved t-SNE or Laplacian Eigenmaps, the dimension of the data will initially be reduced to the initial dimensions parameter value. This is proposed by van der Maaten in order to accelerate the computation. Furthermore, as listed in table 1, the list consist of two momentum parameters. The final momentum is greater than the initial momentum which ensures the solution to find a global optimum rather than a local minimum. Another important parameter is the iteration number for which the algorithm has to stop "lying" about the P-values. This parameter is responsible for creating more spacings between clusters such that clusters are able to move around easier in order to find a better global solution. In their paper it is called "early exaggeration".

Parameter	Value	Description
perplexity	500	the number of effective neighbors
initial dimensions	30	PCA reduces the data to initial number of dimensions
max_iter	1000	maximum number of iterations T
epsilon	500	learning rate η
momentum	0.2	initial momentum $\alpha(t)$
final_momentum α	0.8	final momentum $\alpha(t)$
min_gain	0.01	minimum gain
mom_switch_iter	250	iteration at which momentum is changed
stop_lying_iter	100	iteration at which lying about P-values is stopped

Table 1: Cost and optimization function parameter settings for the experiments

3.5 Evaluation methods

The breast cancer data set contains only 0.09% of cancerous regions. Due to the highly skewed data set, the data set has to be transformed such that each train and test set contains enough malignant as benign cases. This prevents that the classification algorithm is either trained on benign or malignant examples. The data set is divided into 5 subsets with the notion that all region from one person should be in either one of the subsets. The five training and test sets are generated as follows. First, all of the malignant regions were selected and are transferred into a new set. Then, for each of the malignant region the similarity to each benign region was computed, this is done by calculating the Euclidean distances between these two regions. As result, that the two most similar malignant and non-malignant regions are found. These two regions are transferred to a new training set. For each training set, an internal cross validation had been used to find the best parameters C and γ [4] to test the testing data. Moreover, a weight has been attached to each class, such that mistakes on malignant cases are penalized much heavier than benign cases.

According to the first approach, t-SNE will be compared with the Laplacian Eigenmaps visualizations. In order to gain insight whether visualizing data is more useful than classifying the data without reducing the dimensionality, we will compare the performance measures obtained from SVM classification of all three methods. The performance measures that we employ are:

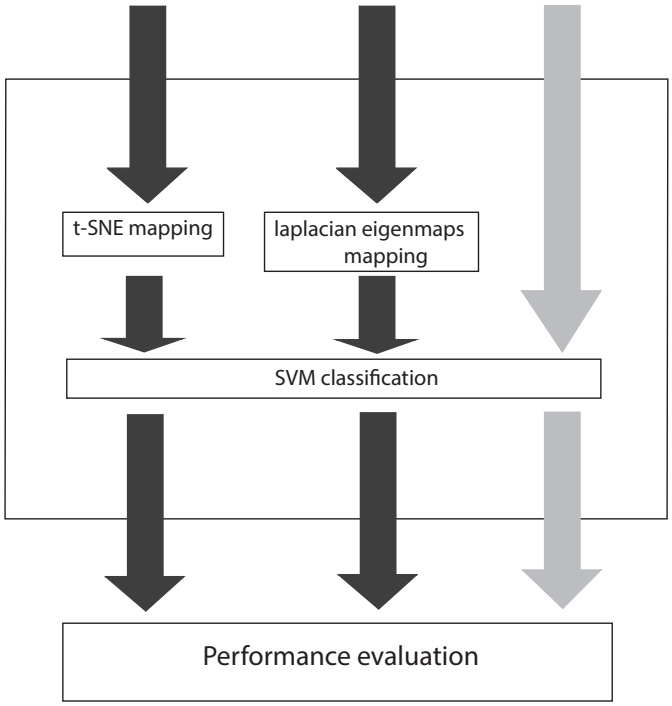


Figure 4: work flow outline of the second approach : the black arrows represents the data that first will be visualized and then be classified. For the grey colored arrow the data which will be classified without reducing the dimensionality. By performing these three experiments for each training and test set, we can compare the results and see whether visualizing the data has a positive contribution to the performance

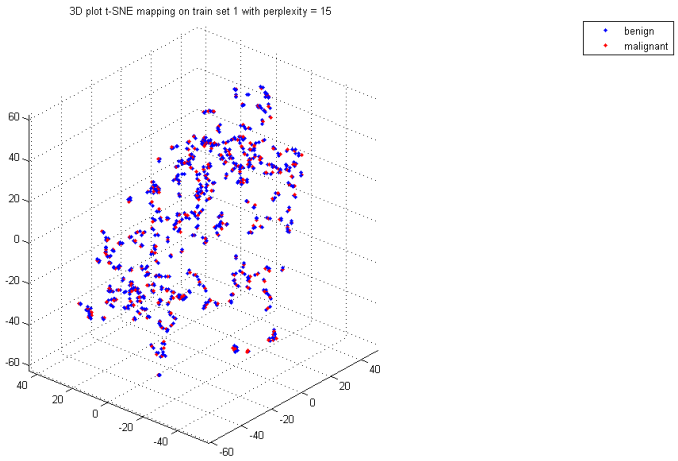
- sensitivity = $\frac{TP}{TP+FN}$
- specificity = $\frac{TN}{TN+FP}$
- accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

4 Results

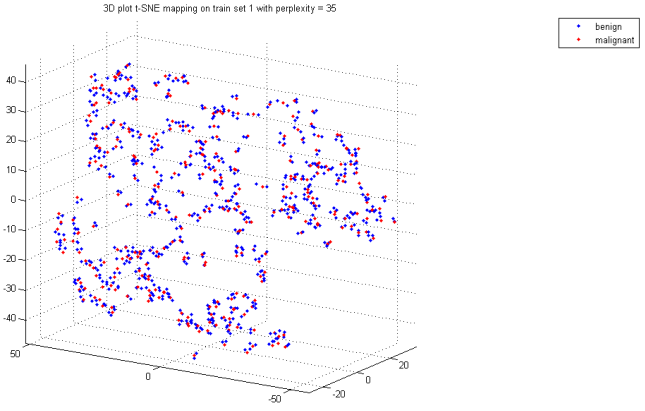
4.1 Visualizations by t-SNE

In this section, the results will be shown and described. In figure 5 visualizations t-SNE with different perplexity values are presented. According to the paper by van der Maaten, the perplexity value should be between 5 and 50. In this figure, the two visualizations by t-SNE with perplexity value of respectively 15 and 35 represented in 3-dimensional space. Reducing the data set to 3 dimensions leads to more insightful visualizations than in 2-dimensional space. This becomes helpful when a lot of data points are covered by other data points. However, as shown figure 5a the data points are still hardly separated in contrast to figure 5b for which the perplexity has increased, as result that data points are more separated from each other. In figure 6a, the visualization by t-SNE with perplexity value of 50 is shown in 2 dimensions which makes it easier to interpret the visualization. Although increasing the perplexity value leads to better visualizations, the malignant and benign cases are still not clustered well. As mentioned, increasing the perplexity value

is able to separate the data points which are extremely similar, moreover, this leads to bigger groups of data points shown in figure 6b which is clearly not the case in when the data set is visualized with a perplexity value of 50.

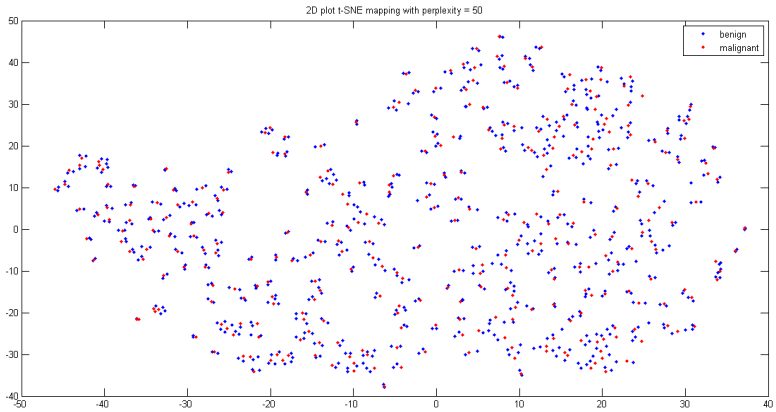


(a) cost function parameter perplexity is set to 15

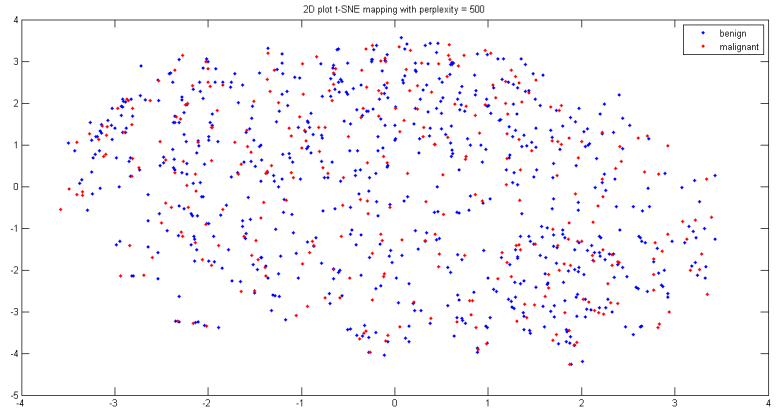


(b) cost function parameter perplexity is set to 35

Figure 5: The breast cancer data set is reduced to 3 dimensions



(a) cost function parameter perplexity is set to 50



(b) cost function parameter perplexity is set to 500

Figure 6: The breast cancer data set is reduced to 2 dimensions

4.2 Comparison with Laplacian Eigenmaps visualizations

In figure 7, the visualization by Laplacian Eigenmaps are shown with default value 12 for k nearest neighbors. The visualizations are not well clustered either. And the Laplacian eigenmaps visualizations consist of bigger clusters for which the benign and malignant cases mixed together.

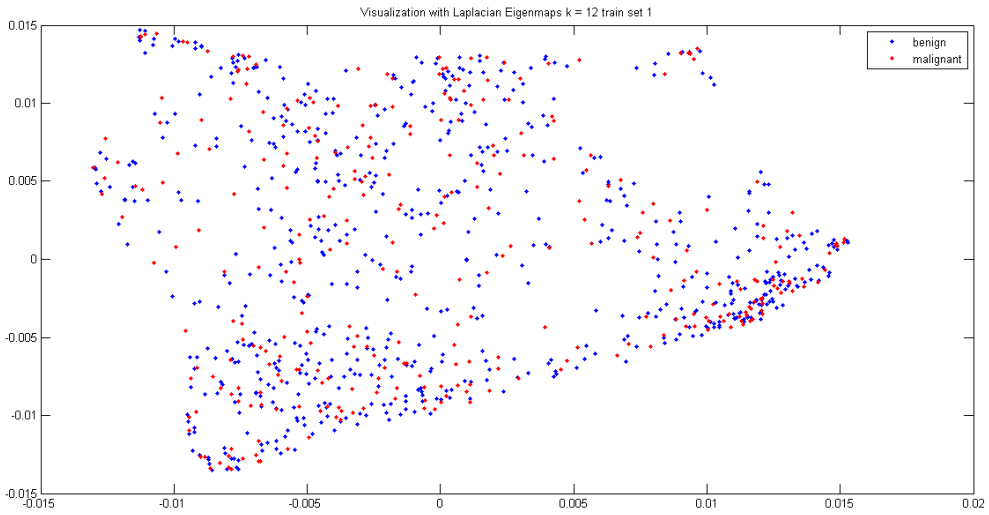
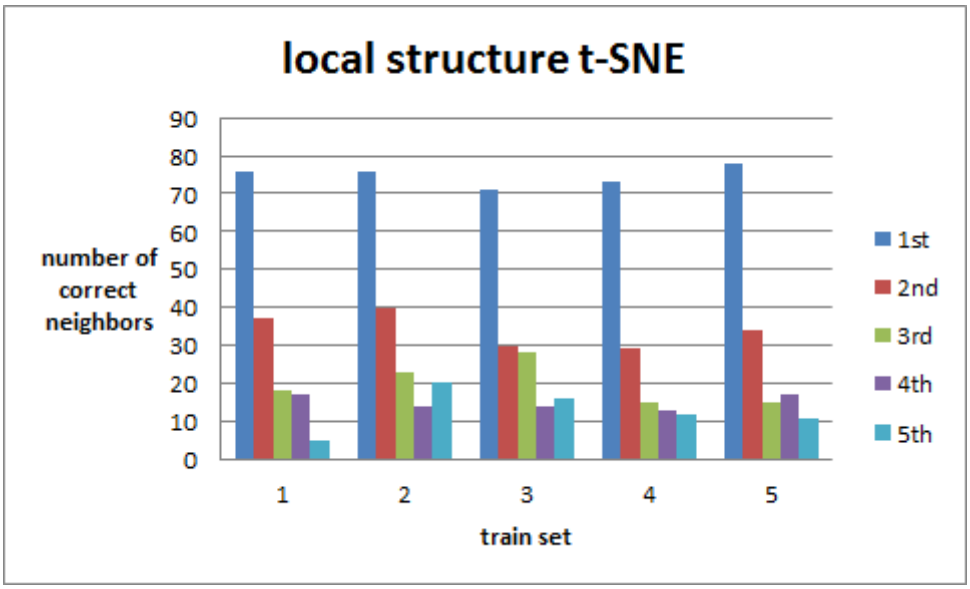


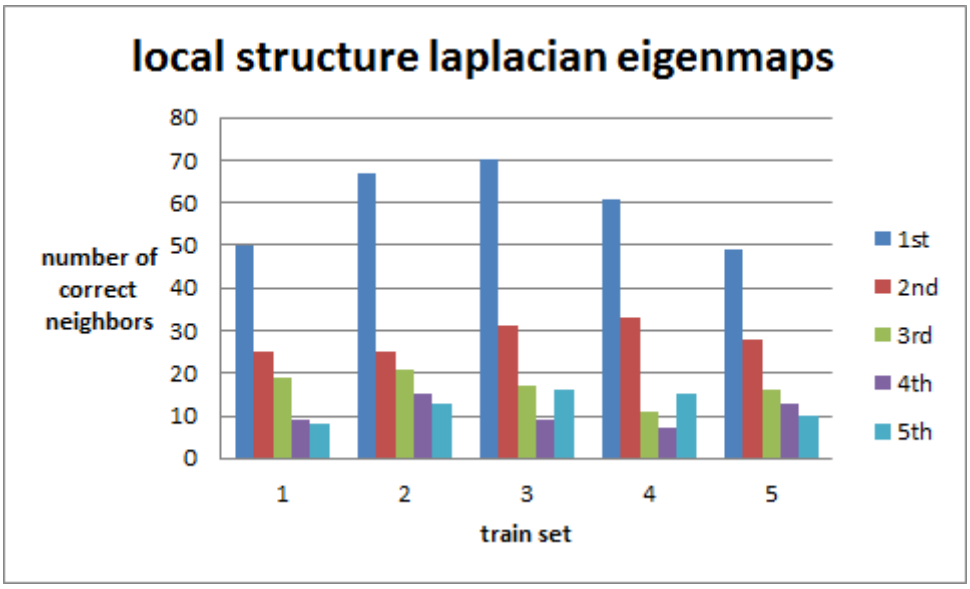
Figure 7: k nearest neighbors set to 12

4.3 Local structure

The main purpose of this visualization technique is reduce the dimensionality such that the global as well as the local structures are preserved. This means that the first nearest neighbor for a data point in the high dimensional space should also be modeled as the first nearest neighbor for that data point in the lower dimension, the second nearest neighbor in the dimension as the second nearest neighbor in the lower dimension, etc. Figure 8 shows the number of correct neighbors for the first five nearest neighbors .



(a) the number of correct neighbors for each train set



(b) the number of correct neighbors for each train set

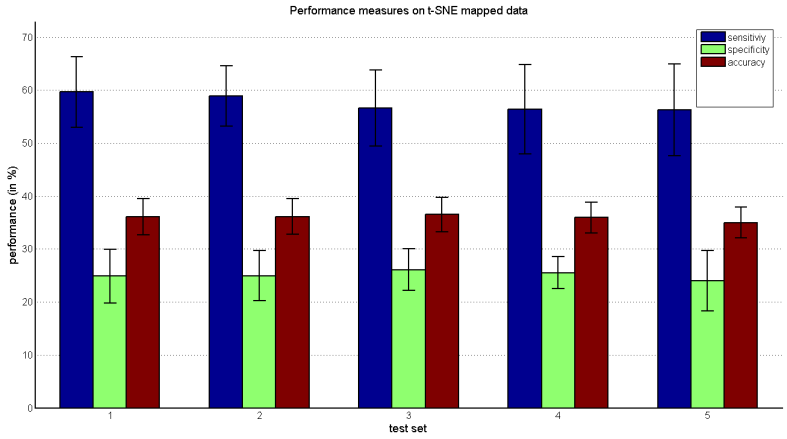
Figure 8: Compare t-SNE with Laplacian Eigenmaps in sense of preserving the local structure

The figure clearly shows that t-SNE is able to preserve the local structure much better than the Laplacian Eigenmaps. Moreover, the differences between data sets are also larger for the Laplacian eigemaps visualizations which can be due to the nearest neighbors that are selected using the k nearest neighbor algorithm is more affected by the distances

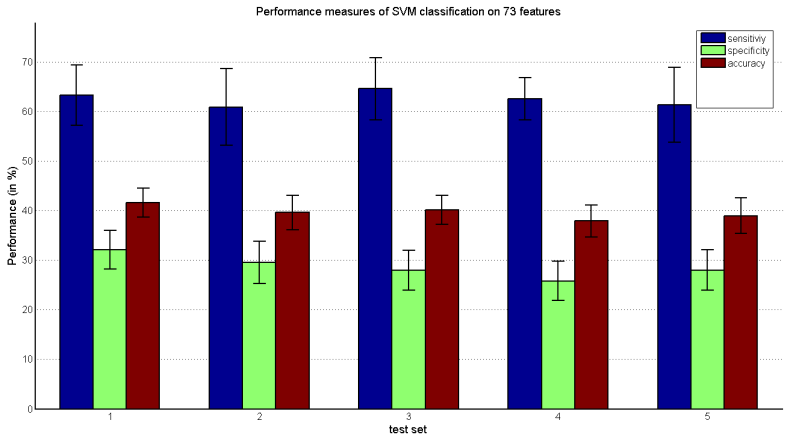
between data points, while in contrast, the perplexity value that has been used by t-SNE shows more consistency.

4.4 SVM classification

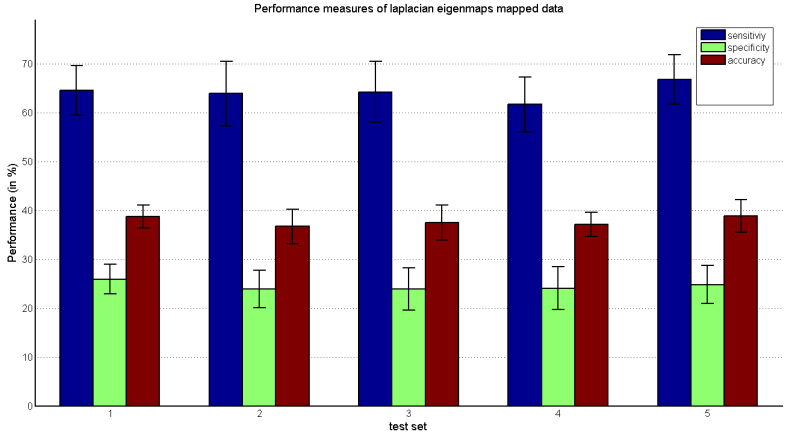
As mentioned in the previous section, three performance measures are used for comparing the SVM classifications. The classification results of both Laplacian Eigenmaps and the unvisualized data are better than on t-SNE test data.



(a) SVM classification on t-SNE data



(b) SVM classification on unvisualized data



(c) SVM classification on Laplacian Eigenmaps data

Figure 9: SVM classification

	sensitivity	specificity	accuracy
t-SNE	0.576	0.251	0.359
Original data	0.625	0.287	0.397
Laplacian Eigenmaps	0.643	0.245	0.378

Table 2: Averages across the five data sets

For the second approach, the results are shown in table 2. The average across the five data sets for each performance measure is calculated. As result that statistically significant differences in the average were found for specificity, sensitivity and accuracy between t-SNE classification performances and the original data with a 95% confidence intervals. This means that visualizing data using t-SNE does not perform better than classification without the data being visualized. Significant differences were also found for sensitivity and accuracy between the t-SNE and Laplacian Eigenmaps visualization. Thus, no significant difference for specificity was found. In the sense of medical diagnostics, using t-SNE might lead to more healthy people be correctly classified as not having breast cancer than using Laplacian Eigenmaps.

5 Conclusion and Discussion

The research question is that we want to know whether using t-SNE visualization technique is able to give useful information from the breast cancer data set. We have used two approaches to explore the applicability of t-SNE. For the first approach we compared the visualization by t-SNE with Laplacian Eigenmaps. Both visualization methods were not able to separate the two categories in an interpretable way. However, the local structure of the data was preserved better by using t-SNE than Laplacian Eigenmaps.

For the second approach, we have compared the classification performances of t-SNE with Laplacian Eigenmaps and original data. Both Laplacian Eigenmaps and the original data performed significantly better on the classification task. Significant differences were found for sensitivity and accuracy. Thus, visualizing the data does not improve the classification performances.

These differences can be caused by the SVM parameters C and γ which are not obtained from the internal cross-validation, but these parameter values are finally manually chosen. The reason to choose the values manually is because of the problems that we have encountered.

Several reasons may have caused the failure of visualizing the breast cancer data using t-SNE in an interpretable way. Firstly, the data set consist of very few number of malignant cases which makes it difficult to train the classifier well, which may be the reason why the unvisualized data did perform poorly either. Secondly, not all data has been used to training and testing, moreover the employed benign and malignant are really similar, which can be defined as the "hard" cases.

The results from the experiments did not show that t-SNE is an applicable technique for the breast cancer data set. Neither the visualization contained well separated clusters nor the performances of the SVM classification on the test data were promising.

References

[1] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems 14*, pages 585–591. MIT Press, 2001.

[2] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15:1373–1396, 2003.

[3] N. Bushati, J. Smith, J. Briscoe, and C. Watkins. An intuitive graphical visualization technique for the interrogation of transcriptome data. *Nucleic Acids Res*, 39(17):7380–9, 2011.

[4] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[5] J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-seperated clusters. *Journal of Cybernetics*, 3:32–57, 1973.

[6] I Gashi, V Stankovic, C Leita, and O Thonnard. An experimental study of diversity with off-the-shelf antiVirus engines. In *NCA09, 8th IEEE International Symposium on Network Computing and Applications, July 9-11, 2009, Cambridge, MA USA*, Cambridge, UNITED STATES, 07 2009.

[7] M. L. Giger. Computer-aided diagnosis of breast lesions in medical images. *Computing in Science and Engg.*, 2(5):39–45, 2000.

[8] G. Hinton and S. Roweis. Stochastic neigbor embedding. *Advances in Neural Information Processing Systems*, 15:833–840, 2002.

[9] R. Hupse and N. Karssemeijer. Use of normal tissue context in computer-aided detection of masses in mammograms. *IEEE Transactions on Medical Imaging*, 28:2033–2041, 2009.

[10] R. Hupse and N. Karssemeijer. The effect of feature selection methods on computer-aided detection of masses in mammograms. *Physics in Medicine and Biology*, 55:2893–2904, 2010.

[11] Imaginis. How mammography is performed: Imaging and positioning. `http://www.imaginis.com/mammography/how-mammography-is-performed-imaging-and-positioning-2`.

[12] A. Jamieson, M. Giger, K. Drukker, H. Li, Y. Yuan, and N. Bhooshan. Exploring nonlinear feature space dimension reduction and data representation in breast CADx with laplacian eigenmaps and t-SNE. *Medical Physics*, 37(1):339–351, 2010.

[13] A. Jamieson, M. Giger, K. Drukker, and L. L. Pesce. Enhancement of breast CADx with unlabeled data. *Medical Physics*, 37(8):4155–4172, 2010.

[14] N. Karssemeijer and G.M. te Brake. Detection of stellate distortions in mammograms. *Medical Imaging, IEEE Transactions on*, 15(5):611–619, 1996.

[15] World Health Organization. Breast cancer: prevention and control. http://www.who.int/cancer/detection/breastcancer/en/, Januari 2012.

[16] A. Platzer. Visualization of SNPs with t-SNE. *PLoS ONE*, 8(2):e56883, 2013.

[17] P. J. Rousseeuw. Silhouettes - a graphical aid to the interpretation and validation od cluster-analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

[18] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *SCIENCE*, 290:2323–2326, 2000.

[19] J.W. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, 18(5):401–409, 1969.

[20] G. M. te Brake and Nico Karssemeijer. Single and multiscale detection of masses in digital mammograms. *IEEE Trans. Med. Imaging*, 18(7):628–639, 1999.

[21] J. B. Tenenbaum, V. De Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

[22] L. van der Maaten. Matlab toolbox for dimensionality reduction (v0.8.1 - march 2013). http://homepage.tudelft.nl/19j49/Matlab_Toolbox_for_Dimensionality_Reduction.html.

[23] L. van der Maaten. t-distributed stochastic neighbor embedding. http://homepage.tudelft.nl/19j49/t-SNE.html.

[24] L. van der Maaten and G. Hinton. Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, 2008.