TECHNICAL UNIVERSITY OF CRETE, GREECE

SCHOOL OF ELECTRONIC AND COMPUTER ENGINEERING

Data Classification and Mapping in Optical Dynamic Contrast-Enhanced Imaging of Cervical Neoplasia

Eleni Chiou

Thesis Committee

Associate Professor Costas Balas (ECE)

Professor Michalis Zervakis (ECE)

Associate Professor Michail G. Lagoudakis (ECE)

Chania, Jule 2014

Abstract

The main purpose of this study is the evaluation of temporal feature extraction methods from normal and abnormal cervical epithelium images acquired by Dynamic Contrast Enhanced Optical Imaging (DCE-OI). Dynamic optical data were recorded *in vivo* during agent-tissue interaction inducing a transient tissue whitening phenomenon, known as acetowhitening. The degree and duration of the optical phenomenon is associated with the lesion grade. The available data, observed on 64 patients and confirmed by in total 371 biopsy samples, include all possible cases of neoplasia grades. In order to reduce the data dimensionality and extract valuable information, an extensive number of feature extraction methods, including Wavelet Transform (WT), Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA), Piecewise Aggregate Approximation (PAA), Adaptive Piecewise Aggregate Approximation (APAA) and Symbolic Aggregate Approximation (SAX), have been implemented and (cross) validated with a 1-NN classifier. The results indicate that using a subset of the entire feature set, WT, PCA and KPCA methods present similar or better performance compared to using the entire feature set. Also, selection of the best features extracted from PCA and WT demonstrates better performance for some classification cases. Regarding the execution times, PCA presents lower execution time than the other methods. Thus, high discrimination performance between the various stages of cervical neoplasia can be achieved.

Acknowledgements

First, I would like to thanks Professor Costas Balas for trusting me with this interesting topic and for his guidance during the course of this thesis.

Next, I would like to thank George Epitropou for his help and guidance throughout this thesis and all the members of Opto-electronics laboratory for creating a great working environment.

Many thanks to my friends Nikos, Marina, Katerina, Paoki, Nikolas, Billy, Christos, Xanthia and Bouklas, for the amazing moments we have together the last years.

Last, but not least, I would like to thank my family for the support and encouragement all these years.

Contents

List of Figures

List of Tables

LIST OF TABLES

Chapter 1

Introduction

1.1 Cervical Cancer

Cervical cancer is the fourth most common cancer affecting women worldwide. In 2012, 528.000 new cases were diagnosed. The problem is most common in the developing countries where about 83% of all new cases appear and cervical cancer is the leading cause of cancer-related death among women. In almost all cases, the main cause is the changes in the structure of DNA caused by the Human Papillomavirus (HPV) [1]. HPV is actually a group of more than 100 related viruses. The infections with some types of sexually transmitted HPV do not cause noticeable symptoms and resolve spontaneously. However, other types of HPVs, called high-risk HPVs, may progress slowly to abnormal and possibly precancerous lesions, also known as Cervical Intraepithelial Neoplasia (CIN). CIN may be categorized into grades I, II and III depending upon the severity of abnormality. High-grade dysplasia (CIN II, CIN III) carry higher probability of progressing to invasive cancer rather than Low-grade dysplasia (CIN I), which progress slowly to high-grade lesions. Thus, the early detection of dysplasia is very important for effective treatment.

Currently, Pap test is the most common method for cervical cancer screening. Cells collected from a woman's cervix are spread on a microscope slide for examination. The cells are evaluated for abnormalities, specifically for pre-cancerous and cancerous changes. Pap test has led to a significant reduction in the mortality from cervical cancer, mainly in the developed countries. However, the accuracy of this screening method remains significantly low (50%) because of sampling and reading errors[2].

Women with abnormal Pap tests are referred to colposcopy which determines the location of the most severe dysplastic region for biopsy sampling. Colposcopic examination allows the colposcopist to visually identify abnormal tissue and take directed biopsies for further pathological examination. Before colposcopy, $3\% - 5\%$ acetic acid solution is applied to cervix to stain the abnormal areas. However, conventional colposcopy is subjective as it mainly depends on the visual prowess and the experience of the physician[3, 4].

On the other hand, optical diagnostic techniques have attracted much interest recently as they offer the possibility of objective diagnostics both in vitro and in vivo[5]. The key idea between all these different technologies is the exploitation of the light-tissue interaction phenomena in order to provide information for the biochemical and micro-structural changes occurring in neoplastic growth. In other words, as the grade of a neoplastic tissue advances, complex biological processes take place, which induce local variations in the refraction index and fluctuations in the traveling population of photons as they are scattered or absorbed in the tissue. Although certain limitations exist depending on the implementation, this techniques can provide substantial diagnostic information. Granting that a validated and positive quantitative correlation or pattern between the physics phenomena and the biological changes is established, new methodologies can improve, automate and decrease the cost of screening and detection.

1.2   Thesis Contribution

The main purpose of this study is the discrimination between the different stages of cervical neoplasia using data acquired by Dynamic Contrast Enhanced Optical Imaging (DCE-OI). Dynamic optical data were recorded *in vivo* during agent-tissue interaction inducing a transient tissue whitening phenomenon, known as acetowhitening.

We applied several feature extraction methods on this data in order to obtain valuable information for the classification problem. More specifically, a wide range of feature extraction techniques, including Wavelet Transform (WT), Principal Component Analysis(PCA), Kernel Principal Component Analysis(KPCA), Piecewise Aggregate Approximation(PAA), Adaptive Piecewise Aggregate Approximation(APAA), and Symbolic Aggregate Approximation(SAX), were implemented and validated using a k-nn classifier.

We applied the above mention techniques and evaluated their performance using both the entire sets of features and a subset of them. The results indicate that using a selected subset of the entire feature set we can achieve similar or better performance. Except of the classification performance, a comparison in terms of computational time was attempted.

Taking into account the classification performance of 1-NN classifier using each of the proposed feature extraction methods as well as the execution time of each method we selected three of them to generate pseudo-color maps representing the stages of disease. The mapping can provide valuable, real-time information which could assist medical personnel in the diagnosis of cervical cancer.

1.3 Thesis Summary

In Chapter 2 we discuss the methods applied in this study. More specifically, we analyze six methods for feature extraction and dimensionality reduction used to extract valuable features from the original data. In addition a simple supervised classification algorithm used in this work is presented.

In Chapter 3 we discuss the problem of discrimination between the various cervical tissue types and present information about the data used for the evaluation of the methods.

In Chapter 4 we describe the various cases which was examined and the procedure which was followed to implement and evaluate the classification performance using the proposed feature extraction methods.

In Chapter 5 we present results about the impact of feature extraction methods to the classification performance. We examine the classification performance using both the entire set of the extracted features and a subset of it. In addition, results related to the classification performance using features obtained from the combination of three features extraction methods are presented.

In Chapter 6 we present maps generated after the classification process of image stacks obtained from patients with High and Low grade disease and no evidence of disease as well. The generated pseudo-color map, with different colors representing different tissue types, is overlaid onto the color image of the cervix providing valuable information about the stage of cervical neoplasia.

Finally, in Chapter 7 we summarize the conclusions we were guided towards and the possible future research directions on the problem.

Chapter 2

Theoretical Background

In pattern recognition, *classification* is the procedure of detecting the class in which an individual observation belongs, based on a set of inherent characteristics known as *features* [6]. A set of N observations $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, called *training set*, are used to tune the parameters of a classifier. The class of each observation can be expressed by a *label* l. Applications for which the training set consists of labeled observations are known as *supervised learning* problems. In other types of problems known as *unsupervised learning* problems, the training set consists of a set of observations without any corresponding label values. In this study, the labels of training observations are known and the classification problem constitutes a supervised classification problem. For most applications, the classification problem is complicated enough because of the high dimensionality of input observations. Thus, pre-processing to reduce the dimensionality of them is required. Feature extraction and feature selection constitute two common techniques in data pre-processing [7, 8]. *Feature extraction* is the transformation of the original input data by projecting them onto a new feature space where the classification problem can be solved easier. *Feature selection* is the process of selecting a subset of the original features in order to build a better classifier. Often the above mentioned pre-processing techniques are combined in order to reduce the dimensionality of the original dataset. The rest of this chapter covers the theoretical background concerning the dimensionality reduction and classification methods used in this study.

2.1 Feature Extraction

2.1.1 Principal Component Analysis

Principal Component Analysis (PCA) constitutes a widely known technique for feature extraction and dimensionality reduction. It is a statistical procedure which reduces the dimension of the data by projecting them onto a lower dimensional linear space, known as the principal subspace, such that the variance of the projected data is maximized [9].

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K]^T$ be a $K \times D$ observation matrix, where K is the number of observations and D the dimensionality of each observation $\mathbf{x}_k$, for $k = 1 \ldots K$. As mentioned above, the goal of PCA is to project the data onto a principal subspace of dimensionality $N \leq D$, maximizing the variance of the projected data.

Consider the projection of $\mathbf{x}_k$, where $k = 1 \ldots K$, onto a one-dimensional space described by a direction vector $\mathbf{u}_1 \in \mathbb{R}^D$. The projection of $\mathbf{x}_k$ is given by

$$\mathbf{x}_{k \perp u_1} = \frac{\mathbf{u}_1^T \mathbf{x}_k}{\|\mathbf{u}_1\|} \tag{2.1}$$

while the mean value of the projected data is given by

$$\bar{x}_{\perp u_1} = \frac{\mathbf{u}_1^T \bar{\mathbf{x}}}{\|\mathbf{u}_1\|}, \text{ where } \bar{\mathbf{x}} = \frac{1}{k} \sum_{k=1}^{K} \mathbf{x}_k \tag{2.2}$$

The variance of the projected data is given by

$$
\begin{aligned}
\sigma^2_{\perp u_1} &= \frac{1}{K} \sum_{k=1}^{K} (\mathbf{x}_{k \perp u_1} - \bar{x}_{\perp u_1})^2 \\
&= \frac{1}{K} \sum_{k=1}^{K} \left(\frac{\mathbf{u}_1^T \mathbf{x}_k}{\|\mathbf{u}_1\|} - \frac{\mathbf{u}_1^T \bar{\mathbf{x}}}{\|\mathbf{u}_1\|} \right)^2 \\
&= \frac{1}{K} \frac{1}{\|\mathbf{u}_1\|} \sum_{k=1}^{K} (\mathbf{u}_1^T (\mathbf{x}_k - \bar{\mathbf{x}}))^2 \\
&= \frac{1}{K} \frac{1}{\|\mathbf{u}_1\|} \sum_{k=1}^{K} \mathbf{u}_1^T (\mathbf{x}_k - \bar{\mathbf{x}}) \mathbf{u}_1^T (\mathbf{x}_k - \bar{\mathbf{x}}) \\
&= \frac{1}{K} \frac{1}{\|\mathbf{u}_1\|} \sum_{k=1}^{K} \mathbf{u}_1^T (\mathbf{x}_k - \bar{\mathbf{x}}) (\mathbf{x}_k - \bar{\mathbf{x}})^T \mathbf{u}_1 \\
&= \frac{1}{\|\mathbf{u}_1\|} \mathbf{u}_1^T \mathbf{S} \mathbf{u}_1, \text{ where } \mathbf{S} = \frac{1}{K} \sum_{k=1}^{K} (\mathbf{x}_k - \bar{\mathbf{x}}) (\mathbf{x}_k - \bar{\mathbf{x}})^T
\end{aligned}
\tag{2.3}
$$

The maximization of the projected variance with respect to $\mathbf{u}_1$ is achieved by solving the following optimization problem

$$\max \mathbf{u}_1^T \mathbf{S} \mathbf{u}_1$$
$$\text{subject to } \mathbf{u}_1^T \mathbf{u}_1 = 1 \tag{2.4}$$

Introducing a Lagrangian multiplier to enforce the above constrain, we obtain the following objective function

$$\mathbf{u}_1^T \mathbf{S} \mathbf{u}_1 - \lambda_1 (1 - \mathbf{u}_1^T \mathbf{u}_1) \tag{2.5}$$

Setting to zero the gradient of the Equation 2.5 with respect to $\mathbf{u}_1$, we have that

$$\mathbf{S} \mathbf{u}_1 = \lambda_1 \mathbf{u}_1 \tag{2.6}$$

Since $\mathbf{S}$ is a positive semi-definite matrix, it can be diagonalized by its orthonormal eigenvectors as follows

$$\mathbf{S} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \tag{2.7}$$

where $\mathbf{\Lambda} = diag[\lambda_1, \ldots, \lambda_D]$ a diagonal matrix of the non-negative, ordered eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_D$ of $\mathbf{S}$ and $\mathbf{U} = [\mathbf{u}_1, \ldots \mathbf{u}_D]^T$ an orthogonal matrix containing the corresponding eigenvectors. Thus, the Equation 2.6 can be defined as an eigenvalue problem for the $D \times D$ matrix $\mathbf{S}$ with eigenvalues λ_d and eigenvectors $\mathbf{u}_d$. The desired vector $\mathbf{u}_1$ is an eigenvector of the covariance matrix $\mathbf{S}$ and λ_1 the corresponding eigenvalue. The Equation 2.6 can be written as follows

$$\mathbf{u}_1^T \mathbf{S} \mathbf{u}_1 = \lambda_1, \text{ where } \mathbf{u}_1^T \mathbf{S} \mathbf{u}_1 \text{ the variance of projected data} \tag{2.8}$$

According to Equation 2.8 the variance of the projected data is maximized when $\mathbf{u}_1$ is the eigenvector corresponding to the largest eigenvalue λ_1.

As regards the general N-dimensional projection space, it is proven by induction that the optimal liner subspace, maximizing the variance of the projected data is defined by the n eigenvectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\}$ of the data covariance matrix S corresponding to the N largest eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$.

The projection of the original data onto the N dimensional space is given by

$$\mathbf{Y} = \mathbf{X} \mathbf{U}_N^T, \tag{2.9}$$

where $\mathbf{U}_N$ a $D \times N$ matrix containing the eignevectors of $\mathbf{S}$, corresponding to the N largest eigenvalues and Y the projection of data onto the N-dimensional subspace.

Finally, selecting the number of eigenvectors spanning the N-dimensional subspace is a trade-off between dimensionality reduction and retained variance. The percentage of variance retained by keeping N eigenvectors is given by the following equation

$$rv = \frac{\sum_{n=1}^{N} \lambda_n}{\sum_{d=1}^{D} \lambda_d} \times 100 \qquad (2.10)$$

The retained variance is a useful indicator since the larger the retained variance, the less the loss of information.

2.1.2 Kernel Principal Component Analysis

Kernel Principal Component Analysis (KPCA) is a non-linear dimensionality reduction technique which constitutes an extension of the conventional PCA. Like in PCA, KPCA seeks to perform a projection of the original data into a lower dimensional space, maximizing the variance of projected data. However, KPCA performs dimensionality reduction, after mapping the data onto a high dimensional feature space through a non-linear transformation, using the kernel trick [10].

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K]^T$ be a $K \times D$ observation matrix, where K is the number of observations and D the dimensionality of each observation $\mathbf{x}_k$, for $k = 1 \ldots K$. Suppose that the data are centered and have zero mean, so that $\frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_k = 0$. As mentioned in Section 2.1.1, the principal components are the eigenvectors satisfying

$$\mathbf{S}\mathbf{u}_d = \lambda_d \mathbf{u}_d, \qquad (2.11)$$

where $\mathbf{S} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_k \mathbf{x}_k^T$ the $D \times D$ covariance matrix of the centered data.

Consider a non-linear transformation $\phi(x)$, mapping the data onto a high dimensional feature space F of dimensionality M so that

$$\begin{aligned} \phi : R^D &\to F, \\ \mathbf{x}_k &\mapsto \phi(x_k) \end{aligned} \qquad (2.12)$$

The principal components in feature space F are the eigenvectors satisfying

$$\mathbf{C}\mathbf{v}_m = \lambda_m \mathbf{v}_m, \ \ \text{m=1} \ldots \text{M} \qquad (2.13)$$

where $\mathbf{C} = \frac{1}{K}\sum_{k=1}^{K}\phi(\mathbf{x}_k)\phi(\mathbf{x}_k)^T$ the $M \times M$ covariance matrix of the center data in feature space F and $\mathbf{v}_m$, λ_m the eigenvectors and eigenvalues of $\mathbf{C}$. Since the eigenvectors $\mathbf{V} = \{\mathbf{v_1}\dots\mathbf{v_M}\}$ lie in the span of $\phi(\mathbf{x}_1)\dots\phi(\mathbf{x}_K)$, there are two important properties. Firstly, there exist coefficients a_{mk} such that

$$\mathbf{v}_m = \sum_{k=1}^{K} a_{mk}\phi(\mathbf{x}_k) \tag{2.14}$$

and secondly we can obtain the equivalent equation

$$\phi(\mathbf{x}_i)^T\mathbf{C}\mathbf{v}_m = \phi(\mathbf{x}_i)^T\lambda_m\mathbf{v}_m, \ i=1\dots K \tag{2.15}$$

Combining the Equations 2.14 and 2.15, we obtain

$$\frac{1}{K}\sum_{k=1}^{K}\phi(\mathbf{x}_k)\phi(\mathbf{x}_i)^T\sum_{i=1}^{K}a_{mi}\phi(\mathbf{x}_i)\phi(\mathbf{x}_k)^T = \lambda_i\sum_{k=1}^{K}a_{mk}\phi(\mathbf{x}_k)\phi(\mathbf{x}_i)^T \tag{2.16}$$

Expressing the Equation 2.16 in terms of kernel function, we obtain

$$\frac{1}{K}\sum_{k=1}^{K}k(\mathbf{x}_k,\mathbf{x}_i)\sum_{i=1}^{K}a_{mi}k(\mathbf{x}_i,\mathbf{x}_k) = \lambda_m\sum_{k=1}^{K}a_{mk}k(\mathbf{x}_k,\mathbf{x}_i) \tag{2.17}$$

Defining a $M \times M$ matrix $\mathbf{K}$ we obtain

$$\mathbf{K}^2\mathbf{a}_m = \lambda_m N\mathbf{K}\mathbf{a}_m \tag{2.18}$$

where $\mathbf{a}_m$ K-dimensional columns vectors. Since $\mathbf{K}$ is positive semi-definite, we can define the following eigenvalue problem

$$\mathbf{K}\mathbf{a}_m = \lambda_m N\mathbf{a}_m \tag{2.19}$$

which gives all solutions $\mathbf{a}_m$ for the equation 2.18.

As regards the condition of normalization for the coefficients $\mathbf{a}_m$, it is obtained by requiring that the corresponding eigenvectors $\mathbf{v}_m$ in the feature space F be normalized, so that

$$\mathbf{v}_m^T\mathbf{v}_m = 1 \tag{2.20}$$

Combining Equation 2.14, 2.20 the condition that the coefficients $\mathbf{a}_m$ have to satisfy is the following.

$$(\mathbf{a}_m^T\mathbf{a}_m)\lambda_m = 1 \tag{2.21}$$

So far we have assumed that the data in the feature space F have zero mean. However, it is not the general case. Thus, it is required the formulation of Gram matrix for the general case. Consider $\widetilde{\phi}(\mathbf{x}_n) = \phi(\mathbf{x}_n) - \frac{1}{N}\sum_{j=1}^{N}\phi(\mathbf{x}_j)$ the centered data in the feature space F. The elements $\widetilde{K}_{nm}$ of Gram matrix is given by

$$\widetilde{K}_{nm} = \widetilde{\phi}(\mathbf{x}_n)\widetilde{\phi}(\mathbf{x}_n)^T \tag{2.22}$$

From Equation 2.22 it is obtained that the Gram matrix $\widetilde{K}$ given by

$$\widetilde{\mathbf{K}} = \mathbf{K} - \mathbf{1}_N\mathbf{K} - \mathbf{K}\mathbf{1}_N + \mathbf{1}_N\mathbf{K}\mathbf{1}_N, \tag{2.23}$$

where $\mathbf{1}_N$ a $N \times N$ matrix with all entries eqaul to $1/N$.

Thus, in the general case the eigenvalue problem has the following form

$$\widetilde{\mathbf{K}}\widetilde{\mathbf{a}}_m = \widetilde{\lambda}_m N\widetilde{\mathbf{a}}_m \tag{2.24}$$

where $\widetilde{\mathbf{a}}_m$, $\widetilde{\lambda}_m$ the eigenvlaues and eigenvectors of $\widetilde{\mathbf{K}}$.

As in Equation 2.21, the normalization condition for $\widetilde{\mathbf{a}}_m$ is obtained by requiring

$$(\widetilde{\mathbf{a}}_m^T\widetilde{\mathbf{a}}_m)\widetilde{\lambda}_m = 1 \tag{2.25}$$

where $\widetilde{\mathbf{a}}_m$ the expansion coefficients of eigenvectors $\widetilde{\mathbf{v}}_m$.

Having solved the eigenvalue problem, the projection of an observation $\mathbf{x}$ onto an eigenvector $\mathbf{v}_i$ in F is given by

$$\begin{aligned}
\mathbf{y_i} &= \phi(\mathbf{x})^T\mathbf{v}_i, \text{ where } i = 1 \ldots N \\
&= \sum_{n=1}^{N} a_{in}\phi(\mathbf{x})^T\phi(\mathbf{x}_n) \\
&= \sum_{n=1}^{N} a_{in}k(\mathbf{x}, \mathbf{x}_n)
\end{aligned} \tag{2.26}$$

2.1.3 Piecewise Aggregate Approximation

Piecewise Aggregate Approximation (PAA) is a simple method for time series feature extraction and dimensionality reduction [11]. The basic idea of this method is the representation of time series by the mean values of equal sized segments.

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K$ a set of K time series of dimensionality D. As mentioned above, the goal of PAA is to represent the data by $N \leq D$ segments. A time series $\mathbf{x}_k$ of

dimensionality D is represented in N dimensional space by a vector $\bar{\mathbf{x}}_k = [\bar{x}_{k1} \ldots \bar{x}_{kN}]$. The element $\bar{x}_{ki}$ of $\bar{\mathbf{x}}_k$ is obtained by

$$\bar{x}_{ki} = \frac{N}{D} \sum_{j=\frac{D}{N}(i-1)+1}^{\frac{D}{N}i} x_{kj} \tag{2.27}$$

To sum up, dimensionality reduction is performed dividing the data onto N equal sized segments. Then, the mean value of the data-points falling in each segment is calculated. The N mean values are the data representation onto the N dimensional space.

2.1.4 Adaptive Piecewise Aggregate Approximation

Adaptive Piecewise Aggregate Approximation (APAA) is a method for time series feature extraction and dimensionality reduction similar to PAA [12]. The basic idea of this method is the representation of time series by the mean value of segments of varying length such that the individual reconstruction errors are minimized.

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K$ a set of K time series of dimensionality D. As mentioned above, the goal of APAA is the representation of the data by $N \leq D$ segments. This method requires the recording of two numbers per segment. The first number record the mean value of data-points falling in each segment while the second number records the right endpoint of the corresponding segment. Thus, the time series representation onto the N dimensional space is given by

$$\bar{\mathbf{x}}_k = [(\bar{x}_{k1}, r_{k1}) \ldots (\bar{x}_{k\frac{N}{2}}, r_{k\frac{N}{2}})], \tag{2.28}$$

where $\bar{x}_{ki} = mean(x_{r_{k(i-1)}+1} \ldots x_{r_{ki}})$ the mean value of data-points falling in the ith segment, r_{ki} the right point of the ith segment and $k = 1 \ldots K$.

The algorithm finds a representation for the time series $\mathbf{x}_k$ converting the problem into a wavelet compression problem for which there are optimal solutions. More specifically, the algorithm computes the wavelet coefficients $\mathbf{w}_k$ of the original time series $\mathbf{x}_k$ and retains the $M \leq \frac{N}{2}$ largest coefficients. Then, the approximation of the time series $\mathbf{x}_k$ is reconstructed using the retained coefficients. The reconstructed time series using the retained coefficients is one optimal representation for the original time series. The values of each segment in the reconstructed time series and the right endpoint of each segment constitute the APAA representation of the original time series.

2.1.5 Symbolic Aggregate Approximation

Symbolic Aggregate Approximation(SAX) is a method for time series feature extraction and dimensionality reduction which allows a time series of arbitrary length to be reduced to a sequence of symbols [13].

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K$ be a set of K time series of dimensionality D. In order to represent the time series with a symbol sequence of size N, it uses the PAA which was mentioned in Section 2.1.3 to produce N equal sized segments represented by their mean value. Having transformed the time series using PAA, a further transformation is applied to obtain a discrete representation. Assuming that the normalized time series have highly Gaussian distribution, the algorithm defines regions in the y-axis mapping each segment mean value to a specific symbol. More specifically, given that the normalized time series have highly Gaussian distribution, the algorithm simply determines the "breakpoints" B which produce a equal-sized areas under Gaussian curve. The values of the breakpoints for specific equal-sized areas are obtained by the following Table.

	$a = 3$	$a = 4$	$a = 5$	$a = 6$	$a = 7$	$a = 8$	$a = 9$	$a = 10$
β_1	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
β_2	0.43	0	-0.25	-0.43	-0.57	-0.67	-0.76	-0.84
β_3		0.67	0.25	0	-0.18	-0.32	-0.43	-0.52
β_4			0.84	0.43	0.18	0	-0.14	-0.25
β_5				0.97	0.57	0.32	0.14	0
β_6					1.07	0.67	0.43	0.25
β_7						1.15	0.76	0.52
β_8							1.22	0.84
β_9								1.28

Table 2.1: A lookup table that contains the breakpoints that divide a Gaussian distribution in an arbitrary number (from 3 to 10) of equiprobable regions.

2.1.6 Wavelet Transform

Wavelet Transform (WT) is a useful tool for signal processing which overcome the limitations of Fourier Transform (FT) [14]. FT decomposes a signal in complex exponential

functions at different frequencies extracting valuable information in the frequency domain. However, FT has a important drawback related to the loss of time information after the transformation as projecting the signal on complex exponential leading to good frequency analysis, but no time localization. Unlike to FT, who decomposes the signal into a basis of complex exponential functions, WT decomposes the signal over a set of shifted and scaled wavelets. This difference allows to the WT to perform multiresolution analysis, which means that different frequencies are processed with different way.

Consider a signal $x(t)$ and a basis function $\psi(t)$, called mother wavelet, the Continious Wavelet Transform(CWT) of the signal $x(t)$ given by

$$Wx(u,s) = \frac{1}{\sqrt{2}} \int_{-\infty}^{+\infty} x(t)\psi^* \left(\frac{t-u}{s} \right) dt \tag{2.29}$$

2.1.6.1 Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) make use of the continuous wavelets with discrete shifted and scaled factors. The wavelet transform is then evaluated at discrete scales and translations. The discrete scale is expressed as $s = s_0^i$, where i is integer and $s_0 = 2$ is a fixed dilation step while the discrete shifted factor is expressed as $\tau = s_0^i k$, where k is an integer.

The DWT transform can be consider a multiresolution representation based on the differences of information available at two successive resolutions 2^j and 2^{j+1}. Consider $x(t)$ a signal of length $N = 2^J$. The output signal will also have length N. The output has $\frac{N}{2}$ values at the highest resolution and $\frac{N}{4}$ values at the next resolution, and so on, that is, the frequency resolution is low at the high frequencies and high at the low frequencies, whereas the time resolution is high at the higher frequencies and low at the lower frequencies. Let $N = 2^J$, and let the number of frequencies, or resolutions, be J we are considering $J = \log N$ levels. Therefore, the frequency index j varies as $1 \ldots J$ corresponding to the scales $2^1, \ldots, 2^J$. If we imagine these data sequences stacked on top of one another, then they constitute a hierarchical pyramid structure with $\log N$ levels. The original signal $x(n)$ is at the bottom or zero level. At j^{th} level, the signal sequence is obtained from the data sequence (approximate coefficients) in the $(j-1)^{th}$. The equation to compute a 1-D wavelet transform at level j using orthogonal filters of length L given

by

$$\alpha_j(i) = \sum_{m=0}^{L-1} \alpha_{j-1}(2i - m)h(m) \quad \text{where,} 1 \leq j \leq \log N, 0 \leq i \leq N_j - 1 \tag{2.30}$$

$$b_j(i) = \sum_{m=0}^{L-1} \alpha_{j-1}(2i - m)l(m) \quad \text{where,} 1 \leq j \leq \log N, 0 \leq i \leq N_j - 1 \tag{2.31}$$

where, $\alpha_j(i)$, $b_j(i)$ the ith approximate and detail coefficients at level j, h, l the low and high pass filter obtained from the chosen wavelet, $N_j = \frac{N}{2^j}$ the number of approximate coefficients at level j.

Regarding the wavelet function, various families can be used in DWT. In this study, we use the Haar wavelet transform because of the high performance reported in a previous study on the same data [14]. The Haar wavelet function is given by

$$\psi(t) = \begin{cases} 1, & \text{if } 0 \leq t \leq \frac{1}{2} \\ -1, & \text{if } \frac{1}{2} \leq t \leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{2.32}$$

and its scaling function $\phi(t)$ as:

$$\phi(t) = \begin{cases} 1, & \text{if } 0 \leq t \leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{2.33}$$

The Haar transform is the simplest of the wavelet transforms and can be considered as averages and differences between every two adjacent values of the input data.

2.2 Classification

2.2.1 k - Nearest Neighbor Classifier

The k nearest neighbor algorithm is a simple method for classification. Given a new data-point, k nearest neighbor classifies it based on the labels of the set of data-points, which are closest to it and called nearest neighbors.

Let a set of labeled observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K$ and an unlabeled observation $\mathbf{x}'$. The k nearest neighbor classifier assigns the observation $\mathbf{x}'$ the label which is the most frequent among the k nearest neighbors. The nearest neighbors are defined measuring the distance of the observation $\mathbf{x}'$ from each of the label observations $\mathbf{x}_i$ where $i = 1 \ldots K$. The most common distance metric for the k nearest neighbor classifier is the Euclidean distance, which is given by

$$d_i = \|\mathbf{x}' - \mathbf{x}_i\|, \tag{2.34}$$

where d_i is the Euclidean distance of the unlabeled observation $\mathbf{x}'$ from each of the labeled observations $\mathbf{x}_i$.

Chapter 3

Problem Specification and Related Work

In this diploma thesis, we have encountered the problem of the discrimination between the various stages of cervical neoplasia based on the aceto-whitening effect kinetics.

The data used in this thesis is captured using the digital colposcope DySIS [15]. DySIS is an imaging device which perform tissue imaging using a 1024×768, 8 bit/channel digital color charge coupled device (CCD) video camera. The camera is interfaced with a dual-core microprocessor computer, through a fire-wire (IEEE-1394) cable for data processing and display. In addition, systems calibration is performed by using a Ba2SO4 calibration plate, in order to ensure device-independent and reproducible imaging. An embedded image registration algorithm is used in order to correct the misalignment of the images which is provoked by tissue contractions and micro-movements of patients.

DySIS imaging device records the aceto-whitening effect, capturing an image stack after the application of acetic acid solution to the cervix. Acetic acid solution is used in conventional colposcopic examination as a contrast agent. It interacts with the abnormal tissues, provoking transient alterations to their light scattering properties. The degree and the duration of the latter is associated with the degree of the lesion and called aceto-whitening effect. In Figure 3.1 depicted the evolution of the aceto-whitening effect for high grade tissue types.

After the application of the contrast agent, a series of images is captured successively during the evolution of aceto-whitening effect. From the green channel of the captured image stack, the Intensity of the back scattered light as a function time is extracted for
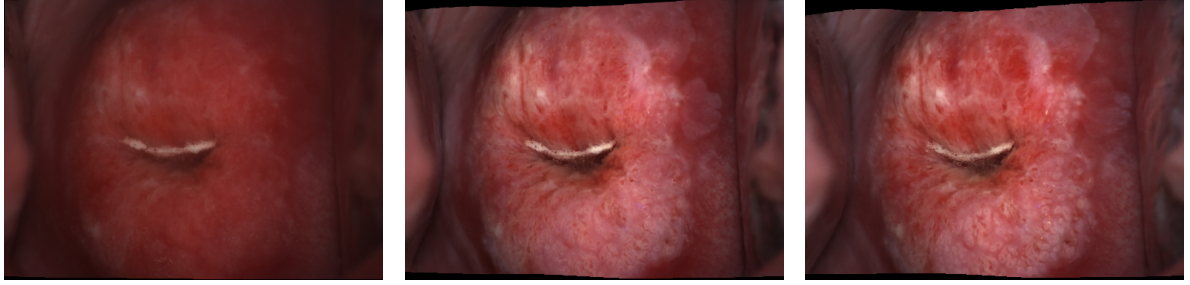
(a) Captured image at 7th sec   (b) Captured image at 70th sec   (c) Captured image at 175th sec

Figure 3.1: Depiction of acetowhitening effect at spesific time instances

every image pixel, expressing the temporal characteristics of the aceto-whitening effect. The green channel has been selected for the monitoring of the aceto-whitening effect due to its great dynamic range and high S/N ratio. The obtained image stack pixels, which represent a time sequence, can be processed in order to be calculated a series of parameters. After the processing of the curves a pseudo-color map is generated with different colors corresponding to different values of parameters, allowing the visualization of the spatial distribution of parameters which express quantitatively the aceto-whitening effect kinetics.

Tissue type	Patients	Curves	Percentage
Normal	12	72	19%
Inflammation	7	42	11%
HPV	8	65	18%
CIN1	9	50	13%
CIN2	5	26	7%
CIN3	22	110	30%
Cancer	1	6	2%
Total	64	371	100%

Table 3.1: Information about the curves of the first dataset

3.1 Data Description

In this study, we use two datasets. The first dataset consists of 371 Intensity of the Back Scattered Light (IBSL) vs time curves. These curves obtained from 64 patients who underwent colposcopic examination with automated capturing of an image stack in the clinics of Hammersmith Hospital and St Mary's Hospital in London, United Kingdom and Alexandra Hospital in Athens, Greece. More specifically, an image was captured before the application of acetic acid solution as reference image. After the application of 3% acetic acid solution, a series of images were captured automatically with a frequency of 1 image every 7 seconds from the 7th second to the 84th second and with a frequency of 1 image every 10 seconds until the 234th second. The final image was captured in the 240th second. Thus, each image stack (Figure 3.2) consists of 29 images and can be viewed as a 3-dimensional $M \times N \times K$ matrix , where M and N are the spatial dimensions of each image and K is the number of captured images. Each pixel of an image stack corresponds to a curve which represents the intensity of the backscattered light in different time points. From the available image stacks a total of 371 curves were extracted from 73 tissue points which were subsequently biopsied. Each tissue point corresponds approximately to five curves. The extracted curves represent seven tissue types, covering all the stages of the disease as well as the normal case (Figure 3.3). More specifically 72 curves correspond to Normal tissues, 42 curves to Inflammation, 65 curves to HPV, 50 curves to CIN1, 26 curves to CIN2, 110 curves to CIN3 and 6 curves of Cancer tissues. The above information about the curves is summarized into the Table 3.1.

As regards the second dataset, it consists of image stacks which are captured during the colposcopic examination. The image stacks are discriminated in three main categories depending on the stage of cervix neoplasia. More specifically, there are seven image stacks from high grade clinical cases, two image stacks from low grade clinical cases and two image stacks from normal clinical cases. In this case the image stack consists of fewer captured images, since the final images were captured in the 185th second. More specifically, after the application of acetic acid a series of images were captured automatically with a frequency of 1 image every 7 seconds from the 7th second to the 84th and with a frequency of 1 image every 10 seconds until the 185th second.

The image stacks can be viewed as an $M \times N \times K$ matrix where $M \times N$ the resolution of each image. In this case the resolution of the images is 768×1024, which means that
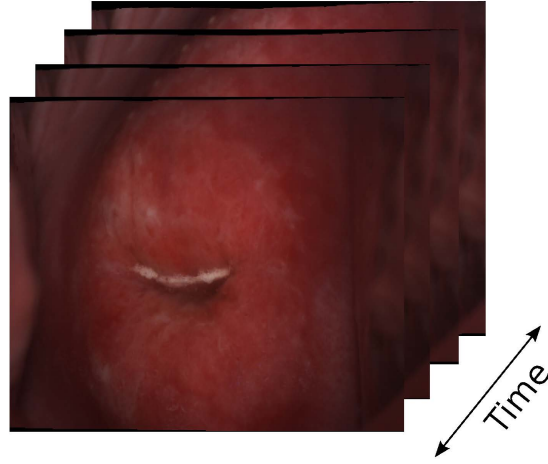
Figure 3.2: Image stack visualization
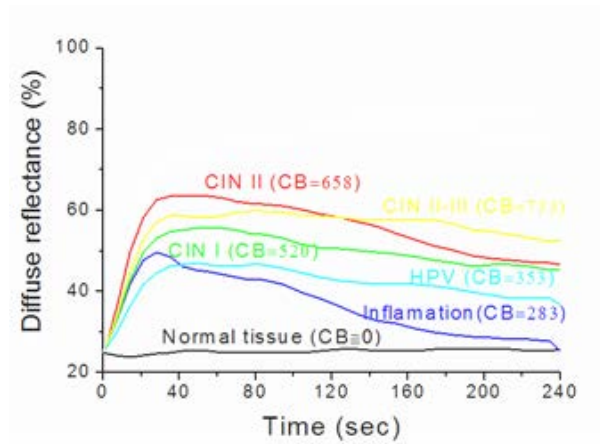
Figure 3.3: Characteristic curves of all the possible classes

there are 786.432 IBSL vs time extracted curves for each one of the above mentioned clinical cases.

3.2 Influential Work

The optical imaging techniques for cervical cancer screening require the development of image processing techniques to automate the interpretation of colposcopic image stream. In this section, we give a short overview of the most notable techniques of image processing used to assess the acetowhitening effect kinetics quantitatively and accurately.

Park et al. [16] proposed an diagnostic tool for the automated identification of cervical neoplasia from digital images. The image analysis was performed in two stages. Firstly, patterns with similar optical characteristics were clustered together More specifically, a variant of the k-means algorithm was used to segment the areas which present different characteristics. After the clustering, classification algorithms were used to discriminate the type of tissues(normal, low grade, high grade, cancer tissues) containing in the clustered areas. Five features, such as the intensity value of red, green and blue channels; the ratio of intensity of the green to red channel; and the changes in grayscale intensity values, were selected as the most relevant for the classification. Given these five features they proposed an ensemble classifier consisting of a linear classifier with Euclidean distance, a linear classifier with Mahalanobis distance, 7- nearest neighbor classifier and a support vector machine with a linear kernel. The proposed method reached a performance with sensitivity and specificity 79% and 88% to distinguish high grade from low grade and normal cases.

Li et al. [17] developed a system for automated diagnosis of cervical neoplasia using only two cervical images captured before and after the application of acetic acid respectively. The system calculated the opacity index to discriminate high grade cases from low grade and normal cases. They determine the opacity index by subtracting the pre-acetic-acid from the post-acetic-acid image and applying clustering algorithms. The sensitivity and specificity of the proposed method is 94% and 87% respectively.

Acosta-Mesa et al. [18] proposed a simple classification approach to discriminate automatically normal and abnormal tissues based on the intensity value of each pixel over time. Firstly, the intensity value over time was fitted using a polynomial function. Then, a k-NN classification algorithm was used over the entire length of the aceto-whitening temporal patterns. Euclidean distance was used as similarity distance and various values of k was examined. For $k = 20$ the algorithm had the better performance with sensitivity and specificity of 71% and 59%.

Acosta-Mesa et al. [19] proposed a second classification strategy to classify the tissue type based on the intensity value over time. Firstly, they reduced the dimensionality of temporal curves calculating approximations of them through discretization techniques. Piecewise Linear Approximation (PLA) and Piecewise Slope Approximation (PSA) were the two discretization techniques which were used. After the discretization stage, the PLA and PSA representation were used in combination with a Naive Bays classifier. The PLA, PSA methods reached a performance with sensitivity 67%, 61% and specificity 76%, 70% respectively . In this pilot study the parameters for the discretization techniques were selected experimentally.

In a third study Fragoso et al. [20] presented the results from the two above mentioned approaches [18, 19]with a new classification approach. More specifically, they compared the performance of three classifcation methods, namely kNN, Naive Bayes and C4.5. In addition, in this work they used the PLA and PSA discretization techniques. However, in this study the parameters of the discretization techniques was optimized as a single parameter using evolutionary programming. The sensitivity of the proposed techniques was between $53\% - 71\%$ while the specificity was between $59\% - 80\%$. Assessing the results authors supported that the low performance due to the similarity of normal and abnormal epithelium as regards the temporal pattern. Thus, they proposed the combination between temporal patterns and characteristics of the cervical epithelium to perform better classification performance.

Wu et al. [21] studied the performance of multivariate statistics algorithms in terms of their ability to separate accurately the different types of cervical tissues based on acetowhitening effect kinetics. More specifically, they study the classification performance of Principal Component Analysis (PCA) and Support Vector Machine (SVM) for discrimination between CIN lesions based on the kinetics of acetowhitening process. PCA was used to reduce the dimension of the acetowhitening kinetics curves. More specifically, they kept the most informative PCs and projected the original onto the feature space, reducing the dimensionality of the data. The new data representation was used by a SVM classifier with a radial basis (RBF) kernel. They also provided results from the SVM algorithm using the original data. Applying these two classification approaches, they conducted two trials. In the first trial, they focused on the separation of CIN lesions and non-CIN tissues. In the second trial they focused on the separation of high-grade (CIN2/3) lesions form other tissue types. In the first trial both of the two classification

approaches achieved similar performance with a sensitivity of 95% and a specificity of 96%. In the second trial the performance of the classification approach based on the original data is slightly better than the one based on the two PC scores. More specifically, the sensitivity is 91% for both of the two classification approaches, but the specificity is 90%and 87% respectively, implying the new data representation led to loss of information.

Park et al. [22] proposed a domain-specific automated image analysis framework for the discrimination of the several tissue types of the cervix. In this study, they include domain-specific diagnostic features in a probabilistic manner using conditional random fields (CRF). They proposed a conditional random field model which includes both the optical characteristics of each tissue type and the diagnostic relationships between neighboring regions. Firstly, they design a k-means algorithm to extract anatomical feature and segment the image into tissue type regions. Given the diagnostic features extracted from each segmented image region, the classify the tissue in each region as normal or abnormal using an automated image classification algorithm using a CRF-based classier. The diagnostic performance of the proposed approach reached a performance with sensitivity 70% and specificity 80% in detecting neoplastic tissues.

Chapter 4

Our Approach

The main purpose of this study is the design of a model which will provide objective discrimination between the various stages of cervical neoplasia. Thus, we have to evaluate the impact of the feature extraction methods, described in Section 2.1, on the performance of a 1-NN classifier in order to explore which of them is more beneficial for the classification problem.

We examine two major cases and evaluate the impact of the above mentioned feature extraction methods. In the first case, the performance of the 1-NN classifier using the entire feature set is evaluated while in the second case the performance using a subset of the entire feature set is examined.

In order to examine the impact of the feature extraction methods we use the first dataset which was described in Section 3.1. As shown in the Table 3.1 the dataset includes 371 observations. Each observation is labeled and belongs to one of the seven classes mentioned in Section 3.1. In addition, each observation consists of 29 features, corresponding to the intensity of the back scattered light in specific time instances. The discrete wavelet transform operates on feature vectors whose length is an integer power of two. Thus, we apply cubic spline interpolation to original observations so that the new feature vectors contain 32 features. Furthermore by applying interpolation to the original curves we achieve the smoothing of them [14]. After the interpolation, each observation consists of 32 features. The feature extraction methods described above, namely WT, PCA, KPCA, PAA, APAA and SAX are applied to the first dataset and the sets of the extracted features are obtained. KPCA and SAX have a parameter which has to be selected properly. More specifically, for KPCA, we have to define the standard deviation

σ of the Gaussian kernel while for SAX we have to define the alphabet size. The optimal parameters are selected with respect to classification performance obtaining from the mean AUC score of 100 repeated 10 fold validation.

The classification performance of 1-nearest neighbor classifier is evaluated for various classification cases. The performance is measured in terms of accuracy, sensitivity, specificity and Area Under the ROC Curve (AUC) using 10-fold cross validation. The final results for each classification task constitute the average results of 100 repeated 10-fold cross validation.

After the performance evaluation of the feature extraction methods using the entire feature set, we examine the performance of the methods using a subset of the entire set of features. Keeping a subset of the entire set of features, we achieve to reduce the computational cost maintaining high performance.

Margariti [14] showed that a subset of the WT can achieve similar performance with the one obtained by using the entire set of coefficients. More specifically, the results of her study showed that we can achieve same performance by keeping only 5 of the 32 coefficients. Based on these results, we keep 5 coefficients of the entire feature sets and compare the performance of the feature extraction methods. As regards the other feature extraction methods, the feature subset consists also of 5 features for fairness reasons.

In order to select a subset of the Wavelet Coefficients (WCs), Sequential Backward Elimination (SBE) is used [14]. SBE is a commonly used search method in feature subset selection [7]. The selection procedure starts with the entire set of features and at each step the sub-optimal subset is found by removing the feature whose absent least decreases a performance measure p. The procedure terminates when a specific criterion is satisfied. In this case, AUC score constitutes the performance measure p and the termination criterion is the number of features to be 5. The AUC score for each feature subset is the mean AUC value of 10 repeated 10-fold-validation. In each step of (SBE), the algorithm drops the feature which yields the lower mean AUC score. Consequently, for each of 13 classification cases which examined before, a feature subset consisting of 5 WCs is extracted using the (SBE) algorithm and the performance of the method is evaluated. Regarding the PCA and KPCA, the percentage of retained variance among the data constitutes the criterion for keeping the most important principal components. More specifically, it is known that the first principal components retain as much of the variance in the data as possible, thus the first five principal components are kept. The

retained variance keeping 5 principal components is approximately 99% for each examined classification case. In PAA and SAX, we set the segments number of the new time series representation to 5. In PAA and SAX the new 5-dimensional time series representation is approximated by segmenting the sequences into 5 equal-length sections and recording the mean value of these sections. However, in SAX a further transformation is applied to obtain a discrete representation of the time series. Finally, in APAA we define the segments number of the new time series representation to 3. In APAA two numbers are required per segment, which mean that the new time series representation consist of 6 features. However, subtracking the last feature which represents the right endpoint of each segment and is the same for each time series, we obtain 5 features.

So far, we have selected the number of retained PCs of PCA and KPCA based on the number of selected features. Next, we will examine whether we can use less features achieving the same performance. In addition, we will examine the relation between the percentage of total variation that the selected PCs contribute and the classification performance. We assume that the most discriminative information is captured by the largest variance in the feature space. Since the direction of the largest variance encodes the most information this is likely to be true. However, there are cases where the discriminative information actually resides in the directions of the smallest variance, such that PCA could greatly hurt classification performance. In this case, the classification performance increase as the total variation increase.

Furthermore, we present another approach which tries to improve the classification performance of k-NN combining PCA, KPCA and WCs. The basic idea behind combination of PCs and WCs is rather intuitive. The added 5 PCs exported form both PCA and KPCA are selected so that they cover most variance and present high AUC score. The added 5 WCs are selected so that they present high AUC score for each classification case.

In order to evaluate the performance of this approach the below procedure is followed. The dataset is divided into training and test sets. PCA, KPCA and WT are applied independently to training data, producing PCA, KPCA and WT features, correspondingly. Then, the original data in training set is mapped onto the lower-dimensional spaces and the WCs are selected. Thus, three transformed training sets are produced, two of which contains PCs instead of original features and the other selected WCs. Then, these trans-

formed datasets are merged, so that the resulting training sets contains PCA, KPCA, WT features and class labe from the original training set.

After the combination of the three different types of features, we examine the performance of the above approach using a Subset of the entire set of Combined Feature (SCF). Firstly, we follow the procedure which was described above to extract a set of PCs and WCs. Then, we try to extract a subset of the entire set of features achieving similar classification performance with the one obtained by using the entire set of coefficients.

To be able to evaluate the learnt classifier, the test set should also be transformed to the same format. This is done in a similar way, so that test set is transformed independently with PCA, KPCA and WT models, and PCs and WCs are constructed. Then, PCs are merged with WCs and class labels from the original training set.

In order to select a subset of the entire set of PCs and WCs, SBE is used. Thus, for each of 13 classification cases which were examined before, a feature subset consisting of 5 SCF is extracted using the SBE algorithm and the performance of the method is evaluated.

As we mention above, the main purpose of this study is to find a classification model which will discriminate the various tissue types of cervix, based on information provided by an image stack and will generate a pseudo-color map with different colors representing different tissue types. Such a classification model has to offer a time-efficient way to classify objectively the various tissue types. Thus, we present the execution time of classification using each of the four feature extraction methods which present high classification performance. The results have been generated on a quad core processor of 2.4 Ghz and with 6 GB RAM memory. Table 5.33 shows results related to the execution time of each methods. It is noted that PCA presents better performance than the other methods.

Chapter 5

Performance Evaluation

5.1 Performance Evaluation using the Entire Set of Features

In this section, we present the impact of the feature extraction to the performance to 1-NN classifier. A 1-NN is applied at the entire set of features extracted using the above mention feature extraction methods and the performance of classification is evaluated.

In Tables 5.1- 5.13 the results related to the performance of each feature extraction method for various classification cases is presented. The results indicate that using the entire extracted feature set all the proposed methods present similar performance expect for the SAX which gives the least promising results. More specifically, SAX method has an AUC score $10\% - 20\%$ lower than the other feature extraction methods for various classification cases. The other feature extraction methods present similar AUC score as using the raw data. In most classification cases the AUC score is quite high, however there some classification cases where the discrimination of the various classes is quite difficult.

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	96.79(0.46)	96.90(0.46)	95.25(0.67)	98.32(0.65)
Wavelet	**32**	96.82(0.46)	96.92(0.46)	95.41(0.58)	98.22(0.64)
PCA	**32**	96.85(0.47)	96.95(0.48)	95.38(0.59)	98.31(0.67)
KPCA	**32**	96.57(0.44)	96.64(0.45)	95.46(0.56)	97.66(0.75)
PAA	**16**	96.32(0.53)	96.39(0.52)	95.22(0.70)	97.41(0.65)
APAA	**32**	96.16(0.45)	96.20(0.46)	95.68(0.58)	96.66(0.69)
SAX	**16**	69.89(1.00)	70.42(0.98)	62.36(1.36)	77.40(1.21)

Table 5.1: Classification between High versus Low grade of neoplasia

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	97.04(0.92)	96.97(0.91)	97.54(1.17)	96.54(1.28)
Wavelet	**32**	96.96(0.87)	96.91(0.83)	97.38(1.41)	96.55(1.07)
PCA	**32**	97.01(0.91)	96.96(0.89)	97.46(1.27)	96.57(1.02)
KPCA	**32**	97.18(0.72)	97.16(0.66)	97.52(1.28)	96.88(0.67)
PAA	**16**	96.96(0.90)	96.90(0.94)	97.44(1.17)	96.48(1.30)
APAA	**32**	95.15(1.03)	95.34(1.01)	93.62(1.52)	96.66(1.14)
SAX	**16**	78.83(1.45)	79.29(1.50)	75.32(1.97)	82.34(2.06)

Table 5.2: Classification between CIN1 versus HPV

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	85.07(1.68)	85.28(1.59)	87.42(1.80)	82.74(2.76)
Wavelet	**32**	85.15(1.45)	85.34(1.45)	87.40(1.67)	82.88(2.36)
PCA	**32**	85.37(1.39)	85.55(1.33)	87.66(1.48)	83.05(2.09
KPCA	**32**	84.91(1.85)	85.09(1.74)	87.08(2.33)	82.71(2.74)
PAA	**16**	84.43(1.48)	84.57(1.42)	85.80(1.67)	83.10(2.28)
APAA	**32**	88.92(1.40)	89.00(1.44)	89.52(2.07)	88.38(1.95)
SAX	**16**	79.90(1.65)	79.67(1.61)	77.58(2.04)	82.17(2.61)

Table 5.3: Classification between CIN1 versus Inflammation

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	95.92(1.32)	96.26(1.25)	97.57(1.59)	94.24(2.09)
Wavelet	**32**	95.97(1.28)	96.29(1.20)	97.46(1.42)	94.48(1.88)
PCA	**32**	95.92(1.38)	96.26(1.31)	97.58(1.58)	94.21(2.12)
KPCA	**32**	95.72(1.52)	96.11(1.43)	97.63(1.57)	93.76(2.53)
PAA	**16**	92.46(1.42)	92.90(1.39)	94.58(1.81)	90.29(2.16)
APAA	**32**	93.77(1.20)	93.70(0.97)	93.49(0.78)	94.02(2.40)
SAX	**16**	85.33(1.25)	84.17(1.30)	80.00(1.78)	90.62(1.06)

Table 5.4: Classification between HPV versus Inflammation

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	90.09(1.03)	90.18(0.95)	90.61(0.95)	89.56(1.67)
Wavelet	**32**	90.26(1.05)	90.34(0.99)	90.67(0.99)	89.88(1.60)
PCA	**32**	90.01(1.15)	90.12(1.04)	90.62(0.89)	89.40(1.87)
KPCA	**32**	90.26(1.01)	90.29(0.95)	90.44(0.97)	90.06(1.59)
PAA	**16**	90.87(1.08)	90.78(1.02)	90.44(1.07)	91.26(1.77)
APAA	**32**	90.19(1.21)	89.99(1.11)	89.06(1.10)	91.34(2.03)
SAX	**16**	79.44(1.33)	78.63(1.31)	74.99(1.76)	83.88(1.92)

Table 5.5: Classification between Normal versus CIN1

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	84.51(1.00)	84.52(0.97)	81.50(1.44)	87.50(1.31)
Wavelet	**32**	84.62(1.07)	84.61(1.07)	81.54(1.43)	87.65(1.51)
PCA	**32**	84.58(1.18)	84.60(1.17)	81.57(1.37)	87.61(1.68)
KPCA	**32**	84.56(1.08)	84.58(1.09)	81.66(1.35)	87.48(1.60)
PAA	**16**	82.56(1.11)	82.57(1.09)	79.96(1.39)	85.15(1.59)
APAA	**32**	83.61(1.26)	83.63(1.25)	80.48(1.89)	86.75(1.39)
SAX	**16**	74.29(1.29)	74.33(1.29)	62.49(1.88)	86.07(1.85)

Table 5.6: Classification between Normal-Inflammation vs HPV-CIN1

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	97.39(0.64)	97.22(0.45)	97.01(0.36)	97.76(1.15)
Wavelet	**32**	97.42(0.60)	97.22(0.45)	96.99(0.38)	97.84(1.01)
PCA	**32**	97.46(0.65)	97.27(0.48)	97.05(0.41)	97.86(1.15)
KPCA	**32**	97.38(0.57)	97.19(0.40)	96.96(0.36)	97.82(1.03)
PAA	**16**	96.24(0.69)	96.56(0.44)	96.95(0.40)	95.52(1.34)
APAA	**32**	96.52(0.59)	96.85(0.44)	97.24(0.43)	95.80(1.08)
SAX	**16**	70.96(1.52)	73.51(1.42)	76.48(1.67)	65.44(2.46)

Table 5.7: Classification between CIN2/3 vs CIN1

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	88.42(1.08)	92.41(0.62)	81.92(1.98)	94.94(0.51)
Wavelet	**32**	88.26(1.06)	92.32(0.57)	81.65(2.02)	94.89(0.48)
PCA	**32**	88.25(0.98)	92.27(0.49)	81.69(1.97)	94.82(0.42)
KPCA	**32**	88.28(1.01)	92.34(0.61)	81.65(1.95)	94.91(0.56)
PAA	**16**	87.75(1.13)	92.24(0.56)	80.42(2.22)	95.09(0.50)
APAA	**32**	85.35(1.15)	91.08(0.57)	75.99(2.28)	94.71(0.48)
SAX	**16**	66.04(1.10)	83.71(0.56)	37.15(2.13)	94.92(0.48)

Table 5.8: Classification between Normal versus all the others classes

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	82.36(1.69)	94.30(0.54)	66.83(3.20)	97.81(0.45)
Wavelet	**32**	82.39(1.75)	94.33(0.62)	67.02(3.20)	97.81(0.49)
PCA	**32**	82.13(1.71)	94.22(0.57)	66.48(3.33)	97.77(0.44)
KPCA	**32**	82.11(1.46)	94.25(0.58)	66.36(2.74)	97.81(0.49)
PAA	**16**	80.65(1.60)	93.31(0.62)	64.24(3.01)	97.02(0.49)
APAA	**32**	85.03(1.35)	94.80(0.53)	72.45(2.61)	97.65(0.44)
SAX	**16**	69.62(1.26)	90.09(0.41)	43.19(2.46)	96.08(0.36)

Table 5.9: Classification between Inflammation versus all the others classes

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	92.30(1.35)	95.07(0.64)	88.02(2.59)	96.57(0.48)
Wavelet	**32**	92.02(1.16)	94.94(0.56)	87.52(2.23)	96.52(0.45)
PCA	**32**	91.94(1.19)	94.92(0.56)	87.38(2.24)	96.53(0.44)
KPCA	**32**	92.15(1.24)	95.00(0.62)	87.68(2.35)	96.56(0.47)
PAA	**16**	89.70(1.14)	93.70(0.62)	83.51(2.14)	95.87(0.58)
APAA	**32**	87.79(0.96)	92.20(0.60)	80.98(1.82)	94.58(0.61)
SAX	**16**	73.54(1.70)	88.10(0.64)	51.08(3.32)	95.96(0.44)

Table 5.10: Classification between HPV versus all the others classes

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	85.59(1.52)	92.01(0.59)	76.80(2.97)	94.37(0.46)
Wavelet	**32**	85.54(1.29)	92.05(0.52)	76.62(2.49)	94.46(0.42)
PCA	**32**	85.54(1.48)	91.99(0.57)	76.70(2.92)	94.38(0.47)
KPCA	**32**	85.50(1.32)	92.05(0.50)	76.54(2.63)	94.47(0.43)
PAA	**16**	85.32(1.30)	91.86(0.56)	76.38(2.52)	94.27(0.49)
APAA	**32**	86.32(1.51)	93.29(0.57)	76.78(2.99)	95.86(0.46)
SAX	**16**	65.71(1.45)	87.15(0.67)	36.36(2.76)	95.06(0.59)

Table 5.11: Classification between CIN1 versus all the others classes

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	81.84(1.47)	95.72(0.41)	65.88(2.54)	97.97(0.37)
Wavelet	**32**	81.89(1.81)	95.76(0.48)	65.77(3.35)	98.02(0.42)
PCA	**32**	82.09(1.76)	95.71(0.41)	66.27(2.73)	97.93(0.42)
KPCA	**32**	82.13(1.61)	95.70(0.47)	66.42(2.73)	97.91(0.44)
PAA	**16**	81.30(1.97)	95.19(0.46)	65.15(3.62)	97.46(0.44)
APAA	**32**	81.27(1.69)	94.99(0.46)	65.35(2.81)	97.22(0.43)
SAX	**16**	64.08(2.24)	91.59(0.53)	32.19(4.00)	96.07(0.43)

Table 5.12: Classification between CIN2 versus all the others classes

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	91.39(0.67)	93.00(0.55)	87.43(1.21)	95.35(0.60)
Wavelet	**32**	91.49(0.68)	93.12(0.54)	87.49(1.21)	95.50(0.53)
PCA	**32**	91.44(0.75)	93.10(0.56)	87.35(1.41)	95.52(0.51)
KPCA	**32**	91.23(0.78)	92.82(0.57)	87.32(1.49)	95.13(0.55)
PAA	**16**	90.67(0.87)	92.44(0.66)	86.33(1.57)	95.02(0.60)
APAA	**32**	91.40(0.76)	92.94(0.58)	87.62(1.41)	95.19(0.54)
SAX	**16**	71.82(1.24)	76.35(0.97)	60.67(2.16)	82.96(0.86)

Table 5.13: Classification between CIN3 versus all the others classes

5.2 Performance Evaluation using a Subset of Features

In the Tables 5.14- 5.26 the results related to the performance of a 1-NN in combination to specific feature extraction methods for various classification tasks is presented. The results indicate that using a subset of the entire feature set, Wt, PCA and KPCA present similar performance. Same performance is achieved using the entire feature set. APAA and PAA present slightly worse performance while SAX present the worst performance having an AUC score $10\% - 20\%$ lower than the others.

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	96.83(0.43)	96.94(0.44)	95.33(0.59)	98.32(0.64)
Wavelet	**5**	95.54(0.59)	95.55(0.60)	95.28(0.74)	95.79(0.89)
PCA	**5**	96.20(0.38)	96.26(0.39)	95.26(0.55)	97.12(0.60)
KPCA	**5**	95.80(0.59)	95.84(0.61)	95.21(0.63)	96.38(0.88)
PAA	**5**	92.68(0.54)	92.60(0.55)	93.78(0.76)	91.59(0.84)
APAA	**5**	92.82(0.67)	92.78(0.67)	93.22(0.76)	92.41(0.96)
SAX	**5**	71.23(1.16)	72.26(1.12)	56.76(1.93)	85.69(1.34)

Table 5.14: Classification between High versus Low grade of neoplasia

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	97.04(0.83)	96.98(0.83)	97.56(1.26)	96.54(1.19)
Wavelet	**5**	96.72(0.95)	96.64(0.92)	97.20(1.56)	96.22(1.10)
PCA	**5**	96.86(0.80)	96.83(0.75)	97.16(1.45)	96.57(0.69)
KPCA	**5**	96.09(0.83)	96.13(0.79)	95.66(1.33)	96.49(1.03)
PAA	**5**	96.86(0.93)	96.78(0.93)	97.40(1.35)	96.31(1.24)
APAA	**5**	95.34(0.86)	95.53(0.78)	93.78(1.61)	96.88(0.71)
SAX	**5**	75.84(1.21)	76.36(1.24)	71.80(1.76)	79.86(1.8)7

Table 5.15: Classification between CIN1 versus HPV

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	85.30(1.51)	85.45(1.47)	87.48(1.79)	83.02(2.57)
Wavelet	5	84.79(1.82)	84.78(1.77)	85.02(2.63)	84.50(2.90)
PCA	5	85.10(1.59)	85.27(1.57)	87.40(1.85)	82.74(2.47)
KPCA	5	84.91(1.63)	85.11(1.67)	87.14(2.38)	82.69(2.03)
PAA	5	82.57(1.35)	82.70(1.35)	84.24(1.71)	80.86(2.03)
APAA	5	81.42(1.75)	81.51(1.74)	82.24(2.11)	80.64(2.61)
SAX	5	68.37(2.20)	67.70(2.33)	60.90(3.81)	75.79(1.89)

Table 5.16: Classification between CIN1 versus Inflammation

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	95.95(1.35)	96.24(1.35)	97.40(1.73)	94.45(1.92)
Wavelet	5	94.97(1.29)	95.50(1.25)	97.42(1.54)	92.55(1.71)
PCA	5	96.21(1.29)	96.54(1.22)	97.82(1.32)	94.57(2.14)
KPCA	5	94.88(1.43)	95.04(1.43)	95.60(1.96)	94.17(1.93)
PAA	5	90.74(1.47)	91.31(1.37)	93.34(1.56)	88.17(1.96)
APAA	5	88.04(1.14)	88.55(1.14)	90.42(1.53)	85.67(1.79)
SAX	5	83.83(1.21)	83.28(1.16)	81.40(1.24)	86.19(2.00)

Table 5.17: Classification between HPV versus Inflammation

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	90.16(1.03)	90.22(0.96)	90.58(0.94)	89.70(1.57)
Wavelet	5	91.05(1.32)	91.12(1.21)	91.49(1.11)	90.60(2.28)
PCA	5	89.85(0.87)	89.93(0.82)	90.33(0.99)	89.36(1.45)
KPCA	5	90.39(0.92)	90.57(0.86)	91.31(1.02)	89.50(1.54)
PAA	5	87.67(1.25)	87.66(1.17)	87.64(1.16)	87.70(2.00)
APAA	5	86.61(1.30)	87.23(1.26)	90.06(1.59)	83.16(2.15)
SAX	5	67.91(1.73)	65.14(1.74)	52.51(2.46)	83.32(2.51)

Table 5.18: Classification between Normal versus CIN1

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	84.60(1.03)	84.62(1.05)	81.58(1.48)	87.63(1.48)
Wavelet	5	85.33(0.90)	85.31(0.92)	86.80(1.24)	83.84(1.11)
PCA	5	83.39(1.07)	83.40(1.05)	80.74(1.16)	86.04(1.68)
KPCA	5	82.87(1.19)	82.88(1.17)	81.16(1.21)	84.59(1.87)
PAA	5	79.86(1.06)	79.86(1.06)	76.76(1.44)	82.93(1.57)
APAA	5	78.34(1.19)	78.31(1.19)	79.35(1.62)	77.29(1.83)
SAX	5	67.57(1.23)	67.61(1.21)	51.89(1.69)	83.19(1.73)

Table 5.19: Classification between Normal-Inflammation vs HPV-CIN1

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	97.31(0.69)	97.17(0.48)	97.01(0.37)	97.60(1.24)
Wavelet	5	97.55(0.52)	97.61(0.33)	97.68(0.31)	97.42(1.04)
PCA	5	96.38(0.50)	96.62(0.35)	96.90(0.34)	95.86(0.95)
KPCA	5	96.35(0.77)	96.78(0.56)	97.29(0.52)	95.42(1.40)
PAA	5	94.14(0.84)	95.15(0.52)	96.33(0.41)	91.94(1.62)
APAA	5	92.15(0.72)	93.97(0.59)	96.10(0.59)	88.20(1.19)
SAX	5	76.87(1.78)	76.32(1.34)	75.68(1.41)	78.04(3.34)

Table 5.20: Classification between CIN2/3 vs CIN1

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	88.19(1.14)	92.27(0.63)	81.53(2.13)	94.86(0.47)
Wavelet	5	87.92(0.99)	93.02(0.54)	79.58(1.88)	96.25(0.49)
PCA	5	88.37(1.03)	92.26(0.59)	81.97(1.97)	94.74(0.54)
KPCA	5	87.81(1.09)	91.50(0.59)	81.75(2.01)	93.85(0.47)
PAA	5	83.95(0.99)	90.13(0.56)	73.85(1.85)	94.05(0.51)
APAA	5	84.08(1.39)	89.38(0.70)	75.44(2.63)	92.74(0.53)
SAX	5	63.30(1.14)	82.27(0.65)	32.26(2.13)	94.31(0.60)

Table 5.21: Classification between Normal versus all the others classes

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	82.23(1.53)	94.30(0.52)	66.45(2.91)	97.85(0.44)
Wavelet	**5**	88.18(1.56)	95.01(0.60)	79.29(2.98)	97.02(0.55)
PCA	**5**	82.38(1.65)	94.54(0.61)	66.71(3.17)	98.09(0.47)
KPCA	**5**	82.11(1.56)	94.51(0.51)	66.10(3.05)	98.13(0.43)
PAA	**5**	77.33(1.25)	92.64(0.45)	57.36(2.35)	97.14(0.42)
APAA	**5**	76.21(1.40)	91.16(0.61)	56.98(2.42)	95.52(0.60)
SAX	**5**	59.54(1.19)	88.33(0.43)	22.24(2.33)	96.76(0.38)

Table 5.22: Classification between Inflammation versus all the others classes

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	91.92(1.24)	94.89(0.58)	87.28(2.43)	96.51(0.46)
Wavelet	**5**	90.83(1.37)	94.35(0.62)	85.37(2.76)	96.25(0.44)
PCA	**5**	91.19(1.22)	94.44(0.68)	86.15(2.21)	96.20(0.57)
KPCA	**5**	90.88(1.20)	94.38(0.59)	85.43(2.33)	96.28(0.52)
PAA	**5**	87.13(1.19)	92.57(0.69)	78.82(2.23)	95.49(0.59)
APAA	**5**	85.44(1.34)	91.71(0.64)	75.80(2.54)	95.08(0.47)
SAX	**5**	67.54(1.31)	85.24(0.62)	40.31(2.56)	94.78(0.55)

Table 5.23: Classification between HPV versus all the others classes

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	85.66(1.36)	92.05(0.59)	76.92(2.58)	94.41(0.47)
Wavelet	**5**	86.36(1.51)	92.81(0.45)	77.54(3.09)	95.19(0.36)
PCA	**5**	84.12(1.49)	91.57(0.53)	73.92(2.97)	94.31(0.42)
KPCA	**5**	84.20(1.52)	91.55(0.56)	74.14(3.00)	94.26(0.42)
PAA	**5**	81.40(1.22)	90.02(0.52)	69.60(2.38)	93.20(0.47)
APAA	**5**	79.34(1.37)	90.55(0.55)	64.00(2.73)	94.68(0.50)
SAX	**5**	55.22(1.29)	84.22(0.60)	15.52(2.58)	94.93(0.62)

Table 5.24: Classification between CIN1 versus all the others classes

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	82.25(1.91)	95.75(0.42)	66.42(3.23)	97.97(0.40)
Wavelet	**5**	83.58(1.95)	95.72(0.53)	69.62(3.56)	97.68(0.48)
PCA	**5**	81.90(1.91)	95.25(0.44)	65.92(3.19)	97.46(0.40)
KPCA	**5**	81.04(2.08)	95.13(0.42)	64.81(3.84)	97.41(0.38)
PAA	**5**	77.86(2.14)	94.78(0.41)	58.35(3.28)	97.52(0.36)
APAA	**5**	79.86(2.03)	94.74(0.46)	62.65(3.72)	97.15(0.41)
SAX	**5**	53.10(1.14)	92.12(0.32)	7.81(2.01)	98.48(0.31)

Table 5.25: Classification between CIN2 versus all the others classes

	feat	AUC %(std)	Acc%(std)	Sens%(std)	Spec%(std)
none	**32**	91.29(0.75)	92.98(0.56)	87.14(1.44)	95.45(0.56)
Wavelet	**5**	89.80(0.70)	91.88(0.55)	84.70(1.24)	94.90(0.53)
PCA	**5**	90.76(0.70)	92.56(0.51)	86.33(1.32)	95.19(0.48)
KPCA	**5**	90.55(0.74)	92.25(0.57)	86.38(1.44)	94.72(0.65)
PAA	**5**	87.09(0.82)	88.64(0.67)	83.27(1.38)	90.90(0.62)
APAA	**5**	86.59(0.81)	88.75(0.67)	81.26(1.50)	91.91(0.75)
SAX	**5**	65.65(1.21)	76.56(0.88)	38.85(2.21)	92.46(0.70)

Table 5.26: Classification between CIN3 versus all the others classes

5.3 Relation between Retained Variance and Classification Performance

Tables 5.27- 5.30 present the classification performance and the percentage of total variation contribute retained Principal Components (PCs) extracted from PCA technique and KPCA respectively. More specifically, we present the AUC score for various classification cases the percentage of total variation contribute m retained PCs, where $m = 1 \ldots 5$. In addition, the AUC performance using all the available PCs is presented in the final column of the Tables. The results indicate that using the five first PCs we can achieve the same or even better performance as by using the entire feature set. In addition, it is

indicated that in some cases similar performance is achieved using the first three or four PCs.

Case	# PCs	Variation(%)	AUC(%)	AUC(%) (32 features)
	1	0.91	67.49	
High	**1,2**	0.97	78.48	
vs	**1,2,3**	0.99	94.26	96.76
Low	**1,2,3,4**	0.99	95.93	
	1,2,3,4,5	0.99	96.22	
	1	0.88	77.80	
CIN1	**1,2**	0.97	91.53	
vs	**1,2,3**	0.99	96.15	96.88
HPV	**1,2,3,4**	0.99	96.35	
	1,2,3,4,5	0.99	96.99	
	1	0.83	66.98	
CIN1	**1,2**	0.96	83.01	
vs	**1,2,3**	0.99	81.80	85.05
Inflammation	**1,2,3,4**	0.99	85.18	
	1,2,3,4,5	0.99	85.34	
	1	0.91	77.48	
HPV	**1,2**	0.97	91.64	
vs	**1,2,3**	0.99	88.94	96.09
Inflammation	**1,2,3,4**	0.99	95.05	
	1,2,3,4,5	0.99	96.16	
	1	0.92	70.95	
Normal	**1,2**	0.98	82.70	
vs	**1,2,3**	0.99	88.85	90.00
CIN1	**1,2,3,4**	0.99	90.24	
	1,2,3,4,5	0.99	89.70	
Normal	**1**	0.91	63.29	
Inflammation	**1,2**	0.98	75.09	
vs	**1,2,3**	0.99	78.80	84.78
HPV	**1,2,3,4**	0.99	83.37	
CIN1	**1,2,3,4,5**	0.99	83.62	
	1	0.91	82.37	
CIN2/3	**1,2**	0.98	90.86	
vs	**1,2,3**	0.99	93.61	97.37
CIN1	**1,2,3,4**	0.99	96.44	
	1,2,3,4,5	0.99	96.46	

Table 5.27: Retained PCs using PCA and AUC score for various Classification Cases

Case	# PCs	Variation(%)	AUC(%)	AUC(%) (32 features)
	1	0.90	53.10	
Normal	**1,2**	0.97	69.87	
vs	**1,2,3**	0.99	86.44	88.17
All	**1,2,3,4**	0.99	88.59	
	1,2,3,4,5	0.99	88.45	
	1	0.90	61.50	
Inflammation	**1,2**	0.97	74.12	
vs	**1,2,3**	0.99	77.70	82.35
All	**1,2,3,4**	0.99	80.90	
	1,2,3,4,5	0.99	82.28	
	1	0.90	61.04	
HPV	**1,2**	0.97	76.97	
vs	**1,2,3**	0.99	86.36	92.01
All	**1,2,3,4**	0.99	90.45	
	1,2,3,4,5	0.99	90.91	
	1	0.90	66.24	
CIN1	**1,2**	0.97	75.58	
vs	**1,2,3**	0.99	78.86	85.51
All	**1,2,3,4**	0.99	84.19	
	1,2,3,4,5	0.99	84.46	
	1	0.90	61.40	
CIN2	**1,2**	0.97	69.02	
vs	**1,2,3**	0.99	79.79	81.82
All	**1,2,3,4**	0.99	81.72	
	1,2,3,4,5	0.99	82.15	
	1	0.90	66.24	
CIN3	**1,2**	0.97	72.86	
vs	**1,2,3**	0.99	88.57	91.41
All	**1,2,3,4**	0.99	90.71	
	1,2,3,4,5	0.99	90.62	

Table 5.28: Retained PCs using PCA and AUC score for One versus All Classification Cases

Case	# PCs	Variation(%)	AUC(%)	AUC(%) (32 features)
High **vs** **Low**	**1**	0.86	67.03	
	1,2	0.92	78.40	
	1,2,3	0.97	86.82	96.50
	1,2,3,4	0.99	94.14	
	1,2,3,4,5	0.99	95.83	
CIN1 **vs** **HPV**	**1**	0.85	77.41	
	1,2	0.94	90.36	
	1,2,3	0.98	95.76	97.04
	1,2,3,4	0.99	95.98	
	1,2,3,4,5	0.99	96.06	
CIN1 **vs** **Inflammation**	**1**	0.81	66.76	
	1,2	0.95	82.40	
	1,2,3	0.97	82.64	85.06
	1,2,3,4	0.98	81.93	
	1,2,3,4,5	0.99	84.75	
HPV **vs** **Inflammation**	**1**	0.87	77.31	
	1,2	0.94	91.95	
	1,2,3	0.98	89.50	95.91
	1,2,3,4	0.99	89.58	
	1,2,3,4,5	0.99	94.71	
Normal **vs** **CIN1**	**1**	0.89	70.96	
	1,2	0.95	83.02	
	1,2,3	0.97	82.43	90.24
	1,2,3,4	0.99	88.90	
	1,2,3,4,5	0.99	90.49	
Normal **Inflammation** **vs** **HPV** **CIN1**	**1**	0.88	64.13	
	1,2	0.95	75.09	
	1,2,3	0.98	76.31	84.60
	1,2,3,4	0.99	78.90	
	1,2,3,4,5	0.99	83.11	
CIN2/3 **vs** **CIN1**	**1**	0.86	81.58	
	1,2	0.94	90.72	
	1,2,3	0.98	90.76	97.34
	1,2,3,4	0.99	93.49	
	1,2,3,4,5	0.99	96.14	

Table 5.29: Retained PCs using KPCA and AUC score for various classification cases

Case	# PCs	Variation(%)	AUC(%)	AUC(%) (32 features)
Normal **vs** **All**	**1**	0.85	53.47	
	1,2	0.92	70.15	
	1,2,3	0.96	75.20	88.10
	1,2,3,4	0.99	86.17	
	1,2,3,4,5	0.99	87.74	
Inflammation **vs** **All**	**1**	0.90	60.98	
	1,2	0.97	74.00	
	1,2,3	0.99	77.58	82.16
	1,2,3,4	0.99	81.01	
	1,2,3,4,5	0.99	82.36	
HPV **vs** **All**	**1**	0.91	60.69	
	1,2	0.97	77.09	
	1,2,3	0.99	86.24	92.02
	1,2,3,4	0.99	90.39	
	1,2,3,4,5	0.99	91.10	
CIN1 **vs** **All**	**1**	0.85	67.15	
	1,2	0.92	76.11	
	1,2,3	0.96	77.72	85.63
	1,2,3,4	0.99	79.52	
	1,2,3,4,5	0.99	83.90	
CIN2 **vs** **All**	**1**	0.86	61.81	
	1,2	0.92	68.77	
	1,2,3	0.96	72.39	82.18
	1,2,3,4	0.99	79.69	
	1,2,3,4,5	0.99	81.42	
CIN3 **vs** **All**	**1**	0.85	66.89	
	1,2	0.92	72.65	
	1,2,3	0.96	80.19	91.24
	1,2,3,4	0.99	88.79	
	1,2,3,4,5	0.99	90.55	

Table 5.30: Retained PCs using KPCA and AUC score for "one vs all" classification cases

5.4   Combination of PCA, KPCA, and WT features

In Tables 5.31- 5.32 the results related to the performance of PCA, WT, SCF for various classification tasks is presented. The results indicate that using a subset of combined PCs and WTs we can achieve better classification performance than using individual feature of PCA and WT for various classification cases and especially for cases where the discrimination of various classes is quite difficult.

5.5   Execution Time

Table 5.33 shows results related to the execution time of each feature extraction methodm mentioned above. It is noted that PCA presents better performance.

Method	Execution Time
Wavelet	22 min
PCA	11.2 sec
KPCA	36.07 sec
PAA	70 sec
APAA	40 min

Table 5.33: Execution Time

45

Case	Feat	AUC	Accuracy	Sensitivity	Specificity
High	**SCF**	97.27(0.54)	97.26(0.56)	97.39(0.57)	97.15(0.84)
vs	**PCA**	96.25(0.38)	96.31(0.39)	95.35(0.49)	97.15(0.55)
Low	**Wavelet**	96.26(0.62)	96.33(0.62)	95.29(0.82)	97.24(0.86)
CIN1	**SCF**	97.04(0.74)	96.96(0.75)	97.60(0.94)	96.46(0.99)
vs	**PCA**	96.96(0.83)	96.92(0.78)	97.30(1.40)	96.63(0.61)
HPV	**Wavelet**	96.31(0.90)	96.16(0.81)	97.50(1.62)	95.12(0.88)
CIN1	**SCF**	90.75(2.12)	90.79(2.10)	91.26(2.46)	90.24(3.20)
vs	**PCA**	85.21(1.19)	85.38(1.20)	87.34(1.82)	83.05(1.80)
Inflammation	**Wavelet**	86.40(1.97)	87.42(1.93)	86.68(2.29)	86.12(3.14)
HPV	**SCF**	95.95(1.25)	96.30(1.27)	97.63(1.54)	94.24(1.83)
vs	**PCA**	95.84(1.21)	96.21(1.20)	97.55(1.59)	94.12(2.13)
Inflammation	**Wavelet**	94.97(1.16)	95.47(1.08)	97.45(1.40)	92.40(2.11)
Normal	**SCF**	92.54(1.18	92.80(1.18	94.00(1.41)	91.08(1.69)
vs	**PCA**	90.08(0.93)	90.16(0.86)	90.62(0.93)	89.48(1.57)
CIN1	**Wavelet**	91.14(1.17)	91.44(1.03)	92.76(0.93)	89.54(2.18)
Normal					
Inflammation	**SCF**	88.16(1.13)	88.18(1.14)	86.53(1.55)	89.82(1.58)
vs	**PCA**	85.44(1.19)	83.45(1.16)	85.71(1.42)	86.17(1.69)
HPV	**Wavelet**	86.18(0.94)	87.18(0.96)	82.68(1.32)	90.65(1.26)
CIN1					
CIN2/3	**SCF**	98.60(0.53)	98.13(0.43)	97.60(0.49)	99.60(0.98)
vs	**PCA**	96.43(0.64)	96.67(0.41)	96.96(0.30)	95.90(1.22)
CIN1	**Wavelet**	97.62(0.59)	97.65(0.36)	97.69(0.26)	97.54(1.13)

Table 5.31: Classification performance using WCs, PCs, SCF

Case	Feat	AUC	Accuracy	Sensitivity	Specificity
Normal	**SCF**	90.36(1.06)	93.37(0.63)	85.46(1.96)	95.27(0.57)
vs	**PCA**	87.54(0.84)	92.32(0.49)	80.29(1.56)	94.74(0.48)
All	**Wavelet**	88.08(1.12)	93.14(0.62)	79.76(2.12)	96.36(0.48)
Inflammation	**SCF**	90.27(1.40)	95.41(0.50)	83.57(2.78)	96.92(0.44)
vs	**PCA**	82.43(1.62)	94.55(0.56)	66.76(3.12)	98.10(0.45)
All	**Wavelet**	87.99(1.46)	94.97(0.52)	78.81(2.86)	97.04(0.47)
HPV	**SCF**	91.11(0.97)	94.61(0.47)	85.66(1.95)	96.51(0.46)
vs	**PCA**	91.04(1.13)	94.40(0.49)	85.83(2.30)	96.22(0.43)
All	**Wavelet**	91.22(1.06)	94.52(0.54)	86.09(1.98)	96.31(0.43)
CIN1	**SCF**	86.70(1.20)	93.03(0.50)	78.04(2.36)	95.36(0.44)
vs	**PCA**	84.14(1.41)	91.58(0.47)	73.94(2.88)	94.33(0.41)
All	**Wavelet**	84.73(1.43)	90.81(0.61)	76.40(2.74)	93.06(0.52)
CIN2	**SCF**	88.23(2.20)	95.28(0.48)	80.58(4.14)	96.77(0.40)
vs	**PCA**	81.59(1.95)	95.20(0.45)	66.12(3.36)	97.40(0.42)
All	**Wavelet**	82.79(2.16)	95.05(0.44)	68.69(3.92)	96.73(0.40)
CIN3	**SCF**	90.84(0.79)	92.43(0.62)	86.93(1.41)	94.74(0.61)
vs	**PCA**	90.72(0.76)	92.54(0.55)	86.25(1.45)	95.19(0.50)
All	**Wavelet**	89.05(0.81)	90.92(0.62)	84.47(1.53)	93.63(0.64)

Table 5.32: Classification performance using WCs, PCs, SCF

Chapter 6

Mapping and Visualization of the Results

In this Chapter we present maps generated after the classification process. The generated pseudo-color map with different colors representing different tissue types, is overlaid onto the color image of the cervix providing valuable information about the stage of cervical neoplasia.

Taking into account the classification performance of 1-NN classifier using each of the proposed feature extraction methods as well as the execution time of each method we select three of them, namely PCA, WT and the combination of PCs and WCs as described in Section 5.4, to generate pseudo-color maps representing the stages of disease. WT, PCA obtain similar classification performance. However the execution time of PCA is lower, which constitutes an important advantage of this method. As regards the combination of PCs and WCs, it presents slightly better results in terms of classification performance.

In order to generate pseudo-color maps, we use the dataset distributed in Chapter 3 and consists of images stacks. As it is referred in Chapter 3 after the application of acetic acid a series of images were captured automatically with a frequency of 1 image every 7 seconds from the 7th second to the 84th and with a frequency of 1 image every 10 seconds until the 185th second. The image stacks can be viewed as an $M \times N \times K$ matrix where $M \times N$ is the resolution of each image. In this case the resolution of the images is 768×1024, which means that there are 786.432 extracted curves from each image stack corresponding to clinical cases.

The generation of pseudo-color map which provides valuable information about the stage of cervical neoplasia requires the classification of 786.432 extracted curves which were described above. However, before the classification the training of the classifier is required. In the training stage we use the first dataset which was described in Chapter 3 and consists of 371 labeled curves. During the training stage a simple multiclass classification model is build constructing one 1-NN for each pair of classes. In our case, where the there are $c = 7$ different classes, $c(c-1)/2 = 21$ different 1-NN classifiers are trained to distinguish the samples of one class from the samples of another class. For each one of the 21 classifiers different features are extracted and used in the training stage. After the training, the classification stage follows. Each of the unknown patterns are classified from the different trained classifiers and the decision about the class win which the pattern belong is done according to the maximum voting.

Figures 6.5- 6.10 help us visualize the classification results through the mapping procedure. The available images that we use for the mapping depict one case of high grade neoplasia, one case of low grade neoplasia and one case of no evidence of disease. The visualization of classification for the rest clinical cases can be found in Appendix A. For each patient we will present six images, the first two frames are the images before and at 185 sec after the acetic acid application, respectively. The third and fourth image present the mapping for distinguishing each class versus all the others by using the five selected WT features and the five first PCs, respectively. The fifth image shows the mapping that is produced by the combination of WCs, PCs. Finally, the last image illustrates the mapping that is produced by current method.

Figure 6.1 illustrates the color used to describe each class. More precisely, when the class is Normal no color is used for the mapping of the pixel. In the Inflammation case the color blue is used, while in the HPV the cyan. CIN1 is mapped with green color, CIN2 with yellow, CIN3 with orange and finally Cancer with red. Figure illustrates The colormap used by DySIS device is slightly different than in our case. More precisely, for the pixels that are classified as normal, no color or blue is used, for the low grade cases the colors green and red are chosen (red represents the more severe cases) and for the high grade yellow and white (white represents the more severe cases).

The visualization of classification for clinical cases verifies the high performance of the proposed methods. More specifically, patient 71 was diagnosed with high grade cervical neoplasia. Figures 6.3 and 6.4 indicate that the three mapping strategies present the
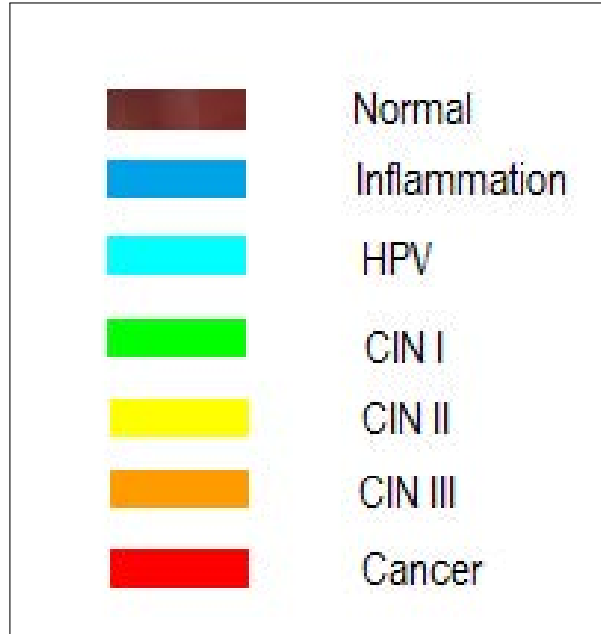
Figure 6.1: Legend of colormap

same behavior, classifying correctly the critical regions. A few differences between them can been seen in the classification of some non-biopsy areas. However, they are classified as low grade cases by each strategy.

Patient 77 was diagnosed with normal grade cervical neoplasia. Figures 6.5 and 6.6 show that the three mapping strategies present high performance since each one of them classifies all the image pixels as low grade.

Patient 12 does not present any evidence of disease. Figures 6.8 and 6.9 illustrate the fact that the tree classification strategies classify the majority of the pixels as normal. However there is a region where the tissues areclassified as HPV using 5 WCs and as Inflammation using 5 PCs.
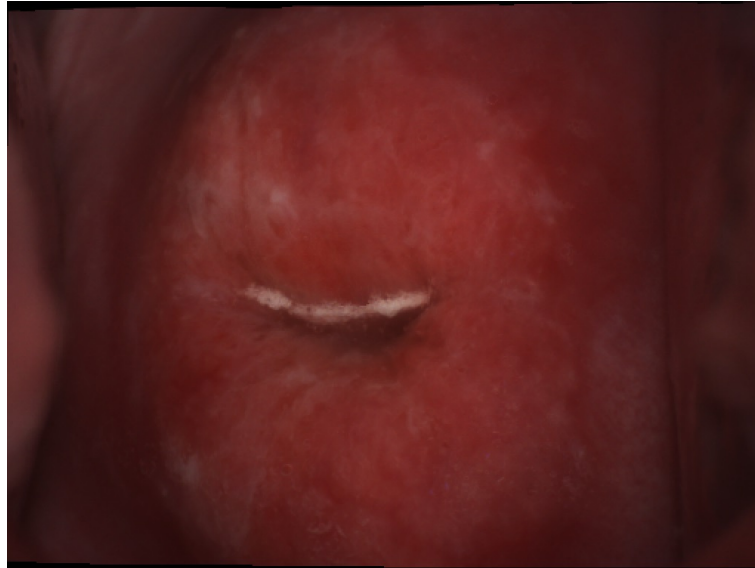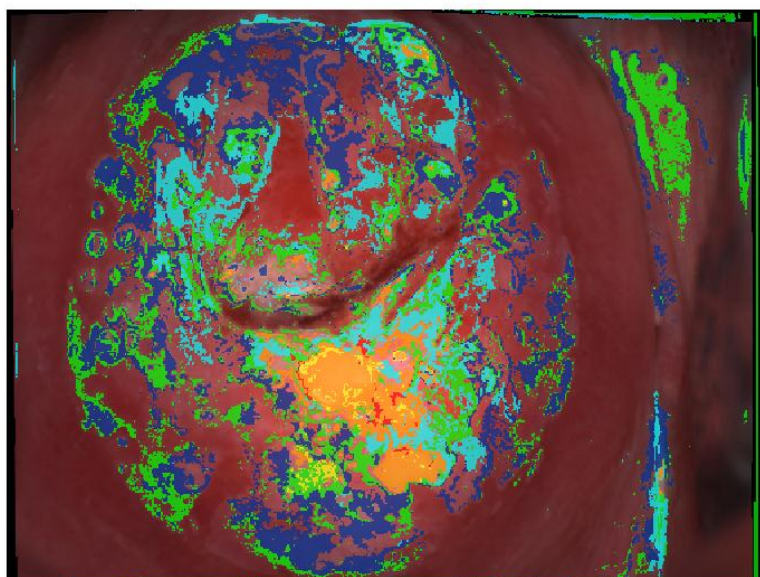
(a) Pre acetic acid image
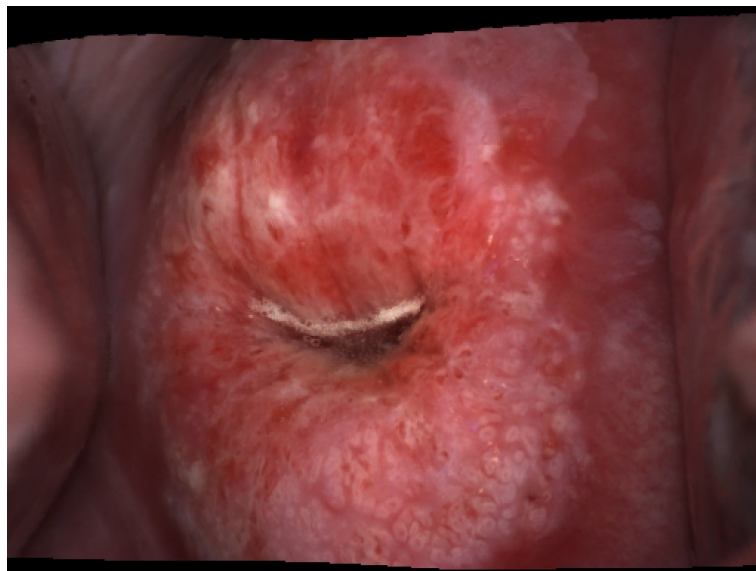
(b) Post acetic acid image

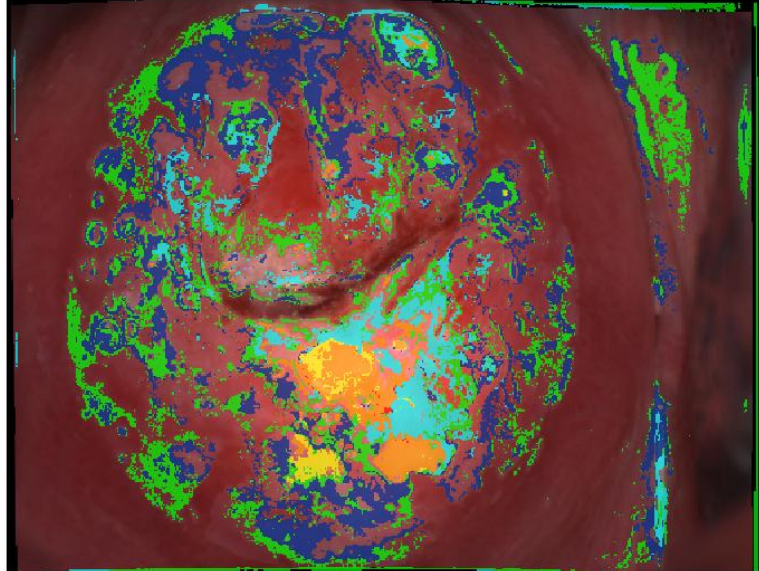Figure 6.2: Pre and Post acetic acid image for patient 71 with high grade cervical lesion
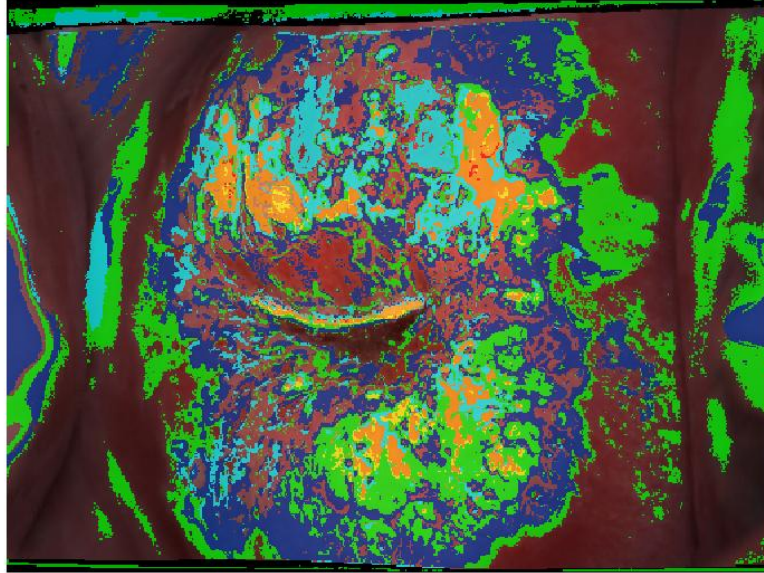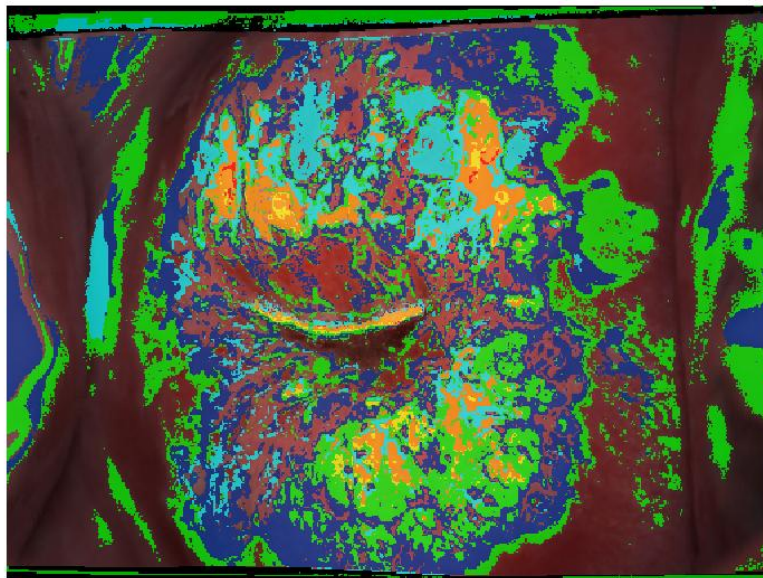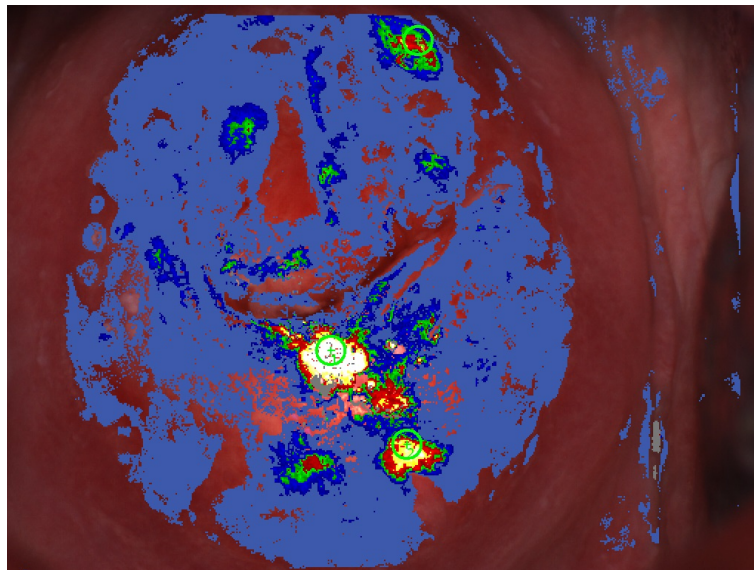
(a) Mapping with 5 WCs

(b) Mapping with 5 PCs

Figure 6.3: Mapping for patient 71 with high grade cervical lesion
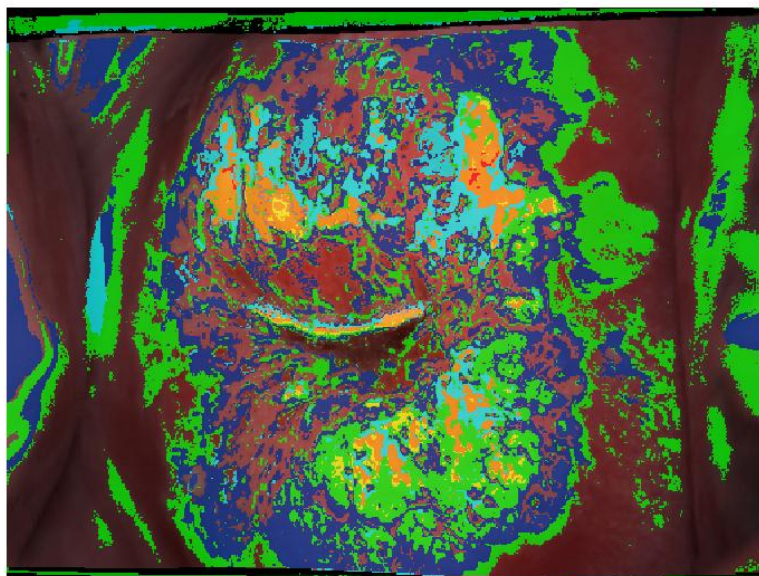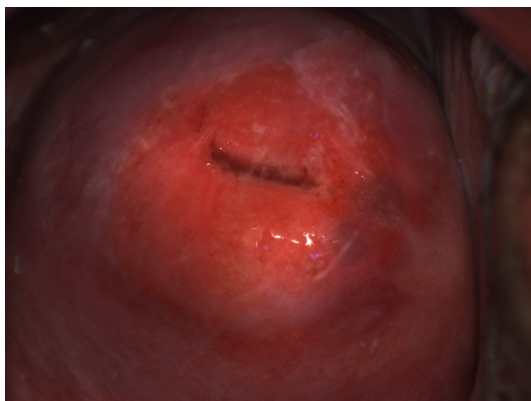
(a) Mapping with combination of 5 PCs, WCs
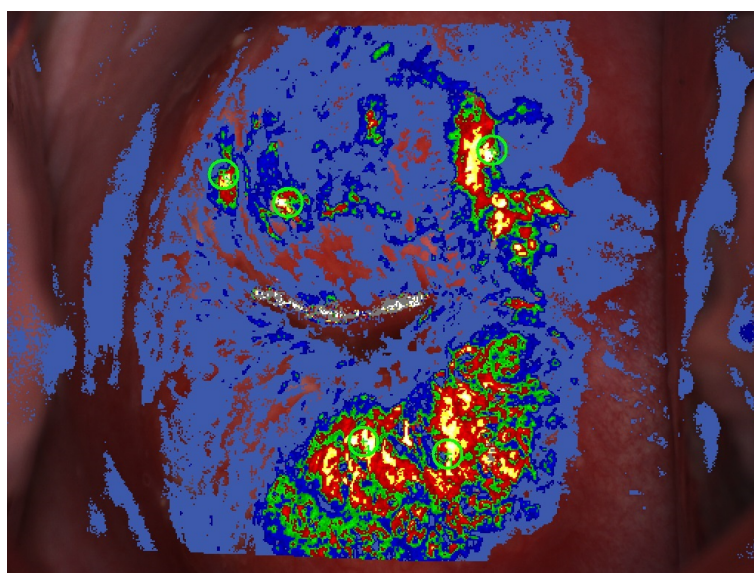
(b) Mapping with DySIS method

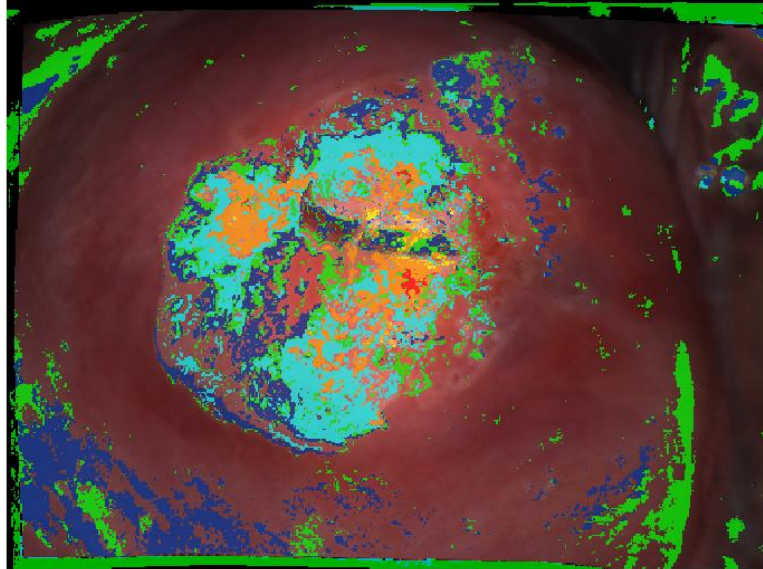Figure 6.4: Mapping for patient 71 with high grade cervical lesion
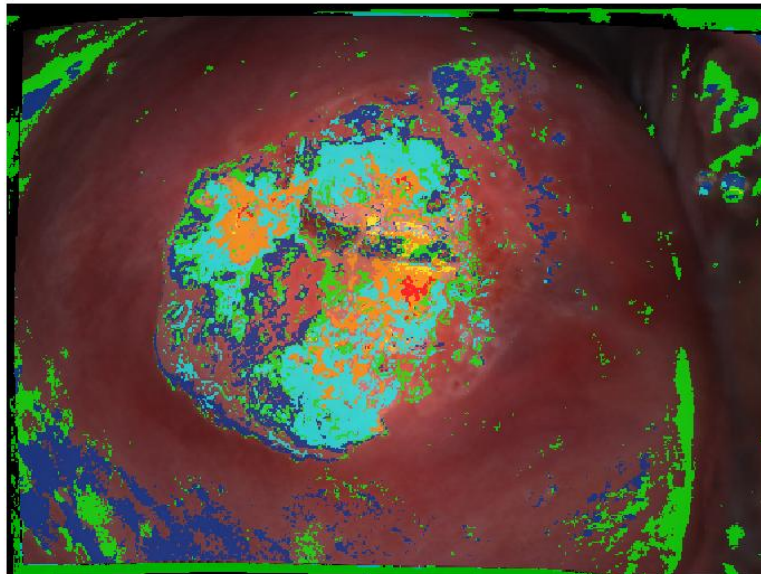
(a) Pre acetic acid image

(b) Post acetic acid image

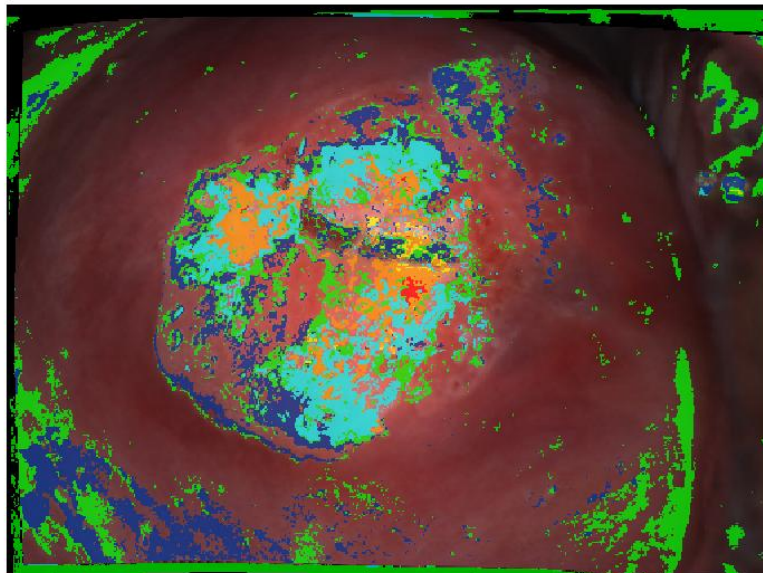Figure 6.5: Pre and Post acetic acid image for patient 77 with low grade cervical lesion
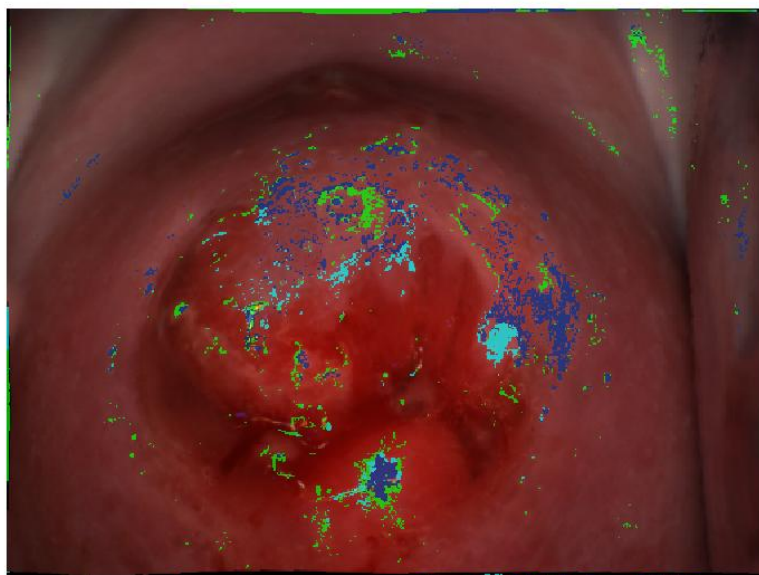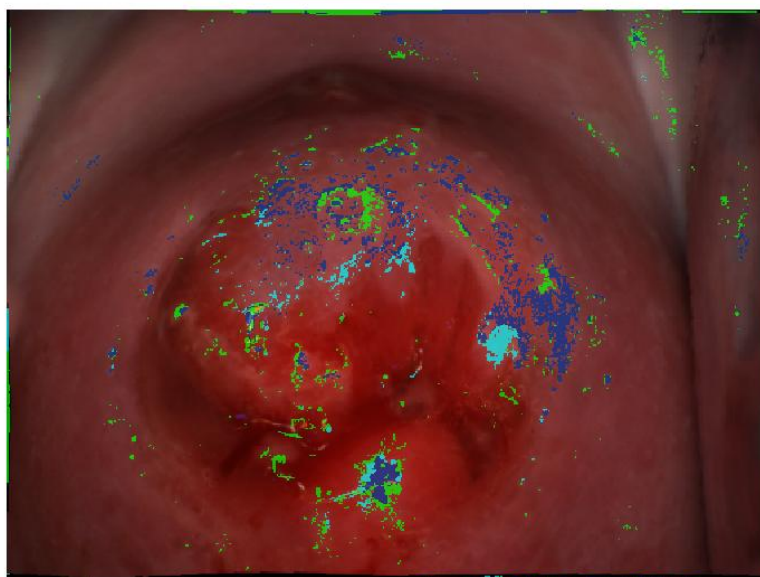
(a) Mapping with 5 WCs

(b) Mapping with 5 PCs

Figure 6.6: Mapping for patient 77 with low grade cervical lesion
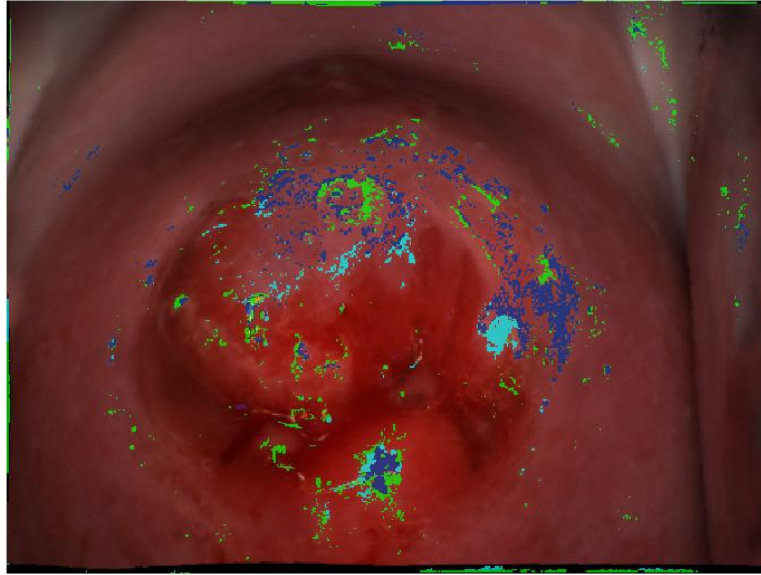
(a) Mapping with combination of 5 PCs, WCs
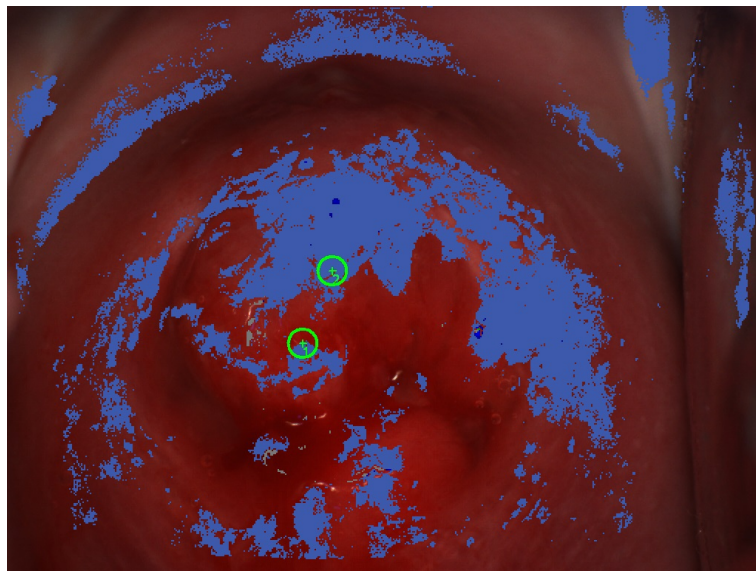
(b) Mapping with DySIS method

Figure 6.7: Mapping for patient 77 with low grade cervical lesion
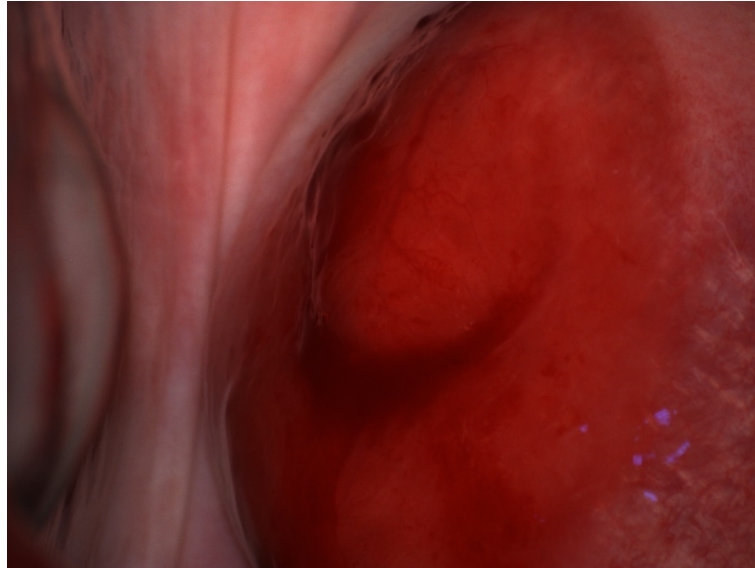
(a) Pre acetic acid image
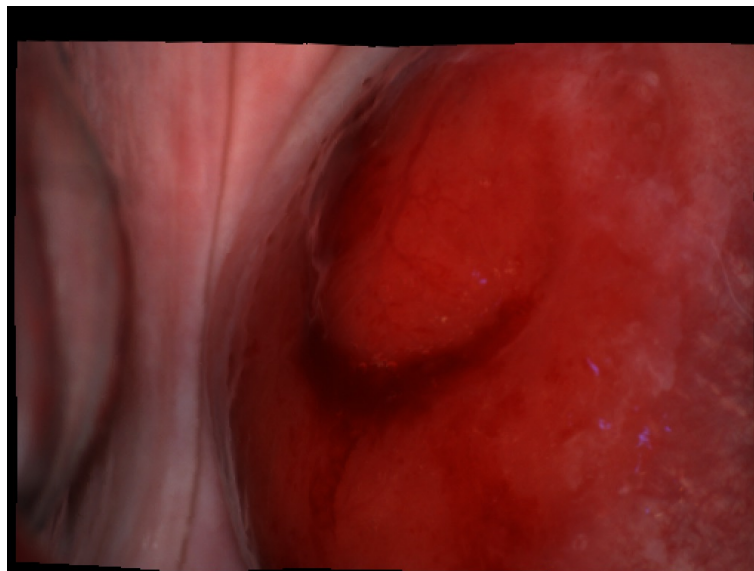
(b) Post acetic acid image

Figure 6.8: Pre and Post acetic acid image for patient 12 with no evidence of disease
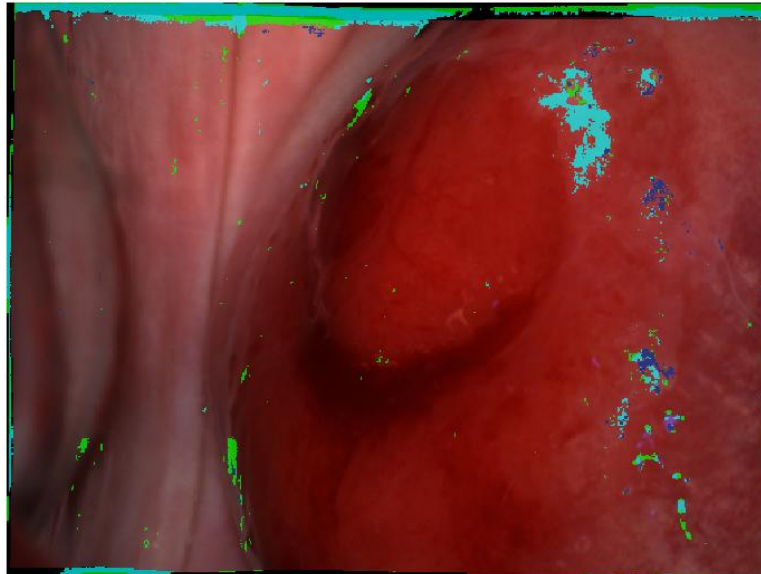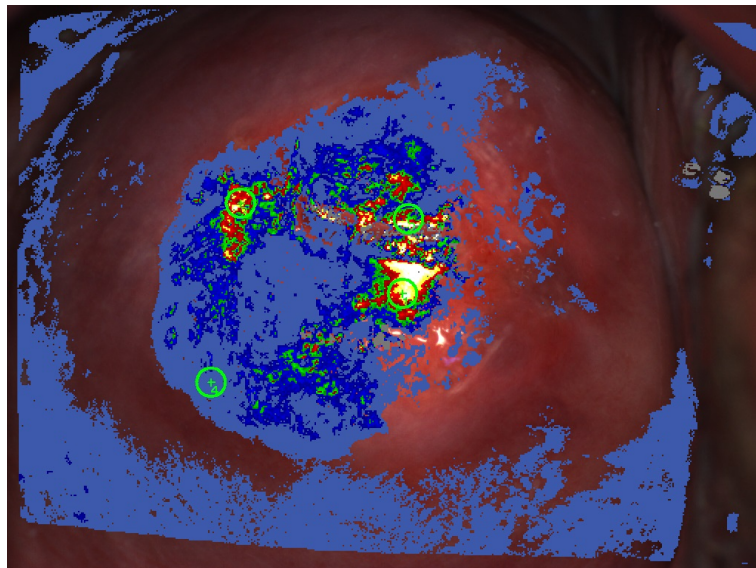
(a) Mapping with 5 WCs
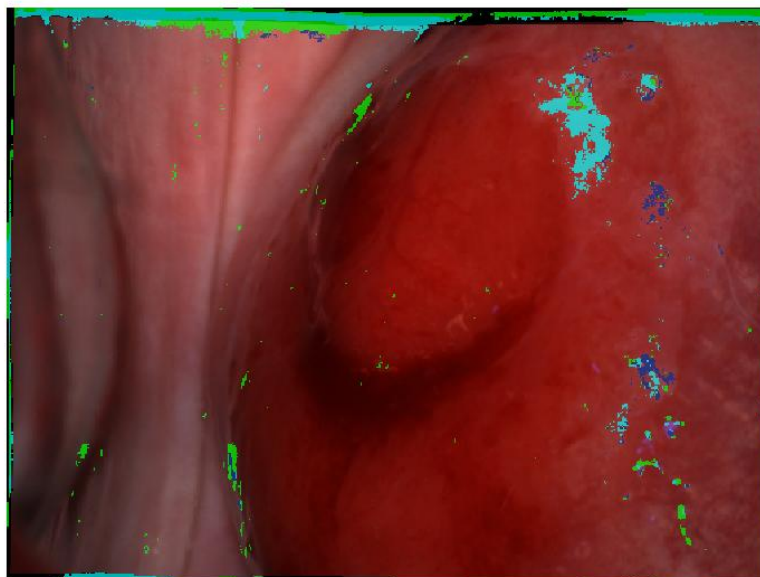
(b) Mapping with 5 PCs

Figure 6.9: Mapping for patient 12 with no evidence of disease
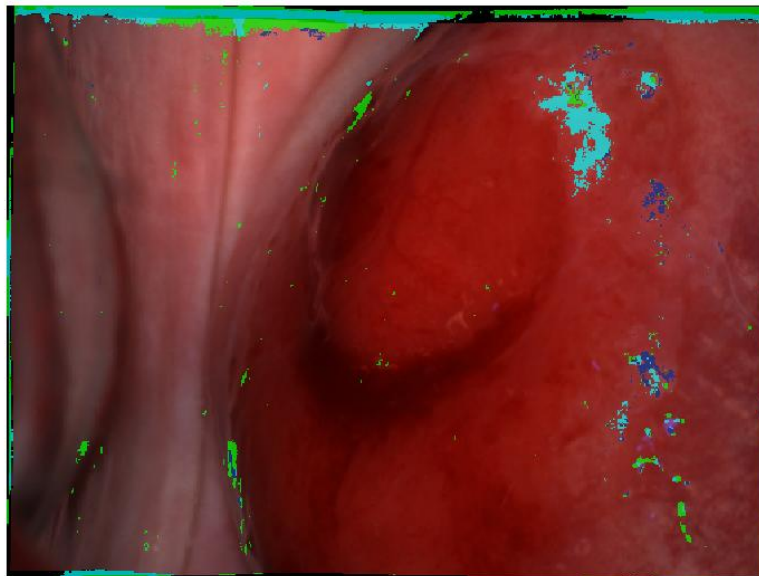
(a) Mapping with combination of 5 PCs, WCs
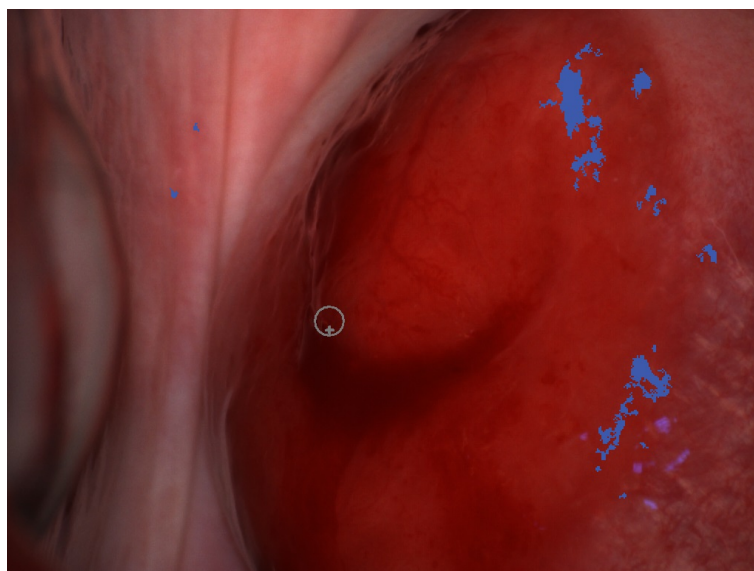
(b) Mapping with DySIS method

Figure 6.10: Mapping for patient 12 with no evidence of disease

Chapter 7

Conclusion and Future Work

The main purpose of this study was the evaluation of temporal feature extraction methods from normal and abnormal cervical epithelium images acquired by Dynamic Contrast Enhanced Optical Imaging (DCE-OI) In order to reduce the data dimensionality and extract valuable information, various feature extraction methods including, WT, PCA, KPCA, PAA, APAA, SAX were implemented and evaluated using 1-nn classifier. The results indicate that using a subset of the entire feature set, WT, PCA and KPCA methods present similar or better performance compared to using the entire feature set. Also, selection of the best features extracted from PCA and WT demonstrates better performance for some classification cases. Regarding the execution times, PCA presents better performance than the other methods. Thus, high discrimination performance between the various stages of cervical neoplasia and Normal cases as well can be achieved with low computational cost.

PCA, WT and combination of these were also used for the mapping process of image stacks obtained from patients with High and Low grade disease and no evidence of disease as well. The mapping can provide valuable, real-time information which can assist medical personnel in the diagnosis of cervical cancer.

Regarding the future work, different and more sophisticated classification algorithms can be implemented in order to improve the performance of classification. In addition, the above mentioned methods have to be evaluated using larger datasets. Finally, classification strategies which combine temporal and spatial information could contribute to more accurate results.

Appendix A

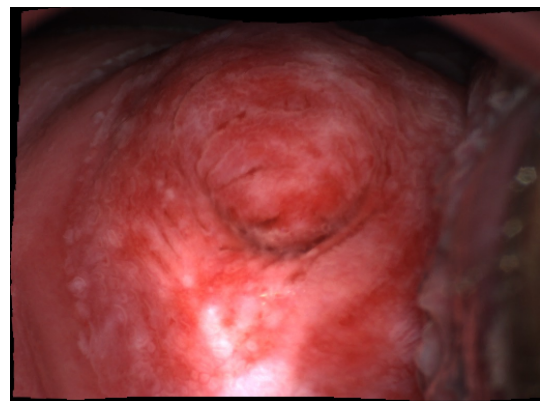# Mapping and Visualization

High Grade Cervical Neoplasia

Patient 103

Figures A.1- A.3 illustrate the classification results for patient 103 who was diagnosed with high grade cervical neoplasia.
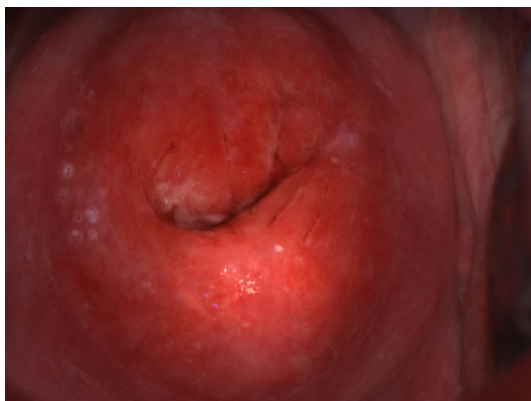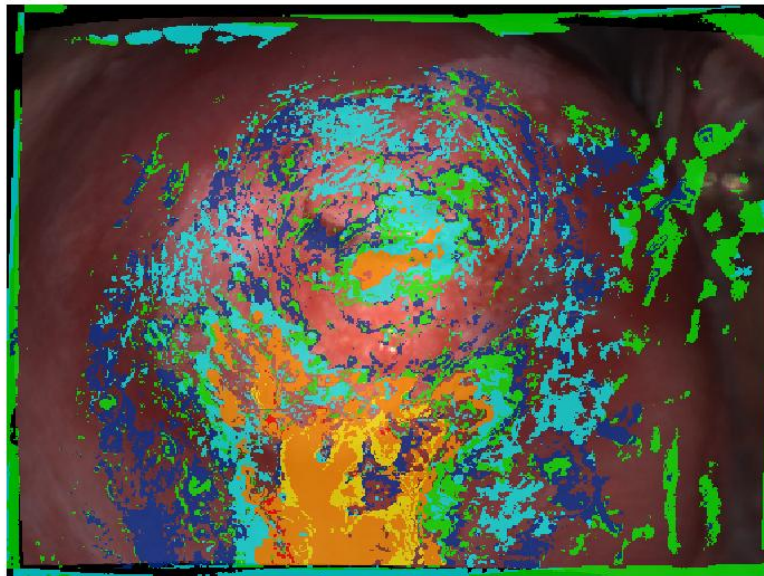

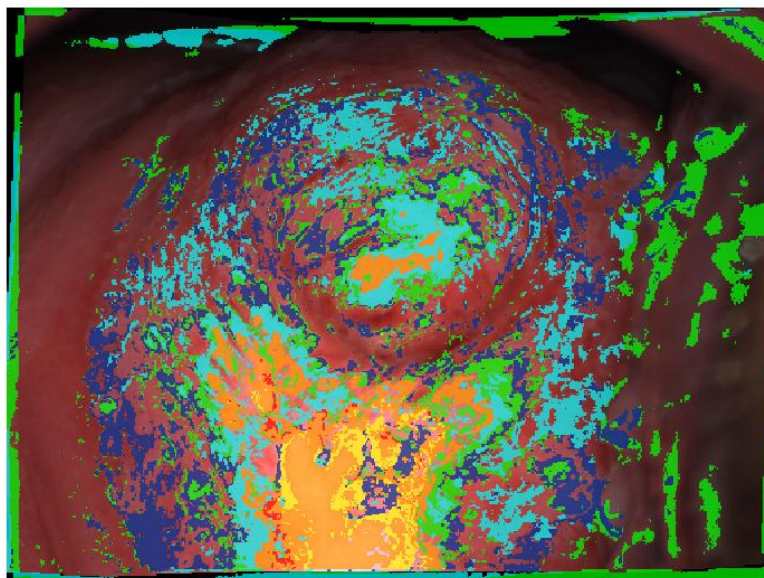(a) Pre acetic acid image   (b) Post acetic acid image

Figure A.1: Pre and Post acetic acid image for patient 103 with high grade cervical lesion
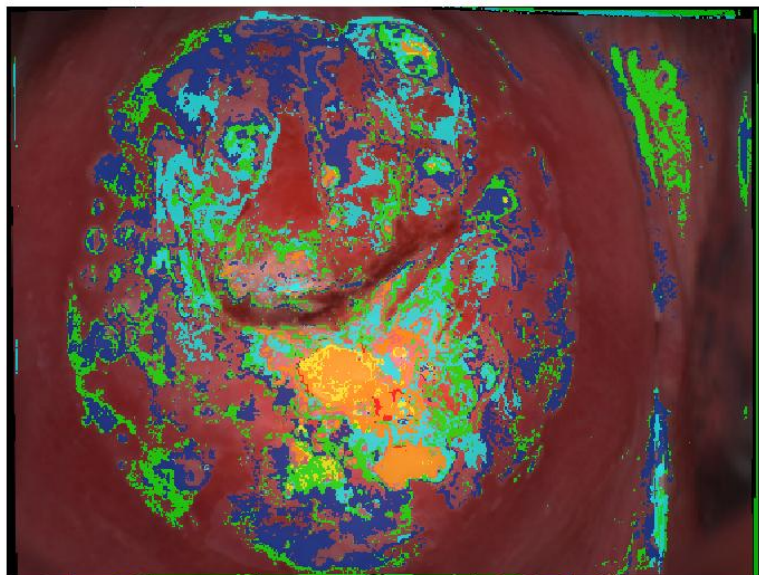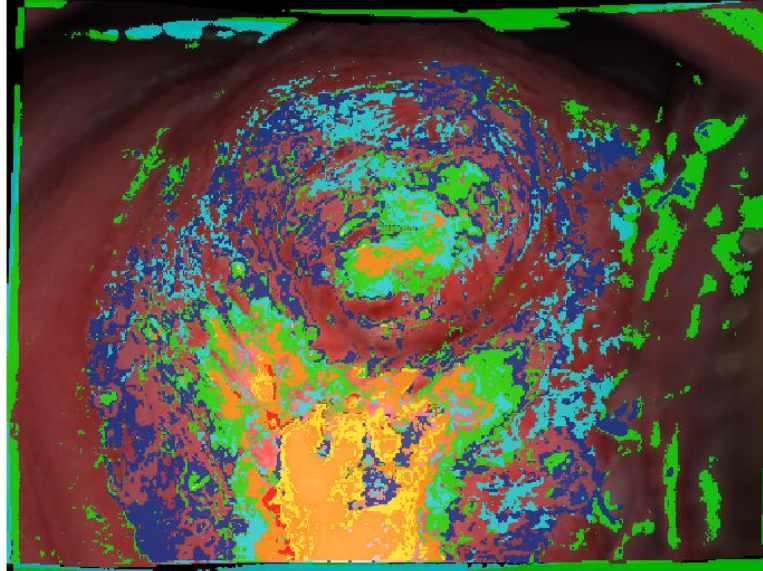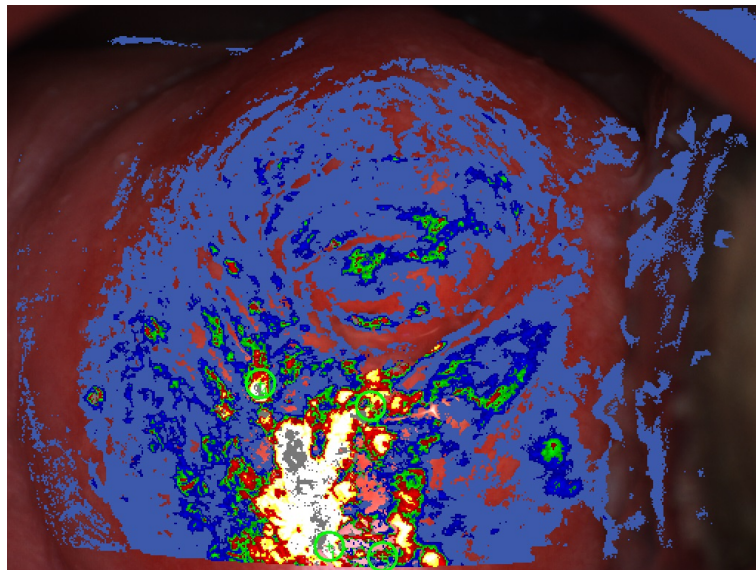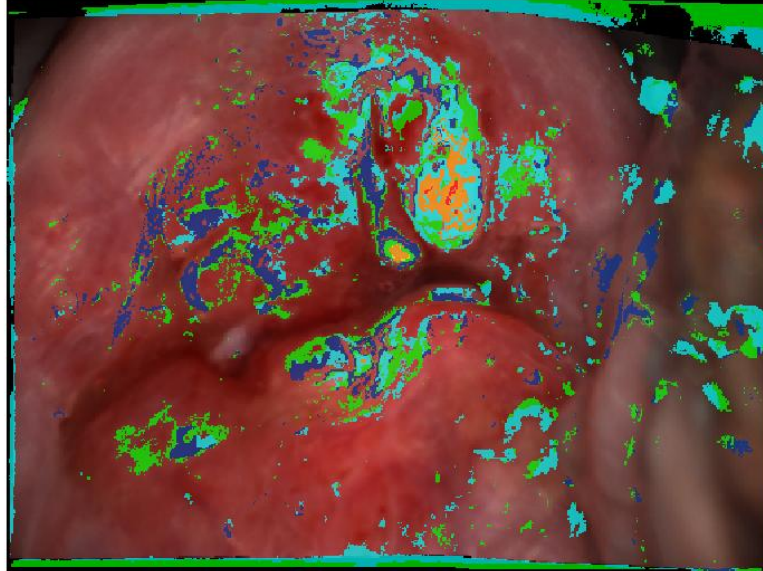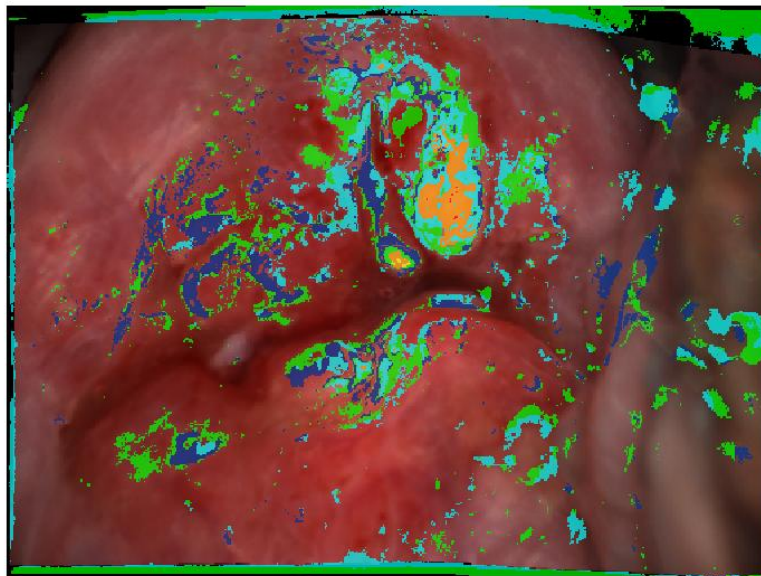
(a) Mapping with 5 WCs

(b) Mapping with 5 PCs

Figure A.2: Mapping for patient 103 with high grade cervical lesion

(a) Mapping with combination of 5 PCs, WCs

(b) Mapping with DySIS device

Figure A.3: Mapping for patient 103 with high grade cervical lesion

Patient 106

Figures A.4- A.6 illustrate the classification results for patient 106 who was diagnosed with high grade cervical neoplasia.

(a) Pre acetic acid image   (b) Post acetic acid image

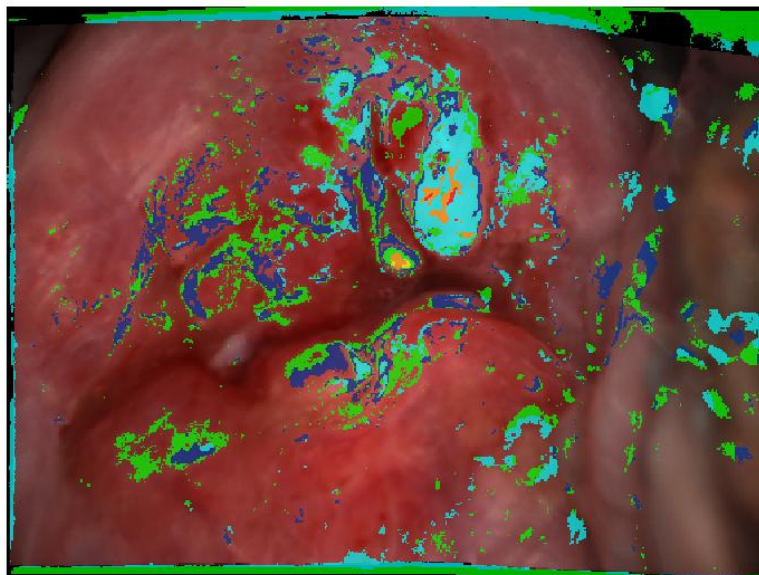Figure A.4: Pre and Post acetic acid image for patient 106 with high grade cervical lesion

(a) Mapping with 5 WCs

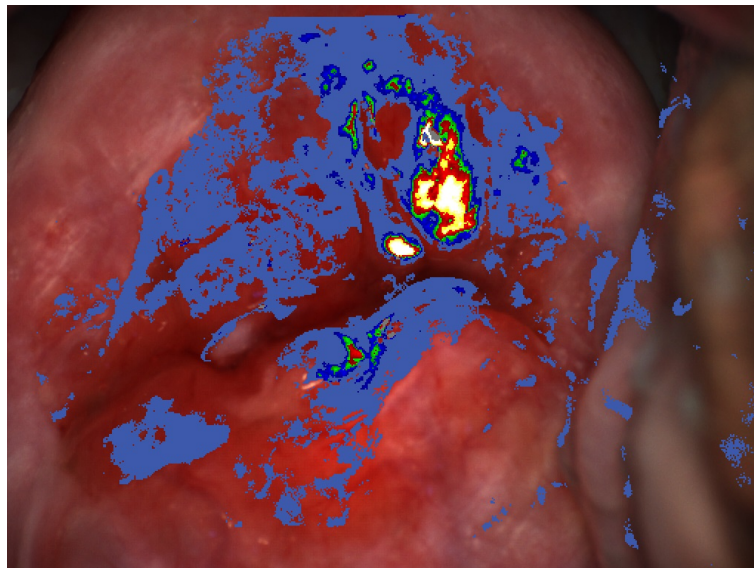(b) Mapping with 5 PCs

Figure A.5: Mapping for patient 106 with high grade cervical lesion

(a) Mapping with combination of 5 PCs, WCs

(b) Mapping with DySIS device

Figure A.6: Mapping for patient 106 with high grade cervical lesion

Patient 107

Figures A.7- A.9 illustrate the classification results for patient 107 who was diagnosed with high grade cervical neoplasia.
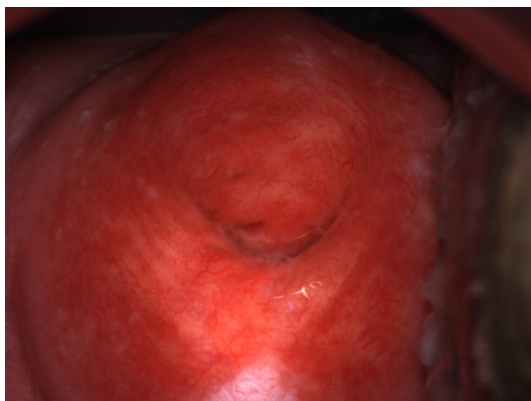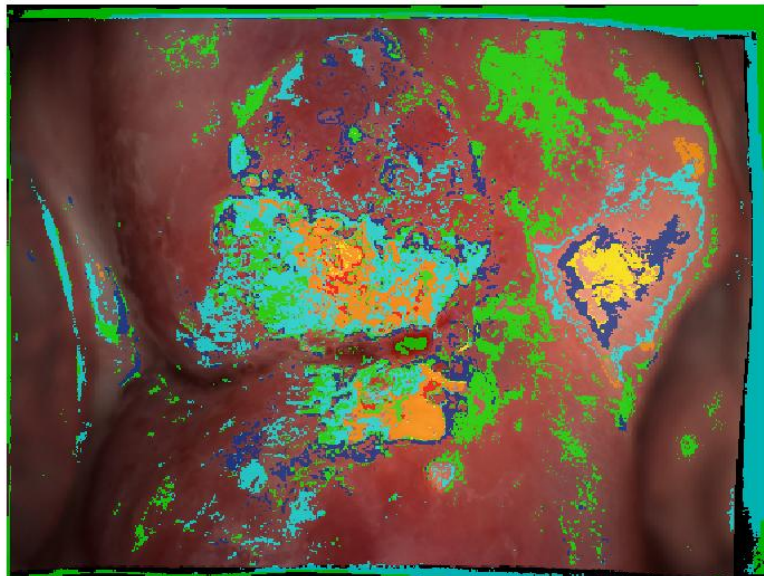

(a) Pre acetic acid image

(b) Post acetic acid image

Figure A.7: Pre and Post acetic acid image for patient 107 with high grade cervical lesion

(a) Mapping with 5 WCs

(b) Mapping with 5 PCs

Figure A.8: Mapping for patient 107 with high grade cervical lesion

(a) Mapping with combination of 5 PCs, WCs

(b) Mapping with DySIS device

Figure A.9: Mapping for patient 107 with high grade cervical lesion

Patient 110

Figures A.10- A.12 illustrate the classification results for patient 110 who was diagnosed with high grade cervical neoplasia.
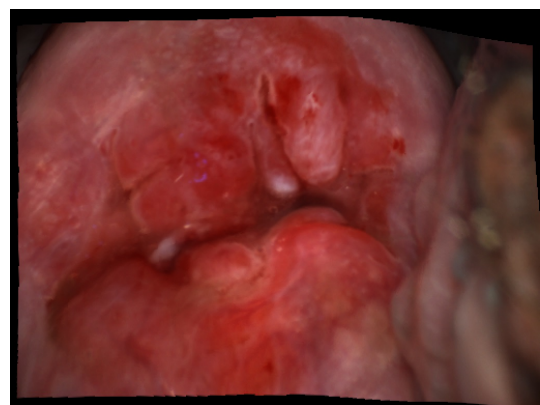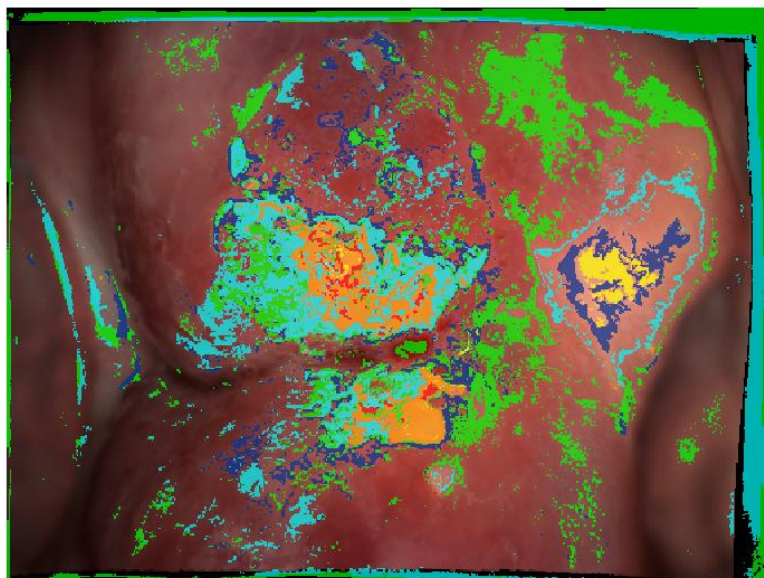

(a) Pre acetic acid image   (b) Post acetic acid image

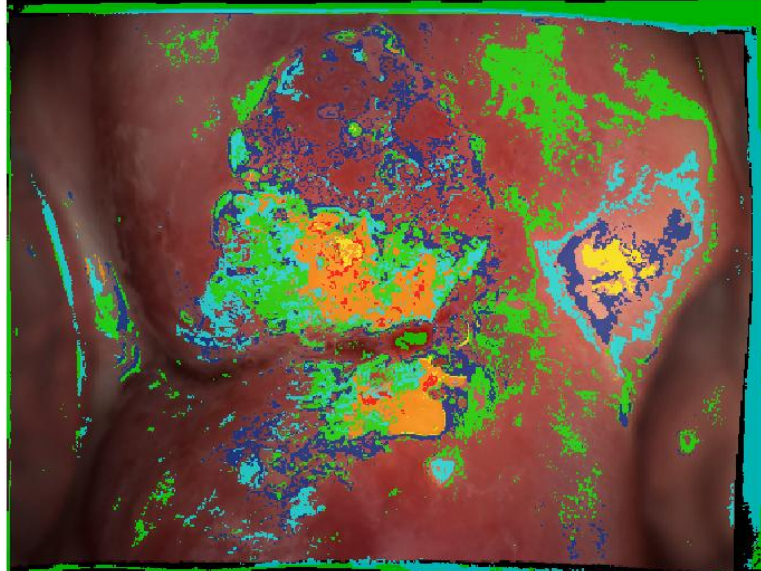Figure A.10: Pre and Post acetic acid image for patient 110 with high grade cervical lesion
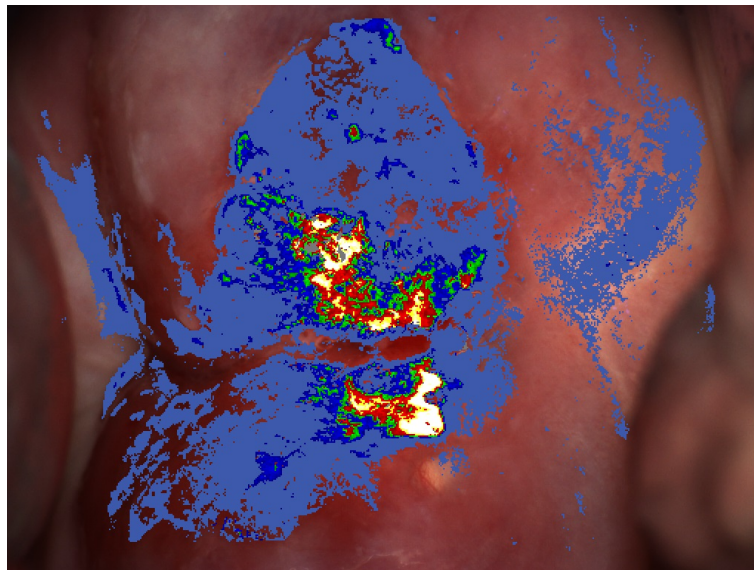
(a) Mapping with 5 WCs

(b) Mapping with 5 PCs

Figure A.11: Mapping for patient 110 with high grade cervical lesion
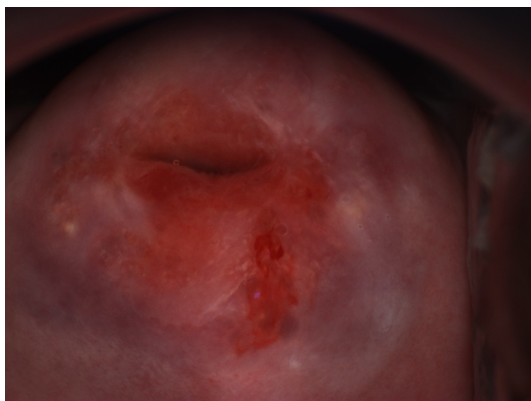
(a) Mapping with combination of 5 PCs, WCs

(b) Mapping with DySIS device

Figure A.12: Mapping for patient 110 with high grade cervical lesion

Patient 114

Figures A.13- A.15 illustrate the classification results for patient 110 who was diagnosed with high grade cervical neoplasia.

(a) Pre acetic acid image   (b) Post acetic acid image

Figure A.13: Pre and Post acetic acid image for patient 114 with high grade cervical lesion

(a) Mapping with 5 WCs

(b) Mapping with 5 PCs

Figure A.14: Mapping for patient 114 with high grade cervical lesion

(a) Mapping with combination of 5 PCs, WCs

(b) Mapping with DySIS device

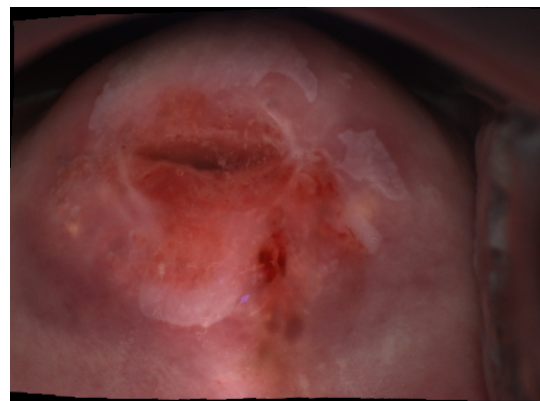Figure A.15: Mapping for patient 114 with high grade cervical lesion

Low Grade Cervical Neoplasia

Patient 86

Figures A.16- A.18 illustrate the classification results for patient 86 who was diagnosed with high grade cervical neoplasia.
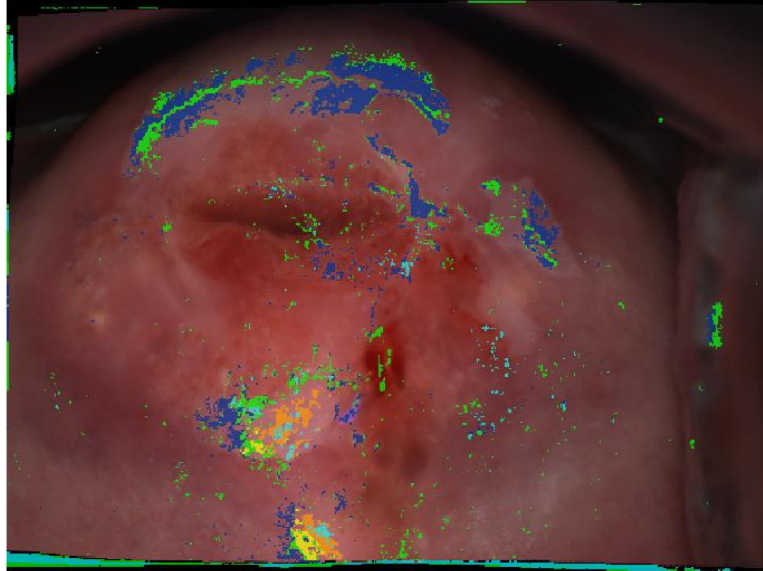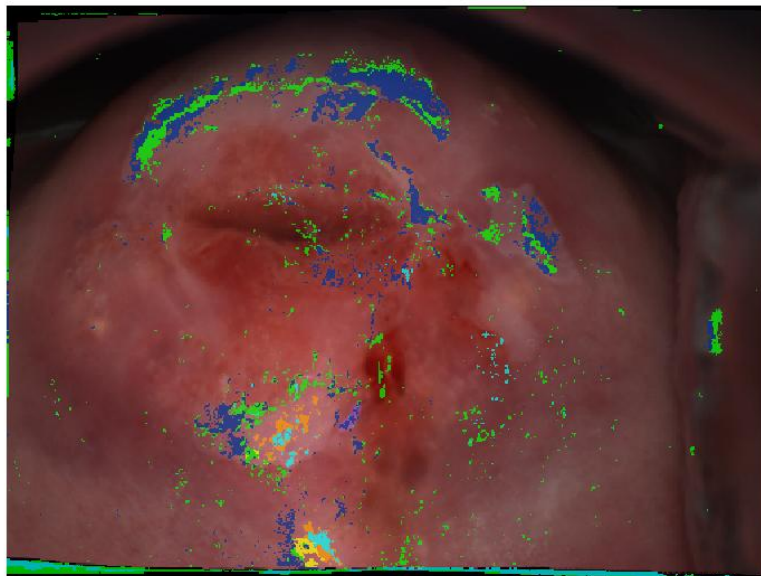

(a) Pre acetic acid image   (b) Post acetic acid image

Figure A.16: Pre and Post acetic acid image for patient 86 with high grade cervical lesion
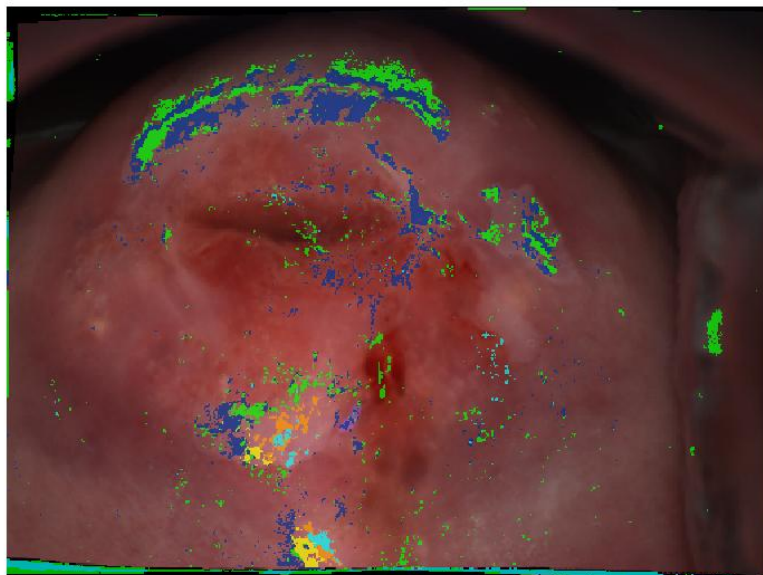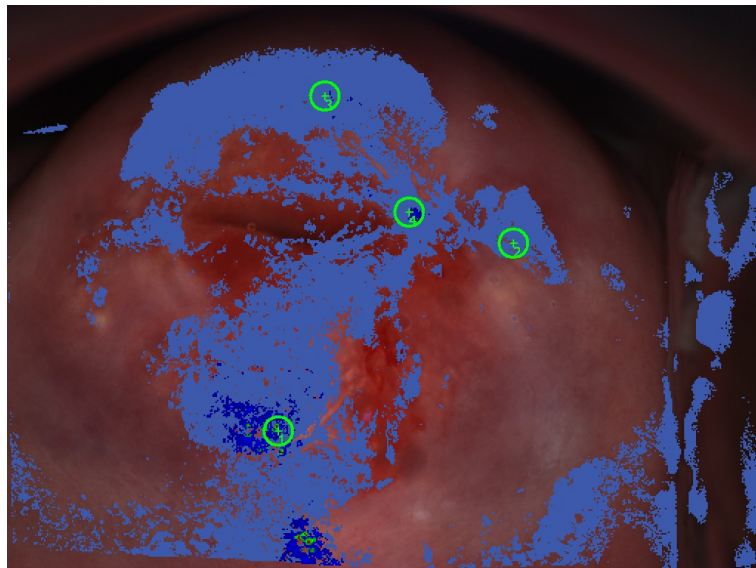
(a) Mapping with 5 WCs

(b) Mapping with 5 PCs

Figure A.17: Mapping for patient 86 with high grade cervical lesion

(a) Mapping with combination of 5 PCs, WCs

(b) Mapping with DySIS device

Figure A.18: Mapping for patient 86 with high grade cervical lesion

No evidence of neoplasia

Patient 38

Figures A.19- A.21 illustrate the classification results for patient 38 who does not present evidence of disease.
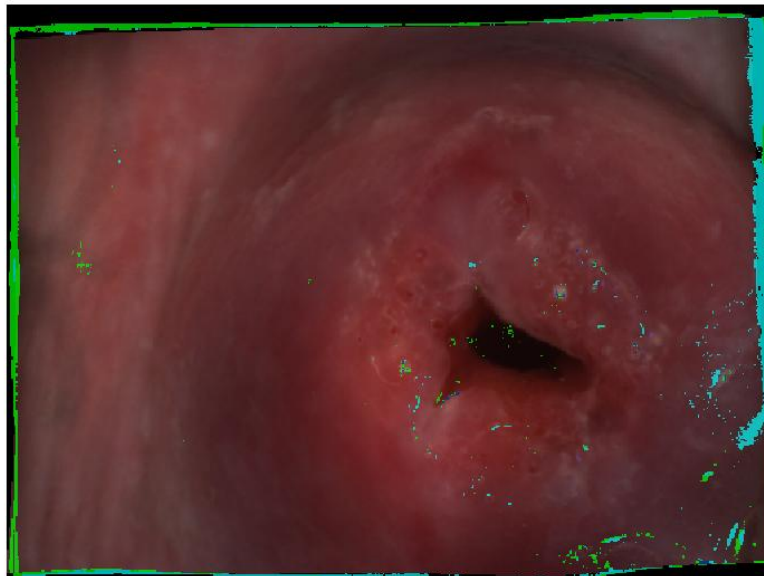

(a) Pre acetic acid image   (b) Post acetic acid image

Figure A.19: Pre and Post acetic acid image for patient 38 with high grade cervical lesion

(a) Mapping with 5 WCs

(b) Mapping with 5 PCs

Figure A.20: Mapping for patient 38 with high grade cervical lesion

(a) Mapping with combination of 5 PCs, WCs

(b) Mapping with DySIS device

Figure A.21: Mapping for patient 38 with high grade cervical lesion

References

[1] A. N. Burchella, R. L. Winerb, S. Sanjosec, and E. L. Francoa, "Chapter 6: Epidemiology and transmission dynamics of genital hpv infection," *HPV Vaccines and Screening in the Prevention of Cervical Cancer*, vol. 24, no. 3, pp. 52–61, 2008. 1

[2] K. Nanda, D. C. McCrory, E. R. Myers, L. A. Bastian, V. Hasselblad, J. D. Hickey, and D. B. Matchar, "Accuracy of the papanicolaou test in screening for and follow-up of cervical cytologic abnormalitiesa systematic review," *Annals of Internal Medicine*, vol. 132, no. 10, pp. 810–819, 2000. 1

[3] E. H. Hopman, F. J. Voorhors, P. Kenemans, C. J. Meyer, and T. J. Helmerhorst, "Observer agreement on interpreting colposcopic images of cin," *Gynecologic Oncology*, vol. 58, no. 2, pp. 206–209, 1995. 2

[4] S. B. Cantor, M. Cardenas-Turanzas, D. D. Cox, E. N. Atkinson, G. M. Nogueras-Gonzalez, J. R. Beck, M. M. Follen, and J. L. Benedet, "Accuracy of colposcopy in the diagnostic setting compared with the screening setting," *Obstetrics and Gynecology*, vol. 111, no. 1, pp. 7–14, 2008. 2

[5] I. M. Orfanoudaki, D. Kappou, and S. Sifakis, "Recent advances in optical imaging for cervical cancer detection," *Archives of Gynecology and Obstetrics*, vol. 284, no. 5, 2008. 2

[6] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006. 5

[7] H. Liu and H. Motoda, *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Norwell, MA, USA: Kluwer Academic Publishers, 1998. 5, 26

REFERENCES

[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2Nd Edition)*. Wiley-Interscience, 2000. 5

[9] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, pp. 417–441 and 498–520, 1933. 6

[10] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, pp. 1299–1319, July 1998. 8

[11] E. J. Keogh and M. J. Pazzani, "A simple dimensionality reduction technique for fast similarity search in large time series databases," *Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Current Issues and New Applications*, pp. 122–133, 2000. 10

[12] K. Chakrabarti, E. Keogh, S. Mehrotra, and M. Pazzani, "Locally adaptive dimensionality reduction for indexing large time series databases," *ACM Trans. Database Syst.*, vol. 27, pp. 188–228, June 2002. 11

[13] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, (New York, NY, USA), pp. 2–11, ACM, 2003. 12

[14] V. Margariti, "Cervical cancer diagnosis through the wavelet analysis of the acetowhitening process," master thesis, Technical University of Crete, Greece, 2010. 12, 14, 25, 26

[15] C. Balas, "A novel optical imaging method for the early detection, quantitative grading, and mapping of cancerous and precancerous lesions of cervix," *Biomedical Engineering, IEEE Transactions*, vol. 48, pp. 96–104, Jan. 2001. 17

[16] S. Young Park, M. Follen, A. Milbourne, H. Rhodes, A. Malpica, N. MacKinnon, C. MacAulay, M. K. Markey, and R. Richards-Kortum, "Automated image analysis of digital colposcopy for the detection of cervical neoplasia," *Journal of Biomedical Optics*, vol. 13, no. 1, p. 10, 2008. 21

[17] W. Li, S. Venkataraman, U. Gustafsson, J. C. Oyama, D. G. Ferris, and R. W. Lieberman, "Using acetowhite opacity index for detecting cervical intraepithelial neoplasia," *Journal of Biomedical Optics*, vol. 14, no. 1, p. 10, 2009. 21

[18] H. G. Acosta-Mesa, N. Cruz-Ramireza, and R. Hernandez-Jimenez, "Aceto-white temporal pattern classification using k-nn to identify precancerous cervical lesion in colposcopic images," *Computers in biology and medicine*, vol. 39, no. 9, pp. 778–84, 2009. 21, 22

[19] H. G. Acosta-Mesa, N. Cruz-Ramirez, K. Gutierrez-Fragoso, R. E. Barrientos-Martinez, and R. Hernandez-Jimenez, *Assessing the Possibility of Identifying Precancerous Cervical Lesions Using Aceto-White Temporal Patterns*. InTech. 22

[20] K. Gutierrez-Fragoso, H. G. Acosta-Mesa, N. Cruz-Ramirez, and R. Hernandez-Jimenez, "Automatic classification of acetowhite temporal patterns to identify precursor lesions of cervical cancer," *Journal of Physics: Conference Series*, vol. 475, no. 1, p. 10, 2013. 22

[21] T. Wu, T.-H. Cheung, S.-F. Yim, and J. Y. Qu, "Clinical study of quantitative diagnosis of early cervical cancer based on the classification of acetowhitening kinetics," *Journal of Biomedical Optics*, vol. 15, no. 2, p. 7, 2010. 22

[22] S. Park, D. Sargent, R. Lieberman, and U. Gustafsson, "Domain-specific image analysis for cervical neoplasia detection based on conditional random fields," *Medical Imaging*, vol. 30, pp. 867 – 878, Jan. 2011. 23