UNIVERSITY OF AGDER

Pattern Recognition Based Prediction of the Outcome of Radiotherapy in Cervical Cancer Treatment

.

By

Mohammad Yasar and Vimala Nunavath

Supervisor
Ole Christoffer Granmo

This Master's Thesis is carried out as a part of the education at the University of Agder and is therefore approved as a part of this education..

University of Agder, [2011]
Faculty of [Engineering and Science]
Department of [Information and Communication Technology]

Abstract

Cervical Cancer is one the most common cancers amongst women. Every year almost 300 Norwegian women are diagnosed with cervical cancer. It is the 5th most deadly cancer type amongst women in the world. Estimates show that there are approximately 473,000 cases of cervical cancer in 2008 and 253,500 deaths per year. As we can see from the statistics, cervical cancer is a very severe and common type of cancer which costs many human lives every year. Therefore any progression in prognostication of this disease is very essential to treatment of its patients.

Our task in this project was to analyze contrast enhanced MR imaging data from 78 patients. This data was recorded after a certain period of time after the patients received radiotherapy. The data was collected after a median time of 48 months for each patient. The outcome of the treatment and propagation of the contrast medium in to the blood vessels (in tumor region) was recorded.

The main focus of this project was to model spatial patterns in the Cervix Cancer data set using hidden Markov models (HMM) in one of the machine learning techniques can be used to predict the outcome of radiotherapy treatment of the cervical cancer patients based on identified patterns with given data samples.

To find the unobserved (hidden) patterns, we have used hidden Markov models on the dataset to find hidden patterns in the data. These models show the distribution of the outcome of the treatment, grouped by the similarities between properties of the contrast medium in the blood vessels.

Our research shows that hidden Markov models are not feasible for this dataset. It was not possible to retrieve any information with high enough accuracy to be able to predict outcome of radiotherapy treatment.

Preface

This master thesis is submitted in partial fulfillment of the requirements for the degree Master of Science in Information and Communication Technology at the University of Agder, Faculty of Engineering and Science in Grimstad, Norway. The project is provided by Institute for Cancer Research at Oslo University hospital in cooperation with Department of Medical Physics at Oslo University hospital. The main research activities in Department of Medical Physics are how to improve non-invasive detection, treatment and treatment evaluation of the cancer. This work was carried out under the supervision of Associate Professor Ole -Christoffer Granmo at the University of Agder, Norway.

First of all, we want to thank our supervisor Ole-Christoffer Granmo for his great assistance, help and inspiration throughout the project period. He introduced us to this interesting field of medical science and has given his valuable support and feedback during the entire project period.

Secondly, we would also like to thank Erlend Andersen and Eirik Malinen for proposing such an interesting and innovative problem for the project. We also wish to thank Heidi Lyng for providing images of the Cervical Cancer patients.

Last but not the least, we would like to thank our families for supporting, helping and believing in us. Without their support and understanding, we have never made this achivement of this study possible.

Grimstad, May 2011.
Mohammad Yasar and Vimala Nunavath

Contents

List of Figures

List of Algorithms

List of Tables

1 Introduction

Cancer has become the second leading cause for the death in both men and women in the world. In next few decades, this disease is expected to become the leading reason for death. Therefore, it is necessary to find the ways to predict the outcome of the treatment of the disease to save at least few more lives from death [1]. There are many types of cancers existing in both men and women. For example breast cancer, lung cancer, cervical cancer, etc. Our focus will be the last one mentioned.

A patient can be diagnosed with cervical cancer by using a technique called *DESMRI* – Dynamic Contrast Enhanced Magnetic Resonance Imaging. It is a medical test used by physicians to diagnose and treat different diseases. A contrast medium is injected into the blood vessels and a magnetic field is then used to produce detailed pictures of any internal body organs. These images can be examined on a computer monitor. After the patient receives cervical cancer treatment DESMRI is used again to check whether the cancerous area is completely cured or there is some residual.

In this project we have post-treatment DESMRI data from 78 patients. Our goal is to use this data to help predict the outcome of the treatment before the patient is treated. The prediction can be then used to estimate whether a patient will have any effects of the treatment or not.

After a patient is diagnosed with cervical cancer, the tumor data is collected using DESMRI. The tumor area is divided in to thousands of voxels (a volume entity with xyz-coordinates). The behavior of the contrast medium in each voxel is recorded. This behavior is represented by three attributes: amplitude of the contrast medium and its rate of flow in and out of a voxel. The resulting data is presented in a large continuous data table.

Due to amount and complexity of the data, it is nearly impossible for a human analyst to analyze it manually. Data analysis can be done on existing and newly appearing data. Therefore it is necessary to take automated computer system's help to find the patterns in unstructured data [2], and give meaning to these large sets of unstructured data and make it easier for us to understand it.

There are so many machine learning techniques are present to perform the data analysis. For example artificial neural networks, genetic algorithms, fuzzy sets, support vector machines, rough sets, wavelet filters and statistical transforms, Genetic Programming, Bayesian Networks and hidden Markov models, etc.

In this thesis, we present how hidden Markov models are used to show the interdependencies of the voxel with other coordinates in medical decision making, in particular diagnosis, (prognostic) prediction and treatment selection. A HMM model that is developed to assist clinicians in the diagnosis and selection of Radiotherapy treatment for patients with cervical cancer and we will show how the HMMs are tested using validation algorithm called Leave-One-Out-Cross validation.

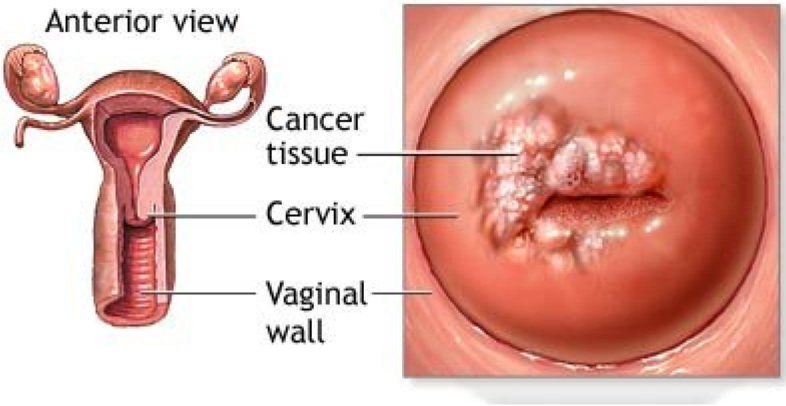

Figure 1: Cervical cancer

1.1 Background and motivation

1.1.1 Cervical Cancer

Cervical cancer is a one type of cancers which is formed in tissues of the cervix (the organ connecting the uterus and vagina). It is usually a slow-growing cancer that is mostly caused by HPV infection (Human Papilloma Virus). [3]

It is the third most common cancer worldwide, and 80% of cases occur in the developing world. It is the leading cause of death from cancer among women in developing countries, where it causes about 190,000 deaths each year. Rates of the disease were highest in Central America, sub-Saharan Africa, and Melanesia (M Parkin, International Agency for Research on Cancer, personal communication, July 2000).

When compared to other Nordic countries, Norway is bit better with the decline in mortality rate despite organized screening programs were introduced in Nordic countries in 1960-70s. Norway has not witnessed any improvement in the prognosis of cervical cancer patients from mid 70s [4].

It is demonstrated by Epidemiological studies that human papilloma virus test is the most essential independent risk factor for rising of both cervical dysplasia and invasive cancer. The relative risk is insignificant when it is associated with traditional factors like sexual behavior. There are only some infected individuals who actually develop cervical cancer, despite researchers have identified HPV as the primary cause of cervical cancer. Other environmental and host factors can also influence the progression of HPV infection to high-grade squamous intraepithelial lesions (HSIL) and cervical cancer.

Women at high risk are advised to have follow-up check-ups and treatments to make sure that significant subset of women are treated. It is essential to identify the host determinants of viral persistence to understand the mechanisms of tolerance, and it also can lead to the development of tests in high-risk individuals. [5, 6, 7, 8, 9]

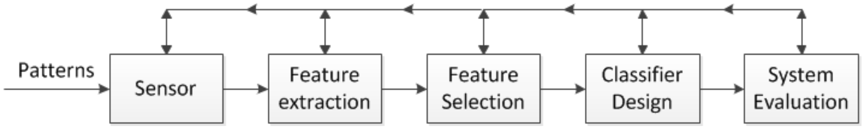

Figure 2: Different steps of a typical CAD system for

Unlike many cancers, cervical cancer can also be prevented. Primary prevention of cervical cancer through preventing human papillomavirus (HPV) infection, a sexually transmitted agent that causes cervical cancer, will contribute to reducing cancer mortality. Primary prevention of HPV infection is more challenging than prevention of most other sexually transmitted infections. HPV-infected women generally are asymptomatic, HPV is transmitted easily, and no therapies eliminate the underlying infection.

The development of a vaccine against HPV is under investigation, but vaccination as a means of primary prevention is years away. Secondary prevention involves using relatively cheap screening and treatment technologies that can detect dysplasia before it progresses to invasive cancer.

Detection of cancer is an important area of research in the community of image processing and pattern recognition and early detection is the key to reduce the death rate in cervical cancer. If the cancers are detected in primitive stage, better treatment can be provided in time. It is however essential that the early detection must be with accurate and reliable diagnosis.[10] For better efficiency and accuracy of diagnoses and treatment, the techniques in image processing and pattern recognition are extensively used. These techniques can also be used to the analysis and recognition of cancer, evaluation of treatment and the prediction of the development of the cancer [1].

1.1.2 Pattern recognition

Detection of the presence of particular signal features is relied on the conventional methods of monitoring and diagnosing the disease. In the past 10-15 years, Computer aided-diagnosis (CAD) approaches for automated diagnostic systems have been developed in order to solve the problem of large number of patients in intensive care units and continuous observation. These systems function by transforming the qualitative diagnostic criteria into a qualitative feature classification problem.[11]

The above picture explains how the different stages are followed by the design of a classification system. It is evident from the feedback arrows that these stages are not independent, but they are interrelated. Based on the results, one may go back to redesign the earlier stages so as to improve the overall performance [11]

The classification is defined as a basic task in data analysis. Pattern recognition is required in this analysis for the construction of a classifier, i.e. a function that assigns a class label to instances described by a set of attributes. The main problem in machine learning is the induction of classifiers from data sets of pre-classified instances. [12]

Pattern recognition is "*the act of taking in raw data and taking an action based on the category of the pattern*"

Pattern recognition is very central in data mining because data mining is largely dependent on effective pattern recognition to extract meaningful information from unstructured data. Pattern recognition techniques are very important in various fields, including computer sciences, psychology, ethology, cognitive science and medical science. In medical science it forms basis for computer-aided diagnoses (CAD).

Pattern recognition is a vast field, where the goal is to find structures and meaning in enormous amount of unstructured data. The main purpose of pattern recognition in cancer diagnosis is to solve the pattern classification dilemma where a set of input features are used to determine if a patient has a particular disorder .There are several algorithms and techniques exist. For example, supervised and unsupervised learning. In addition to using the existing algorithms we will also try to improve and tweak them to better fit our needs. In this process we are might find more efficient ways to locate structures and patterns in similar data sets as we will be working with.

There has been lot of research done on, how Pattern Recognition applications are used on different types of cancer diagnosis by analyzing the data and to detect the outcome of the treatment on various types of cancers. [1].

1.1.3 Medical diagnosis

Despite the humans know that they can't analyze the different situations of the real world as isolated facts, they keep on trying to describe them in terms of patterns of related facts. These relations become implicit as they indicate to the same object. It is sometimes essential to explicitly connect these characteristics for finding a relation [13].

In order to understand and observe the power of perception of human beings, we need to be adopted to carry out pattern processing activities.The subjectivity of the specialist is the central and significant issue in medical diagnosis. In specific pattern recognition activities, it is noted that experience of the professional is closely connected to the final diagnosis. It is witnessed that the result is not depended on a systematized solution, but on the interpretation of the patient's signal.

In medical technology, medical diagnostic decision support systems are playing a vital role and have become established components. The decision characteristics of the diseases are used to diagnose future patients with the help of the concept of the medical technology as an inductive engine. It is necessary to utilize a number of CAD approaches for the improvement in the accuracy of diagnosis. Several significant approaches like ANNs and BNs are proposed for cervical cancer diagnosis and prognostic risk evaluation [11].

It can be given as an example in case of *balance disorders diagnosis.* In this diagnosis, the signal corresponding to the ocular movement of the patient has to be analyzed and this signal gives a pattern named nystagmus. The frequency of this pattern decides the type of lesion with the help of different tests. This pattern is also relatively connected to the type of signal and it varies from patient to patient [14].

In case of *cell count*, something similar patterns occur. In common perspective, every histological sample has special associated normal values and which characterize the cell populations forming it. These values will permit the specialist to do a first classification of the tissue under examination as normal or pathological [13].

It is very important to use methods that define proportion of cells with as high objectivity as possible to get more accurate and diagnosis and prognosis. Any such method that counts the similar cells should be able to distinguish between normal and pathological samples [13].

It is important to not use tools that implement a specific algorithm, but instead use a tool that can adapt according to the data of the problem. Hidden Markov models are very useful for this type of analysis because they are capable of finding the pattern in the data with help of an expert, as well as generalizing the information in the data, and showing us the complex relations in it.

1.1.4 Hidden Markov models

A hidden Markov model is a stochastic model that can be used to find hidden statistical properties of the observed data. It creates an accurate model of the source data and then can be used to simulate the source. Machine learning techniques that implement HMM have been successfully applied problem of various types. These include optical character recognition (OCR), speech recognition and bioinformatics.

Bioinformatics is a vast field where machine learning techniques can be utilized to analyze DNA sequences, proteins, genes and mutations, as well as medical image analysis [29]. In our thesis we will use various algorithmic and statistical methods to help us predict the outcome of radiotherapy treatment of cervical cancer patients. Our main focus will be statistical model-based approach using HMMs.

HMMs are gaining increased popularity among informatics researchers because of its efficiency and accuracy. Many freely available software tools implement hidden Markov models.

Hidden Markov models can useful in many different scenarios. For example for decision problems where the timing of events is crucial or where the crucial events may occur more than once. Using traditional decision trees for these types of settings usually require over simplification of the problem and unrealistic assumptions. In HMMs a patient is always in one of the finite number of states (See Figure 3). These states are called Markov states. All event in Markov models are represented by transitions from one state to another. [31]

We will explore how data from MRI (Magnetic Resonance Imaging) scans can be used to predict the presence of cancer in a patient after a treatment by using hidden Markov models. This results can also be helpful when deciding

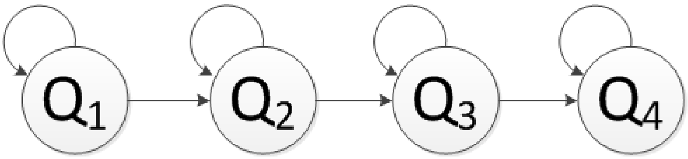

Figure 3: A simple hidden Markov model

whether a cancer patient should receive radiotherapy treatment or not. After modeling the HMMs, we will train and test our models by using leave-one-out cross-validation.

1.1.5 Leave-one-out cross validation

When evaluating statistical models cross-validation is usually used for testing quality of the models. It is especially useful when we do not have extremely large dataset because then we an uses all examples for both training and testing [32]. There are different types of cross-validation techniques. Among them there are k-fold cross-validation and leave-one-out cross-validation. When using K-fold cross-validation the data is divided into k disjoint subsets of same size. For *K* experiments *K-1* folds are used for training and the rest for testing.

Leave-one-out cross-validation is the more extreme form of k-fold cross-validation because we use all of the observations for both training and testing. As the name suggest, for each experiment one example is left-out. The computational cost can be very high for large datasets but it can be the best option for cross-validation if computational cost is not a big issue or the dataset is not too large.

1.2 Thesis definition

We formulate the thesis definition in the following manner: *"The purpose of this thesis is to see how hidden Markov models can be used to model spatial patterns in the cervical cancer data. To analyze the cervical cancer data, hidden Markov models are modeled and tested. The main effort should be to model and test the hidden Markov models on the dataset that is used in the prediction of the outcome of the Radiotherapy treatment".*

1.3 Research questions

In our research we will answer the following questions:

- How hidden Markov models can be used to model spatial patterns in the cervical cancer data?

The question is an overall problem statement, and covers the central research element i.e., to model spatial patterns in the cervical cancer data. The question also defines a method that should be applied to the problem. The method is the Hidden Markov Model (HMM) which is developed by Baum and Petrie [41], Baum and Eagon [42], Petrie [43] and Baum [44]. The HMM is especially used to find hidden statistical properties of the observed data. After modeling the HMMs, we will train and test our models by using leave-one-out cross-validation.

- Can these models be used to predict outcome of radiotherapy treatment with high enough accuracy?

After modeling the HMM on given cervical data, We will explore how data from MRI (Magnetic Resonance Imaging) scans can be used to predict the presence of cancer in a patient after a treatment. These results can also be helpful to decide whether a cancer patient should receive radiotherapy treatment or not. we want to predict the outcome in high enough accuracy so that it can be useable in real life.

1.4 Literature Review

In our initial finding, there has not been done any research in this particular type of cervical cancer data using hidden Markov models. Therefore we will only mention two research papers that are somewhat related to our thesis. We have chosen two papers for this review.

"Hidden Markov models in Bioinformatics with Application to Gene Finding in Human DNA", [29] in this paper researchers used hidden Markov models to find the genes in human DNA and examined how knowledge of the molecular mechanism of gene transcription guides the design of gene finding HMMs. After modeling the HMMs, they trained and tested them by using Baum-Welch Algorithm.

"Multistate Markov models for disease progression with classification error",[33] in this paper researcher tried to present general hidden Markov model for simultaneously estimating transition rates and probabilities of stage misclassification. To do the classification and to estimate the transition rates, Research used Baum-Welch Algorithm.

The above mentioned two papers are not directly relate to our problem, but mostly utilized to know about hidden Markov models (HMMs). In these papers, they have used HMMs to analyze their data. The same algorithm is applied in this thesis as well.

"Data mining approach to cervical cancer patients' analysis using clustering techniques" [2], this paper is somewhat similar to our project but they have analyzed the data using K-means clustering technique and the data used is based on patients' demographics.

We have documented all our references in the Bibliography.

1.5 Claim

We claim that hidden Markov models can be used to find the hidden patterns and properties in the cervical cancer data in an efficient way. By efficient we mean results that are fast to produce while at the same time being meaningful in this context. We also claim that these results can be helpful for medical experts and professionals in assistance and prediction of radiation therapy for women with cervical cancer.

1.6 Limitations and key Assumptions

1.6.1 Limitations

We have data from only 78 patients. Cancer is a very common disease. Therefore data from only 78 persons might not be enough to conduct any research with high accuracy.

Different approaches to analyze the data can affect the results. We will not be able to test all of these approaches due to time span of this thesis.

1.6.2 Key Assumptions

Our main assumption is that the dataset provided is representative for all population of cervical cancer patients. We are furthermore assuming that all the data we have is recorded correctly. We will not be doing any error detection of the values in the dataset except for removing the extreme values on both ends of the scale.

1.7 Contribution to the knowledge

The results from this project will contribute with new knowledge in both computer and medical sciences, especially the later one where little to no research has been done for knowledge discovery in MR data from cervical cancer using pattern recognition.

Our work can help in prognostication based on dynamic contrast enhanced MR (DESMR) imaging. A positive result can also help in identification of cervical cancer patients with high risk of treatment failure, and treatment selection and treatment planning of cervical cancer patients.

1.8 Target Audience

The target audience of this thesis is anyone that has interests within the machine learing techniques and cervical cancer treatment. Particularly, people that are interested in the usage of HMM in cervical cancer decision making. Medical experts, HMM experts and other people interested in this field may find this thesis interesting.

For the reader to fully understand the concepts and reasoning behind the solution, some knowledge of machine learning and a good understanding of common basic elements within the field of computer science is recommended. In addition there are several elements of probability theory mentioned in the sections explaining Bayesian networks. This means the reader should have some fundamental knowledge of probability theory, since we not intend to give a comprehensive introduction to probability theory in the thesis.

1.9 Report outline

The rest of the thesis is organized as follows: In Chapter 2 gives brief overview of what cervical cancer is, how Radiotherapy is done, what MRI is, what Pattern recognition is. Chapter 2 and 3 are intended to supply background information and also to be independent from each other. In Chapter 2 we also discussed how much pattern recognition is useful in the medical diagnosis. Chapter 3 is intended to present theoretical background of hidden Markov models (HMMs). i.e., what HMMs are, how HMMs are used in Bioinformatics, main problems involved in HMMs and solutions for them, how the parameters are estimated using using Baum-Welch algorithm. Chapter 4 represents the previous research in the fields of cervical cancer and hidden Markov models individually. Chapter 5 gives sufficient background and details.. Chapter 6 In Chapter 7 we discuss our main findings and some additional aspects of the HMMs. Chapter 8 is intended to wrap up, provide a conclusion and suggest interesting aspects that may be pursued in further research.

2 Theoretical Background

In this chapter we will describe details about different topics and techniques relevant to our project. Starting with the terminology used in this report we will give detailed information about cervical cancer, its risk factors, symptoms, treatment and prognosis. We will continue with data mining, pattern recognition, its uses, supervised and unsupervised learning along with various algorithms that can be used for pattern recognition. Theoretical background of the machine learning technique - hidden Markov models - we are using in our research is gone through in depth in a separate chapter, Chapter 3.

2.1 Terminology

There are certain terms and subjects repeated in the report are not defined as standard terminology. These terms and subjects will thoroughly be explained below. This section is written to make it easy for the readers to understand the medical terms which we have used in our report.

DESMRI – Dynamic Contrast Enhanced Magnetic Resonance Imaging is a medical test used by physicians to diagnose and treat different diseases. It uses magnetic field to produce detailed pictures of any internal body organs. These images can be examined on a computer monitor. In this project we have data collected by using DESMR images from 78 patients.

Voxel – is a volume entity. It represents a 3-dimensional element on a regular grid. It is characterized by x, y, z coordinates as well as other properties we want to study. The values are Hounsfield units in CT-scans and give the opacity of material to X-rays [7]. In our dataset each voxel is attributed by values related to the behavior of contrast medium.

Residiv – represents residual of cervical cancer after radiotherapy treatment. In this report, it is sometimes mentioned as outcome or class. It has two values, '0' or '1' which can be represented as class. If a treatment is successful then it means the patient data belongs to class '0' otherwise opposite [8].

Amplitude – total amount of contrast medium in a voxel [8].

K_{el} – transfer rate of the contrast medium out of the blood vessels [8].

K_{ep} – transfer rate of the contrast medium in to the blood vessels [8].

2.2 Cancer: an Introduction [34]

Human body is made up of some hundreds and millions of living cells and they orderly grow, divide and die in each and every stage of human life. In the early stage of human life, it is evident that normal cells are divided faster and make the person to grow. Once the person has become an adult, majority of human cells are divided only for the replacement of worn-out or dying cells resulted by injuries.

Cancer is generally caused with the irregular growth or out-of-control growth of abnormal cells. Most of the cancers start as the the growth of abnormal cells are not controlled or prevented.

In human body, normal cells continue to grow gradually and at a certain stage they are divided and soon died, whereas cancer cells continue to grow and form new abnormal cells and invade other tissues. A cancer cell is made, because the normal cell grows out of control and invade other tissues. It is also because of damage of DNA of normal cell. In a normal cell, when DNA is damaged, the cell can either repair the damage or die . This is not the case in connection with a cancer cell. In cancer cells, the damaged DNA is neither repaired nor dead, but gradually keep on making new cells that make the body worse and dangerous. These cells are not at all useful for the growth of human body. Every new cancer cell has the same damaged DNA as the first cancer cell has.

DNA damage can be caused because of environmental effects or mistakes that happen when the normal cell is reproduced. It is sometimes evident and clearly found that this is caused by toxics and harmful habits like drinking alcohol and smoking cigarettes. In general, cancer cells have the tendency to penetrate to other parts of the body and where they start to grow and form new cancer cells which ultimately lead to form tumors. Sometimes this process happens in blood circulation as well, which will cause other organs to get these cancer cells through the blood. It is not obvious in certain cancers like leukemia that cancer cells cause the tumors. Cancer can spread to any part in the human body, but it is always named where it began, but not where it is spread. If a cancer is caused to the tissue of lungs and spread to the breast then it is named as lung cancer but not as breast cancer. This process of penetration of cancer cells from one organ to another is called metastasis.

Every cancer has its own characteristics and has to be treated in different way. Despite the cause is the same as cancer cells, they grow at different levels and should be treated differently. Treatment should be to aim at targetting the particular type of cancer. No single treatment for all cancers. As stated earlier, cancer cells cause the tumors, but all tumors are not cancerous and such tumors are called benign tumors. Sometimes, benign tumors also give problems because they grow very large in size by causing press and pressure to neighboring healthy organs and tissues. Benign tumors are not able to invade other tissues and they are not life threatening. As per the records from World Health Organization, cancer is the main cause for 13% (7.4 millions) of all deaths worldwide in 2004. This rate is expected to rise up to 12 million deaths in 2030 (35)

2.2.1 Cervical Cancer

Cervical cancer is a cancer type formed in tissues of the cervix (the organ connecting the uterus and vagina). It is usually a slow-growing cancer that is mostly caused by HPV infection (Human Papilloma Virus).

2.2.1.1 Risk factors and causes of cervical cancer A risk factor is something that may increase the chances of developing a certain disease. Some women with certain type of risk factors are more likely to get cervical cancer than others.

There are many different risk factors that can increase the risk of developing cervical cancer. These risk factors can act together and increase the risk of getting cancer. Some of the major risk factors of cervical cancer are listed below.

- HPV infection
- Lack of regular Pap tests
- Smoking
- Weakened immune system
- Having man sex partners
- Using birth control pills for a long time
- Having many children

Symptoms

As mentioned earlier, cervical cancer is a slow-growing cancer. Therefor early stages of it may not cause any symptoms. As the tumor grows larger, women may notice one or more of the following symptoms:

- Abnormal vaginal bleeding
  - Between regular menstrual periods
  - After sexual intercourse, douching or a pelvic exam
  - After going through menopause
  - Menstruation that last longer than before
- Increased vaginal discharge
  - Pelvic pain
  - Pain during sex [9]

Infections or other health problems may also cause these symptoms.

2.2.1.2 Staging Cervical cancer is staged by the International Federation of Gynecology and Obstetrics (FIGO) staging system. It is based on clinical examination, and not surgical findings. It is only allowed to do certain types of diagnostic tests to determine a stage. These tests include palpation, inspection, colposcopy, endocervical curettage, hysteroscopy, cystoscopy, proctoscopy, intravenous urography, and X-ray examination of the lungs and skeleton and cervical conization [9].

The TNM staging system divides cervical cancer in 4 different stages (I to IV) with each stage having two sub stages (A and B).

Cervical cancer 5 Year Survival Rates by Stage [13, 14, 15]

"**Stage IA:** This is micro-invasive or very early cervical cancer. The five-year survival rate ranges from 96 to 99 percent. Treatment options for stage IA include surgery.

Stage IB: In this stage, the cancer is visible without the use of a microscope. Five-year survival rates for this stage of cervical cancer are 80 to 90 percent. Common treatments include surgery, chemotherapy and radiation.

Stage II: In stage II, cancer has spread outside the uterus to adjacent tissue, but has not reached the lower third of the vagina or all the way to the lateral wall of the pelvis. Five-year survival is 65 to 69 percent. Common treatment for stage II cervical cancer includes surgery, radiation and chemotherapy.

Stage III: Stage III cervical cancer indicates that the cancer has advanced beyond the parameters for stage II or has caused changes in the kidney. Five-year survival is 40 to 43 percent. Common treatments include chemotherapy and radiation.

Stage IV: Stage IV is the last stage of cervical cancer. In this stage the cancer has left the pelvis and affected more distant organs. The five-year survival rate for this stage of cancer is 15 to 20 percent. Types of treatment include chemotherapy and radiation". [13, 14, 15]

2.2.1.3 Treatment Stage I is usually treated by hysterectomy. It involves removal of the uterus. Part of the vagina and the lymph nodes are also removed. Women who want to remain fertile can get local surgical procedure such as a loop electrical excision procedure (LEEP) or cone biopsy.

Early stages (I and IIA less than 4 cm) are usually treated with radical hysterectomy and removal of the lymph nodes or radiation therapy. Radiation therapy is given as external beam radiotherapy to the pelvis and brachytherapy (internal radiation). Patients treated with surgery who have high risk features found on pathological examination are given radiation therapy with or without chemotherapy in order to reduce the risk of relapse.

Larger early stage tumors (IB and IIA more than 4 cm) can be treated with radiation therapy and cisplatin-based chemotherapy, hysterectomy or cisplatin chemotherapy followed by hysterectomy.

Advanced stage tumors (IIB-IVA) are treated with radiation therapy and cisplatin-based chemotherapy. [9]

2.2.1.4 Prognosis Prognosis of cervical cancer depends on the stage of the cancer. Patients that get treatment on early stages of the cancer, have survival 5 year survival rate of 92 %. The overall 5-year survival rate for all five stages is about 72%. [9]

According to the International Federation of Gynecology and Obstetrics, survival improves when radiotherapy is combined with cisplatin-based chemotherapy. If the cancer spreads to other parts of the body, prognosis drops radically. That is because treatment of local tumor is usually more effective than treatments of the whole body with radiation or chemotherapy.

Regular evaluation of the patient after treatment is vital. Early detection of recurring cervical cancer can be successfully treated with surgery, radiation, chemotherapy, or a combination of these. 35 % of patients of cervical cancer have persistent or recurrent cancer after treatment.

2.2.2 DCEMRI (Dynamic Contrast Enhanced Magnetic Resonance Imaging)

Magnetic Resonance is a procedure where the identification and characterization of the tumors is performed prior to the surgery. In this process, it is feasible and easy to get the images of the high level of soft tissue contrast. With the help of these images, the cancer can be identified and treated in time. "Dynamic Magnetic resonance imaging (MRI) is a diagnostic study that makes pictures of organs of the body using magnetic field and radio frequency pulses that cannot be felt. Dynamic contrast enhanced-magnetic resonance imaging (DCE-MRI) uses faster imaging and contrast material (a substance used to make specific organs, blood vessels, or tumors easier to see) that is given by vein. DCE-MRI gives extra information which is not available with the regular MRI. The regular MRI only shows pictures of the tumor while the DCE-MRI also gives information about the blood vessels of the tumor" [37, 61].

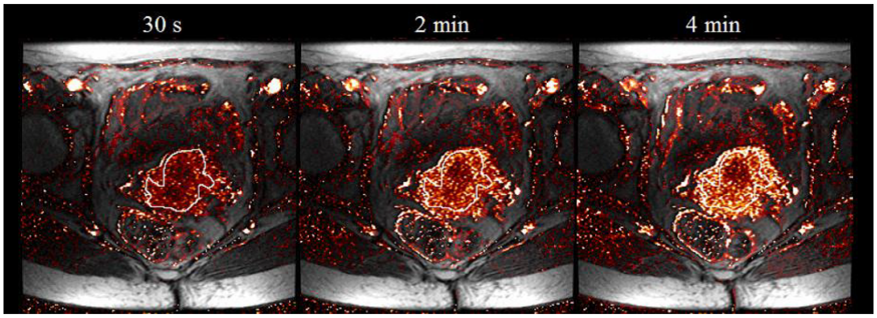

Figure 4: A Dynamic Contrast Enhanced MR (DESMR) image series showing the contrast enhancement pattern (red) of a patient with cervix cancer at different instances after contrast injection. The white contour is the macroscopic tumor.

2.2.3 Radiotherapy

Radiotherapy [62, 38, 39] is one of the procedures to treat the cancer where a high speed ionizing radiation is used to hit the DNA of cancer cells. This is also called radiation therapy. Damaged DNA of cancer cell is the target to kill and prevent the cell from growing and dividing. A Radiation Oncologist is a specialist in radiation treatment. There are two types of radiotherapy i.e external radiation and internal radiation. External radiation is performed by using a machine through which radiation comes, it is also a local treatment and it targets cancer cells only in the specified area. In internal radiation procedure, a small radioactive material is implanted into or near the tumor in order to kill the cancer cells directly. There are also some patients who need to be treated under both methods.

External radiation is generally performed on patients who do not need to be admitted in hospital for many days. They visit the hospital when the radiation therapy is actually needed. However, in internal radiation therapy, the patient has to stay in the hospital for several days, because a radioactive implant is fixed in the cancerous organ of the body. This implant can be temporary or permanent. The level of radiation is quite high in this method.[38]

Radiation therapy can cause to many side effects depending on the dose of radiation and organ of the body that is treated. Skin reactions, dizziness and loss of appetite are common side effects. Sometimes it can also cause to the reduction of white blood cells. [38]

In spite of few disadvantages, radiotherapy is considered as an effective treatment for certain cancers like uterine cervical cancer. [38, 39]. “However, the prognosis of advanced cervical cancer is not satisfactory, because the 5-year survival rate after radiation therapy of patients with stage II, III, and IV is reported to be about 70%, 50% and 30%, respectively”[39, 63, 64, 65]. After radiation therapy, it has also been a challenging issue in the improvement of the prognosis in patients with uterine cervical cancer. Based on some reports[63, 64, 65], there are also results and patterns of failures of radiation therapy for advanced cervical cancer. [39]

2.3 Pattern Recognition

Pattern recognition is commonly categorized into two categories: supervised and unsupervised learning. In supervised learning it is assumed that the provided dataset has correctly labeled inputs and outputs. The learning algorithm then tries to generalize the data as much as possible. On the other hand, unsupervised learning algorithms assumes unlabeled data and try to find patterns that can be used to predict the correct output. Recently, a combination of these two techniques has also been explored. It is called semi-supervised learning and uses a combination of labeled and unlabeled data.

Note that procedures that give same type of output in supervised and unsupervised learning may have different terms associated with the procedure. For example when talking about grouping of data, the term used in supervised learning is classification while it is called clustering in unsupervised learning.

An example from input data for which an output is generated, is usually known as an instance or a vector of features. These vectors define descriptions of all characteristics of an instance. It can for example be the coordinates in a multidimensional plane along with other attributes attached to the instance. These vectors can then be manipulated by using standard vector manipulations, for example dot product or angle between two vectors. Vector features are usually categorical, ordinal, integer-valued or real-valued. The first too are often grouped together. Many pattern recognition algorithms work only on categorical data.

There are many algorithms for patterns recognition. Their implementation depends of whether the learning is supervised or unsupervised. Some of the most used classifiers in supervised learning are Naive Bayes classifier, decision trees and k-nearest-neighbor. While k-means clustering and hierarchical clustering is used in unsupervised learning. Several of pattern recognition algorithms use statistical interference to find the best label for any given instance. Instead of just finding a single best label these algorithms can find N-best labels with associated probabilities.

For complex dataset sometimes it is necessary to apply transformation techniques to the raw feature vectors. These techniques include feature extraction which tries to reduce dimensionality of the feature vector, and feature selection which ignores irrelevant features.

2.3.1 Supervised and Unsupervised learning

The goal in supervised learning is to produce a classifier from input data. This is done by providing a set of inputs and outputs to build a model. While in unsupervised learning the main goal is to an internal representation of the statistical structure of the observations. In other words, one tries to determine how to do a task correctly in supervised learning, and to learn how the data is organized with unsupervised learning where the learner is input with unlabeled examples. [12]

All the observations are supposed to be caused by hidden variables, in unsupervised learning. That means that observations are at the end of a casual chain. Usually the probability of undefined input values is left out. This model is not needed as long as the input values are available. But it is not possible to understand anything about the outputs if some of input values are missing. In case of missing input values the inputs can also be modeled. Then they will be considered as hidden variables.

Figure above shows the difference between the causal structure of supervised (Figure 1a) and unsupervised (Figure 1b) learning. It is also possible to have a combination of the two, where both input observations and latent variables are supposed to have produced the output observations.

It is possible to learn larger and more complex models with unsupervised learning than supervised learning, because the aim of supervised learning is to find the relation between two sets of observations. The complexity of such an algorithm increases exponentially and the sets are almost impossible to process. Therefore it is almost impossible to learn from complex datasets with supervised learning.

Contrary to supervised learning, the learning can proceed hierarchically into more and more abstract levels of representation of the inputs. The complexity increases only linearly when moving from one step in the hierarchy to the next. Therefor it is much more feasible to use unsupervised learning for complex datasets, where the connection between inputs and outputs is very complex.

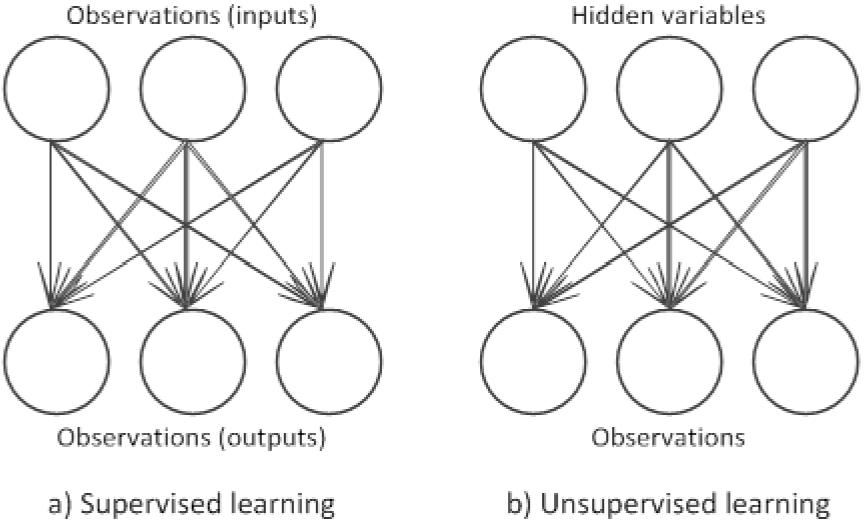

Figure 5: Difference between structure of supervised and unsupervised learning

This is illustrated in the figure below. Hidden variables in the top levels of abstraction are the causes for both sets and pass on the dependencies between inputs and outputs. The model is built upwards from both sets of observations so that in higher levels of abstraction the gap between inputs and outputs is easier to bridge. [12]

2.3.2 Usage

It is a matter of fact that practicing medicine is laden with ambiguity and confusion with many questions. Sometimes, physician are confused to take the right and accurate decision while treating the patients like in diagnosis, tests to perform, treatment to choose and etc. In spite of these uncertainties, physicians keep on making their efforts to get the excellent outcomes. Medical practice is conventionally guided by the pragmatic experience and observation like in anecdotes, case reports and clinical trials. With the help of health-maintenance organizations and group of physicians, a new source of population data is available. [40]

In the world of technology, computers are playing a significant role in medical and healthcare sector. Computer based medical tools give detailed and accurate clinical information and necessary help to determine how the computer based medical information should be used. It is not surprised to say sometimes that physicians and expertise have to depend and contact the computer-based medical programs to ascertain the proper advice in taking the right decisions in medical processes. Of late, Expert systems in Artificial Intelligence are playing a significant role in helping and replacement of human task. These systems give

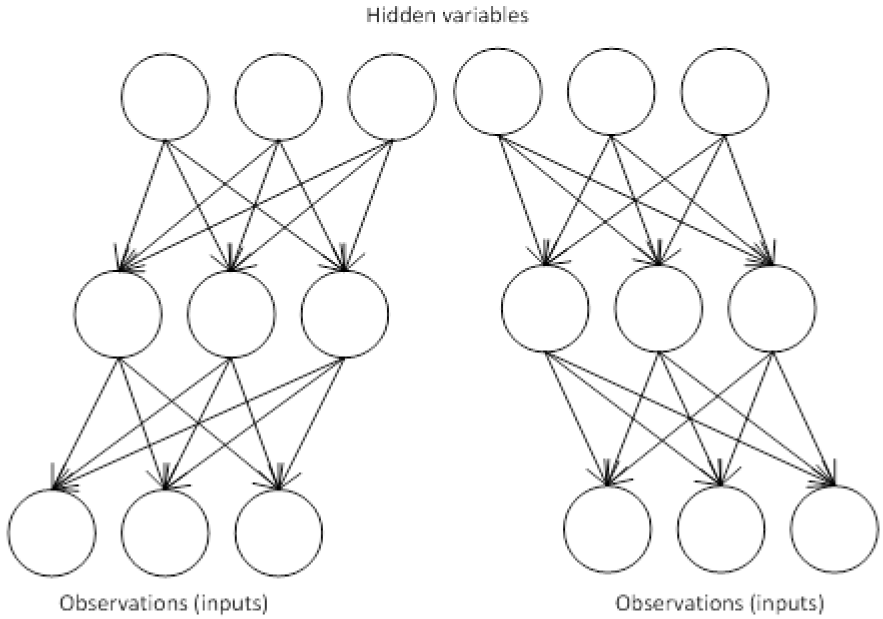

Figure 6: shows how unsupervised learning can be used to remove the gap between input and output observations

correct information and results where human experience and knowledge is not reliable.[40]

“Although there are domains where tasks can be specified by logic rules, other domains are characterized by inherent uncertainty. Probability was not taken into account, for some time, as a reasoning method for expert systems trying to model uncertain domains, because the computational requirements were considered too expensive. At the end of the 80s, Lauritzen and Spiegelhalter [14] shown that these difficulties can be overcome by exploiting the modular character of the graphical models associated with the so-called probabilistic expert systems”,[40] that in this work we call hidden Markov models.

Pattern recognition can be applied in several different types of data. It is the basis of computer-aided diagnosis (CAD) in medical sciences. CAD aids physicians to in diagnosis and treatment of patients.

Other applications of pattern recognitions include text classification (can be used to detect spam in e-mails), speech recognition, handwriting recognition and image recognition. Handwriting recognition can for example be used to automatically sort post based on the postal code written on the envelopes. Image recognition can be used by security departments to find suspicious persons.

3 Hidden Markov models

3.1 Introduction

"In 1940, HMMs were first studied but they could not be extensively applied in application. The Theory of HMM was firstly developed by Baum and Petrie [41], Baum and Eagon [42], Petrie [43] and Baum [44]. Due to the remarkable development in the field of computer science in the last 20 years, the utilization area of HMMs has broadly increased. The application area of HMMs are estimation of state space, the algorithms that would make the models usable and carrying out methods which are based on similarity. In the last years, statistical deduction studies were done by Leroux [45], Bickel and Ritov [46], Bickel, Ritov and Ryden [47] and Fuh [48]". [16]

In the fields of Engineering, if the behavior of the real-world objects and processes are distinguished by using models then it is a great interest of problem. Observable outputs are generally produced by processes.[23]

Modelling a process provides several objectives: Let us consider an example, if the theoretical background of a process is known then it is very easy to predict the behavior of the process plus that allows us to know how to build a process to get a desired output.[23]

There are two classes of models are available to characterize an objects behavior excellently: First one is, *Deterministic models.* In this type of model,the output dignal has some known parameters that are deterministic (like being a sine wave). This indicate that the randomness is not involved and it is always possible to determine its current state by virtue of model's parameters and its previous states. so it is obligatory only when the parameters are estimated to distinguish the process (for example, the frequency of the sine wave). *In Statistical models,* a process is described in terms of its random variables, i.e., its statistical properties. Let us consider an example, a process which is described by the mean considered to be Gaussian. Hidden Markov Models (HMM) are fall in this type of class.[23]

Hidden Markov Models theory has been widely used in many application like science, engineering, and many other areas(speech recognition, optical character recognition, machine translation, bioinformatics, computer vision, finance and economics, and in social science).

Before introducing Hidden Markov Models, Markov chains and markov processes are explained to clear term confusion.

3.2 Markov chains and Markov processes

A Markov chain is a discrete-time process. When the state of the past and the present is given, the future behavior of process will only depend on the present, but not on the past states. Markov chains and Markov processes are two vital groups of stochastic processes. Markov process can be considered as the continuous-time version of a Markov chain. It means the future behavior of the process can not be exactly forcasted or assumed. There are many queuing models which are also considered as markov processes. In this chapter, it is aimed to center on the characteristics needed for the modelling and analysis of queuing problems. [49]

3.2.1 Markov processes

A Markov process is a specific incidence of stochastic process. In this process, each state is connected to a finite set of states. Distribution of the probability of a state is specifically based only on the last states, but not on all other states.[30]. The outcome at any step is based only on the outcome of the earlier step. The probabilities are generally constant. [50]

3.2.1.1 Discrete Time markov processes Let us assume that a system which is in one of a set of N distinct states and it is listed by {1,2,...,N}. As per the set of probabilities connected with the state, in a regular interval, with discrete times the system goes through a change of state and which also indicates the same state and a repetition. The time instants are denoted as $t = 1,\ 2, ...$ The actual state at time t is indicated as q_t. Specification of the prevalent state at time t is required for a complete probabilistic description of the system and also all prior states. [51]

The below equation represents the first order Markov chain, so the probabilistic dependence can be restricted to only the previous state.

$$P\left(q_t \mid q_{t-1} = i,\ q_{t-2} = k, ..., q_1 = 1\right) = \left(q_t \mid q_{t-2} = i\right) \tag{1}$$

In addition, the processes which are at right-hand side of (1)are to be considered only if they are independent of time. The state-transition probabilities $a_{ij}$are thus obtained.

$$a_{ij} = P\left(q_t \mid q_{t-1} = i\right), \quad 1 \leq i,\ j \leq N \tag{2}$$

The state-transition probabilities have the below properties, because the a_{ij} follows the stochastic constraints.

$$a_{ij} \geq 0, \quad \forall\, j,\, i \tag{3}$$

and

$$\sum_{j=1}^{N} a_{ij} = 1, \quad \forall i \tag{4}$$

Each state represents to an observable action, at every instant of time, since the output of the process is the set of states. Therefore, this stochaastic process is called an *Observable Markov model.*

Below is explained a simple three state Hidden Markov Model.

Let us consider that a day's weather is noticed as follows:

- State 1: rain or snow
- State 2: cloudy
- State 3: sunny

It is emphasized that the weather on day t is indicated by any one of the three states given. The matrix A of state-transition probabilities a_{ij} is:

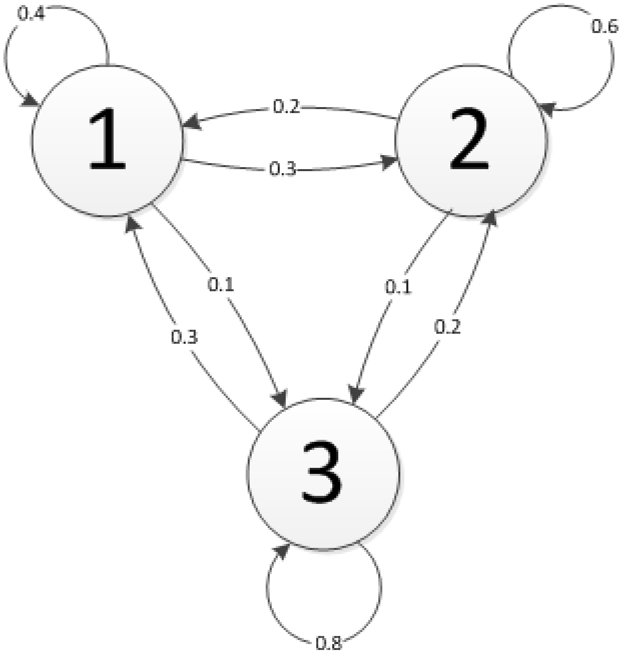

Figure 7: Markov model of the weather

$$A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

The questions regarding the behavior of the weather patterns over time can be solved by the above Markov model. For example, “What is the probability that the weather for eight consecutive days is sun-sun-sun-rain-rain-sun-cloudy-sun?”:

Observation sequence O can be represented as $O = (3, 3, 3, 1, 1, 3, 2, 3)$ and the probability distribution over the initial state πas

$$\pi_i = P(q_1 = i), \quad 1 \leq i \leq N \tag{5}$$

Then we can directly determine $P(O \mid Model)$, assuming we assign $\pi = (0, 0, 1)$:

$$\begin{aligned} P(O \mid Model) &= P(3, 3, 3, 1, 1, 3, 2, 3 \mid Model) \\ &= P(3)\,P(3 \mid 3)\,P(3 \mid 3)\,P(1 \mid 3)\,P(1 \mid 1)\,P(3 \mid 1)\,P(2 \mid 3)\,P(3 \mid 2) \\ &= \pi_3 \cdot (a_{33})^2 \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} = \underline{1.536 \cdot 10^{-4}} \end{aligned}$$

3.2.1.2 Limitations of Discrete time Markov processes If the output is connected to the state, Markov models cannot model random, in this model each state relates to a deterministically observable event. These models are very limited for many problems of interest like speech recognition [51]

3.2.2 Markov chain

A Markov chain is a first-order Markov process. The probability distribution of a state is depended only on the earlier state, but not on other states at a given time. When the present state is given, it is not necessary to consider past states, because in Markov process, the probability of the next state is depended only on the present state. Transition probabilities are considered as a finite set of possible states and transitions among those states are controlled by a set of conditional probabilities of the future state only when the present state is given. As transition probabilities are not dependent on time, Markov chains are also called as homogeneous and stationary.[30].

In order to give an example for a DNA sequence, the 'time' is considered as the position along the sequence, The consecutive transitions from one state to the next state, given an initial state, produce a time-evolution of the chain. It is thus extensively governed by a sequence of states which prior states are taken randomly.

3.3 Definition of hidden Markov models

We have now established the fact that HMM is a very powerful statistical method of finding characteristics in the observed data. We also know enough about Markov chains and processes to be able to define hidden Markov models. We can define a hidden Markov model by five elements: [16]

1. The number of hidden states, N in the model. Individual states are denoted as $S = \{s_1, s_2, \ldots\ldots, s_N\}$, and the state at time t as q_t.

2. The number of distinct observation symbols, M per state, the discrete alphabet size. These symbols represents the physical output of the system that we want to model. Individual symbols are denoted as $V = \{v_1, v_2 \ldots . v_M\}$. All observations are independent from the the previous ones. They are only dependent on the system state at the time of observation. [52]

3. The transition probability distribution matrix, $A = \{a_{ij}\}$. Dimension of A is $N \times N$. For

$$a_{ij} = P[q_{t+1} = (S_j | q_t = S_i)], \quad 1 \leq i, j \leq N \tag{6}$$

   a_{ij} also fulfills the following two conditions:

$$a_{ij} \geq 0, \;\; i, j = 1, 2, \ldots.., N \tag{7}$$

$$\sum_{j=1}^{N} a_{ij} = 1, \; i = 1, 2, \ldots.., N \tag{8}$$

   The state transition probabilities are independent of the observations and do not change over time. [53]

4. The observation symbol probability distribution matrix of state j, $B = \{b_j(k)\}$where:

$$b_j(k) = P[v_k at \;\; t \mid q_t = S_j], \; 1 \leq j \leq N, \quad 1 \leq k \leq M \tag{9}$$

   Dimension of B is $N \times M$ and it fulfills the following two conditions.

$$b_j(k) \geq 0, \quad 1 \leq j \leq N, \quad 1 \leq k \leq M \tag{10}$$

$$\sum_{k=1}^{M} b_j(k) = 1, \quad 1 \leq j \leq N \tag{11}$$

5. The initial state distribution matrix, $\pi = \{\pi_i\}$ where:

$$\pi_i = P[q_i = S_i], \quad 1 \leq i \leq N \tag{12}$$

A HMM is specified by two model parameters (N and M) and three probability measures (A, B and π). For convenience, we use the notation $\lambda = (A, B, \pi)$.

After defining a HMM using the five variables N, M, λ it can be used to generate an observation sequence:

$$O = O_1, O_2, \ldots\ldots.., O_T \tag{13}$$

where each observation O_t , is one of the observation symbols from V and T is the number of observations in the sequence [21].

HMMs are mostly defined by models having the state and measurements inside discrete set and discrete time [54]. They can be classified as continuous or discrete [27].

3.4 Basic problem and algorithms for HMMs

There are three main types of problems in the real-world application of HMMs, related to the evaluation, decoding and learning of the models [30]. We will take a deeper look into the most common algorithms used to solve these problems.

1. Evaluation problem: Probability of a given model generating a given sequence of observations is computed. The most common algorithms are:
    - The forward algorithm: Probability of emission distribution is computed starting at the beginning of the sequence.
    - The backward algorithm: Probability of emission distribution is computed starting at the end of the sequence.
2. Decoding problem: Find the most likely hidden states when a model and a set of observations are given. The most common algorithms are:
    - Viterbi algorithm: Find the sequence of internal states which has the highest probability.
    - Posterior decoding: For each position find the internal state with the highest probability.
3. Learning problem: Find an optimal model when a set of observations is given. The most common algorithms are:
    - Viterbi training algorithm: Uses the Viterbi algorithm recursively to find the optimal model.
    - Baum-Welch algorithm: Uses posterior decoding algorithm recursively to find the optimal model.

3.4.1 THE EVALUATION PROBLEM

The likelihood of a hidden Markov model is given by

$$P(E \mid \lambda) = \sum P\left(E \mid S; \lambda\right) \cdot P\left(S \mid \lambda\right), \tag{14}$$

where E is sequence of emissions and λ is a hidden Markov model.

Direct computation of the above sum for all the N^L possible sequences S of internal states has too high complexity. Therefore the forward and the backward algorithms are used. The complexity of these algorithms is $O(N^2L)$.

3.4.1.1 The Forward Algorithm We define auxiliary variables called forward variables, φ_k where $\varphi_k\left(u\right) = P\left(e_1, \ldots, e_k; s_k = u \mid \lambda\right)$ defines the probability of observing a partial sequence of emissions $e_1 \ldots e_k$ and a state $s_k = u$ at time k. See Algorithm 1 for details.

Algorithm 1 The forward algorithm

Initialization	$\varphi_1\left(u\right) = \pi_u b_u\left(e_1\right)$
Recursion$(for\, 1 \leq k \leq L)$	$\varphi_{k+1}\left(u\right) = b_u\left(e_{k+1}\right)\sum\limits_{v}\varphi_k\left(v\right).a_{v,u}$
Termination	$P\left(E \mid \lambda\right) = \sum\limits_{u}\varphi_L\left(u\right)$

3.4.1.2 The Backward Algorithm We define auxiliary variables called backward variables,β_k where $\beta_k\left(u\right) = P\left(e_{k+1}, \ldots, e_L \mid s_k = u; \lambda\right)$ defines the probability of observing a partial sequence of emissions $e_{k+1}, \ldots, e_L$ and a state $s_k = u$ at time k. See Algorithm 2 for details.

3.4.2 The Decoding Problem

In decoding problems we must find the most likely hidden states when a model and a set of observations is given. Two different algorithms are used for the two most common problems of this type: Viterbi algorithm and posterior decoding.

3.4.2.1 Viterbi Algorithm The Viterbi algorithm can be used to solve the decoding problem like the following:

Suppose a model λ and a sequence E of observed states are given. Find the sequence $S^*$of internal states that maximizes the probability $P(E, S \mid \lambda)$.

$$S^* \equiv argmax_S(P(E, S \mid \lambda)) \tag{15}$$

This can be achieved by following the steps in Algorithm 3.

3.4.2.2 Posterior Decoding Posterior decoding can be used to solve the decoding problem like the following:

Suppose a model λ and a sequence E of observed states are given. For each k in u, possible internal states, find the most probable internal state s_k^*.

Algorithm 2 The backward algorithm

Initialization	$\beta_L(u) = 1$
Recursion$(for\, L > k \geq 1)$	$\beta_k(u) = \sum_v \beta_{k+1}(v).a_{v,u}.b_v(e_{k+1})$
Termination:	$P(E \mid \lambda) = \sum_u \beta_1(u).\pi_u.b_u(e_1)$

Algorithm 3 Viterbi algorithm

Initialization	$\gamma_1(u) = b_u(e_1).\pi_u$ $\psi_1(u) = 0$
Recursion $(for 1 < k \leq L)$	$\gamma_k(u) = b_u(e_k).max_v(\gamma_{k-1}(v).a_{v,u})$ $\psi_k(u) = argmax_v(\gamma_{k-1}(v).a_{v,u})$
Termination	$P^* = max_v(\gamma_L(v))$ $s_L^* = argmax_v(\gamma_L(v))$
Backtracking $(for L > k \geq 1)$	$s_k^* = \psi_{k+1}\left(s_{k+1}^*\right)$

The forward and backward algorithm are used to find forward, φ and backward, β variables which are used to find the probability of each possible internal state. s_k^* is the internal state with highest probability, for each position of the sequence.

$$P\left(s_k = u \mid E\right) = \frac{\varphi_k(u).\beta_k(u)}{P(E \mid \lambda)}, \quad 1 < k \leq L \tag{16}$$

$$s_k^* = argmax_u\left(\varphi_k(u).\beta_k(u)\right), \quad 1 < k \leq L \tag{17}$$

3.4.3 THE LEARNING PROBLEM

We know the set of possible internal states, the set of possible external states, and a number of sequences of emissions. We hypothesize that the emissions originate from the same underlying HMM, and more specifically that each sequence of external states has been emitted from an associated sequence of internal states following the laws of the model.

The learning problem is to find the model when we have the possible sets of internal and external states and a number of emission sequences. In other words we want to determine the transition and emission probabilities. Let:

$E_j \equiv \left(e_k^j, k = 1, \ldots, L^j\right), \quad 1 \leq j \leq R$ be the given sequences of emissions, and

$S^j \equiv \left(s_k^j, k = 1, \ldots, L^j\right), \quad 1 \leq j \leq R$ the sequences of internal states.

After setting a stopping criteria and an initial guess for emission and transitions probabilities the probabilities and improved iteratively improved. The algorithm terminates when the stopping criteria is met. Viterbi training algorithm and Baum-Welch are most commonly used for learning problems.

3.4.3.1 Viterbi Training Algorithm When using Viterbi training algorithm, Viterbi decoding algorithm is used to derive the most internal state sequence with highest probability for each of the observations. Number of emissions and transitions is estimated using these most probable sequences. Model parameters can then be recalculate using these counts. See Algorithm 4 for details.

3.4.3.2 Baum-Welch Algorithm In Baum-Welch algorithm the posterior decoding algorithm is used to derive probability distribution of all the internal states. Number of emissions and transitions is estimated using the probability distributions. Model parameters can then be recalculate using these counts. See Algorithm 5 for details.

3.5 Advantages and disadvantages of HMMs

As most machine learning techniques hidden Markov models possess certain strengths and weaknesses. Even though HMMs can be very powerful it has

Algorithm 4 Viterbi training algorithm

Initialization	Initial guess	$A, B, \prod$
	Pseudocounts (the values to be added to the frequency counts)	$\widetilde{A}, \widetilde{B}$
Recursion	Calculate the most probable internal state sequence S^jusing the Viterbi decoding algorithm for each of the sequences	
	Calculate the observed frequency counts of transitions and of emissions. $A^{\hat{}}$ and $B^{\hat{}}$	$\hat{a}_{u,v} = \sum_j \sum_k \delta\left(u, s_k^j\right).\delta\left(v, s_{k+1}^j\right)$ $\hat{b}_u\left(x\right) = \sum_j \sum_k \delta\left(u, s_k^j\right).\delta\left(x.e_k^j\right)$ where $\delta$is the usual Kronecker delta.
	Calculate the regularized frequency counts	$\overline{A} = \hat{A} + \widetilde{A}$ $\overline{B} = \hat{B} + \widetilde{B}$
	Update the matrices A and B	$a_{u,v} = \frac{\overline{a}_{u,v}}{\sum_w \overline{a}_{u,w}}$ $b_u\left(x\right) = \frac{\overline{b}(x)}{\sum_y \overline{b}_u(y)}$
	Apply a similar updating to $\prod$	
Termination	Stop when the maximum number of iterations is reached or the convergence is too slow	

Algorithm 5 Baum-Welch algorithm

Initialization	Initial guess	$A, B, \prod$
	Pseudocounts (the values to be added to the frequency counts)	$\widetilde{A}, \widetilde{B}$
Recursion	calculate backward and forward coefficients using forward and backward algorithms for each sequence	
	Calculate the observed frequency counts of transitions and of emissions. $A\hat{}$ and $B\hat{}$	$\hat{a}_{u,v} = \sum_j \frac{1}{p\left(E^j \mid \lambda\right)} \sum_{k+1} \varphi_k^j\left(u\right).a_{u,v}$ $\hat{b}\left(x\right) \quad = \quad \sum_j \frac{1}{p\left(E^j \mid \lambda\right)} \sum_{k-1} \varphi_k^j\left(u\right)$ $.\beta_k^j\left(u\right).\delta\left(x.e_k^j\right)$ where δ is the usual Kronecker delta
	Calculate the regularized frequency counts	$\overline{A} = \hat{A} + \widetilde{A}$ $\overline{B} = \hat{B} + \widetilde{B}$
	Update the matrices A and B	$a_{u,v} = \frac{\overline{a}_{u,v}}{\sum_w a_{\bar{u},w}}$ $b_u\left(x\right) = \frac{\bar{b}\left(x\right)}{\sum_y \bar{b}_u\left(y\right)}$
	Apply a similar updating to $\prod$	
Termination	Stop when the maximum number of iterations is reached or the convergence is too slow	

some weaknesses which the designer should be aware of . In this section we will highlight the main advantages and disadvantages of HMMs.

3.5.1 Advantages

We will present 4 of the main advantages of HMMs. These are statistical grounding, modularity, incorporation of prior knowledge, and model transparency [58]. We will go through them briefly one-by-one.

3.5.1.1 Statistical grounding By statistical grounding we mean that we can use the stable and solid base of statistical mathematics to implement HMMs. This allows a lot of freedom in implementation while still being in sensible constraints of mathematics. Statistical analysis of individual steps of the process can also be obtained. These analysis can help us find weakness of a certain step.

3.5.1.2 Modularity Multiple HMMs can be easily combined together. This can be achieved by simply creating transitions from states of one HMM to states of another HMM. The probability of the newly created states can be set manually or a training set can be used to set them for the new model. This property of HMMs can be used to split complex training data into easily understandable models.

3.5.1.3 Incorporation of Prior Knowledge We can incorporate prior knowledge when designing and testing a HMM. This can be achieved in various ways: by adding previous knowledge into the HMM architecture, by using prior knowledge to constrain the model, or by initializing the model with prior knowledge before the training. These methods can be used alone or in combination to get a more stable and accurate training of data.

3.5.1.4 Model Transparency It is very easy to present the trained data visually using graphical representation. These graphs can be easily interpreted by human experts who can extract valuable information from them.

3.5.2 Disadvantages

We will present 3 of the main disadvantages of HMMs. These are the Markov Principle, speed and Over-fitting [58]. We will go through them briefly one-by-one.

3.5.2.1 The Markov Principle HMMs are first order Markov chains, which means that a pattern given a time t is only dependent on the pattern at *t-1*. That means that strict HMMs cannot get more precision in the model unless more states are added.

3.5.2.2 Speed HMM algorithms are very computational costly because they have to go through all the paths in the model. They are slow to process especially when the dataset is very large. By using effective and clever programming techniques the time can be decreased to an acceptable level.

3.5.2.3 Over-fitting Over-fitting occurs when a model is trained a way that avoids or limits generalization. This usually happens when the training data is very small. This can be avoided by using hold-out validation.

4 Related work

As mentioned in the first chapter there has not been done any research in this particular type of cervical cancer data using hidden Markov models. We will therefore divide previous research for cervical cancer and hidden Markov models individually.

4.1 Cervical Cancer

There has been lot of machine learning techniques like ANN, Bayesian Networks applied on cervical cancer. We will mention some of the most relevant among those.

"*Survival prediction using artificial neural networks in patients with uterine cervical cancer treated by radiation therapy alone*" [39]

Objective: In this paper, they evaluated the usefulness of artificial neutral networks (ANN's) for the survival prediction in patients with uterine cervical cancer treated by radiotherapy.

Methods: They used data from 134 patients with uterine cervical cancer treated by combined external and high dose-rate remote after loading Intracavitary radiotherapy between 19 78 and 1993. The ANNs were trained using the data from 67 randomly selected patients. Using the trained ANN's they predicted the 5-year survival in the remaining 67 patients, and compared it with the known 5-year survival. The performance of the ANNs was evaluated using a receiver operating characteristic (ROC) curve curve and was compared using the area under the ROC curve (Az).

Results: When fundamental factors, such as age, performance status, hemoglobin, total protein, International Federation of Gynecology and Obstetrics (FIGO) stage, and historical type were used as inputs in the ANNs. When the historical grading of radiation effect determined by periodic biopsy, the examination was used in addition to the fundamental factors. When the cytological grading of radiation effect by the periodic smear was used in addition to the fundamental factors, which was not significantly different from that when only the fundamental factors were used.

Conclusion: ANNs allowed them to evaluate the importance of prognostic factors, and make it possible to predict the survival of each patient. Using ANNs, the combination of histological grading of radiation effect determined by periodic biopsy examination, in addition to the fundamental factors, is the most effective for prediction of survival in patients with uterine cervical cancer.

"*Bayesian Model Combination and Its Application to Cervical detection*" [10]

Objective: In this paper, Researchers proposed a method to combine several models using a Bayesian approach. The method selects the most relevant attributes from several models, and produces a Bayesian classifier that has a high accuracy for cervical . They applied this method for the diagnosis of precursor lesions of cervical cancer.

Methods: They used 1055 sample data labeled by an expert. They used a holdout testing procedure, with approx. 2/3 of the data for training and 1/3 for testing (800 for training and 255 for testing). They developed a methodology to combine several models using a Bayesian approach. The method selects the most relevant attributes from several models, and produces a Bayesian classifier that has a high accuracy for cervical . Based on conditional information measures, the method eliminates irrelevant variables, and joins or eliminates dependent variables; until an optimal Bayesian classifier is obtained.

Results: They evaluated a methodology in the classification of different regions of colposcopy videos. Firstly, some image sequences were classified with the help of an expert. Secondly, based on the training cases, the time series were obtained and described using the 3 models. Finally, the model combination methodology was applied and a Bayesian classifier was generated that combines the 3 models. This combined classifier was tested with other image sequences, different to the ones for training.

Conclusion: Researchers applied above mentioned method for the analysis of colposcopy images for diagnosis of cervical cancer. For this they used the parameters from three mathematical models that characterize the temporal evolution of each pixel in the image. Each model has different number of parameters, in total 11 attributes, all continuous. There are three classes. Their method produces a very simple and efficient classifier with an accuracy of 95%.

4.2 Hidden Markov models

"*Computer-Aided Prostate in Ultrasono-graphic-images*" [59]

Objective: Prostate cancer is one of the most frequent cancers in men and a major cause of mortality in developed countries. Detection of the prostate carcinoma at an early stage is crucial for a successful treatment. In this paper, a system for computer-aided, TRUS-based detection of prostate cancer was presented

Methods: In this paper, Researchers used images from urology department in Spain where a corpus of 4944 images was acquired from 1648 biopsy sessions (3 images per session) involving 301 patients (5 to 6 biopsies per patient). On these images, they used K-NN and HMM classification to predict the malignancy of a region around a pixel in a TRUS image of a previously unknown patient. The training process of HM models was carried out using the well-known instance of the EM algorithm called backward-forward or Baum-Welch re-estimation.

Results: Based on classification schemes, HMM performed slightly better than k-NN when working with gray maps, however, no significant difference is found between both classifiers when working with Spatial Gray Level Dependence Matrices (SGLDM).

Conclusion: A system for computer-aided, transrectal ultrasonography-based detection of prostate cancer had been presented. The aim of the system was to help an expert decide where to perform a biopsy. Two classifiers based on k-Nearest Neighbors and hidden Markov models are compared.

"*Application of Hidden Markov models to Gene Prediction in DNA*" [60]

Objective: Identification or prediction of duplicate sequence from within genomic regions has been a major rate limiting step in the search of genes. In this paper, a HMM is developed to detect the gene structure.

Methods: In this paper, Researchers develop a hidden Markov model (HMM) to represent the degeneracy features of splicing junction donor sites in eucaryotic genes. The HMM system is fully trained using an expectation maximization algorithm and the system performance is evaluated using the 10-way crossvalidation method. Researchers used only the local information in their research for the donor classification.

Results: The modeled HMM system correctly classified more than 95% of the candidate sequences into the right categories. More than 91% of the true donor sites and 97% of the false donor sites in the test data were classified correctly.

Conclusion: A hidden Markov model had been developed to detect the structure of the genes. the aim of the the modeled HMM was to represent the degeneracy features of splicing junction donor sites in eucaryotic genes. the HMM system was completely trained by using an expectation maximization algorithm and the performance was evaluated by using 10-way cross validation method.

5 Solution

In this chapter we will present our solution and results. First, we will go through the requirements of this project. Then we will move on to the implementation where we talk about selection of technology for our data analysis, show the preprocessing of the data and how we used the selected technology to process the data. We will end this chapter with presenting our results.

The requirement of this project was to use Bayesian networks to analyze and extract any useful patterns from the provided dataset. To train and classify the data our main priority should be to use existing technology instead of reinventing the wheel.

5.1 Implementation

We chose to use continuous hidden Markov models (HMM) which is a Markov model with hidden states. It can be seen as the simplest form of Bayesian Networks. We tested several implementations of HMM in various programming languages but found that those were not designed for over specific needs. Those rejected implementations include UDMHMM (C++), Jahmm (Java), HMM Toolbox (Matlab) and Mendel HMM Toolbox (Matlab).

Finally, we decided to use General hidden Markov model library (GHMM) which is a C library implementing basic and extended hidden Markov models. Due to limitations of discrete HMMs only in GHMM we were not able to use its graphical user interface. Therefore we used an additional Python-wrapper library to modify and utilize GHMM for our purpose.

5.1.1 Dataset pre-processing

We were given post-treatment data of 78 patients where 46 where were cured and 32 were not cured. Total number of examples was 868,810 and the ratio between data from cured and not cured patients was 0.43 and 0.57. Table 1 shows the first few lines of the dataset. We removed the irrelevant columns and information (see Section 7.1 for a discussion). This was accomplished through a Python-script.

Table 1: Shows the first 20 lines from the original dataset.

```
# name: mm001,residiv:        0,stime:       66.7870
#          x,          y,          z,   Amplitude,         kep,          kel
         119         114           1     1.11368,     1.97132,     0.000000
         120         114           1     1.34914,     1.87449,     0.000000
         121         114           1     1.85973,     2.29584,     0.000000
         122         114           1     2.02015,     2.10357,     0.000000
         123         114           1     1.77714,     1.41152,     0.000000
         124         114           1     1.52227,     1.22585,     0.000000
         125         114           1     1.36091,     2.03624,     0.000000
         126         114           1     1.64858,     2.07266,     0.000000
         127         114           1     1.74830,     1.85514,     0.000000
         128         114           1     1.67203,     1.66085,     0.000000
         114         115           1     1.02867,     1.18023,     0.000000
         115         115           1     1.08240,     1.17235,     0.000000
         116         115           1     1.30867,     1.34240,     0.000000
         117         115           1     1.44298,     1.70905,     0.000000
         118         115           1     1.36391,     1.68453,     0.000000
         119         115           1     1.15545,     1.53698,     0.000000
         120         115           1     1.24364,     1.67675,     0.000000
         121         115           1     1.70529,     1.94580,     0.000000
         122         115           1     1.69368,     1.62077,     0.000000
         123         115           1     1.66562,     1.38886,     0.000000
```

Name and stime were removed. K_{ep} and K_{el}were also removed due to their inaccuracy. This means that we are only considering the coordinates of a voxels and the related amplitude of the contrast medium. The data is then grouped into two subsets: one containing data from cured (residiv = 0) patients and the other for non-cured (residiv = 1). The amplitude for all the voxels in each subset is then collected by traversing the data in a zigzag fashion. It is visualized in Figure 8 and represents tumor area for a patient. The 3 layers represents z-coordinate while the x- and y-coordinates are represented by the grid on each layer. The red arrows show alternating direction of the traversal. The direction is changed for each row as well as for each layer. When one layer is processed from top to bottom, the next is processed from bottom to top.

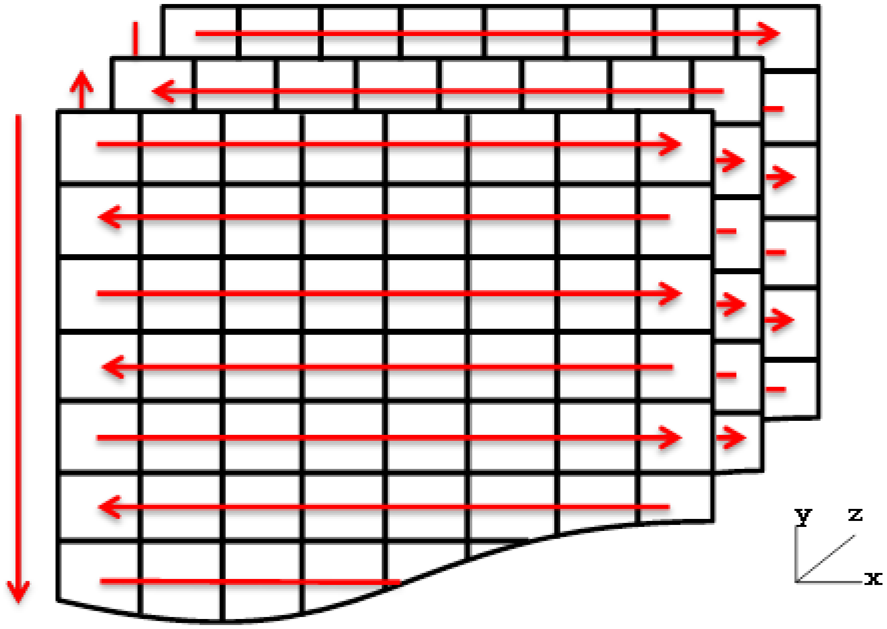

Figure 8: Shows how the data is traversed row by row.

The resulting data are two arrays of arrays, one each of the two groups of patients. Each of the sub-arrays represents one patient, and contains zigzagged values of amplitude. These two arrays are then used for training and classification of the data.

5.1.2 Training and classification

We set number of hidden states, N between 2 and 10. For each N we used a $N \times N$ transition matrix, A, and two transitions with probability of 0.5. One transition to the state itself and the second one to the neighboring state.

$$A = \begin{bmatrix} 0.5 & 0.5 & 0 & \dots & 0 \\ 0 & 0.5 & 0.5 & \dots & 0 \\ 0 & 0 & 0.5 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 \end{bmatrix} \tag{18}$$

The probability distribution, B is set to $N \times 2$ matrix since we only have 2 possible observation.

$$B = \begin{bmatrix} random(5) & random(25) \\ random(5) & random(25) \\ \vdots & \vdots \end{bmatrix} \tag{19}$$

while the initial probability is a $1 \times N$ matrix where $N_{1j} = {}^1\!/_n$

$$\pi_i = \begin{bmatrix} {}^1\!/_n & {}^1\!/_n & \cdots \end{bmatrix} \tag{20}$$

We create two model, M_0 and M_1 by using Gaussian distribution and λ (3 matrices mentioned above).

One random example is left-out from either of the two groups, Seq_0 and Seq_1 of the patients. Then, we use Baum-Welch algorithm and log-likelihood for each group of patients to train and learn the two models. We find log-likelihood, L_0 and L_1 using both groups of data for each of the models. We then use a merged sorted list of differences, D between them ($L_0 - L_1$for M_0 and $L_1 - L_0$for M_1) to find minimum error rate, E_{min} and threshold, T. Note that while merging the list we keep track of which model each element in D belongs to.

Initially, we set $E_{min} = 1.0$, $T = 0$ and assume that one group of patients is always classified correctly, $E_0 = 0$ while the other always incorrectly, $E_1 = l_{seq1}$ *(length of* Seq_1). Error is defined as:

$$E = \frac{E_0 + E_1}{l_{seq0} + l_{seq1}}$$

This calculation is repeated for all elements in D. Threshold is set to the current element of D if $E < E_{min}$. E_0 is increased by 1 if the the element belongs to E_0 while E_1 is decreased by 1 if the the element belongs to E_1.

After determining threshold we find log-likelihood of both models using the ignored example, i as the test data. We denote this as L_{0i} and L_{1i}. We use difference between them to determine whether the classification is correct or not.

The classification is correct if $L_{0i} - L_{1i} < T$ and the ignored example belonged to Seq_0, or if $L_{1i} - L_{0i} < T$ and the ignored example belonged to Seq_1.

The above procedure is repeated 1000 times for $2 \leq N \leq 8$. For each of the 1000 iteration we store values of E_{min} and the classification, c (0 for correct and 1 for incorrect classification). These pairs are sorted by ascending order of E_{min}. We calculate the accuracy of the top x percents of the classifications.

$$x\epsilon(1, 2, \ldots 9, 10, 20, \ldots 90, 100)$$

The accuracy of the classification is defined as:

$$\frac{1}{10x} \sum_{n=0}^{x} n \tag{21}$$

The accuracy for each x is plotted in a diagram. These diagrams are used to evaluate the results by comparing the guessed error rate, E_g. The results are presented in the next subsection of this chapter.

$$E_g = \frac{min\left\{l_{seq0}, l_{seg1}\right\}}{l_{seq0} + l_{seq1}} \tag{22}$$

6 Results

We analyzed data from 78 patients where 46 where were cured and 32 were not cured after radiotherapy. This gives us $E_g = {}^{31}/_{77} \approx 0.406$ or $E_g = {}^{32}/_{77} \approx 0.413$ by using Equation 22.

On the next pages we present results produced after 1000 iterations of HMM with leave-one-out cross-validation. The number of hidden states ranges between *2-8*. Results for each state are prestsented in a diagram. Note, that the diagrams are in logarithmic x-axis. (See Chapter 5.1.2for details about calulations of the diagrams).

The accuracy of the iterations with least error rate was very low. It was even lower than E_g most of the times. Table 2 shows 20 iterations with least error rate for two hidden states. The accuracy of the prediction is only 0.15 which is very low considering our initial estimates of HMMs accuracy.

Table 3 is used for plotting the accuracy of prediction in Figure 9. Similar tables are used to plot diagrams for other states.

We can observe in Figure 9-15 that the accuracy of prediction is usually increasing for when increasing number of iterations are included. But in those cases the error rate is too high to consider those classifications.

An accuracy of 0.6 for the 10 % iterations would have been a good result. But we never achieved that in our experiments.

Table 2: The top 20 iterations and the corresponding prediction

Error Rate	Prediction
0.285714285714	1
0.298701298701	0
0.298701298701	0
0.298701298701	0
0.298701298701	0
0.298701298701	0
0.298701298701	0
0.311688311688	0
0.311688311688	0
0.311688311688	0
0.311688311688	0
0.311688311688	0
0.311688311688	0
0.311688311688	1
0.311688311688	0
0.311688311688	1
0.311688311688	0
0.311688311688	0
0.311688311688	0
0.311688311688	0

Table 3: The results for 2 states.

Iterations	Accuracy
1 %	0,1000
2 %	0,1500
3 %	0,1667
4 %	0,2000
5 %	0,1800
6 %	0,2167
7 %	0,2143
8 %	0,2250
9 %	0,2778
10 %	0,3100
20 %	0,4300
30 %	0,4600
40 %	0,4675
50 %	0,4520
60 %	0,4267
70 %	0,3986
80 %	0,3888
90 %	0.3989

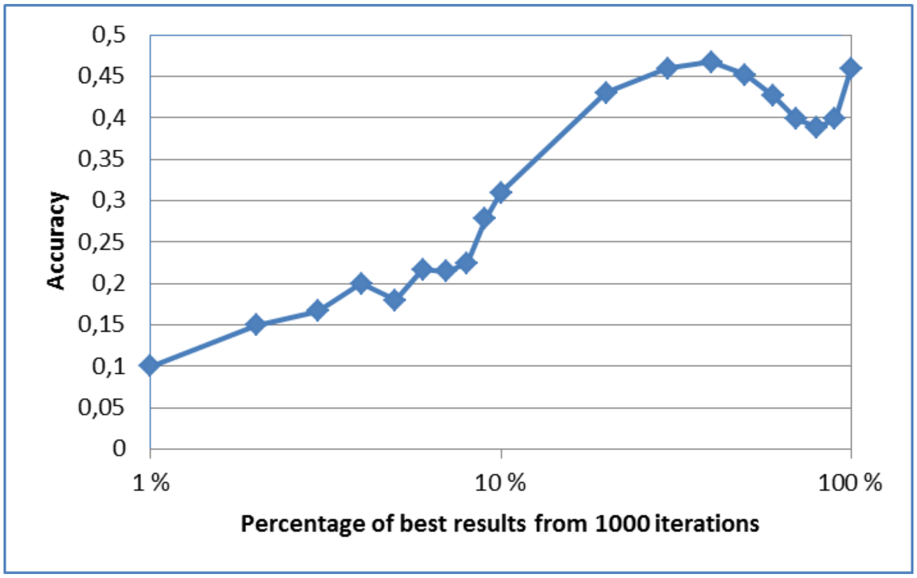

Figure 9: Results for 2 states

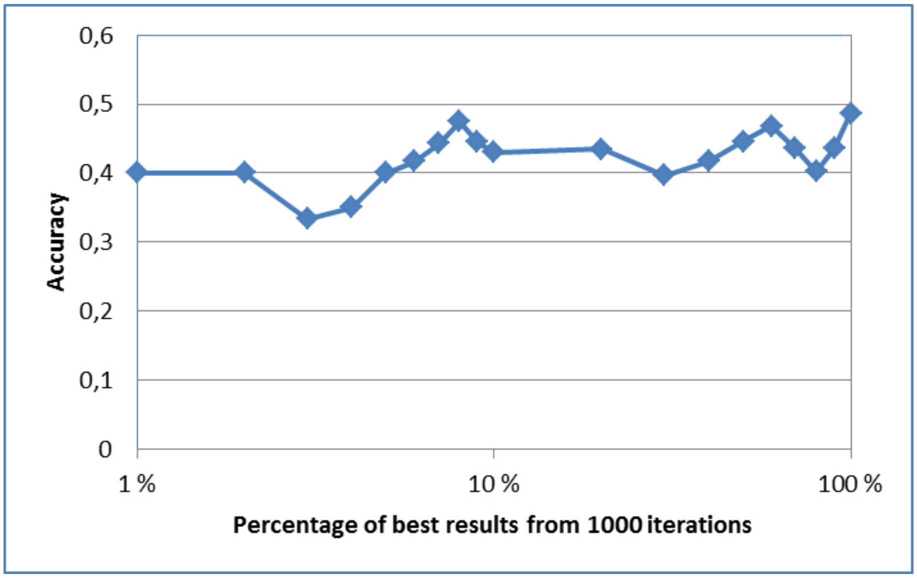

Figure 10: Results for 3 states

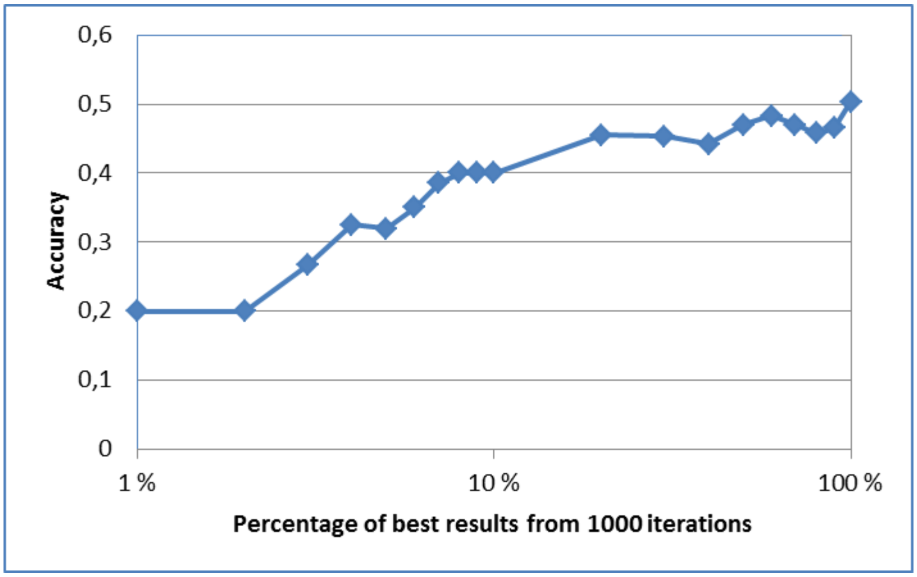

Figure 11: Results for 4 states

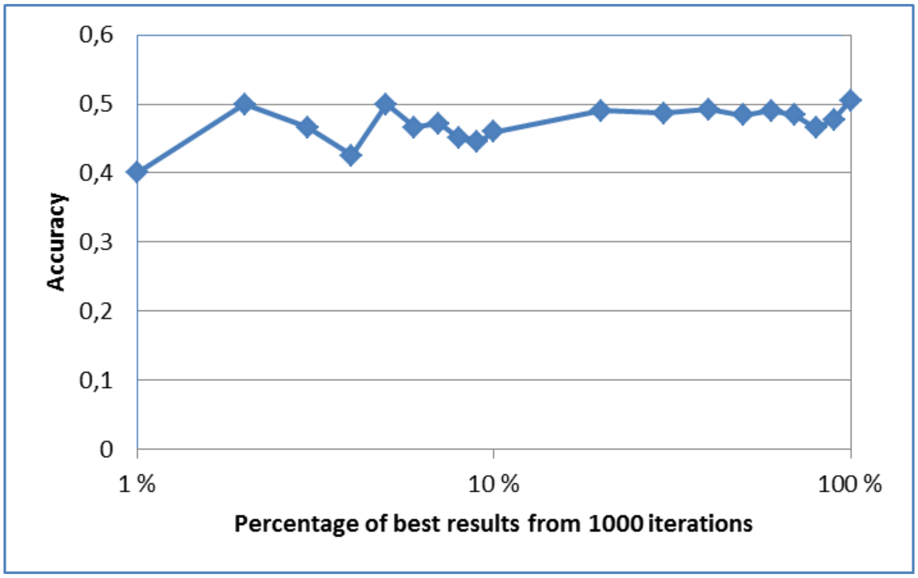

Figure 12: Results for 5 states

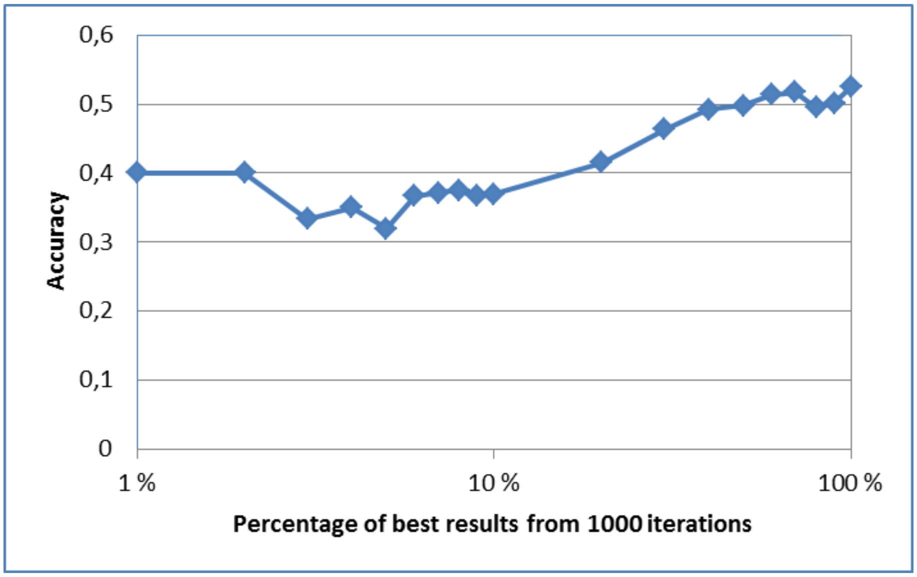

Figure 13: Results for 6 states

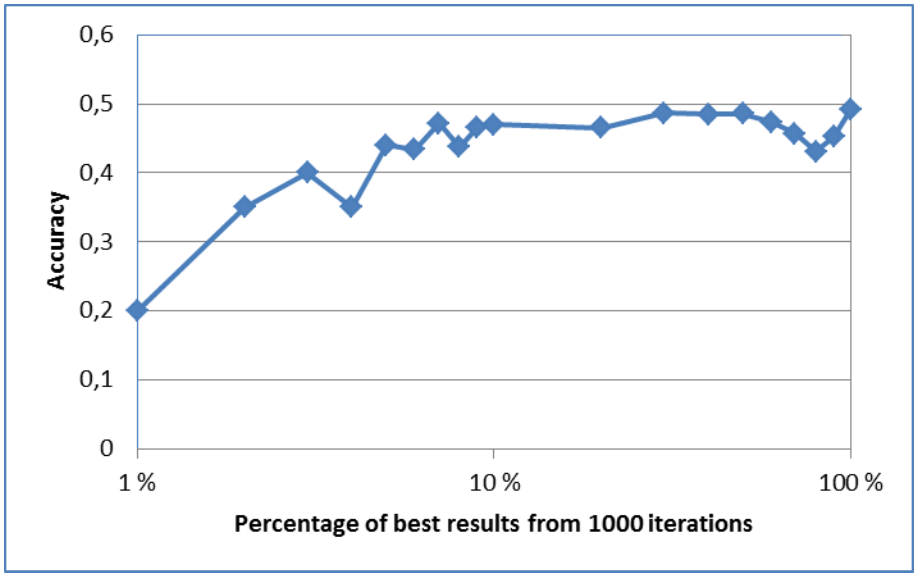

Figure 14: Results for 7 states

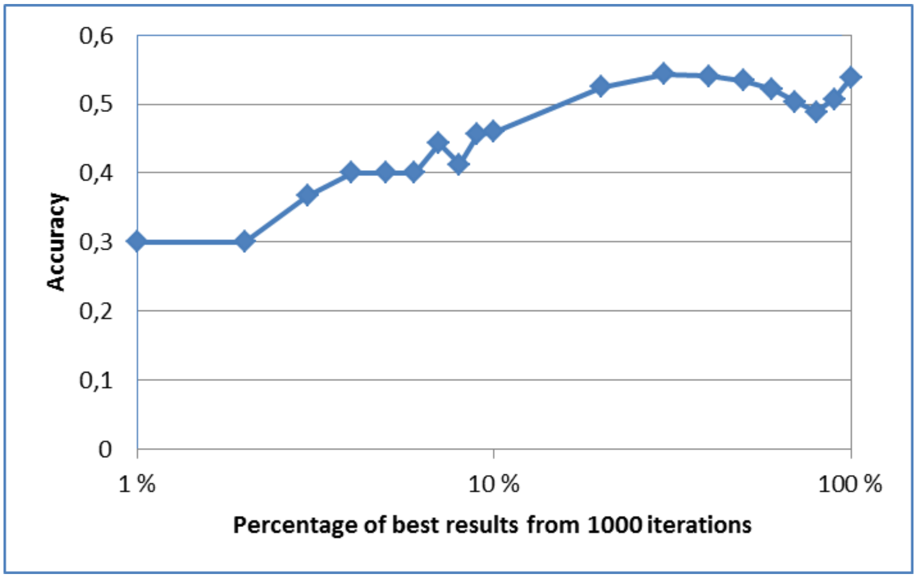

Figure 15: Results for 8 states

7 Discussion

Our aim for this project was to analyze a large set of post-treatment cervical cancer data. The dataset had data from dynamic contrast enhanced MR imaging along with the outcome of the treatment. We wanted to find any relations between contrast medium data and the outcome of the treatment by applying different unsupervised learning algorithms. Our initial research showed that hidden Markov model was the best suitable algorithm for this dataset.

7.1 Dataset

We were given post-treatment data of 78 patients where 46 where were cured and 32 were not cured. Total number of examples was 868,810 and the ratio between data from cured and not cured patients is *0.43* and *0.57*. In other words 57 % of the data is from 41 % of the not-cured patients. This means that most of the non-cured patients had large tumor areas than the 46 cured patients. In other words, patients with relatively large tumors are less likely to get cured by radiotherapy.

Another problem we faced was the extreme values in the sample sets. The training of data usually resulted in non-converging models when those values were included. Luckily, examples with extreme values were only a fraction of the whole dataset therefore the problem with non-converging models was easily overcome by removing those examples.

We chose to only focus on one property of the voxel, Amplitude. Flow of contrast medium in and out of the blood vessels, K_{ep} and K_{el} were ignored. It was because these variables were not recorded well enough or the values were too similar and small for all the examples. A less erroneous recording would have been used to a more accurate classification.

7.2 Hidden Markov models

Our thorough evaluation of the HMM on the data set, using a wide range of states, classification thresholds, and learning runs, shows that the HMM is able to find patterns that clearly discriminates between the two classes appearing in the training set. However, when leave-one-out cross-validation is introduced, this discriminative power disappears completely, leaving the HMM unable to beat even a majority classifier. This indicates that linear scanning of the voxels, using HMM for learning and classification, being a reasonable approach, clearly is unsuccessful in practice on our data set.

The question we set out to answer by our research were:

- How hidden Markov models can be used to model spatial patterns in the cervical cancer data?

We implemented hidden Markov models using Baum-Welch algorithm and leave-one-out cross-validation. Considering the size of the dataset our implementation is very efficient for small number of hidden states.

- Can these models be used to predict outcome of radiotherapy treatment with high enough accuracy?

The simple answer to this question is unluckily “No”. Even though we did extensive initial research showed that HMM would be able to find hidden patterns in our dataset but that was not the case when we used leave-one-out cross-validation. An accuracy of 0.6 for the 10 % iterations would have been a good result. But we never achieved that in our experiments.

Our initial claim of being able to find hidden patterns in the dataset is therefore rendered useless but we can conclude that we need a more sophisticated approach based on dynamic Bayesian networks to find hidden patterns, if any, in the dataset.

8 Conclusion

In this project our job was to analyze data from dynamic contrast enhanced MR imaging data recorded from cervical cancer patients after receiving radiotherapy treatment. We wanted to find any relations between the outcome of the treatment and the behavior of the amplitude in the tumor area. And whether we could use that relation to predict outcome of a patient before the treatment.

We used hidden Markov models to explore and analyze the dataset. Our thorough evaluation of the HMM on the dataset using a wide range of states, classification thresholds, and learning runs, showed that the HMM is able to find patterns that clearly discriminates between the two classes appearing in the training set. However, when leave-one-out cross-validation is introduced, this discriminative power disappears completely, leaving the HMM unable to beat even a majority classifier. This indicates that linear scanning of the voxels, using HMM for learning and classification, being a reasonable approach, clearly is unsuccessful in practice on our data set.

Our approach of using linear traversing of the dataset and using HMM for learning and classification was not successful. As further work on this dataset we would like to suggest a more sophisticated approach based on dynamic Bayesian networks and taking more of the spatial relationships into account. This approach may reveal any potential hidden patterns in the dataset.

References

[1] Abarghouei, A.A., Ghanizadeh, A., Sinaie S., & Shamsuddin, S.M. (2009). "A survey of Pattern Recognition applications in cancer diagnosis". *International Conference of Soft Computing and Pattern Recognition.*

[2] Thangavel, K., Jaganathan, P.P., & Easmi, P.O. (2006). "Data mining approach to cervical cancer patients' analysis using clustering technique". *Asian journal of Information Technology*, 5(4), 413- 417.

[3] J. M. Walboomers, M. V. Jacobs, M. M. Manos, F. X. Bosch, J4A. Kummer, K. V. Shah, P. J. Snijders, J. Peto, C. J. Meijer, and N. Munoz,"Human papillomavirus is a necessary cause of invasive cervical cancer worldwide," J Pathol., vol. 189, pp. 12–29, 1999

[4] T Bjøre, and ø Kravdal, "Reproductive factors and prognosis of uterine cervical cancer in Norway", *British Journal of cancer*, vol 74, pp.1843-1846,1996.

[5] D. D. Davey and R. J. Zarbo, "Human papillomavirus testing—Are you ready for a new era in cervical cancer screening?," *Arch. Pathol. Lab. Med.*, vol. 127, pp. 927–929, 2003

[6] F. X. Bosch, A. Lorincz, N. Munoz, C. J. Meijer, and K. V. Shah, "The causal relation between human papillomavirus and cervical cancer," *J. Clin. Pathol.*, vol. 55, pp. 244–65, 2002.

[7] H. zur Hausen, "Papillomaviruses causing cancer: Evasion from host-cell control in early events in carcinogenesis," *J Nat. Cancer Inst.,* vol. 92, pp. 690–698, 2000.

[8] F. P. Perera, "Environment and cancer: Who are susceptible?," *Science,vol.* 278, pp. 1068–73, 1997.

[9] E. S. Calhoun, R. M. McGovern, C. A. Janney, J. R. Cerhan, S. J. Iturria,D. I. Smith, B. S. Gostout, and D. H. Persing, "Host genetic polymorphism analysis in cervical cancer," *Clin. Chem.,* vol. 48, pp. 1218–1224, 2002

[10] Miriam.M, Luis.E.S, Hector.G.Acosta and Nicandro.C, "Bayesian Model Combination and Its Application to Cervical ", J.S. Sichman et al. (Eds.): IBERAMIA-SBIA 2006, LNAI 4140, pp. 622–631, 2006, *Springer-Verlag Berlin Heidelberg 2006.*

[11] Alireza.O and Bita.S, (2010) "Machine learning Techniques to diagnose Breast Cancer"*, Journal of 5th International symposium on Health Informatics and BioInformatics.*

[12] Friedman.N, Dan.G and Moises.G, (1997) "Bayesian Network Classifiers".*Journal of Machine learning*, 29,131-163

[13] Lic.Laura.L and Ing.A.De.G, “Pattern Recognition in Medical Images Using Neutral Networks”.

Lanzarini, Vargas, Estelrrich and De Giusti,“Real Time Analysis of the Nystagmus and Movement Patterns in Balance Disturbances”. *19th International Conference Information Technology Interfaces.* Croatia. 1997.

[15] V.M. Mantyla, Discrete hidden Markov models with application to isolated user-dependent hand gesture recognition, Finland: *Valtion Teknillinen Tutkimuskeskus Publications,* pp. 35, 2001.

[16] Ersoy oz, “Algorithms of Hidden Markov model and a Prediction Method on Product Preferences”. *Journal of computing*, Vol II. (2010)

[17] B. Haubold, T. Wiehe, Introduction to Computational Biology An Evolutionary Approach, *Basel-Switzerland: Birkhauser Verlag,* pp. 114, 2006.

[18] Ries LAG, Harkins D, Krapcho M, Mariotto A, Miller BA, Feuer EJ, Clegg L, Eisner MP, Horner MJ, Howlader N, Hayat M, Hankey BF, Edwards BK (eds). *SEER Cancer Statistics Review*, 1975-2003, National Cancer Institute. Bethesda, MD, Acessed 21 Nov 2006.

[19] "What Are the Key Statistics About Cervical Cancer?." 08 Aug 2006. *American Cancer Society.* Accessed 21 Nov 2006.

[20] "Stage Information." Cervical Cancer (PDQ®): Treatment . 18 May 2006. *National Cancer Institute.* Accessed 21 Nov 2006.

[21] L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[22] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, “Biological sequence analysis”. *Cambridge Univ.* Press, 1998.

[23] Migul Simoes, (2010). “A Hidden Markov model for cancer progression”. *Master thesis in Computer science*

[24] X. Huang, A. Acero, and H.W. Hon, Spoken Language Processing A Guide Theory, Algorithm, and System Development,United States of America: *Prantice Hall*, Inc., pp. 375, 2001.

[25] L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition”, *Proceedings Of The IEEE*,vol. 77, no. 2, pp. 257–286, February 1989.

[26] R.J. Elliott, L. Aggoun, J.B. Moore, Hidden Markov models Estimation and Control, *Second Printing, United States of America: Springer-Verlag*, pp. 3, 1997.

[27] H. Xue and V. Govindaraju, “Hidden Markov models Combining Discrete Symbols and Continuous Attributes in Handwriting Recognition”, *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol.28, no.3, pp. 458-462, March 2006.

[28] Tone.B,Steinar ø.T, and Gry B.S, “Incidence, Survival and mortality in cervical camcer in Norway, 1956-1990”, *European Journal of cancer*, vol 29,issue 16,pp. 2291-2297, May 1993.

[29] Kaleigh Smith, “Hidden Markov models in Bioinformatics with Application to Gene Finding in Human DNA”, *Machine learning project*, pp. 308-761, January 2002.

[30] Valeria De.F, Flippo.A.P, and V.Parisi, “Hidden Markov models in Bioinformatics”, *Journal of current Bioinformatics*, vol II, pp. 49-61, 2007.

[31] Sonnenberg FA and Beck JR, “Markov models in medical decision making: a practical guide ”, *Journal of Medical decision making*, vol 13(4), pp.322-338, December 1993.

[32] Cawley, Gavin C,Talbot, and Nicola.L.C, “Gene Selection in cancer classification using sparse logistic regression with bayesian regularization”, *Journal of Oxford university press*, Vol 22, PP. 2348-2355(8), October 2006.

[33] Christopher H.Jackson, Linda D.S, Simon G.Thompson, Stephen W.D, and Elisebeth C, “Multistate Markov models for disease progression with classification error”, *Journal of the Royal statistical society*, pp.193-209, December 2002.

[34] ”American Cancer Society”, http://www.cancer.org/cancer/cervicalcancer/detailedguide/cervical-cancer-what-is-cancer.

[35] ”Cancer Facts”, world health organization, 2011. http://www.who.int/mediacentre/factsheets/fs297/en/

[36] ”Global and Regitional Statistics : Cervical cancer and HPV”, 2009, *international toolkit.* http://www.womeningovernment.org/files/file/prevention/toolkit/Global%20Statistics.pdf.

[37] ”Memorial Sloan-Kettering Cancer Center”, National Institutes of Health (NIH), January 2011. http://clinicaltrials.gov/ct2/show/NCT00581906.

[38] ”Definition of Radiotherapy”, http://www.medterms.com/script/main/art.asp?articlekey=12172

[39] Takashi. O, Kenya Murase, Takashi Fuji, Masashi K and Junpei Ikezoe, “Survival prediction using ANNs in patients with uterine cervival cancer treated by radiotherapy alone ”, *The Journal Society of clinical oncology*, pp.294-300, June 2002.

[40] Basilio.S, I.Inza, and Pedro.L, “ Medical Bayes Networks”, *Lecture notes in Computer science,* Vol 1933|2002, PP.1-49, 2000.

[41] L.E. Baum and T. Petrie, “Statistical inference for probabilistic functions of finite state Markov chains”, *Ann. Math. Statist.,* vol. 37, pp. 1554-1563, 1966.

[42] L.E. Baum and J.A. Eagon, “An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology”, *Bull. Am. Math. Stat.,* vol. 37, pp. 360-363, 1967.

[43] T. Petrie, “Probabilistic functions of finite state Markov chains”, *Ann. Math. Statist.*, vol. 40, pp. 97–115, 1969.

[44] L.E. Baum, “An inequality and associated maximization tech- nique in statistical estimation for probabilistic functions of Markov processes”, Inequalities, vol. 3, pp. 1-8, 1972.

[45] B.G. Leroux, “Maximum likelihood estimation for hidden Mar- kov models”, *Stochastic Process.* Appl., vol. 40, pp.127–143, 1992.

[46] P. Bickel and Y. Ritov, “Inference in hidden Markov models. I. Local asymptotic normality in the stationary case”, Bernoulli, vol. 2, no. 3, pp. 199-228, 1996.

[47] P. Bickel, Y. Ritov and T. Ryden, “Asymptotic normality of the maximum likelihood estimator for general hidden Markov models”, Ann. Statist., vol. 26, no. 4, pp. 1614-1635, 1998.

[48] K8 C.D. Fuh, “SPRT and CUSUM in Hidden Markov models”, *The Annals of Statistics,* vol. 31, no. 3, pp. 942–977, 2003.

[49] ”Markov chains and Markov processes”, http://www.win.tue.nl/~iadan/sdp/h3.pdf.

[50] ”Markov processes”, http://www.math.osu.edu/~husen/teaching/571/markov_1.pdf.

[51] Markus stengel, “Introduction to Graphical Models, Hidden Markov models and Bayesian Networks”, *Department of Information and Computer Sciences.* PP. 441-8580, March 2003.

[52] A. Schliep, B. Georgi, W. Rungsarityotin, I.G. Costa, and A. Schönhuth, “The General Hidden Markov model Library: Analyzing Systems with Unobservable States”, *Forschung und wis- senschaftliches Rechnen: Beitrage zum Heinz-Billing Preis,* Series GWDG-Bericht, pp. 121-135, 2004.

[53] R. Bhar, S. Hamori, Hidden Markov models Applications to Finan- cial Economics, The Netherlands: *Kluwer Academic Publishers*, pp. 17, 2004.

[54] R.J. Elliott, L. Aggoun, J.B. Moore, Hidden Markov models Esti- mation and Control, Second Printing, United States of America: Springer-Verlag, pp. 3, 1997.

[55] L.K. Saul, M. Rahim, "Modeling acoustic correlations by factor analysis", 1998

[56] A.J. Viterbi, "Error bounds f. convolutional codes and asymptotically optim. decod. algor.", 1967

[57] Zoubin Ghahramani, "An Introduction to Hidden Markov models and Bayesian Networks",

[58] Colin cherry, "A general survey of Hidden Markov models in Bioinformatics", University of Alberta, Lecture notes, PP. 1-19.

[59] Rafael Llobet, Alejandro H. Toselli, Juan C. Perez-Cortes, and Alfons Juan, "Computer-Aided Prostate Cancer Detection in Ultrasonographic Images",Springer-Verlag Berlin Heidelberg 2003,F.J. Perales et al. (Eds.): IbPRIA 2003, LNCS 2652, pp. 411–419, 2003.

[60] Michael M. Yin and Jason T. L. Wang, " Application of Hidden Markov models to Gene Prediction in DNA", *Journal of IEEE.*

Glossary of medical terms

- **Human Papillomavirus (HPV)** A group of small non-enveloped DNA viruses infecting epithelia sexually transmitted.
- **Cervical dysplasia:** It is a condition in which the cells of the inner lining of the cervix have precancerous changes.
- **Invasive cervical cancer:** cancer that has spread from the surface of the cervix to tissue deeper in the cervix or to other parts of the body.
- **Biopsy:** Removal and pathologic examination of specimens in the form of small pieces of tissue from the living body.
- **Colposcopy** The examination, therapy or surgery of the cervix and vagina by means of a specially designed endoscope introduced vaginally.
- **HPV DNA testing** DNA probes specific for the identification of human papilloma virus.
- **Papinicolaou smear (Pap smear)** Collection of cell samples from the vagina, cervix, and cervical canal and spread on a glass slide.