Information Systems Department

Faculty of Computer and Information Sciences

Ain Shams University

A Novel Model based on Non Invasive Methods for Prediction of Liver Fibrosis

Thesis submitted as a partial fulfillment of the requirements for the degree of Master of Science in Computer and Information Sciences

By

Mahmoud Moustafa Abdallah Nasr

Information Systems Department,

Faculty of Computer and Information Sciences, Ain Shams University

Under Supervision of

Prof. Dr. Khaled ElBahnasy

Professor at Information Systems Department,

Faculty of Computer and Information Sciences, Ain Shams University

Prof. Sanaa Mohamed Kamal

Professor at Hepatology Department,

Faculty of Medicine, Ain Shams University

Dr. Mohamed Hamdy Mohamed Eleleimy

Associate Professor at Information Systems Department,

Faculty of Computer and Information Sciences, Ain Shams University

2019

Acknowledgment

First and above all, I want to thank Allah the almighty for providing me the opportunity to proceed successfully and the capability to overcoming any challenges and obstacles that have faced me during the preparation of this thesis. This thesis appears in its current form due to the assistance and full support from many people. Therefore, I would like to offer my gratitude and sincere thanks to all of them.

I would like to acknowledge and express my deep thanks to Prof. Dr. Khaled Bahnasy for the continuous encouragement, valuable guidance and his leadership throughout my years of study and through the process of researching and writing this thesis.

I would like to sincerely and deeply thank Dr. Mohamed Hamdy for his patience, understanding, enthusiasm, inspiring instructions and great help and his valuable recommendations that motivated me during rough moments throughout this thesis.

Also, I would like to express my deepest gratitude and appreciation to Dr. Sanaa Kamal for her advices, caring, encouragement and her great valued and supportive supervision.

A special appreciation, many thanks and a great love goes to my family for being always for me and their unfailing support. First of all, I want to warmly thank my parents. Thank you for everything. I cannot forget my sisters' spiritual support and prayers for me and thank Allah who gives me my new born daughter Sofia at the end of this work.

Lastly, I want to thank my wife for her continued support, encouragement and understanding during my research. Many thanks to you, I appreciate all these things. Without her spiritual support, I would not have been able to finish this research.

List of Publications

- M. Nasr, K. El-Bahnasy, M. Hamdy and S. M. Kamal, "A novel model based on non invasive methods for prediction of liver fibrosis," *2017 13th International Computer Engineering Conference (ICENCO)*, Cairo, Egypt, 2017.
- M. Nasr, K. El-Bahnasy, M. Hamdy and Nour zawi, " Alzheimer disorder biomarker Extraction*,"* *2018 11th International Information Systems Conference (INFOS2018)*, Cairo, 2018.

Abstract

Serial liver biopsies are typically the gold standard for diagnosis of liver fibrosis progression. However, it is associated with serious complications, inconvenient to patients, in some cases it causes dying, and it is expensive. The spread and the danger of the hepatitis C virus which infect the liver leading to deadly diseases like liver fibrosis. The challenge is to substitute the liver biopsy (i.e. the invasive procedures) with non-invasive method depending on the computer added tools i.e. a decision support system. The proposed technique is making a rule mining depends on a complete search but not exhaustive to guarantee finding optimal itemset rules, to build a robust and precise decision support system, it depends on new pruning rules to efficiently reduce search and dimensional space. It is not only search 100% of the dataset but also finds all minimal unique rules. This introduces to an optimal itemset rules mining tool. It is employed to resolve this medical diagnosing issue with average accuracy 99.48% for 5-folds cross validation. This accuracy paves the way to utilize classification models as a clinically non-invasive and reliable method to assess the degree of liver fibrosis as a noninvasive method for prediction of liver fibrosis. This technique is applied using other datasets like Alzheimer Disorder as a biomarker extraction tool. Alzheimer’s infection (AD) is the most widely recognized neurodegenerative issue related to dementia in the elderly. Although, initiating events are still unknown, it is clear that AD results from a combining of genetic and environmental risk factors. Diagnosis can be improved by the use of biological measures. However, it takes time (Deterioration of patient condition), the challenge is to save time. The proposed technique is employed to resolve this issue with average accuracy 97.15% for 10-folds cross-validation.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AD: Alzheimer Disorder

AI: Artificial Intelligence.

BN: Bayesian Networks.

CDSS: Clinical Decision Support System.

Conf: confidence.

DSS: Decision Support System.

FP-tree: Frequent-pattern tree.

FAO : Food and Agriculture Organization.

GAs: Genetic Algorithms.

HCV: Hepatitis C Virus.

IDS: intelligent decision systems.

KNN: K-Nearest Neighbor.

LP: Linear projection.

Minconf: minimum confidence.

Minsup: minimum support.

ML: machine learning.

MPR: minimal predictive rules

MUL: Minimal Unique List.

NNs: Neural Networks.

NUL: Not Unique List.

PNU: Pursuit Not Unique.

pp-tree: PrePost tree .

PS-FS-C: Power-Set tree based feature selection-Complete algorithm.

PS-FS-I: Power-Set tree based feature selection –Incomplete.

PS-tree: powerset tree.

RFP: Rate of Fibrosis Progression.

SPADE: Sequential pattern discovery using equivalence classes Algorithm.

SRBC: Subsumption Rule Based Classifier.

Supp: Support.

SVMs: Support vector machines.

Chapter 1:

Introduction

Chapter 1: Introduction

1.1 Overview

Hepatitis C virus (HCV) infection affects more than 170 million people worldwide [**1**]. Egypt has the highest prevalence of hepatitis C in the world with prevalence rates passed 18%, where *Figure 1* shows the HCV Infection prevalence in Egypt during the 1997 to 2018. The Egyptian government has been lunched a national campaign to make Egypt free of HCV. Thus, HCV represents a major public health and economic problem in Egypt [**2**] [**3**] [**4**] [**5**] [**6**] [**7**]. HCV infection is marked by high tendency to persistence and evolution to chronic hepatitis with development of serious consequences such as cirrhosis and liver cancer in some patients [**6**]. To date, there are no indicators or reliable criteria that identify who would develop cirrhosis and when given that the rate of progression of fibrosis is highly variable. The treatment for HCV is both difficult and expensive. Egypt has launched a nationwide government sponsored campaign to treat patients with chronic hepatitis. The large pool of patients and the financial constraints necessitate prioritizing therapy to those most likely to progress rapidly to liver fibrosis [**8**]. Serial liver biopsies are typically the gold standard for diagnosis of liver fibrosis progression [**9**]. Liver biopsies are invasive, associated with serious complications, inconvenient to patients and expensive [**10**]. Recently, several non-invasive serum markers and imaging techniques have emerged as tools for diagnosis of fibrosis. However, to date such biomarkers and imaging procedure have not been adequately validated as reliable alternatives for liver biopsy [**11**], [**12**].

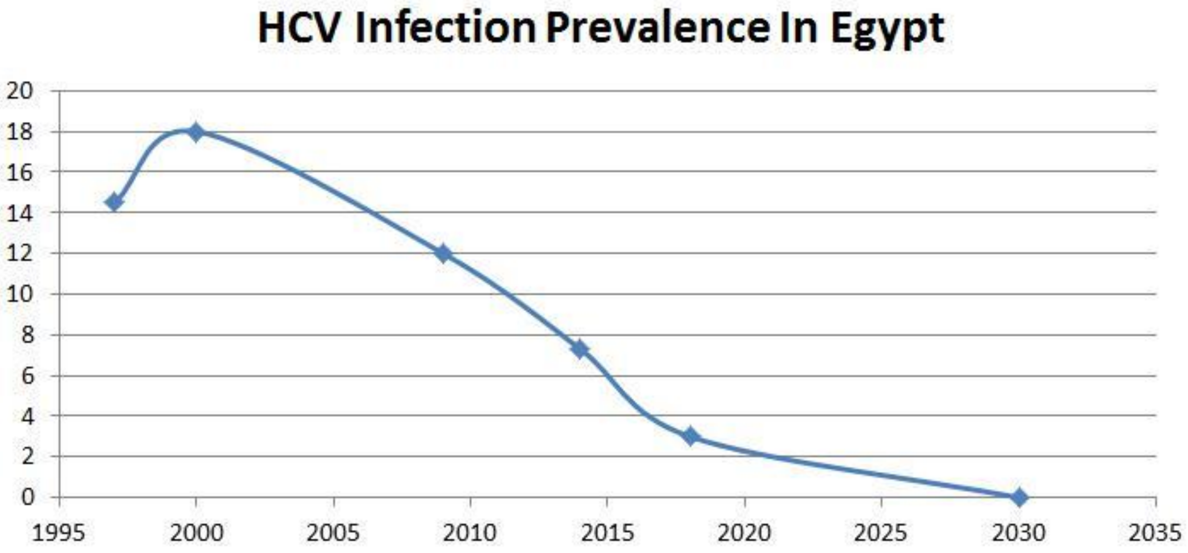

Figure 1 HCV Infection Prevalence in Egypt

Therefore, this work develop, evaluate and validate a prediction model that replaces the invasive techniques, and to be a measurement to liver fibrosis progression. Also, developing and validating computerized clinical decision-support system (CDSS) to support identification of individuals at higher risk of accelerated liver fibrosis progression. The Clinical decision support systems (CDSS) use decision support system theory and technology to assist clinicians in the evaluation and treatment process. Using historical clinical data and the relationship processed by Artificial Intelligence (AI) techniques to aid physicians in their decision making process is the goal of CDSS [**13**], From a computational point of view, by a decision support system (DSS) we understand a computer-based information system assisting the decision making process, and used to solve a large variety of real-life problems. Basically, DSSs are developed to support the solution of unstructured management issues in order to improve the decision-making process [**14**]. Recent advances in artificial intelligence (AI) and statistical learning (SL) enhanced these systems, giving rise to intelligent decision systems (IDS) [**15**].

In this thesis, the study focuses on a new complete search technique but not exhaustive depending on new pruning rules which make searching the dimensional space efficient.

Thesis is organized as follows. In section 2, literature survey about Noninvasive techniques and search strategies will be introduced. In section 3, a study for A Novel Model based on Non Invasive Methods for Prediction of Liver Fibrosis which is consisted of two algorithms: Pursuit Not-Unique algorithm and Subsumption Rule based Classifier. In section 4, presents the experiment results and discussion. In section 5 the conclusion and future work are presented.

1.2 Motivation

In medical field, Liver biopsy is an invasive procedure associated with some complications. Thus, different biomarkers and imaging techniques have been developed for non-invasive diagnosis of liver fibrosis. However, the equivalence of such non-invasive procedures to liver biopsy in diagnosis of liver fibrosis has not been proven. However, there are no tools to predict the risk and rate of liver fibrosis progression accurately like the biopsy. Moreover most of the used computer aided tools are based on heuristic and stochastic search; therefore they don’t guarantee producing or finding the optimal solution.

1.3 Objectives

This thesis introduces to a new rule mining technique based on complete search but not exhaustive, which searches not only 100% of the dimensional space efficiently but also it guaranteed producing all optimal rules combinations in a minimal sizes. As a consequence, these rules promise to build high accurate predictor. Moreover, it provides field experts with knowledge base which can be employed in expert systems or building ontologies for decision support systems.

1.4 Contributions

So this thesis proposes a novel rule mining technique that uses a complete search strategy but not exhaustive. It depends on new pruning rules which efficiently reduce the search space to find all possible minimal unique patterns.

Therefor this thesis introduces to two developed algorithms, where the first algorithm designed to search the dataset to find all possible minimal unique patterns or rules or itemset combinations and the second algorithm is a classifier which use rule evaluation technique to give the most possible accuracy. Therefore, domain expert can be provided with a powerful knowledge base which based on unique rules which also, can build ontologies easily.

The following processes sum up the proposed work:

- Searching dimensional space.
- Pruning rules are deployed to eliminate redundancy to reduce search time.
- Minimal Unique rules extraction.
- Building rule based classifier.
- Cross validation and accuracy measuring.

The output results of hepatitis c virus dataset are validated using cross validation process, where dataset is divided into 5 folds to validate the obtained accuracy which reached 99.48%. and 97.15% for 10-folds cross-validation for Alzheimer’s infection dataset.