

## Accepted Manuscript

Application of data mining techniques and data analysis methods to measure cancer morbidity and mortality data in a regional cancer registry: The case of the island of Crete, Greece

Iraklis Varlamis , Ioannis Apostolakis , Dimitra Sifaki-Pistolla ,  
Neelanjan Dey , Vassilios Georgoulas , Christos Lionis

PII: S0169-2607(16)30920-8  
DOI: [10.1016/j.cmpb.2017.04.011](https://doi.org/10.1016/j.cmpb.2017.04.011)  
Reference: COMM 4409



To appear in: *Computer Methods and Programs in Biomedicine*

Received date: 31 August 2016  
Revised date: 19 March 2017

Please cite this article as: Iraklis Varlamis , Ioannis Apostolakis , Dimitra Sifaki-Pistolla , Neelanjan Dey , Vassilios Georgoulas , Christos Lionis , Application of data mining techniques and data analysis methods to measure cancer morbidity and mortality data in a regional cancer registry: The case of the island of Crete, Greece, *Computer Methods and Programs in Biomedicine* (2017), doi: [10.1016/j.cmpb.2017.04.011](https://doi.org/10.1016/j.cmpb.2017.04.011)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Application of data mining techniques and data analysis methods to measure cancer morbidity and mortality data in a regional cancer registry

### The case of the island of Crete, Greece

Iraklis Varlamis<sup>1\*</sup>, Ioannis Apostolakis<sup>2\*</sup>, Dimitra Sifaki-Pistolla<sup>3\*</sup>, Neelanjan Dey<sup>4</sup>, Vassilios Georgoulas<sup>5\*</sup>, and Christos Lionis<sup>3\*</sup>

<sup>1</sup>Department of Informatics and Telematics, Harokopio University of Athens, Greece

<sup>2</sup>National School of Public Health, Athens, Greece

<sup>3</sup>Clinic of Social and Family Medicine, Department of Social Medicine, University of Crete

<sup>4</sup>Department of Information Technology in Techno India College of Technology, Rajarhat, Kolkata, India

<sup>5</sup>Department of Medical Oncology, University General Hospital of Heraklion, University of Crete, Greece

\*On behalf of the Regional Cancer Registry of Crete, Greece

### Abstract

**Background and Objective:** Micro or macro-level mapping of cancer statistics is a challenging task that requires long-term planning, prospective studies and continuous monitoring of all cancer cases. The objective of the current study is to present how cancer registry data could be processed using data mining techniques in order to improve the statistical analysis outcomes.

**Methods:** Data were collected from the Cancer Registry of Crete in Greece (counties of Rethymno and Lasithi) for the period 1998-2004. Data collection was performed on paper forms and manually transcribed to a single data file, thus introducing errors and noise (e.g. missing and erroneous values, duplicate entries etc.). Data were pre-processed and prepared for analysis using data mining tools and algorithms. Feature selection was applied to evaluate the contribution of each collected feature in predicting patients' survival. Several classifiers were trained and evaluated for their ability to predict survival of patients. Finally, statistical analysis of cancer morbidity and mortality rates in the two regions was performed in order to validate the initial findings.

**Results:** Several critical points in the process of data collection, preprocessing and analysis of cancer data were derived from the results, while a road-map for future population data studies was developed. In addition, increased morbidity rates were observed in the counties of Crete (Age Standardized Morbidity/Incidence Rates ASIR= 396.45±2.89 and 274.77±2.48 for men and women, respectively) compared to European and world averages (ASIR= 281.6 and 207.3 for men and women in Europe and 203.8 and 165.1 in world level). Significant variation in cancer types between sexes and age groups (the ratio between deaths and reported cases for young patients, less than 34 years old, is at 0.055 when the respective ratio for patients over 75 years old is 0.366) was also observed.

**Conclusions:** This study introduced a methodology for preprocessing and analyzing cancer data, using a combination of data mining techniques that could be a useful tool for other researchers and further enhancement of the cancer registries.

**Keywords:** Cancer data, Data mining, Feature selection, Crete, Greece

Download English Version:

<https://daneshyari.com/en/article/4958126>

Download Persian Version:

<https://daneshyari.com/article/4958126>

[Daneshyari.com](https://daneshyari.com)