Person centered prediction of survival in population based screening program by an intelligent clinical decision support system

Reza Safdari¹, Elham Maserat², Hamid Asadzadeh Aghdai³, Amir Hossein Javan Amoli⁴, Hamid Mohaghegh Shalmani⁵

¹ Allied Medical Sciences School, Tehran University of Medical Sciences, Tehran, Iran
² School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran
³ Basic and Molecular Epidemiology of Gastrointestinal Disorders Research Center, Research Institute for Gastroenterology and Liver Diseases, Tehran, Iran
⁴ Islamic Azad University
⁵ Gastroenterology and Liver Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Science, Tehran, Iran

Abstract

Aim: To survey person centered survival rate in population based screening program by an intelligent clinical decision support system.

Background: Colorectal cancer is the most common malignancy and major cause of morbidity and mortality throughout the world. Colorectal cancer is the sixth leading cause of cancer death in Iran. In this survey, we used cosine similarity as data mining technique and intelligent system for estimating survival of at risk groups in the screening plan.

Methods: In the first step, we determined minimum data set (MDS). MDS was approved by experts and reviewing literatures. In the second step, MDS were coded by python language and matched with cosine similarity formula. Finally, survival rate by percent was illustrated in the user interface of national intelligent system. The national intelligent system was designed in PyCharm environment.

Results: Main data elements of intelligent system consist demographic information, age, referral type, risk group, recommendation and survival rate. Minimum data set related to survival comprise of clinical status, past medical history and socio-demographic information. Information of the covered population as a comprehensive database was connected to intelligent system and survival rate estimated for each patient. Mean range of survival of HNPCC patients and FAP patients were respectively 77.7% and 75.1%. Also, the mean range of the survival rate and other calculations have changed with the entry of new patients in the CRC registry by real-time.

Conclusion: National intelligent system monitors the entire of risk group and reports survival rates by electronic guidelines and data mining technique and also operates according to the clinical process. This web base software has a critical role in the estimation survival rate in order to health care planning.

Keywords: Colorectal Cancer, Survival rate, Decision support system, Screening


Introduction

Colorectal cancer (CRC) is a major cause of morbidity and mortality throughout the world (1, 2), with an annual incidence of 1 million cases (3, 4, 5) and an annual mortality of more than 500,000 cases. CRC is the second most common cause of cancer mortality (6) and one of most malignancies cancers in Iran (7, 8). Recent epidemiological studies were demonstrated the increasing incidence trend of CRC in Iran (9, 10). According to consensus, there are more than 3641 new cases of CRC diagnosed in Iran (11). Colorectal cancer is the sixth leading cause of cancer death and annually there are around 2262 deaths from CRC in Iran (11, 12). The overall 5-year survival rate for colorectal cancer patients was 61% (13, 14). Annual incidence of CRC has increased over the next two decades in developing countries (15, 16, 17, 18). Socioeconomic status such as race, ethnicity, geographical and economic characteristics has a significant role on CRC incidence rates, mortality rates, and survival rates (19, 20). More studies are demonstrated that colorectal...
cancer incidence and mortality are reduced with regular screening (21).

Decision Support System (DSS) as a computer-based system improves preventive care services (22) and extracts knowledge from routine care data. Person centered prediction of survival is one of this knowledge (23). Machine-learning services of DSS accurately predict outcomes in CRC. Also health care providers can be used of the DSS tools to facilitate treatment and screening planning (24). Computer-aided decision support systems improve quality of care by providing of the patient specific alert, reminders and recommendations (25, 26). CDSS play critical role in clinical setting by several artificial intelligence methods and techniques (27). Cosine similarity is one the clustering technique (28). Cosine similarity estimate degree of similarity between two items or users. Cosine similarity as a data mining method can be used for measuring similarity rate between two data objects. Similarity is detected distances between features of the objects. High degree of similarity and low degree of similarity was defined by distance interval. All of the features values must be normalized because even one feature is changed dominating the distance calculation (29). Cosine similarity was introduced as the decision rule for query detection in large data sets (30).

In this survey, we used cosine similarity as data mining technique for estimating survival of at risk groups in the screening plan. High volume, high velocity, and high variety information products in the CRC screening plan. However, intelligent approaches as innovation strategy facilitate clustering of data and decision making process. Decisions of this system can be useful for understanding the screening modalities, survival rate of cancer and recommended time intervals of screening. The aim of this study is to predict the survival in the population based screening program by an intelligent clinical decision support system.

Materials and Methods

In the first step, we defined minimum data set (MDS). MDS was approved by experts and reviewing literature. Vital status of the covered population in CRC registry was established by matching with medical record and telephone enquiries. In the second step, MDS were coded by python language and matched with cosine similarity formula. Also MDS such as BMI (Body Mass Index) was calculated by coding method in intelligent system. Sample of coding:

```python
#Calculate BMI
if row[‘Weight’] and row[‘Height’] and row[‘Weight’] != 0 and row[‘Height’] != 0:
    bmi = row[‘Weight’]/(row[‘Height’]/100)**2
else:
    bmi = np.nan
if bmi < 18.5:
    bmi_list.append(1)
elif bmi < 24.9:
    bmi_list.append(2)
elif bmi < 29.9:
    bmi_list.append(3)
else:
    bmi_list.append(4)

persian_bdate = persian_bdate.split(‘-’)
gregorian_bdate = jdate.jd_to_gregorian(jdate.persian_to_jd(int(persian_bdate[0]),
int(persian_bdate[1]), int(persian_bdate[2])))
age = int((time.strftime(‘%Y’)) - gregorian_bdate[0])
if age < 45:
    age_list.append(1)
elif age < 65:
    age_list.append(2)
else:
    age_list.append(3)
```

Cosine similarity implementation in python:

```python
from math import*  
def square_rooted(x):
    return round(sqrt(sum([a*a for a in x])),3)
def cosine_similarity(x,y):
    numerator = sum(a*b for a,b in zip(x,y))
    denominator = square_rooted(x)*square_rooted(y)
    print cosine_similarity([3, 45, 7, 2], [2, 54, 13, 15])
```

Finally, survival rate by percent was illustrated in user interface of intelligent system. Intelligent system was designed in PyCharm environment. PyCharm is an integrated development environment (IDE) applied for programming in Python. This system has rule base environment. Survival rate and risk group of covered population synchronously were presented in the user interface.

Results

Minimum data set related to survival comprise of clinical status, past medical history, life style and socio-demographic information. MDS related to survival of CRC was illustrated in table 1. Main data elements of intelligent system consist demographic information, age, referral type, risk group, recommendation and survival rate. This system estimates survival rate of alive patients by dead patients. Estimation was conducted by similarity. Intelligent system integrates with CRC registry. Information of covered population as a comprehensive database was connected to intelligent system and survival rate estimated for each patient (figure

Gastroenterol Hepatol Bed Bench 2017; 10 (1): 60-65
Person centered prediction of survival by an intelligent clinical decision support system

Stigliano, et al. (37) reported improved prognosis of cancer in patients with HNPCC with a 5 year cumulative survival (94.2%) versus sporadic CRC (75.3%). In this survey, mean range of survival of HNPCC patients was 77.7%. This survival rate is lower rather than to other countries. Survival rate estimation of cancer play key role for health care planning. Survival rate reports can be used for researching base on cancer prevention (40, 41). Reliable research presents the complex mechanisms and pathways in the tumorigenesis and natural history of HNPCC tumors and there aren’t clinical indicators of good prognosis in HNPCC-related colorectal carcinomas. Thus, CRC screening guidelines must be based on the improved expected survival rates of all patients. Also, the mean range of survival of FAP patients were 75.1%. Studies has been shown screening of families with FAP can reduce CRC mortality (42). Recent studies have been shown multidimensional role of guidelines for preventive...
Table 1. Minimum Data Sets related to survival of CRC

<table>
<thead>
<tr>
<th>Core Data Element</th>
<th>Variable</th>
<th>Variable Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socio-demographic Data</strong></td>
<td>Current age and diagnosis age</td>
<td>1 = &quot;&lt; 45&quot; / &quot;45-65&quot;=2 / 3 = &quot; &gt; 65&quot;</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>Female =1/ Male = 2 / Transsexual = 3</td>
</tr>
<tr>
<td></td>
<td>Education level</td>
<td>Illiterate = 1 / Primary school = 2 / High school = 3 / University = 4</td>
</tr>
<tr>
<td></td>
<td>Marital Status</td>
<td>Married = 1 / Single = 2 / Divorced = 3 / Widowed = 4</td>
</tr>
<tr>
<td></td>
<td>Ethnicity</td>
<td>Fars = 1 / Turk = 2 / Kord = 3 / Lor = 4 / Other =1</td>
</tr>
<tr>
<td></td>
<td>Religion</td>
<td>Muslim =1/ Christian=2/ jewish=3/ Zoroastrian=4/ Other=5</td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>City=1/Village =2</td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>18.5&gt;=1/18.5-24.9&gt;=2/25-29.9&gt;=3/30&lt;=4</td>
</tr>
<tr>
<td><strong>Life-style</strong></td>
<td>Exposure to chemical weapons and pollutant-material</td>
<td>Yes=1/NO=0/Unknown=3</td>
</tr>
<tr>
<td></td>
<td>Tobacco Use</td>
<td>Yes=1/NO=0/Unknown=3</td>
</tr>
<tr>
<td></td>
<td>Alcohol Use</td>
<td>Yes=1/NO=0/Unknown=3</td>
</tr>
<tr>
<td></td>
<td>Oral narcotic substance Use</td>
<td>Yes=1/NO=0/Unknown=3</td>
</tr>
<tr>
<td></td>
<td>IV Drug User</td>
<td>Yes=1/NO=0/Unknown=3</td>
</tr>
<tr>
<td></td>
<td>NSAIDs Use</td>
<td>Yes=1/NO=0/Unknown=3</td>
</tr>
<tr>
<td></td>
<td>Tumor size (grad I,II)</td>
<td>20mm&gt;=1/20mm&lt;=2; 35mm=1/35mm&lt;=2</td>
</tr>
<tr>
<td></td>
<td>Tumor Morphology</td>
<td>Adenocarcinoma=1/ Mucinous adenocarcinoma=2/ Signet Ring Cell Carcinoma=3/ Squamous cell carcinoma=4/ Squamous-cell adenocarcinoma=5/ medullary carcinoma=6/</td>
</tr>
<tr>
<td><strong>Tumor Classification</strong></td>
<td>Histological Differentiation Grade</td>
<td>Grade cannot be assessed (undetermined grade) =1/Well differentiated (low grade)=2/Moderately differentiated (intermediate grade) =3/ Poorly differentiated (high grade)=4/Undifferentiated (high grade)=5</td>
</tr>
<tr>
<td></td>
<td>Pt grade</td>
<td>0 = T0/ T1=1/T2=2/ T3=3/ T4=4</td>
</tr>
<tr>
<td></td>
<td>Pn grade</td>
<td>N0=0/N1=1/ N2=2</td>
</tr>
<tr>
<td></td>
<td>Pm grade</td>
<td>M0=0/M1=1</td>
</tr>
<tr>
<td></td>
<td>Pathologic grade</td>
<td>1=I/II-2/III=3/IV=4</td>
</tr>
<tr>
<td></td>
<td>Advancing grade</td>
<td>Primary grade=1/Advanced grade=2</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>Surgery=1/chemotherapy, Radiotherapy =2/Biopsy=3</td>
</tr>
<tr>
<td></td>
<td>HNPPCC</td>
<td>Yes=1/No=0</td>
</tr>
<tr>
<td></td>
<td>FAP</td>
<td>Yes=1/No=0</td>
</tr>
<tr>
<td></td>
<td>Diabetes Mellitus</td>
<td>Yes=1/No=0/Unknown=3</td>
</tr>
<tr>
<td></td>
<td>Hypertension disorder</td>
<td>Yes=1/No=0/Unknown=3</td>
</tr>
<tr>
<td><strong>Past Medical History</strong></td>
<td>IBD</td>
<td>Yes=1/No=0/Unknown=3</td>
</tr>
<tr>
<td></td>
<td>Personal history of CRC</td>
<td>Yes=1/No=0/Unknown=3</td>
</tr>
<tr>
<td></td>
<td>Family history of CRC</td>
<td>Yes=1/No=0/Unknown=3</td>
</tr>
<tr>
<td></td>
<td>Abdominal perforation</td>
<td>Yes=1/No=0</td>
</tr>
<tr>
<td></td>
<td>FUO</td>
<td>Yes=1/No=0</td>
</tr>
<tr>
<td></td>
<td>Obstructive Symptoms</td>
<td>Yes=1/No=0</td>
</tr>
<tr>
<td></td>
<td>Abdominal Pain</td>
<td>Yes=1/No=0</td>
</tr>
<tr>
<td><strong>Symptoms</strong></td>
<td>Anemia</td>
<td>Yes=1/No=0</td>
</tr>
<tr>
<td></td>
<td>Rectal Bleeding</td>
<td>Yes=1/No=0</td>
</tr>
<tr>
<td></td>
<td>Change in Bowel Habit</td>
<td>Yes=1/No=0</td>
</tr>
<tr>
<td></td>
<td>Weakness</td>
<td>Yes=1/No=0</td>
</tr>
<tr>
<td></td>
<td>Weight Loss</td>
<td>Yes=1/No=0</td>
</tr>
<tr>
<td></td>
<td>Survival quality</td>
<td>Normal activity=1/Non normal activity=2/Dead=3</td>
</tr>
<tr>
<td><strong>Vital Quality</strong></td>
<td>Vital status</td>
<td>Alive=1/Dead=0</td>
</tr>
<tr>
<td></td>
<td>Survival time</td>
<td>By month</td>
</tr>
</tbody>
</table>
care services (21). Computer-aided consultation improves real time management (43). Screening guidelines should be integrated with the care process in order to providing suitable patient-specific advice. Also, national intelligent system monitors the entire risk group and report survival rates by electronic guidelines and data mining technique and operates according to clinical process. This web base software has a critical role in optimal interactions between colonoscopy, pathology and laboratory data.

After more than a decade of development of numerous computerized systems, studies on the most effective implementation of intelligent risk assessment systems and guideline base system is still lacking. A big obstacle for implementing of DSS is the difficulty in obtaining accurate and complete data required for decision making. Hence uniform templates for data collecting and standard reporting are essential. Quantitative and qualitative evaluation of data can be assisted to precise estimate of the survival rate.

In conclusion, low survival rate of HNPCC and FAP indicates the extremely urgent needs for health authorities to agreed measures of CRC screening. Early detection and counting of patient care improve expected survival rate of CRC. Identifying an asymptomatic person at risk and needing diagnosis and follow-up treatment could lead to increasing of survival rate.

Also, it is recommended to apply intelligent system for improving quality of care and reduce cost. Decisions of computer base system could be a suitable tool to assist clinicians understanding cancer risk, the screening approach, survival rate, recommended screening time intervals and preferences.

References


27. A Hybrid Case-Based System in Stress Diagnosis and Treatment. Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics; 5-7 January 2012.


