Computer-aided diagnosis of pancreatic and lung cancer

A. Álvaro Núñez Díaz, B. Luis Lancho Tofé

***Abstract* — when we talk about cancer diagnosis the most important thing is early diagnosis to prevent cancer cells from spreading. We may also consider the high cost of diagnostic tests. Our approach seeks to address both problems. It uses a software based on Bayesian networks that simulates the cause-effect relationships and gets the chance of suffering a pancreatic cancer or lung cancer. This software would support doctors and save a lot of time and resources**

***Keywords*— Bayesian networks, pancreatic cancer, lung cancer and computer-aided diagnosis**

I. INTRODUCTION

Pancreatic cancer is the cancer with a higher mortality. Is the fourth leading cause of death in the United States of America. In 2006 were predicted 33,700 cases of which 32,300 were fatal [1]. In Spain (year 2004) the situation is as follows [2]: 2,438 deaths (4% of all cancer deaths in men) and 2098 (5.8% of all cancer deaths in women).

Lung cancer is the most widespread cancer (27.4% in Spain). In USA accounted for more deaths than breast cancer, prostate cancer, and colon cancer combined. In year 2004, 108,355 men and 87,897 women were diagnosed with lung cancer (89,575 men and 68,431 women died).

Both lung cancer and pancreatic cancer persist as a challenge to medicine, because despite advances in diagnostic techniques and therapeutic resources, the statistics reflect the few dressings in relation to the number of patients cared for. This is due largely to the fact that in recent years there have been no meaningful responses to attempts to reduce exposure to carcinogens, in the case of lung cancer, and also that did not improve the detection of the tumour at an early stage, which would have been able to increase the possibility of healing, but especially not yet have an appropriate solution for these pathologies.

In an organization as the Spanish health, oncology institutes are not designed so far to leave in search of any persons concerned. Patients who received are overwhelmingly referrals with confirmed or presumptive diagnosis and usually at a relatively advanced stage of evolution.

Hence the need arises to create a web application that calculates instantly and reliably the likelihood of suffering from cancer. This reliability will ultimately depend on the human factor, since it will be the oncologists who through his experience and statistical data to determine the values of probability tables that use the system to infer the likelihood of final diagnosis. The medical knowledge is stored in a bayesian network (see Fig. 6) that models the relationships between symptoms, risk factors, etc.

The project aims to provide an early warning system and determine the likelihood of developing now or in the future some of the, previously mentioned, types of cancer, according to risk factors, symptoms and other diseases.

Given the seriousness of these types of cancer and the known fact that a large percentage of cases are discovered in advanced stages of the disease, a tool aimed to inform and to alert the population is a great help. This type of tool can be further enhanced in the current social context, where a growing sector of the population has access to the Internet, from where access to this software can be easy and instantaneous, increasing considerably the potential utility of the application for physicians and patients.

Besides the usefulness purely informative, this project would have a preventive character because:

- If the application concludes that a person is at risk of cancer due to certain risk factors, this person could increase their awareness of the problem and how to avoid them.

- A sufficient risk of suffering from cancer could push the individual to carry out periodic reviews, which could lead to earlier detection of the disease.

II. CANCER DIAGNOSIS

A. Pancreatic cancer

Pancreatic cancer is presented with an incidence of 8 to 10 cases per 100,000 inhabitants per year in many industrialized countries of Europe [1]. This incidence is higher than that recorded a few years ago [2]. In Spain, the incidence of pancreatic cancer has risen dramatically in the second half of the twentieth century, with an increase in the same close to 200% [3]. Pancreatic cancer is the fourth leading cause of cancer deaths in the United States, both men and women [1,4,5]. Until the last evolutionary stages, when the majority of diagnoses are introduced, pancreatic cancer is a disease of evolution silent and insidious. Since the disease is usually diagnosed when it is no longer confined to the pancreas, has a survival rate at 1 year of

15-20%, and when survival is valued at 5 years, that figure drops to 2-4% [1,6]. Only when the pancreatic cancer is small (less than 2 cm.) sits at the head of the organ, and there are no lymph node metastases and invasion of neighbouring organs, can be expected survival rates of 20% to 5 years.

Manuscript created May 27, 2008. A. Álvaro Núñez Díaz is now studying at the Pontifical University of Salmanca (e-mail: alvand9@gmail.com).

B. Luis Lancho Tofé is now studying at the Pontifical University of Salmanca (e-mail: elluiso_7@hotmail.com).

B.Lung cancer

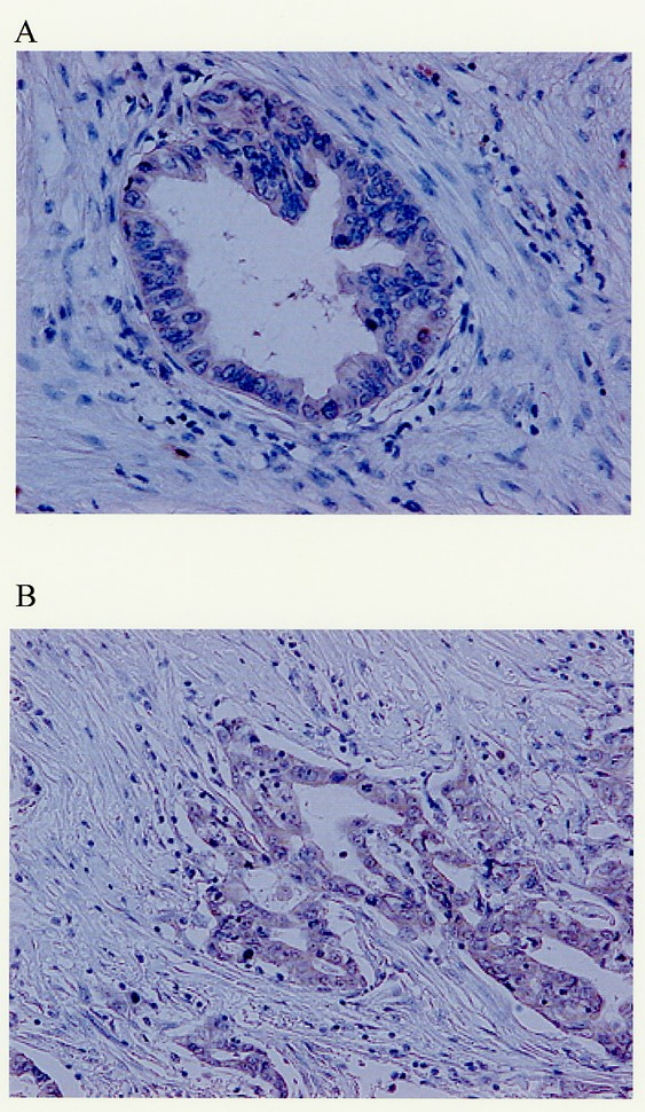

Fig. 1 Immunohistochemical detection of KIT and SCF expression in pancreatic cancer. (A) KIT-positive pancreatic cancer cells (× 200). (B) SCF-positive pancreatic cancer cells (× 200).

[Credit: Yasuda et al. Molecular Cancer 2006 5:46 doi:10.1186/1476-4598-5-46]

It is a very aggressive and deadly cancer, most patients die before the first year after diagnosis. It is the most common cancer in men over 35 countries, especially developed or industrialized countries [22], including Spain, being the leading cause of cancer death in both men and women. More people die of lung cancer than of colon cancer, breast and prostate combined. It represents 12.5% of all malignant tumors in Spain. [23]

The number of cases has increased since the beginning of the century, doubling every 15 years. The incidence has increased nearly 20 times between 1940 and 1970. At the beginning of the century, it was considered that most were metastatic lung tumors, and that cancer was rare primitive.

III.SYMPTOMS, RISK FACTORS AND PATHOLOGIES

A risk factor is anything that affects the likelihood of developing a disease. Having a risk factor, or even several factors, does not mean that a person can contract the disease. In addition, many people who develop the disease may have no known risk factor [9].

A.Pancreatic cancer

Over the years it has investigated the possible existence of risk factors that influence the emergence of pancreatic cancer. In this sense, risk factors have been invoked since demographic factors to personal terms with a greater or lesser risk could be related to pancreatic cancer [10].

It is unknown the cause of emergence of pancreatic cancer; it is more common in smokers and in the obese people and almost 1/3 of cases is due to cigarette smoking. There is controversy over whether diabetes type 2 is a risk factor. In addition, it is known that a small number of cases are related to symptoms that are transmitted through families [11].

Now we are going to show a list of symptoms, risk factors and pathologies related with pancreatic cancer. After, we will develop the most significant facts and will put the range of values used in the bayesian network. We have chosen the range of values depending on the relevancy of the fact in relation with the diagnosis.

Symptoms [11,15,16,17]:

- Weight Loss (Fig. 2)
- Loss of appetite (Fig. 2)
- Abdominal distention (Fig. 3)
- Edema in lower extremities (Fig. 3)
- Backache (Fig. 3)
- Mass palpable abdominal (Fig. 3)
- Abdominal pain (Fig. 3)
- Clots or fatty tissue abnormalities (Fig. 3)
- Uneven texture of fat tissue (Fig. 3)
- Relaxation of the gallbladder (Fig. 3)
- Nausea or vomiting (Fig. 3)
- Diarrhea (Fig. 3)
- Indigestion (Fig. 3)
- Eyes and skin yellowish (Fig. 3)
- Urine of dark colour (Fig. 3)
- Faeces of clear colour (Fig. 3)

Risk factors [11,13,14,19,20,21]:

- Sex (Fig. 4)
- Age (Fig. 4)
- Smoking (Fig. 4)
- Food (Fig. 4)
- Consumption of coffee (Fig. 4)
- Consumption of alcohol (Fig. 4)
- Race (Fig. 4)
- Obesity and physical inactivity (Fig. 4)
- Exposure to asbestos, pesticides, dyes, petroleum (Fig. 4)

Pathologies [10,17,18]:

- Pernicious anemia (Fig. 5)

- Endocrine tumours (Fig. 5)
- Esteatorrea (Fig. 5)
- Anorexia (Fig. 5)
- Ascites (Fig. 5)
- Cystic fibrosis (Fig. 5)
- Chronic pancreatitis (Fig. 5)
- Tonsillectomy (Fig. 5)
- Peptic ulcer surgery (Fig. 5)
- Cholecystectomy (Fig. 5)
- Allergies (Fig. 5)
- Thrombophlebitis (Fig. 5)
- Long-term diabetes (Fig. 5)
- Diabetes mellitus of recent beginning (Fig. 5)

a) **Age**

The risk of pancreatic cancer increases with age. Almost all patients are older than 45 years. More than 90% are over 55 and almost 70% are over 65. The average age at the time of diagnosis is 72 years. (see Fig. 4)

In the bayesian network we used the following range of values*: children, young or adult*

b) **Sex**

Men are slightly more likely to develop pancreatic cancer than women. This difference was most pronounced in the past, but has declined in recent years. Perhaps that difference, at least partly, is due to higher consumption of tobacco. (see Fig. 4)

Range of values: *male or female*

c) **Race**

Black people and Orientals in this order are more likely to develop pancreatic cancer compared to white people. The cause is unknown, but may be due to higher rates of smoking and diabetes among men, and overweight in women of those races. (see Fig. 4)

Range of values: *white, black or oriental*

d) **Tobacco**

Smokers have a risk of two to three times greater of suffering pancreatic cancer. Scientifics consider that this may be due to chemical agents that exist in cigarette smoke. The chemical agents enter in the blood causing damage to the pancreas. It is thought that between 20 and 30 percent of all pancreatic cancer cases are smoking cigarettes. (see Fig. 4)

Range of values: *former smoker, non-smoking, moderate or high*

e) **Diet**

Some studies have found a relationship between pancreatic cancer and a diet high in fats, or that includes the consumption of lots of red meat, pork and processed meats (such as sausage and bacon). Others have discovered that a high consumption of fruits and vegetables can help reduce the risk of pancreatic cancer. But not all studies have made these associations, and the true value of the food is still under study. (see Fig. 4)

Range of values: *good or poor*

f) **Obesity and physical inactivity**

People with overweight are much more likely to develop pancreatic cancer, like those with little physical activity. (see Fig. 4)

Range of values: *mild, moderate or serious*

g) **Diabetes**

The pancreatic cancer is most common in people who suffer from diabetes mellitus. The reason for this relationship is unknown. Most risk is detected in people with diabetes type 2, which usually starts in adulthood and is often linked to overweight and obesity. There is no certainty if there is an increased risk in people with diabetes type 1 (in younger people). (see Fig. 5)

Range of values: *negative or positive*

h) **Chronic pancreatitis**

The chronic pancreatitis is a long-term inflammation of the pancreas. This condition is associated with an increased risk of pancreatic cancer, but most patients with pancreatitis never develop this cancer. Perhaps the main reason for this association is that patients with pancreatitis also are more likely to have other risk factors such as smoking. Very few cases of chronic pancreatitis are caused by inherited gene mutation (see family history). People with this hereditary form of chronic pancreatitis seem to have a high risk for a lifetime of suffering pancreatic cancer (approximately 40 to 75 percent). (see Fig. 5)

Range of values: *negative or positive*

i) **Exposure to working conditions**

The intense exposure at work to certain pesticides, dyes and chemicals used in metal refinery could increase the risk of pancreatic cancer. (see Fig. 4)

Range of values: *negative or positive*

j) **Family history**

It seems that the pancreatic cancer occurs more frequently in some families. Perhaps up to 10% of pancreatic cancers are related to changes inherited DNA (mutations). These changes often increase the risk for other cancers as well. Some examples include:

Mutations in the BRCA2 gene: this mutation also increases the risk of ovarian cancer and breast cancer

Mutations in the p16 gene: it also increases the risk of melanoma.

Mutations of the gene PRSS1: cause severe pancreatitis during an early stage.

Hereditary non polyposis colorectal cancer (HNPCC or Lynch syndrome): also increases the risk of colorectal cancer and endometrial cancer.

Peutz-Heghers Syndrome: also has partnered with polyps in the digestive tract and other cancers.

Scientists have found some of these changes in DNA and can be identified through genetic testing. (see Fig. 4)

Range of values*: negative or positive.*

*k) **Stomach Problems.***

The stomach infection with the bacterium called Helicobacter pylori, which causes ulcers, could increase the risk of pancreatic cancer. Some researchers believe that excessive stomach acid could also increase the risk. (see Fig. 3). Range of values: *mild, moderate or serious*

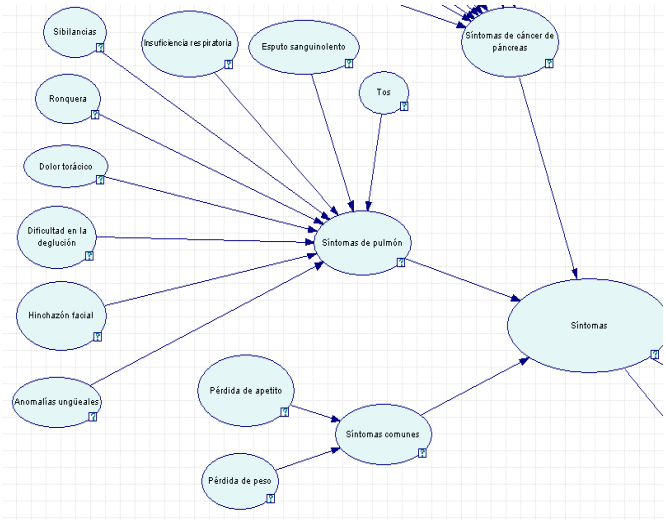

Fig. 2 Detail of the bottom left corner of the bayesian net modeling the patient's symptoms. Both common symptoms and the specific symptoms of lung cancer are included.

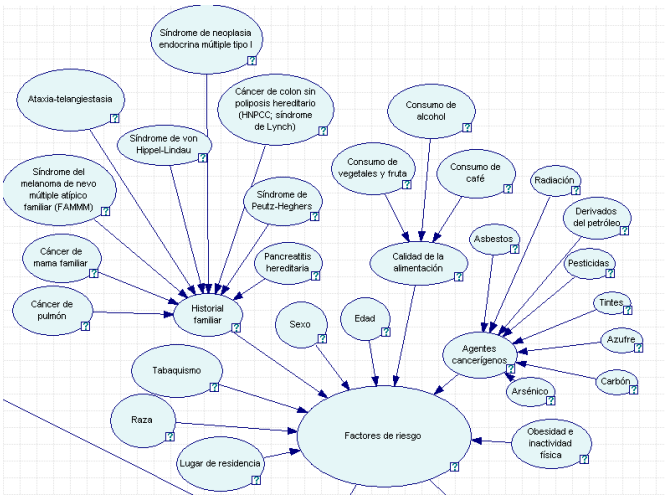

Fig. 4 Detail of the center of the bayesian net. We can observe risk factors grouped in four categories: diseases in the family history, quality of food, exposure to carcinogens and general risk factors. In the last category we have sex, smoking, etc.

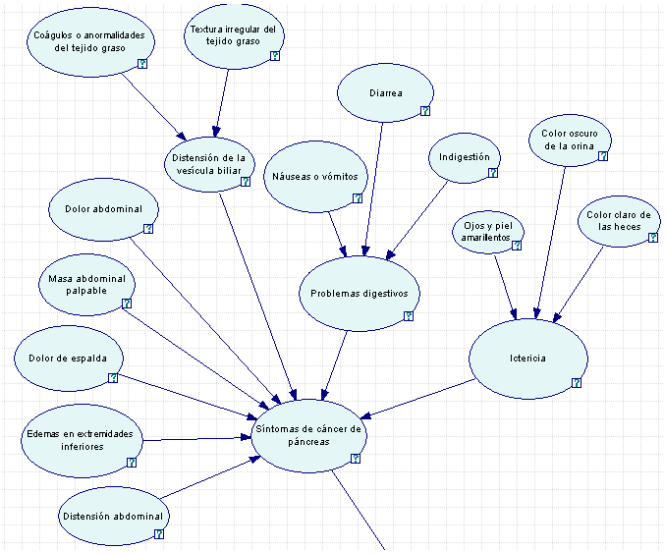

Fig. 3 Detail of the specific symptoms of pancreatic cancer. They are divided in: symptoms related with distension of the gallbladder, symptoms related with digestive problems, symptoms related with jaundice and general symptoms of pancreatic cancer.

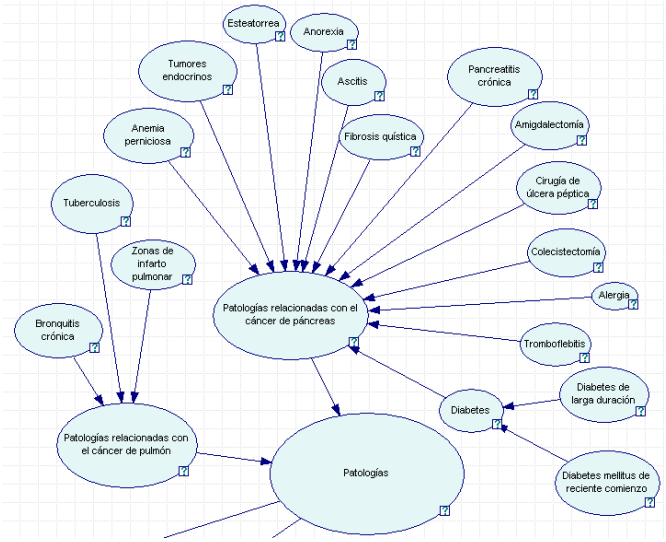

Fig. 5. Detail in the upper right corner. Now we have pathologies. They are divided in two big groups: pathologies related with pancreatic cancer and pathologies related with lung cancer. It is remarkable the two types of diabetes.

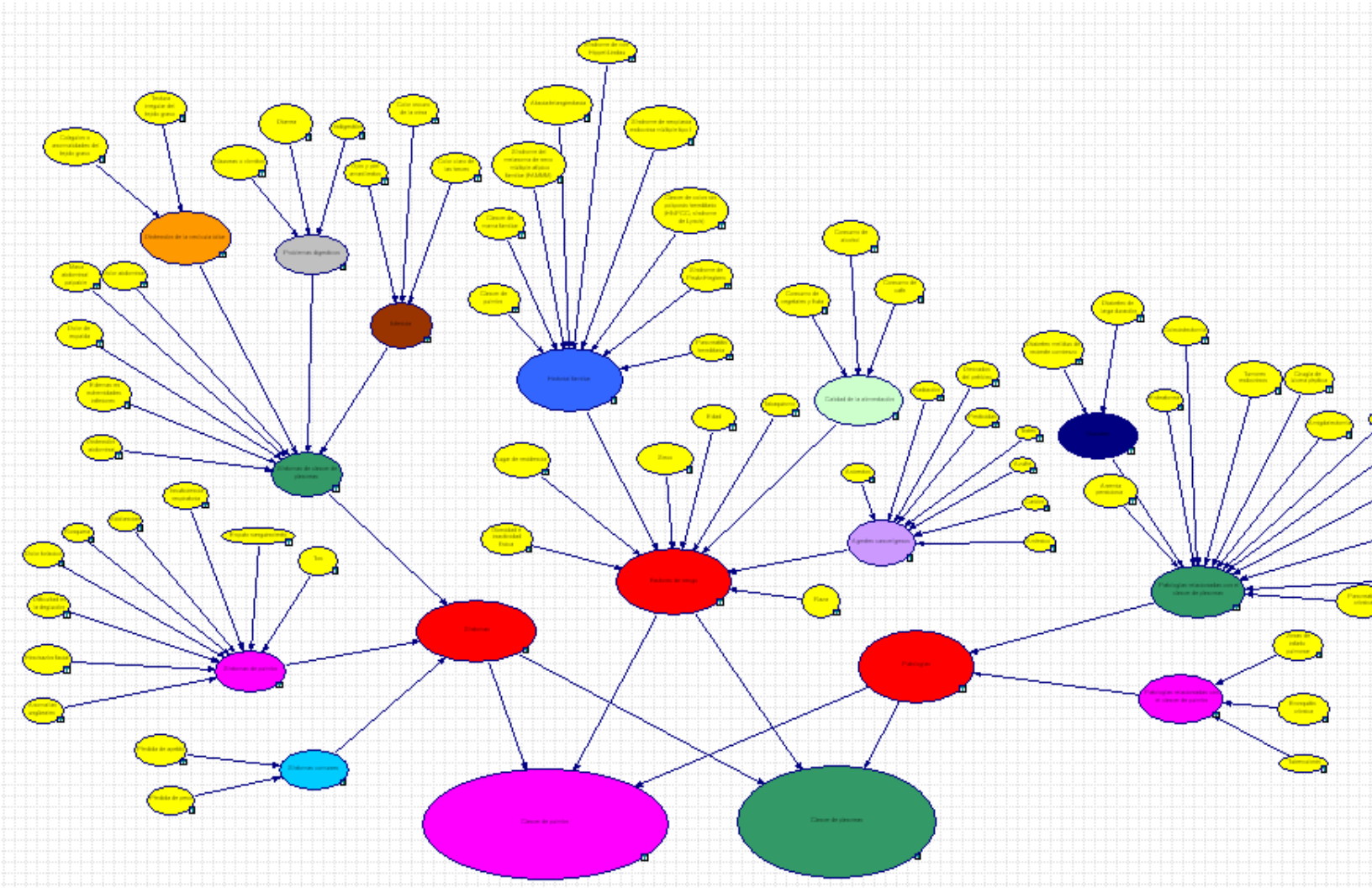

Fig. 6. Overview of the bayesian network that models the system´s medical knowledge. The color code of the nodes is:
- **Yellow**: it expresses the input (risk factors, symptoms and diseases). These nodes have an associated probability value. Depending on the response of the patient (or defaults in case of unknown response) will be established the evidences of the network.
- **Pink**: from left to right, symptoms and pathologies related with lung cancer. Top down, the node of lug cancer that determines the final positive/negative probability (the output)
- **Sea green**: from left to right, symptoms and pathologies related with pancreatic cancer. Top down, the node of pancreatic cancer that determines the final positive/negative probability (the output)
- **Red**: from left to right, the general nodes that determine symptoms, risk factors and pathologies
- **Light green**: node that determines the quality of food knowing consume of alcohol, café, vegetables and fruit
- **Lavender**: node that represents exposure to carcinogens. The node depends on: asbests, radiation, pesticides, dyes, petroleum, sulphur,coal and arsenic
- **Dark blue**: node that represents the incidence of diabetes in the final diagnostic. It depends of two types of diabetes: long-term diabetes and diabetes mellitus of recent beginning
- **Brown**: jaundice. It depends on three symptoms: eyes and skin yellowish, urine of dark colour and faeces of clear colour
- **Grey**: digestive problems. This node depends on three symptoms: nausea or vomiting, diarrhea and indigestion
- **Orange**: distension of the gallbladder. Depends on two symptoms: clots or fatty tissue abnormalities and uneven texture of fat tissue

l) **Consumption of coffee**

Some previous studies indicated that coffee consumption might be a risk factor; more recent studies have been unable to confirm this. (see Fig. 4)

Range of values: *little or much*

m) **Consumption of alcohol**

Most studies have found no relationship between alcohol consumption and pancreatic cancer. However, consumption of alcohol in large quantities may increase the risk of developing diabetes and chronic pancreatitis. (see Fig. 4)

Range of values: *little or much*

B. Lung cancer

Lung cancer is so far the deadliest in the developed world. One of the reasons that make it so deadly is that lung cancer is often not detected until it is in advanced stages.

However, some lung cancers are diagnosed in its early stages as they are detected by tests like bronchoscopy (visualization of the interior of the bronchi through a flexible lighted tube), or sputum cytology (microscopic examination of the cells contained in phlegm that is ejected with cough).

Symptoms:

- Nail anomalies (Fig. 2)
- Weight Loss (Fig. 2)
- Loss of appetite (Fig. 2)
- Facial swelling (Fig. 2)
- Swallowing dificulties (Fig. 2)
- Chest pain (Fig. 2)
- Hoarseness (Fig. 2)
- Wheeze (Fig. 2)
- Shortness of breath (Fig. 2)
- Bloody sputum (Fig. 2)
- Cough (Fig. 2)

Risk factors:

- Sex (Fig. 4)
- Age (Fig. 4)
- Smoking (Fig. 4)
- Food (Fig. 4)
- Consumption of coffee (Fig. 4)
- Consumption of alcohol (Fig. 4)
- Race (Fig. 4)
- Obesity and physical inactivity (Fig. 4)
- Exposure to asbests, pesticides, dyes, petroleum (Fig. 4)

Pathologies:

- Chronic bronchitis (Fig. 5)
- Tuberculosis (Fig. 5)
- Pulmonary infarction zones (Fig. 5)

a) **Tobacco**

The tobacco reaches the alveoli and water-soluble components are absorbed by the mucous, not being absorbed the fat-soluble ones (tar or pitch) that contain carcinogenic polycyclic aromatic hydrocarbons. The pitch is phagocytes by alveolar macrophages and eliminated in the sputum, but not all alveolar macrophages will be eliminated in the sputum, many of them break in their journey towards the glottis leaving the pitch. This will be deposited in the carina (confluence of the bronchi), especially in the more upper and peripherical ones, producing its irritation. The pitch irritates the mucous membrane and destroys the surface of the respiratory epithelium forcing the basement membrane to increase their proliferative capacity. The basement membrane increases both his "turn over" (basal cell hyperplasia, stratification of the epithelium), ending producing squamous cell metaplasia, which will evolve to dysplasia (carcinoma in situ) and finally to anaplasia as carcinoma in situ and invasive carcinoma. [24] . (see Fig. 4)

Range of values: *former smoker, non-smoking, moderate or high*

b) **Sex**

Lung cancer is the most common malignancy in men. The relationship between the sexes nowadays is 4 men for every woman: (4:1). The woman continues to have a lower incidence, but has already been located in some countries in second place after breast cancer, or even in the first place.

Several studies have found that the lung cells of women have a higher chance of contracting cancer when exposed to tobacco. (see Fig. 4)

Range of values: *male or female*

c) **Family history**

If you've suffered a lung cancer, there is an increased risk of having another lung cancer. Brothers and sons of people who have had lung cancer may have a slightly higher risk. If the father and grandfather of a person died from lung cancer, and this person smokes, the most likely cause of his death will be a lung cancer. . (see Fig. 4)

Range of values: *negative or positive*

d) **Exposure to work conditions**

Asbests: This is another risk factor for lung cancer. People who work with asbests have a higher risk of suffering from lung cancer and, if also smoke, the risk increases greatly. Although asbests has been used for many years, Western governments have almost eliminated its use at work and in household products. The rate of lung cancer linked to asbests, mesothelioma, often begins in the pleura.

Radon: Radon is a radioactive gas that is produced by the natural decay of uranium. Radon is invisible and has no taste or smell. This gas can concentrate inside houses and become a potential risk of cancer.

Some workers industry-related to asbests, arsenic, sulfur, (the three Aes) vinyl chloride, hamatita, radioactive materials, nickel chromates, coal products, mustard gas, chloromethyl ethers, gasoline and diesel derivatives, iron, beryllium, etc.., whereas the non-smoker has a probability of 1 suffering from lung cancer, the smoker has 30 or 40, employees of these industries have up to 70 times more at risk. All types of radiation are carcinogenic. The uranium is weakly radioactive, but lung cancer is four times more prevalent among non-smoker miners of uranium mines than in the general population and ten times more prevalent among smoker miners. [25, 26] (see Fig. 4)

Range of values: *negative or positive*

e) **Pollution (place of residence)**

It is conceivable that pollutants from the atmosphere (especially urban) play a role in increasing the incidence of bronchogenic carcinoma today. Lung cancer is most common in the city than in rural areas (1,3-2,3 times higher among men with similar smoking costumes) because:

- Engine gases (cars and motor vehicles in general) and heating systems. The sulfur dioxide is one of the most important carcinogenic substances.
- Particles of pitch pavement of the streets.
- Radioactive Particles.
- The radon gas and natural radioactivity are very abundant in certain geographical areas.

These four factors have an action mechanism like the tobacco. While most authors acknowledge the existence of a small urban factor in the incidence of lung cancer, the main culprit, with overwhelming numerical difference is the tobacco. (see Fig. 4)

Range of values: *field or city*

f) **Diet**

Some studies conclude that diets with little vegetables could increase the risk of lung cancer in people who are exposed to tobacco. It is possible that apples, onions and other vegetables contain substances that offer some protection against lung cancer. It is believed that certain vitamins, especially vitamins A and C, are protective of the bronchial walls, by their ability to inactivate free radicals, carcinogens, or their ability to precisely regulate certain cell functions, across different mechanisms. But it has been shown that β-carotene was ineffective as chemoprevention of lung cancer and also two large studies like the Carotene and Efficacy Trial (caret) and Alpha-Tocopherol, Beta-Carotene (ATBC) Lung Cancer Prevention Study, 18000 and 29000 participants, respectively, showed that consumption of beta-carotene increased (rather than reducing) the risk of contracting lung cancer (18% on average), and particularly for smokers of 40 or more cigarettes daily , Who experienced an increase in the incidence of 42%. (see Fig. 4)

Range of values: *good or poor*

g) **Prior bronchial pathologies**

Chronic bronchitis: It is accepted that can cause lung cancer.

TB: It is a chronic irritation on the lungs parenchyma that leaves a scar that favors the emergence of lung cancer (carcinoma or scar scar on cancer, especially adenocarcinoma).

Areas of pulmonary infarction, inclusions of foreign bodies, idiopathic pulmonary fibrosis (10% dying from cancer bronchogenic), scleroderma and other kinds of scars. (see Fig. 5) Range of values: *negative or positive*

h) **Viruses**

The relation of viruses with lung cancer has two different bases:

In the experimental field, the incidence of bronchial epithelial metaplasia is caused by paramyxovirus.

As for the human pathology, there is a link with bronchus-alveolus carcinoma.

i) **Age**

Mainly this cancer affects people between 55 and 65 years. 80% of cases occur in patients over 50 years of age. Lung cancer is very uncommon in people younger than 40 years. The average age of the patients is 60 years but more cases are being diagnosed in young subjects. (see Fig. 4)

Range of values: *children, young or adult*

IV. THE APPLICATION

B.Description of the system

The main engine of the system is artificial intelligence. Medical knowledge is built through Bayesian networks that simulate disease diagnosis. The system uses the following software: Genie & Smile. (Decision Systems Laboratory, Department of Information Science and Telecommunications and the Intelligent Systems Program, University of Pittsburgh, USA). The project has been developed in the C# language under the platform .NET using the library *smilenet.dll* to access network capacities of the Bayesian networks software previously mentioned.

The network nodes are distributed in three general categories: symptoms, risk factors and diseases. Each one of this general nodes is influenced by several, which are grouped depending on the type of cancer they are related to (pancreatic, lung or both), except in the case of risk factor, which are all considered as global.

Each node contains tables of probability, which reflects the importance of the evidences that each node describes. This probability values must be defined by experts and/or using statistical information. In the case of this prototype, the values have been established according to statistical information found in previous studies. With some reference values, we created a numerical progression from the lower-probability-cases to the higher ones, but this is only an approximation of which it should be. In order to obtain consistent diagnostic values, these tables must be reviewed by an expert who can, according to his knowledge, set more detailed and reliable values for these tables.

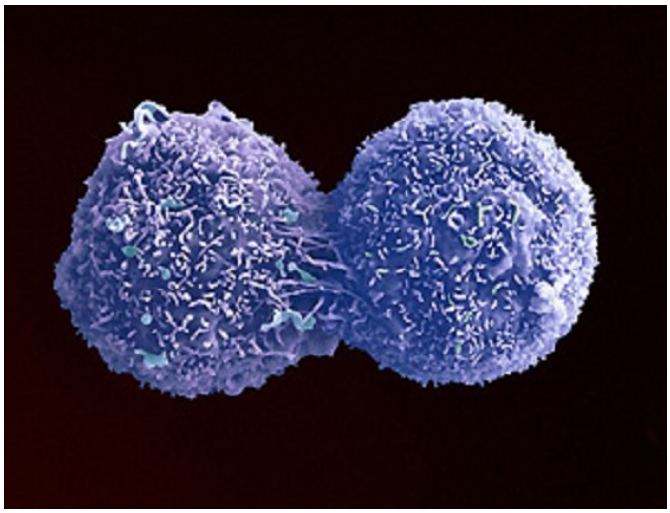

Fig. 7. Colour-enhanced image of a lung cancer cell dividing. [Credit: Anne Weston]

When application is being run, the doctor will ask the patient about the evidences (above nodes of the network) and will insert the appropriate value in each field of the form (each field has as much predetermined answers as rows in its bayesian network's node. The doctor must select one depending on the patient's answer).

If some evidence is unknown, the system will take values of equally likelihood (for example, abdominal distention: <mild, serious> will produce an evidence value of 0.5) in order not to distort the outcome.

Once the form has been completed, the diagnostic can be started. The system will return the likelihood of suffering from lung cancer and pancreatic cancer.

This system has been developed as a web application (which allows easy updates) in order to be adaptable to the complex structure of the modern medical organizations or research groups. Many of these organizations have a distributed architecture which does not allow isolated elements to work properly. The web application allows an expert to modify the bayesian network in the application server (due to new discoveries), either the structure of the bayesian network or the values of the tables, and users will have the latest version available immediately via web (Medical Center Intranet, Internet with VPN, LAN ...).

C.Structure of the application

Each functional element of the object oriented application is implemented by two files: the .aspx file, that shows the interface with the user and the .aspx.cs file, that will control the actions of the functional element. Following this schema we can divide the structure of the application in three parts:

1. **Input**:

a) *AgenteDeInterfaz.aspx*: this file shows a form with fields that should be filled in by the user. The fields are drop down lists with several options depending on the content. For easy understanding, the fields are organized in: symptoms, risk factors, family history, exposure to agents and pathologies. At the end of the form there is a button to make the diagnosis.

b) *AgenteDeInterfaz.aspx.cs*: this file recovers the values of the fields selected by the user and stores them in a web session (a web session is a technical resource for keeping information between different web pages). Then the control flow of the program redirects to the next functional element (the Filter).

2. **Filter**:

a) *AgenteDeFiltrado.aspx:* this file does not show anything. It is needed due to structure of the application.

b) *AgenteDeFiltrado.aspx.cs:* this file creates a new object for managing the bayesian network. At this point we can comment that the bayesian network (see Fig. 6) is a file called 'diagnosticoCancer.xdsl' (this file is created previously with the Genie´s software and represents the medical knowledge of the system) and stored in the root directory of the application. After creating the bayesian network management object we read the network and now we have access to its values. Then we recover the values stored in the web session and we check if the fields have been filled in. If a field is void we set neutral values of likelihood depending on the type of field. We have three types: with two options (0.5), with three options (0.3333) and with four options (0.25). When all checks are finished we write the new values of the network and the control flow of the program redirects to the next functional element (the Output).

3. **Output**:

a) *AgenteRedBayesiana.aspx:* this file shows the final results of the diagnostics. It points out through labels: the positive/negative probability of pancreatic cancer and the positive/negative probability of lung cancer.

b) *AgenteRedBayesiana.aspx.cs:* this file is responsible for the calculations of probability that express the diagnosis. First of all it makes the same work as '*AgenteDeFiltrado.cs'*. But, instead of doing the checks, it sets the values of evidence of the network (for a better understanding see the reference [27]). If a value has not been selected we set some default one. Then we update the beliefs of the bayesian network. And now we make up the calculations of probability with the two final nodes (pancreatic cancer and lung cancer, see Fig. 6) and show the results to the user.

D.Efficiency analysis

In computational complexity theory, big O notation is often used to describe how the size of the input data affects an algorithm's usage of computational resources (usually running time or memory).

The application uses simple instructions and merely carries out checks on the entry of information. As already commented, application filters those data that are not introduced and provides default values. The complexity of these actions is minimal: O(1).

However, the most complex actions are the two 'for' loops that estimate the probability of suffering pancreatic or lung cancer. Each loop has a complexity of O(n).

So, the complexity of the application is O(n), technically it is said linear complexity. Hence we can say that the application is quite efficient.

Of course this analysis only covers our application. For a complete estimate, calculations of Genie´s software on the network should be considered.

For more information read information related to the Bayesian networks, inference, the Bayes theorem ... We suggest as an introduction the following article in the reference [27].

V.CONCLUSION AND FURTHER WORK

Information and communications technologies are an essential tool in the field of health sciences both in research and management. The health system will work better if we introduce them gradually.

This project would produce benefits to pancreatic cancer research (the project is useful as a basis of a Bayesian network more realistic) as well as in a more agile working of medical consultations.

For future developments it would be a chance to add to the Bayesian network results of diagnostic tests (images,

analysis,…), thereby obtaining a complete diagnosis based on symptoms, risk factors, disease and diagnostic tests. The system could, for example, communicate with other medical image processing systems or implement a subsystem of image recognition (using neural networks).

References

[1] PancPro. url: http://astor.som.jhmi.edu/BayesMendel/pancpro.html. Last access: 11/04/07

[2] Mortalidad por cáncer en España. Año 2004. url: http://www.todocancer.com/NR/rdonlyres/8D05F58A-3E73-4B42-B231-102B963BCB41/0/MortPresentacion04.pdf. Last access: 11/04/07

[3] Parker SL, Tong T, Bolden S, Wingo PA. Cancer statistics. Cancer J Clin 1997; 47:5-27.

[4] Warshaw AL, Fernandes-Del Castillo C. Pancreatic carcinoma. N Eng J Med 1992; 326:455-465.

[5] Fernández E, La Vecchia C, Porta M, et al. Trends in pancreatic cancer in Europe, 1955-1989. In J Cancer 1994; 57:786-792.

[6] Gudjonsson B. Cancer of the pancreas: 50 years of surgery. Cancer 1987; 60:2284-2303.

[7] National Cancer Institute. Annual cancer statistics review 1973-1988. Department of Healht and Human Services, Bethesda 1991 (NIH publication no. 91-2789).

[8] Lillemoe KD. Current management of pancreatic carcinoma. Ann Surg 1995; 221:133-148.

[9] ACS::Prevención y factores de riesgo. Cáncer de páncreas. url: http://www.cancer.org/docroot/esp/content/esp_5_1x_prevencion_y_factores_de_riesgo_34.asp. Last access: 05/16/08

[10] Gold EB, Goldin SB. Epidemiology of and risk factors for pancreatic cancer. Surg Oncol Clin North Am 1998; 7:67-91.

[11] MedicinePlus Article. url: http://www.nlm.nih.gov/medicinesplus/spanish/ency/article/000236.htm. Last access: 11/04/07

[12] Diagnostic ultrasound. url: http://drgdiaz.com/eco/capancreas.shtml. Last access: 05/16/08

[13] Cáncer del Páncreas. National Cancer Institute. url: http://www.cancer.gov/espanol/pdq/tratamiento/pancreas/patient. Last access: 05/16/08

[14] MedicinePlus. Pancreatic Cancer. url: http://www.nlm.nih.gov/medlineplus/pancreaticcancer.html. Last access: : 05/16/08

[15] Pancreatitis crónica. url: http://www.aegastro.es/Areas/Pancreas/CaP_pacientes.htm. Last access: 05/16/08

[16] Douglas HD Jr, Kim SY, Meropol NJ. Neoplasms of the exocrine pancreas. En: Holland JF, Frei III E, Bast RC Jr, Kufe DE, Morton DL,Weichelbaum RR (eds). Cancer medicine. Baltimore, William and Wilkins 1997; pp 1989-2017.

[17] Krech RL, Walsh D. Symptoms of pancreatic cancer. J Pain Symptom Management 1991; 6:360-367.

[18] Lynch HT, Smyrk T, Kem SE, et al. Familial pancreatic cancer: A review. Semin Oncol 1996; 23:251-275.

[19] Kauppinen T, Partanen T, Degerth R, et al. Pancreatic cancer and occupational exposures. Epidemiology 1995; 6:498-502.

[20] Ahlgren JD. Epidemiology and risk factors in pancreatic cancer. Semin Oncol 1996; 23:241-250.

[21] Fuchs CS, Colditz GA, Stampher MJ, et al. A prospective study of cigarrette smoking and the risk of pancreatic cancer. Arch Intern Med 1996; 156:2255-2260.

[22] Organización Mundial de la Salud|OMS (2004). Deaths by cause, sex and mortality stratum

[23] Wikipedia [cáncer de pulmón]

[24] Editorial El Mundo online (enero 2004). *El riesgo de cáncer es el mismo con el tabaco 'bajo en alquitrán'*

[25] NCI. url: http://www.cancer.gov/espanol/noticias/atlas

[26] NCI. url: Cancer mortality graphs and maps: http://cancercontrolplanet.cancer.gov/atlas/index.jsp.

[27] Bayesian network. url: http://en.wikipedia.org/wiki/Bayesian_network