MEDICAL EXPERT SYSTEM FOR DIAGNOSIS OF ABDOMINAL DISORDERS

Project Report submitted in partial fulfillment of the requirement for the degree of Bachelor of Technology in

Computer Science & Engineering

Under the supervision of

Ms. RAMANPREET KAUR

By

ABHISHEK GARG

(111218)

To

Jaypee University of Information and Technology

Waknaghat, Solan – 173234, Himachal Pradesh

Certificate

This is to certify that project report entitled "**MEDICAL EXPERT SYSTEM FOR DIAGNOSIS OF ABDOMINAL DISORDERS** ", submitted by **ABHISHEK GARG** in partial fulfillment for the award of degree of Bachelor of Technology in Computer Science & Engineering to Jaypee University of Information Technology, Waknaghat, Solan has been carried out under my supervision.

This work has not been submitted partially or fully to any other University or Institute for the award of this or any other degree or diploma.

Date:

Ms. Ramanpreet Kaur

Assistant Professor (Grade-II)

Acknowledgement

I would like to express the deepest appreciation to HOD(CSE/IT), **Prof. Dr. S P Ghrera** , who has the attitude and the substance of a genius: he continually and convincingly conveyed a spirit of adventure in regard to research and education.

I would like to express my sincere thanks to my guide, **Ms. Ramanpreet Kaur** who gave me the golden opportunity to do this wonderful project on the topic 'Medical Expert System for Diagnosis of Abdominal Disorders', which also helped me in doing a lot of research and I came to know about so many new things. I am highly indebted for the constant supervision as well as for providing necessary information regarding the project.

Date: **Name of the student:** ABHISHEK GARG

Table of Content

Abbreviation and Symbols

1.	**BBN**	Bayesian belief network
2.	**D-map**	Dependency map
3.	**DAG**	Directed acyclic graph
4.	**E**	An evidence node or variable.
5.	**G(V,E)**	The undirected graph containing the set of vertices V and edges E
6.	**P(A)**	The probability table containing the prior probabilities p (a).
7.	**p (a)**	Shorthand for p (A = a). Used where A is obvious from the context
8.	**p (A = a)**	The probability that the random variable A takes the value a as its state.
9.	**P(A|B)**	The probability table containing the conditional probabilities p(a|b)
10.	**p (a|b)**	The conditional probability that A = a, given that B = b.
11.	**CPT**	Conditional Probability Table.
12.	**MSBNx**	Microsoft Belief Network Tool.

List of Tables and Figures

Abstract

Artificial Intelligence (Al) is the area of computer science concerned with the emulation of human thought processes. Efforts in the application of Al methods to intelligent problem solving led to the development of expert systems, systems which perform tasks that require a great deal of specialized knowledge that experts in a particular field acquire from long experience with such tasks. Expert systems are used extensively in many domains ranging from medicine to science and space technology. Many rural communities have an extremely limited access to medical advice. People travel long distances to clinics or medical facilities and there is a shortage of medical experts in most of these facilities. This results in slow service and patients end up waiting long hours without receiving any attention. This problem can be solved by creating a system that can give advice for common conditions such as abdominal diseases. Hence medical expert systems can play a significant role in such cases where medical experts are not readily available. In this, aspects of the design of an intelligent medical system for diagnosis of Common disease that can be detected by expert analysis and research on history. A number of patient cases are selected as prototype. The knowledge acquired from history review and human experts of the specific domain and is used as a base for analysis, diagnosis and recommendations. All the acquired knowledge is represented through hierarchal data chart that combines with production rules and Bayesian network rules. This results in better representation, and facilitates knowledge acquisition and maintenance. Diagnosis is performed through the hierarchal charts, based on patient data and expert analysis. The proposed system will be experimented on various scenarios in order to evaluate its performance.

1. INTRODUCTION

Prior to the advancements of the automation technologies, doctors' works were solely dependent on their instincts and their abilities. The inferences were deduced from the knowledge based on their experiences.

With the advent of AI theories and a considerable boom in the expert and decision support system approaches, the automation of any manual system is inevitable. When it comes to medical sciences, automation of experts is a very critical subject. One of the key factors for the acceptance of these systems in the real world domains is the capability to explain their reasoning.

Understanding technology adoption in emerging regions is challenging given the complex interrelations among socioeconomic factors that affect it directly and indirectly. The issue of impact assessment of technology adoption projects, especially the kind implemented in areas where prior technology has been very limited, is highly problematic and open to many methodological difficulties. Ethnographic evaluations have provided insight into the quality of interactions and into conceptions of technology and its adoption, whereas some quantitative analysis has been useful for high-level abstraction.

One of the essential qualities of real experts is their ability to communicate their knowledge and explain their reasoning. This ability is especially important in the case of clinical systems, not only for tracing the performance during the construction and evaluation of the system, but also for justifying their results when the system is deployed in an operating environment.

The support system gives a broader dimension to a doctor's work and reduces time and space complexities.

There have been multiple attempts in the past to build expert systems that can perform automated uncertainty reasoning and generate user understandable explanations.

In this report I will describe an application of Bayesian networks in the area of diagnosis. Diagnosis has been a crucial application area for AI methodologies due to its high complexity and its requirements for data.

Our mental capacities are so important to our everyday lives and our sense of self that humankind has given itself the scientific name homo-sapiens; man the wise. The field of artificial intelligence (AI) attempts to understand these capacities better known as intelligent entities. AI being a broad topic combines computer science, physiology and philosophy.

Computer-based methods are increasingly used to improve the quality of medical services. Artificial Intelligence (AI) is the area of computer science focusing on creating machines that can engage on behaviours that humans consider intelligent.

The ability to create intelligent machines has intrigued humans since ancient times and today with the advent of the computer and 50 years of research into AI programming techniques, the dream of smart machines is becoming a reality.

Researchers are creating systems which can mimic human thought, understand speech, beat the best human chess player, and countless other feats never before possible.

One of the large scale applications of the field of AI is the development of various decision support and expert systems. Combining expert knowledge and user explanation with automated reasoning in domains with uncertain information poses significant challenges in terms of representation and reasoning mechanisms.

In particular, reasoning structures understandable and usable by humans are often different from the ones for automated reasoning and data mining systems. In modelling intelligent systems for real world applications, one inevitably has to deal with uncertainty. This uncertainty is due to the impossibility to model all the different conditions and exceptions that can underlie a finite set of observations.

Probability theory provides the mathematically consistent framework to quantify and to

compute with uncertainty. In principle, a probabilistic model assigns a probability to each of its possible states. In models for real world applications, the number of states is so large that a sparse model representation is inevitable.

A general class with a representation that allows modelling with many variables are the Bayesian networks. Rules are a convenient and human understandable way to express domain knowledge and build expert systems.

Adding certainty factors to these rules presents one way to deal with uncertainty in rule based expert systems. However such systems have limitations in accurately modelling the domain.

A Bayesian Network, on the other hand, is a probabilistic graphical model that allows accurate modelling of a domain and automated reasoning. But inference in Bayesian Networks is harder for humans to comprehend.

There are multiple frameworks to perform reasoning in domains with uncertain information. An ideal framework must have convenient mechanisms to perform theoretically consistent and automated reasoning. If human users are involved, it must be able to generate user understandable explanations for its predictions.

If the expert system also involves a domain expert, it should have mechanisms to elicit and store the domain knowledge of the expert.

It is well known that experts are most comfortable in specifying their domain knowledge using sets of rules because rules model experts' decision making process naturally.

Research has shown that explanation of the diagnosis in terms of the expert specified rules is much more intuitive to the user than a set of raw values.

Bayesian networks are nowadays well established as a modelling tool for ex- pert systems in domains with uncertainty. Reasons are their powerful but conceptual transparent representation for probabilistic models in terms of a network.

Their graphical representation, showing the conditional independencies between variables, is easy to understand for humans. On the other hand, since a Bayesian network uniquely defines a joint probability model, inference — drawing conclusions based on observations — is based on the solid rules of probability calculus. This implies that the mathematical consistency and correctness of inference are guaranteed. In other words, all assumptions in the method are contained in model, i.e., the definition of variables, the graphical structure, and the parameters.

The method has no hidden assumptions in the inference rules. This is unlike other types of reasoning systems such as e.g., Certainty Factors (CFs) that were used in e.g., MYCIN — a medical expert system developed in the early 1970s. In the CF framework, the model is specified in terms of a number of if-then-else rules with certainty factors.

Furthermore, the CF framework provides prescriptions how to invert and/or combine the if-then-else rules to do inference. These prescriptions contain implicit conditional independence assumptions which are not immediately clear from the model specification and have consequences in their application.

2. PROBLEM IDENTIFICATION

2.1 Medical Expert System

Modern-day medical diagnosis is a very complex process, requiring accurate patient data, a profound understanding of the medical literature and many years of clinical experience. This situation applies particularly to internal medicine, because it covers an enormous range of diagnostic categories. As a result, internal medicine is differentiated in super-specializations.

Diagnosis is a process, by which a doctor searches for the cause (usually a disease) that best explains the symptoms of a patient. The search process is sequential, in the sense that patient symptoms suggest some initial tests to be performed. Based on the outcome of these tests, a tentative hypothesis is formulated about the possible cause(s).

Based on this hypothesis, subsequent tests are ordered to confirm or reject this hypothesis. The process may proceed in several iterations until the patient is finally diagnosed with sufficient certainty and the cause of the symptoms is established.

A significant part of the diagnostic process is standardized in the form of protocols. These are sets of rules that prescribe which tests to perform and in which order, based on the patient symptoms and previous test results.

These rules form a decision tree, whose nodes are intermediate stages in the diagnostic process and whose branches point to additional testing, depending on the current test results. The protocols are defined in each country by a committee of medical experts.

In the majority of the diagnoses that are encountered, the guidelines are sufficiently accurate to make the correct diagnosis. For these “routine” cases, a decision support system is not needed.

A clinical support system is an interactive computer application designed to assist health professionals with decision making, generally determining diagnosis of patient data.

2.2 Types of Expert systems

There are two types of expert systems: rule-based expert systems and knowledge-based expert systems. The main difference between these expert systems is the knowledge representation in the knowledge base. The knowledge representation is more significant in expert system because the approach used to represent knowledge affects the development, efficiency, speed and the maintenance of the system.

2.2.1 Rule based expert system

The rule-based expert system has domain knowledge encoded in the form of rules from a human expert. A rule is a conditional statement that links given conditions to actions. In a rule-based expert system, a knowledge base is usually stored in terms of if-then rules which can be used to reach conclusions. Hence it uses a set of rules to analyse information about a specific class of problems and recommend one or more possible solutions.

2.2.2 Knowledge based expert system

The knowledge-based expert system encodes heuristics and rules into decision making framework. A knowledge-based system uses artificial intelligence techniques in problem solving methods to support human decision making, learning, and action. The knowledge base of expert systems contains both factual and heuristic knowledge. Factual knowledge is the knowledge that is widely shared, typically found in textbooks or journals, and commonly agreed upon by human experts in that particular domain. Heuristic knowledge refers to an experiential, logical and judgmental knowledge used to speed up decision making. Some applications applied by knowledge based expert systems are: medical treatment, production management, knowledge management, financial analysis, etc.

3. REQUIREMENT ANALYSIS

For a working proposed system I identified 10 abdominal diseases with the help of a survey. Several doctors were asked about the symptoms that led to the selected diseases. Thus, a total of 27 symptoms were analyzed.

The chosen symptoms and diseases were transformed into simple questionnaires to obtain the probabilities of the relation of diseases and symptoms with each other. These were then filled up by experts (doctors, medicos, etc).finally the questionnaires were incorporated into conditional probability tables (CPTs) for each disease. [ref B]

Table 1: List of diseases and corresponding symptoms

Acidity	Indigestion Burning sensation Belching Vomiting Fullness
Gastritis	Acidity Upper abdomen pain Gas Nausea Fever
Ameobaisis	Mucus in faeces Left side pain Gastritis
Cholecystitis	Right side pain Gastritis vomiting

Pancreatitis	Upper abdomen left side pain Vomiting Fever
Nephritis	Frequent urination Renal angle pain(left side) Fullness of abdomen Vomiting Fever
Cholitis	Right iliac pain Mucus in faeces constipation vomiting
Appendicitis	Mac burny's point tenderness Constipation Vomiting
Duodenal ulcer	Pain in pre-umbilical region Pain aggrevation after eating Blood in stool Gastritis
Gastric ulcer	Pain in epigastric region Pain aggravation before eating Blood vomiting gastritis

s

3.1 DISEASES TO BE DIAGNOSED

3.1.1. GASTRITIS: Gastritis is an inflammation of the lining of the stomach, and has many possible causes. The main acute causes are excessive alcohol consumption or prolonged use of non steroidal anti-inflammatory drugs (also known as NSAIDs) such as aspirin or ibuprofen. Sometimes gastritis develops after major surgery, traumatic injury, burns, or severe infections. Gastritis may also occur in those who have had weight loss surgery resulting in the banding or reconstruction of the digestive tract. Chronic causes are infection with bacteria, primarily *Helicobacter pylori*, chronic bile reflux, and stress; certain autoimmune disorders can cause gastritis as well. The most common symptom is abdominal upset or pain. Other symptoms are indigestion, abdominal bloating, nausea, and vomiting and pernicious anemia. Some may have a feeling of fullness or burning in the upper abdomen. A gastroscopy, blood test, complete blood count test, or a stool test may be used to diagnose gastritis. Treatment includes taking antacids or other medicines, such as proton pump inhibitors or antibiotics, and avoiding hot or spicy foods. For those with pernicious anemia, B_{12} injections are given, but more often oral B12 supplements are recommended.

3.1. 2. PANCREATITIS: Pancreatitis is inflammation of the pancreas. It has several causes and symptoms and requires immediate medical attention. It occurs when pancreatic enzymes (especially trypsin) that digest food are activated in the pancreas instead of the small intestine. It may be acute, beginning suddenly and lasting a few days, or chronic, occurring over many years. The most common symptoms of pancreatitis are severe upper abdominal burning pain radiating to the back, nausea, and vomiting that is worsened with eating. The physical examination will vary depending on severity and presence of internal bleeding. Blood pressure may be elevated by pain or decreased by dehydration or bleeding. Heart and respiratory rates are often elevated. The abdomen is usually tender but to a lesser degree than the pain itself. As is common in abdominal disease, bowel sounds may be reduced from reflex bowel

paralysis. Fever or jaundice may be present. Chronic pancreatitis can lead to diabetes or pancreatic cancer. Unexplained weight loss may occur from a lack of pancreatic enzymes hindering digestion.

Early complications include shock, infection, systemic inflammatory response syndrome, low blood calcium, high blood glucose, and dehydration. Blood loss, dehydration, and fluid leaking into the abdominal cavity (ascites) can lead to kidney failure. Respiratory complications are often severe. Pleural effusion is usually present. Shallow breathing from pain can lead to lung collapse. Pancreatic enzymes may attack the lungs, causing inflammation. Severe inflammation can lead to intra-abdominal hypertension and abdominal compartment syndrome, further impairing renal and respiratory function and potentially requiring management with an open abdomen (laparostomy) to relieve the pressure.

3.1.3. NEPHRITIS: *Nephritis* comes from Latin, from Ancient Greek νεφρῖτις, from νεφρός "kidney" and -ῖτις, a feminine adjective ending Glomerulonephritis is inflammation of the glomeruli. (When the term "nephritis" is used without qualification, this is often the condition meant). Interstitial nephritis or tubulo-interstitial nephritis is inflammation of the spaces between renal tubules. Nephritis is often caused by infections, toxins, and auto-immune diseases. It can be caused by infection, but is most commonly caused by autoimmune disorders that affect the major organs. For example, those with lupus are at a much higher risk for developing nephritis. In rare cases nephritis can be genetically inherited, though it may not present in childhood (But has been known to be diagnosed in young girls of teen years). Pyelonephritis is inflammation that results from a urinary tract infection that reaches the pyelum (pelvis) of the kidney. Lupus nephritis is an inflammation of the kidney caused by systemic lupus erythematosus (SLE), a disease of the immune system. *Athletic nephritis* is nephritis resulting from strenuous exercise. It may result in proteinuria, hematuria and cylinduria. In most persons these are transient findings that disappear within hours to days after the end of exercise. The findings

generally increase by increasing severity and duration of physical stress. Hematuria after strenuous exercise may also result from march hemoglobinuria, which is caused by trauma to red blood cells, causing hemolysis and resultant release of hemoglobin into the blood. Nephritis is the most common producer of glomerular injury. It is a disturbance of the glomerular structure with inflammatory cell proliferation. This can lead to reduced glomerular blood flow, leading to reduced urine output (oliguria) and retention of waste products (uremia). As a result, red blood cells may leak out of damaged glomeruli, causing blood to appear in the urine (hematuria). Low renal blood flow activates the renin-angiotensin-aldosterone system (RAAS), causing fluid retention and mild hypertension.

Nephritis is a serious medical condition which is the eighth highest cause of human death. As the kidneys inflame, they begin to excrete needed protein from the body into the urine stream. This condition is called proteinuria. Loss of necessary protein due to nephritis can result in several life-threatening symptoms. Most dangerous in cases of nephritis is the loss of protein that keeps blood from clotting. This can result in blood clots causing sudden stroke.

3.1.4. Appendicitis: Appendicitis is a condition characterized by inflammation of the appendix. It is classified as a medical emergency and many cases require removal of the inflamed appendix, either by laparotomy or laparoscopy. Untreated, mortality is high, mainly because of the risk of rupture leading to infection and inflammation of the intestinal lining (peritoneum) and eventual sepsis, clinically known as peritonitis which can lead to circulatory shock. Reginald Fitz first described acute and chronic appendicitis in 1886, and it has been recognized as one of the most common causes of severe acute abdominal pain worldwide. A correctly diagnosed non-acute form of appendicitis is known as "rumbling appendicitis".

The term "pseudo-appendicitis" is used to describe a condition mimicking appendicitis. It can be associated with *Yersinia enterocolitica*. Pain first, vomiting next and fever last has

been described as the classic presentation of acute appendicitis. Since the innervation of the appendix enters the spinal cord at the same level as the umbilicus (belly button), the pain begins stomach-high. Later, as the appendix becomes more swollen and irritates the adjoining abdominal wall, it tends to localize over several hours into the right lower quadrant, except in children under three years. This pain can be elicited through various signs and can be severe. Signs include localized findings in the right iliac fossa. The abdominal wall becomes very sensitive to gentle pressure (palpation). Also, there is severe pain on sudden release of deep pressure in the lower abdomen (rebound tenderness). In case of a retrocecal appendix (appendix localized behind thececum), however, even deep pressure in the right lower quadrant may fail to elicit tenderness (silent appendix), the reason being that the cecum, distended with gas, protects the inflamed appendix from the pressure. Similarly, if the appendix lies entirely within the pelvis, there is usually complete absence of abdominal rigidity. In such cases, a digital rectal examination elicits tenderness in the rectovesical pouch. Coughing causes point tenderness in this area (McBurney's point) and this is the least painful way to localize the inflamed appendix. If the abdomen on palpation is also involuntarily guarded (rigid), there should be a strong suspicion of peritonitis, requiring urgent surgical intervention.On the basis of experimental evidence, acute appendicitis seems to be the end result of a primary obstruction of the appendix lumen (the inside space of a tubular structure). Once this obstruction occurs, the appendix subsequently becomes filled with mucus and swells, increasing pressures within the lumen and the walls of the appendix,

resulting in thrombosis and occlusion of the small vessels, and stasis of lymphatic flow. Rarely, spontaneous recovery can occur at this point.

As the former progresses, the appendix becomes ischemic and then necrotic. As bacteria begin to leak out through the dying walls, pus forms within and around the appendix (suppuration). The end result of this cascade is appendiceal rupture (a 'burst appendix') causing peritonitis, which may lead to septicemia and eventually death.

The causative agents include foreign bodies, trauma, intestinal worms, lymphadenitis, and, most commonly, calcified fecal deposits known as appendicoliths or fecaliths. The occurrence of obstructing fecaliths has attracted attention since their presence in patients with appendicitis is significantly higher in developed than in developing countries, and an appendiceal fecalith is commonly associated with complicated appendicitis. Also, fecal stasis and arrest may play a role, as demonstrated by a significantly lower number of bowel movements per week in patients with acute appendicitis compared with healthy controls. The occurrence of a fecalith in the appendix seems to be attributed to a right-sided fecal retention reservoir in the colon and a prolonged transit time. From epidemiological data, it has been stated that diverticular disease and adenomatous polyps were unknown and colon cancer exceedingly rare in communities exempt from appendicitis.

Also, acute appendicitis has been shown to occur antecedent to cancer in the colon and rectum. Several studies offer evidence that a low fiber intake is involved in the pathogenesis of appendicitis. This is in accordance with the occurrence of a right-sided fecal reservoir and the fact that dietary fiber reduces transit time.

3.1.5. Amoebiasis: Amoebiasis, or Amebiasis, refers to infection caused by the amoeba *Entamoeba histolytica*. The term Entamoebiasis is occasionally seen but is no longer in use; it refers to the same infection. Likewise **amoebiasis** is sometimes incorrectly used to refer to infection with other amoebae, but strictly speaking it should be reserved for *Entamoeba histolytica* infection. Other amoebae infecting humans include:

A gastrointestinal infection that may or may not be symptomatic and can remain latent in an infected person for several years, amoebiasis is estimated to cause 70,000 deaths per year world wide. Symptoms can range from mild diarrhea to dysentery withblood and mucus in the stool. *E. histolytica* is usually

a commensal organism. Severe amoebiasis infections (known as invasive or *fulminant* amoebiasis) occur in two major forms. Invasion of the intestinal lining causes amoebic dysentery or amoebic colitis. If the parasite reaches the bloodstream it can spread through the body, most frequently ending up in the liver where it causes amoebic liver abscesses. Liver abscesses can occur without previous development of amoebic dysentery. When no symptoms are present, the infected individual is still a carrier, able to spread the parasite to others through poor hygienic practices. While symptoms at onset can be similar to bacillary dysentery, amoebiasis is not bacteriological in origin and treatments differ, although both infections can be prevented by good sanitary practices. Amoebiasis is usually transmitted by the fecal-oral route, but it can also be transmitted indirectly through contact with dirty hands or objects as well as by anal-oral contact. Infection is spread through ingestion of the cyst form of the parasite, a semi-dormant and hardy structure found in feces. Any non-encysted amoebae, or *trophozoites*, die quickly after leaving the body but may also be present in stool: these are rarely the source of new infections. Since amoebiasis is transmitted through contaminated food and water, it is often endemic in regions of the world with limited modern sanitation systems, including México, Central America, western South America, South Asia, and western and southern Africa.

Amoebic dysentery is often confused with "traveler's diarrhea" because of its prevalence in developing nations. In fact, most traveler's diarrhea is bacterial or viral in origin. Most infected people, about 90%, are asymptomatic, but this disease has the potential to make the sufferer dangerously ill. It is estimated that about 40,000 to 100,000 people worldwide die annually due to amoebiasis.

Infections can sometimes last for years. Symptoms take from a few days to a few weeks to develop and manifest themselves, but usually it is about two to four weeks. Symptoms can range from mild diarrhea to dysentry with blood and mucus. The blood comes from amoebae invading the lining of the intestine. In about 10% of invasive cases the amoebae

enter the bloodstream and may travel to other organs in the body. Most commonly this means the liver, as this is where blood from the intestine reaches first, but they can end up almost anywhere.

Onset time is highly variable and the average asymptomatic infection persists for over a year. It is theorised that the absence of symptoms or their intensity may vary with such factors as strain of amoeba, immune response of the host, and perhaps associated bacteria and viruses.

In asymptomatic infections the amoeba lives by eating and digesting bacteria and food particles in the gut, a part of the gastrointestinal tract. It does not usually come in contact with the intestine itself due to the protective layer of mucus that lines the gut. Disease occurs when amoeba comes in contact with the cells lining the intestine. It then secretes the same substances it uses to digest bacteria, which include enzymes that destroy cell membranes and proteins. This process can lead to penetration and digestion of human tissues, resulting first in flask-shaped ulcers in the intestine. *Entamoeba histolytica* ingests the destroyed cells by phagocytosis and is often seen with red blood cells inside when viewed in stool samples. Especially in Latin America,a granulomatous mass (known as an amoeboma) may form in the wall of the ascending colon or rectum due to long-lasting immunological cellular response, and is sometimes confused with cancer.

"Theoretically, the ingestion of one viable cyst can cause an infection."

3.1.6.Colitis: In medicine, colitis (pl. colitides) refers to an inflammation of the colon and is often used to describe an inflammation of the large intestine (colon, caecum and rectum).Colitides may be acute and self-limited or chronic, i.e. persistent, and broadly fits into the category of digestive diseases.In a medical context, the label *colitis* (without qualification) is used if:

The aetiology of the inflammation in the colon is undetermined; for example, *colitis* may be applied to *Crohn's disease* at a time when the diagnosis has not declared itself, or The context is clear; for example, an individual with ulcerative colitis is talking about their disease with a physician that knows the diagnosis.

The signs and symptoms of colitides are quite variable and dependent on the etiology (or cause) of the given colitis and factors that modify its course and severity. Symptoms of colitis may include: abdominal pain, loss of appetite, fatigue, diarrhea, cramping, urgency and bloating. Signs may include: abdominal tenderness, weight loss, changes in bowel habits (increased frequency), fever, bleeding (overt or occult)/bloody stools, diarrhea and distension. Signs seen on colonoscopy include: colonic mucosal erythema (redness of the inner surface of the colon), ulcers, bleeding. **Fulminant colitis** is any colitis that becomes worse rapidly. In addition to the diarrhea, fever, and anemia seen in colitis, the patient has severe abdominal pain and presents a clinical picture similar to that of septicemia, where shock is present. About half of human patients require surgery. In horses, the fulminant colitis known as colitis X usually results in death within 24 hours.

Irritable bowel syndrome, a separate disease, has been called spastic colitis. This name may lead to confusion, since colitis is not always a feature of irritable bowel syndrome. Since the etiology of IBS is currently unknown and possibly multifactorial, there may be some overlap in symptoms between IBS and the various forms of colitis. *Indeterminate colitis* is a term used for a colitis that has features of both *Crohn's disease* and *ulcerative colitis*. Indeterminate colitis' behaviour is usually closer to ulcerative colitis than Crohn's disease. *Atypical colitis* is a phrase that is occasionally used by physicians for a colitis that does not conform to criteria for accepted types of colitis. It is not an accepted diagnosis *per se* and, as such, a colitis that cannot be definitively classified. A well-known subtype of infectious colitis is Clostridium difficile colitis,which is informally abbreviated as "c diff colitis".

It classically forms pseudomembranes and is often referred to as *pseudomembranous colitis*, which is its (nonspecific) histomorphologic description. Enterohemorrhagic colitis may be caused by Shiga toxin in *Shigella dysenteriae* or *Shigatoxigenic group* of *Escherichia coli* (STEC), which includes serotype O157:H7 and other enterohemorrhagic *E. coli*. Parasitic infections, like those caused by *Entamoeba histolytica*, can also cause colitis.

How a given colitis is treated is dependent on its etiology. Infectious colitis are usually treated with antimicrobial agents (e.g. antibiotics) such as nifuroxazide. Autoimmune mediated colitis is treated with immune modulators/immune suppressants. Severe colitis can be life-threatening and may require surgery.

Ulcerative colitis can often be treated with changes to one's diet. Although the causes of colitis are not known, the disease can become manifested due to faulty diet. When symptoms such as flatulence, cramping, mucous and bloody discharges and blood on stools begin to occur, the B.R.A.T. diet is recommended. This acronym is derived from bananas, rice, apple sauce and toast. Foods to avoid are dairy products and fried foods.

3.1.7. ACIDITY: Acidity refers to a set of symptoms caused by an imbalance between the acid secreting mechanism of the stomach and proximal intestine and the protective mechanisms that ensure their safety. The stomach normally secretes acid that is essential in the digestive process. This acid helps in breaking down the food during digestion. When there is excess production of acid by the gastric glands of the stomach, it results in the condition known as acidity. However, there are certain types of ulcers where acid secretion is either normal or even low. Acidity is responsible for symptoms like dyspepsia, heartburn and the formation of ulcers (erosion of the lining of the stomach or intestines).

Acidity tends to have a much higher incidence in highly emotional and nervous individuals. It is also more common in the developed and industrialised nations, though a recent increase in incidence has also occurred in the developing countries. Consumption

of Alcohol, highly spicy foodstuffs, non-vegetarian diets, and Non Steroidal Anti-Inflammatory Drugs (NSAID's) also predispose to gastric acidity. The stomach, intestines, and digestive glands secrete hydrochloric acid and various enzymes, including pepsin that break down and digest food. The stomach must also be protected from the same acid and enzymes, or it too can be attacked by the gastric juices. The acid may enter the lower part of the Oesophagus (Gastro-Oesophageal Reflux), due to some weakness in the normal sphincter mechanism that prevents such reflux. This causes heartburn. It commonly occurs after meals and is brought on by excess intra-abdominal pressure like lifting weights or straining.

Ulcers also occur as a result of over secretion of acid. This may happen when there is an imbalance between the digestive juices used by the stomach to break down food and the various factors that protect the lining of the stomach and duodenum (the part of the small intestine that adjoins the stomach). A peptic ulcer is a raw area in the lining of the upper part of the small intestine (duodenal ulcer) or the stomach (gastric ulcer), whose protective mucosal lining has been eroded away by the gastric juices. Duodenal ulcers are three times more common than gastric ulcers. Hydrochloric acid, secreted in the stomach, is one of the factors in the development of ulcers, but is not solely responsible. Acid production in patients with duodenal ulcers tends to be higher than normal, while in those with stomach or gastric ulcers, it is usually normal or lower.

Excessively large amounts of acid secretion occur in certain situations, such as in a condition known as Zollinger-Ellison Syndrome, in which large amounts of secretion are stimulated by tumours located in the pancreas or duodenum. Pepsin is an enzyme that breaks down proteins. Pepsin and hydrochloric acid cause damage to the stomach or duodenum if the stomach's protective system is altered or damaged. The mucous layer, which coats the stomach and duodenum, forms the first line of defence against acid and

pepsin. The body also secretes bicarbonate into the mucous layer, which neutralises the acid.

The defence system also consists of hormone-like substances known as prostaglandins, which help to keep the blood vessels in the stomach dilated, ensuring adequate blood flow. Lack of adequate blood flow to the stomach contributes to ulcers. Prostaglandins are also believed to stimulate bicarbonate and mucous production, which help protect the stomach. If any of these defence mechanisms are deficient, acid and pepsin can attack the stomach lining causing an ulcer. Prevention mainly consists of avoiding the known causative factors like alcohol consumption, spicy foods, drugs like NSAID's, steroids etc. Patients with highly nervous and emotional disposition and those involved in high-stress jobs must be given psychological treatment. Avoiding non-vegetarian diets is also useful in minimising symptoms of acidity. The clinical symptoms and history are very important aspects of diagnosis. Any present and past drug use, especially chronic use of NSAIDs, a history of family members with ulcers, alcohol consumption and smoking, stress assessment and analysis are very useful in determining the cause of the condition. A trial with acid-blocking medication is given with a four-week course of acid-suppressing drugs. In such cases, the symptoms may subside. If symptoms persist, then further testing is needed. Upper Gastrointestinal Endoscopy is done to detect the presence of ulcers. If Zollinger-Ellison Syndrome is suspected, blood levels of gastrin should be measured. Barium Meal studies are also useful as these may show inflammation, active ulcer craters, or deformities and scarring due to ulcers. If an ulcer is present, a precautionary biopsy of the ulcer is usually taken to rule out malignancy as it is not uncommon for a malignancy to manifest as an ulcer.

3.1.8. <u>CHOLECYSTITIS</u>: Cholecystitis is inflammation of the gallbladder, which occurs most commonly due to obstruction of the cystic duct with gallstones (cholelithiasis). Blockage of the cystic duct with gallstones causes accumulation of bile in

the gallbladder and increased pressure within the gallbladder. Concentrated bile, pressure, and sometimes bacterial infection irritate and damage the gallbladder wall, causing inflammation and swelling of the gallbladder. Inflammation and swelling of the gallbladder can reduce normal blood flow to areas of the gallbladder, which can lead to cell death due to insufficient oxygen. Not everyone who has gallstones will go on to develop cholecystitis.

Risk factors for cholelithiasis and cholecystitis are similar and include increasing age, female sex, pregnancy, certain medications, obesity, rapid weight loss, and Native American or Mexican American ethinicity. Females are twice as likely to develop cholecystitis as males. Uncomplicated cholecystitis has an excellent prognosis, however more than 25% of patients require surgery or develop complications. Delayed diagnosis of acute cholecystitis increases morbidity and mortality. Cholelithiasis and cholecystitis may present as a single episode or may recur on multiple occasions. Cholecystitis is often caused by cholelithiasis (the presence of choleliths, or gallstones, in the gallbladder), with choleliths most commonly blocking the cystic duct directly.

This leads to inspissation (thickening) of bile, bile stasis, and secondary infection by gut organisms, predominantly *E. coli* and *Bacteroides* species.

The gallbladder's wall becomes inflamed. Extreme cases may result in necrosis and rupture. Inflammation often spreads to its outer covering, thus irritating surrounding structures such as the diaphragm and bowel.

Less commonly, in debilitated and trauma patients, the gallbladder may become inflamed and infected in the absence of cholelithiasis, and is known as acute acalculous cholecystitis. This can arise in patients with anorexia nervosa, as the lack of stimulation of the gallbladder leads to an infectious process.

Stones in the gallbladder may cause obstruction and the accompanying acute attack. The patient might develop a chronic, low-level inflammation which leads to a chronic cholecystitis, where the gallbladder is fibrotic and calcified.

3.1.9. GASTRIC ULCER: Gastric ulcer, also known as peptic ulcer, is a localized area of erosion in the stomach lining, resulting in abdominal pain, possible bleeding, and other gastrointestinal symptoms. The most common cause of gastric ulcer is a stomach infection associated with the *Helicobacter pylori* (*H pylori*) bacteria. The spread of *H pylori* among humans is not completely understood; it may spread through contaminated food and water. Many people become infected with *H pylori* at a young age, but symptoms most commonly occur in adulthood. In some people, the *H pylori* bacteria cause an infection in the lining of the stomach, which may lead to gastric ulcers. Damage to the stomach lining from stomach acid increases the likelihood that *H pylori* infection will result in a gastric ulcer. Other risk factors for gastric ulcer include alcohol use, tobacco use, and prolonged use of medications such nonsteroidal anti-inflammatory drugs (NSAIDs). Severe illness has also been associated with developing a gastric ulcer.

The signs and symptoms of gastric ulcer can be constant or sporadic, and the disease course varies among individuals. If *H pylori* is the cause, the symptoms will remain as long as the infection is untreated. Some people with gastric ulcers have no symptoms at all, while others may have burning pain, severe nausea, and vomiting.

In the case of *H pylori*-related gastric ulcers, the infection can be treated successfully with antibiotics. For gastric ulcer not related to *H pylori*, antacids or other medications are an effective treatment. You can reduce your risk of *H pylori* bacteria infection by following commonsense hygiene practices such as washing your hands with soap and water prior to preparing food and after handling dirty diapers or using the bathroom.

Other causes of gastric ulcer include agents that can cause inflammation of the stomach lining, including alcohol, tobacco, and medications such as nonsteroidal anti-inflammatory drugs (NSAIDs). Severe illness and radiation therapy have also been associated with gastric ulcers. A history of heartburn, gastroesophageal reflux disease (GERD) and use of certain forms of medication can raise the suspicion for peptic ulcer. Medicines associated with peptic ulcer include NSAID (non-steroid anti-inflammatory drugs) that inhibit cyclooxygenase, and most glucocorticoids (e.g. dexamethasone and prednisolone).

In patients over 45 with more than two weeks of the above symptoms, the odds for peptic ulceration are high enough to warrant rapid investigation by esophagogastroduodenoscopy.

The timing of the symptoms in relation to the meal may differentiate between gastric and duodenal ulcers: A gastric ulcer would give epigastric pain during the meal, as gastric acid production is increased as food enters the stomach. Symptoms of duodenal ulcers would initially be relieved by a meal, as the pyloric sphincter closes to concentrate the stomach contents; therefore acid is not reaching the duodenum. Duodenal ulcer pain would manifest mostly 2–3 hours after the meal, when the stomach begins to release digested food and acid into the duodenum.

Also, the symptoms of peptic ulcers may vary with the location of the ulcer and the patient's age. Furthermore, typical ulcers tend to heal and recur and as a result the pain may occur for few days and weeks and then wane or disappear. Usually, children and the elderly do not develop any symptoms unless complications have arisen.

Burning or gnawing feeling in the stomach area lasting between 30 minutes and 3 hours commonly accompanies ulcers. This pain can be misinterpreted as hunger, indigestion or heartburn. Pain is usually caused by the ulcer but it may be aggravated by the stomach acid when it comes into contact with the ulcerated area. The pain caused by peptic ulcers can be felt anywhere from the navel up to the sternum, it may last from few minutes to

several hours and it may be worse when the stomach is empty. Also, sometimes the pain may flare at night and it can commonly be temporarily relieved by eating foods that buffer stomach acid or by taking anti-acid medication. However, peptic ulcer disease symptoms may be different for every sufferer.

3.1.10. DUODENAL ULCER: A duodenal ulcer is a type of peptic ulcer that occurs in the duodenum, the beginning of the small intestine. Peptic ulcers are eroded areas in the lining of stomach and duodenum, which result in abdominal pain, possible bleeding, and other gastrointestinal symptoms. The most common cause of duodenal ulcer is a stomach infection associated with the *Helicobacter pylori* (*H pylori*) bacteria. Other risk factors for duodenal ulcers include overuse of alcohol, tobacco, and medications such as aspirin and non-steroidal anti-inflammatory drugs (NSAIDs). Severe illness has also been implicated as a risk factor in the development of duodenal ulcer.

The signs and symptoms of duodenal ulcer can be constant or sporadic, and the disease course varies among individuals. If *H pylori* is the cause of the ulcer, the symptoms will remain as long as the infection is untreated. Some people with duodenal ulcers have no symptoms at all, while others may have burning pain, severe nausea, and vomiting.

In the case of *H pylori*-related duodenal ulcer, the infection can be treated successfully with antibiotics. For a duodenal ulcer not related to *H pylori*, antacids or other medications are an effective treatment. You can reduce your risk of *H pylori* infection by following commonsense hygiene practices such as washing your hands with soap and water prior to preparing food and after handling dirty diapers or using the bathroom.

Food passes down the oesophagus (gullet) into the stomach. The stomach makes acid which is not essential, but helps to digest food. After being mixed in the stomach, food passes into the duodenum (the first part of the small intestine). In the duodenum and the rest of the small intestine, food mixes with enzymes (chemicals). The enzymes come from the pancreas and from cells lining the intestine. The enzymes break down (digest)

the food which is absorbed into the body. Your stomach normally produces acid to help with the digestion of food and to kill bacteria. This acid is corrosive so some cells on the inside lining of the stomach and duodenum produce a natural mucous barrier which protects the lining of the stomach and duodenum. There is normally a balance between the amount of acid that you make and the mucus defense barrier. An ulcer may develop if there is an alteration in this balance, allowing the acid to damage the lining of the stomach or duodenum.

Causes of this include the following:

3.1.10.1. Infection with *Helicobacter pylori*

Infection with *Helicobacter pylori* (commonly just called *H. pylori*) is the cause in about 19 in 20 cases of duodenal ulcer. More than a quarter of people in the UK become infected with *H. pylori* at some stage in their life. Once you are infected, unless treated, the infection usually stays for the rest of your life. In many people it causes no problems and a number of these bacteria just live harmlessly in the lining of the stomach and duodenum. However, in some people this bacterium causes an inflammation in the lining of the stomach or duodenum. This causes the defence mucus barrier to be disrupted (and in some cases the amount of acid to be increased) which allows the acid to cause inflammation and ulcers.

3.1.10.2Anti-inflammatory medicines - including aspirin

Anti-inflammatory medicines are sometimes called non-steroidal anti inflammatory drugs (NSAIDs). There are various types and brands. For example: aspirin, ibuprofen, diclofenac, etc. Many people take an anti-inflammatory medicine for arthritis, muscular pains, etc. Aspirin is also used by many people to protect against blood clots forming. However, these medicines sometimes affect the mucus barrier of the duodenum and allow acid to cause an ulcer. About 1 in 20 duodenal ulcers are caused by anti-inflammatory medicines.

4. PROPOSED SYSTEM

Development of a support system for the diagnosis of abdominal disorders on the principles of conditional probability that the system studies and analyses the symptoms to conclude the type of abdominal disorder the person is going through.

The model is captured in a Bayesian Network as a directed a-cyclic graph with the nodes representing symptoms and diseases and the arcs representing the dependencies between them. The diagram illustrates the relations between nodes of how a particular symptom causes a disease.

The main functionality of the application is to list the most probable diseases given the patient-findings (complaints, tests, physical examinations) that are entered. The system is aimed to support diagnosis in general medicine, basically covering abdominal disorders which is a large medical domain with several specializations.

However, a considerable level of detail at which the disease areas are modelled is essential for the system to be of practical use. For this application, this means that the model should contain 1000's of diseases and a factor 10 more of relations between diseases and findings. With such numbers of variables and relations, the standard modelling approach is infeasible.

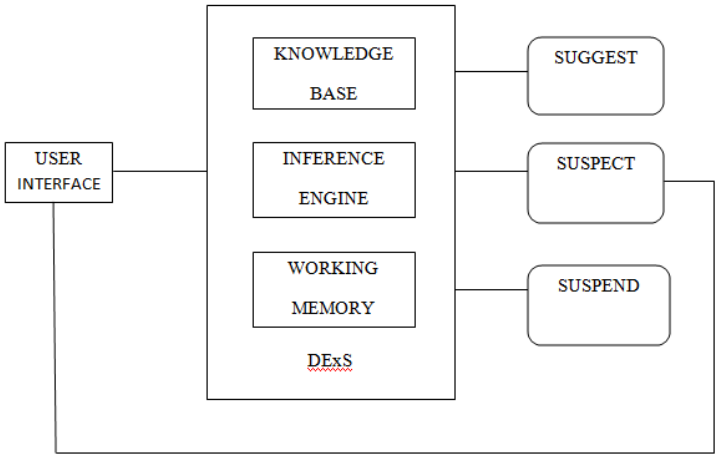


Figure 1: Framework

Probabilistic inference is the problem of computing the posterior probabilities of unobserved model variables given the observations of other model variables. For instance in a model for medical diagnoses, given that the patient has complaints *x* and *y*, what is the probability that he/she has disease *z*?

Inference in a probabilistic model involves summations or integrals over possible states in the model. In a realistic application the number of states to sum over can be very large. In the medical example, the sum is typically over all combinations of unobserved factors that could influence the disease probability, such as different patient conditions, risk factors, but also alternative explanations for the complaints, etc.

In general these computations are intractable. Fortunately, in Bayesian networks with a sparse graphical structure and with variables that can assume a small number of states, efficient inference algorithms exist such as the junction tree algorithm.

I have been developing a clinical support system for abdominal disorders. In this system, patient information, such as age and gender, and findings, such as symptoms, can be entered.

The system then generates patient-specific diagnostic advice in the form of a list of likely diagnoses and suggestions for additional laboratory tests that may be relevant for a selected diagnosis.

The system is intended to support diagnostics in the setting of the outpatient clinic and for educational purposes. Its target users are general internists, super specialists (e.g., gastrologists, gastroenterologists etc.), interns and residents, medical students and others working in the hospital environment.

4.1 Objective

Most of the automation technologies lack the power to reason under uncertainty. The objective is to add intelligence to these technologies by applying Bayes' Theory, Decision Theory and Bayesian Networks.

Detecting diseases at early stage can enable to overcome and treat them appropriately. Identifying the treatment accurately depends on the method that is used in diagnosing the diseases.

Here, aim to develop an application that assists a person to diagnose an abdominal disease on the basis of the symptoms that he selects. The disease with highest probability is the final inference or the result. Dependence on the human expert can be minimized if his/her expertise can be transferred into a computer system.

Enumerated as,

- To help the patient understand his problems.
- To implement the IT in real world problems.
- To assist doctors for various diseases associated with symptoms i.e. to be a home assistant for doctors.

- To assist Medical students working as in Pathological labs.
- To help general practice of doctors, nurses, nursing students and to assist the patients
  as first aid diagnosis.
- To help patients, understand their own problems.

4.2 Bayes' theorem

Bayes' theorem expresses how a subjective degree of belief should rationally change to account for evidence.

To introduce notation, we start by considering a joint probability distribution, or probabilistic model, *P X1 Xn* of *n* stochastic variables *X1 Xn.* Variables *Xj* can be in state *xj.* A state, or value, is a realization of a variable. We use shorthand notation,

$$\mathbf{P(X_1=x_1,.....,X_n=x_n)=P(x_1,.....x_n)}$$

to denote the probability (in continuous domains: the probability density) of variables *X1* in state *x1*, variable *X2* in state *x2* etc.

A Bayesian network is a probabilistic model *P* on a finite directed acyclic graph (DAG). For each node *i* in the graph, there is a random variable *Xi* together with a conditional

$$\mathbf{P(x_1,.....x_n)=\prod_{i=1}^{n} P(x_i|x_{\pi(i)})}$$

probability distribution *P xi xπ i* , where π *i* are the parents of *i* in the DAG.

The joint probability distribution of the Bayesian network is the product of the conditional probability distributions

Since, any joint distribution can be written as

$$\mathbf{P(x_1,.....x_n)=\prod P(x_i|x_{i-1},.....,x_1)}$$

it can be concluded that a Bayesian network assumes

$$\mathbf{P(x_i|x_{i-1},.....,x_1)= P(x_i|x_{\pi(i)})}$$

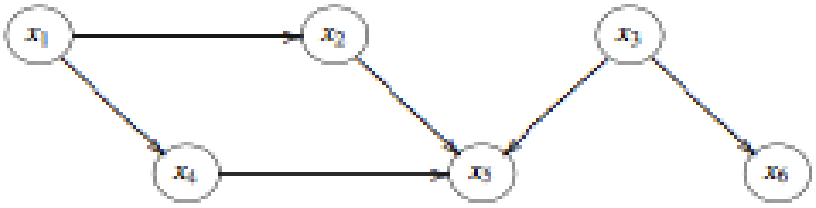


Figure 2:-DAG representing a Bayesian network P (*x1) P(x2 x1)P (x3) P (x4 x1) P (x5 x2 x3 x4)P (x6 x3)*

In other words, the model assumes: given the values of the direct parents of a variable *Xi*, this variable *Xi* is independent of all its other predeceasing variables in the graph.

Since a Bayesian network is a probabilistic model, one can compute marginal distributions and conditional distributions by applying the standard rules of probability calculus.

For instance, in a model with discrete variables, the marginal distribution of variable *Xi* is given by

$$\mathbf{P(x_i)=}\sum_{xi}\dots\sum_{xi-1}\sum_{xi+1}\dots\sum_{xN}\boldsymbol{P(x1\dots.,xN)}$$

Conditional distributions such as *P(xi|xj)* are obtained by the division of two marginal distributions

$$\mathbf{P(x_i|x_j)=\frac{P(xi,xj)}{P(xj)}}$$

Thus, the distribution mathematically used is,

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

4.3 Bayesian Modelling:-

Bayesian Networks Allows a user to model a problem when full knowledge of the domain is either incomplete or uncertain. The construction of a Bayesian network consists of deciding about the domain, what are the variables that are to be modeled, and what are the state spaces of each of the variables. Then the relations between the variables have to be modelled. If these are to be determined by hand (rather than by data), it is a good rule of thumb to construct a Bayesian network from cause to effect.

Start with nodes that represent independent root causes, then model the nodes which they influence, and so on until we end at the leaves, i.e., the nodes that have no direct influence on other nodes. Such a procedure often results in sparse network structures that are understandable for humans.

Often, models are constructed using Bayesian network software such as the earlier mentioned packages. With the use of a graphical user interface (GUI), nodes can be created. The nodes represent the variables in the system.

Typically, variables can assume only values from a finite set. When a node is created, it can be linked to other nodes, under the constraint that there are no directed loops in the network. Finally — or during this process — the table of conditional probabilities are defined, often by educated guesses, and sometimes inferred from data.

A model is captured in Bayesian network as a directed a-cyclic graph whose nodes represent random variables, together with a conditional probability distribution for each node Xi given its parents, P(xi|pa(xi)).The conditional probability for a node without its

parents is just like its prior probability P(xi|ø)=P(xi).

These probability can be obtained from statistical data (such as surveys or database), from the literature on the specific domain or by the judgement of human experts.

The joint probability represented by a Bayesian network is

$$\mathbf{P(x_i,.....,x_n)=\prod_i P(x_i|pa(x_i))}$$

Bayesian Networks, on the other hand, are an example of a probabilistic graphical model that has mechanisms to accurately model the domain's dependencies and perform fully automated reasoning. They have a mathematically consistent way to specify uncertainty in the system.

Bayesian Networks allow a combination of expert knowledge and data where an expert can specify dependencies among nodes or partial network structure and the complete structure and parameters are learned from data. If the expert is not available to provide us with the structure, it can also be learned from the data. Given a network, it is very easy to perform automated reasoning and for the network to make predictions based any a given set of evidence.

In addition it is possible to make predictions about any node in the network based on the evidence. There are multiple algorithms that adapt the network based on additional data so that it can fit the data very well.

Certainty factor based rule based systems can be considered heuristic models as they use an expert's change in belief instead of mathematically consistent probability values. Bayesian Networks are an example of probabilistic graphical models which use graph and probability theory to manage uncertainty in reasoning.

A Bayesian Network is a directed acyclic graph (DAG) whose nodes represent the variables in the problem domain and edges represent direct probabilistic dependencies among nodes.

Lack of edges between nodes corresponds to conditional independence. Each node in a Bayesian Network is associated with a set of probability distributions given by the conditional probability table.

Conditional independence relationships among nodes in Bayesian Networks are determined by the notion of d-separation.

This is how a DAG looks like:

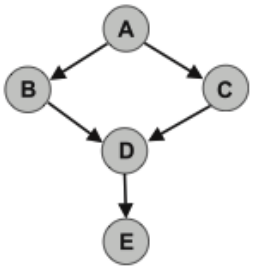


Figure 3: Basic DAG

Additionally, a node is independent of its ancestors given its parents. This means that, we can represent the joint distribution in a more compact way using Bayesian Networks.

Bayesian Networks are a rigorous framework that allows accurate modelling of the domain and provides mechanisms for fully automated reasoning. It is possible to use a Bayesian Network to represent causal relationships between nodes in the domain and then use it to perform diagnostic problem-solving.

4.4 Inference

Inference is the most common operation performed on a Bayesian Network. Probabilistic inference can be defined as computing the conditional probability distribution over values of unobserved nodes given the values of observed nodes.

By taking advantage of conditional probability assumptions, we can perform inference very effectively in most of the cases. Inference can be done by both exact and approximate methods.

A common use of Inference is to perform diagnostic problem solving given the observation of the effects. The most likely cause that results in the observed effects is obtained by inference.

4.5 Learning

It is not always the case that we are provided with both the structure of the Bayesian Network and its probability distribution. In most cases we may not have either of them. Given the data, we can learn both of these.

4.5.1 Parameter Learning

The simplest scenario is that of parameter learning where we have a known structure (possibly from an expert) and the aim is to learn the conditional probability distribution that maximizes the likelihood of training data. The data can be fully or partially observed. For fully observed data, we can use straight forward maximum likelihood estimation to find the parameters. In the case of partially observed data, we can use an EM algorithm to find the MLE of the parameters.

4.5.2 Structure Learning

In structure learning, only the data is available and the structure and parameters of the

network have to be recovered from the data.

4.6 Why using Bayesian Nets???

It has the ability to incorporate prior information(analysis of information from surveys and studies and not just depending upon expert opinion).The use of prior probability distributions represents a powerful mechanism for incorporating information from previous studies and for controlling confounding.

Bayesian models can easily accommodate unobserved variables such as an individual's true disease status in the presence of diagnostic error.

It helps us look at historical data sets to determine the inference and not treat every part as a new or independent problem.

Bayesian Networks allow a user to model a problem when full knowledge of the domain is incomplete or uncertain.

4.7 The 3 Major Modules

- Formation of a KNOWLEDGE BASE.
- The REASONING PROCESS performed by the system to obtain a conclusion.
- The propagation of EVIDENCE.

4.7.1 KNOWLEDGE BASE

The knowledge base contains the rules and an association of compiled data.The expert specifies his domain knowledge using rules and certainty factors. These rules are used to bootstrap a partial Bayesian Network that is then fully learned from the available data.

Using the Bayesian Network, we can perform fully automated reasoning in a consistent way and also generate the rationale for the prediction using rules. The medical knowledge of specialized doctor is required for the development of an expert system.

In the first phase, the medical background of abdominal diseases is recorded through the creation of personal interview with doctors and patients.

The tasks that were carried out during the course of this project were:

1. Studied Bayesian Network rules and methodologies.
2. Conducted several surveys regarding the abdominal diseases.
3. Questionnaires were filled by doctors.
4. Examined various symptoms and causes.
5. Data compiled into CPTs.

4.7.2 REASONING PROCESS

4.7.2.1. Constructed and assembled networks:-

Here, the symptoms and diseases are assembled to form the main network with all the defined dependencies. Initially, smaller networks were designed and constructed that were then put together to form the complete network consisting of 24 nodes in total.

4.7.2.2. Incorporated algorithms like:-

Given a joint probability distribution over variables a set of variables X = X1, X2, . . . , Xn, we can make inferences of the form (Y |Z), where Y ⊂ X is the set of query variables, Z ⊂ X are the evidence variables. The other variables H (those not mentioned in the query) are called hidden variables.

To be clear, X = Y ∪ Z ∪ H. The most naive and expensive way to do inference is to use

the full joint probability distribution and sum out the hidden variables. By the product rule, P (Y |Z)P (Z) = P (Y, Z). (Note that this is true for any distribution. This does not have anything to do with independence.).

So to answer the query P(Y|Z),we can compute P(Y,Z). Note that P(Y,Z)=? P(Y,Z,H). From the view of P(Z) H the full joint probability table, we are summing the probabilities for all of the entries in the table that match the values of the query variables and evidence variables (this includes entries for all of the combinations of values for the hidden variables) and dividing by the sum of all the entries that match the values of the evidence variables.

4.7.2.2.1 *Enumeration algorithm:*

Any conditional probability can be computed by summing terms from the full joint distribution. Since the Bayesian network gives a complete representation of the full joint distribution, it can be written as the products of conditional probabilities from the designed networks.

4.7.2.2.2 *Variable Elimination algorithm:*

The Enumeration algorithm can be improved substantially by eliminating repeated calculations. It works by evaluating expressions in bottom-up orders.

4.7.2.2.3 *Clustering algorithm:*

In a poly tree network, one would need to issue O(n) queries costing O(n) each, for a total of O(n^2) time. Using Clustering algorithms (also known as join tree algorithms), the time can be reduced to O(n).

The basic idea behind clustering algorithm is to join individual nodes of the network to form cluster nodes in such a way that the resulting network is a poly tree.

Thus, used the following formula:

$$P(A|B) = P(A,B)/P(B)$$

From Rules to Bayesian Network Structure

A rule indicates a statistical dependency between the antecedent and the consequent and decides whether the relation between them is causal or diagnostic.

Intuitively, if the expert provides a set of rules, a partial Bayesian Network with the antecedents as the parent nodes and the consequent as the child node can be constructed. If all the rules are causal, the result is a partial causal Bayesian Network.

It is possible that the rule base contains a mix of causal and diagnostic rules. In a typical learning algorithm, parent and child are determined by a statistical independence tests. Here, we assume that the expert's domain knowledge allows him to identify some of the parents and their children nodes in the Bayesian Network.

4.7.3 EVIDENCE

- *MSBNx-Microsoft Belief Network Tool-used for Bayesian Modelling*:

Once a Bayesian Network has been constructed, it can be used for predictions. The method can be used interactively where the user enters observations and is interested in the distribution of specific unobserved variables.

Similarly, the system can be used for classification where given a set of symptoms inference is used to find the most likely value for a target random disease.

Once the system has a distribution of the symptoms, it has to explain the prediction to the user. Doing this purely from the Bayesian Network is very difficult as explanations in terms of the network's probability distribution are exceedingly difficult for users (and even experts) to understand.

On the other hand, knowledge encoded in the rules should be understandable to the expert. As a consequence, the system presented here derives explanation for Bayesian Network predictions using the original rules in the system.

The user provides a set of observations and variables whose distribution he is interested in. The system performs the inference and finds the rules in the rule base whose consequent matches the most likely value of the predicted variable. The system finds the Bayes' factor for each of the matching rules.

A higher value of Bayes' Factor means that the rule is strongly supported by the data and has a higher likelihood of occurrence that its complement.

Although the networks created in this tool can be quite complex, the scope of these software packages obviously has its limitations.

- *GUI in C#*

5. DESIGN

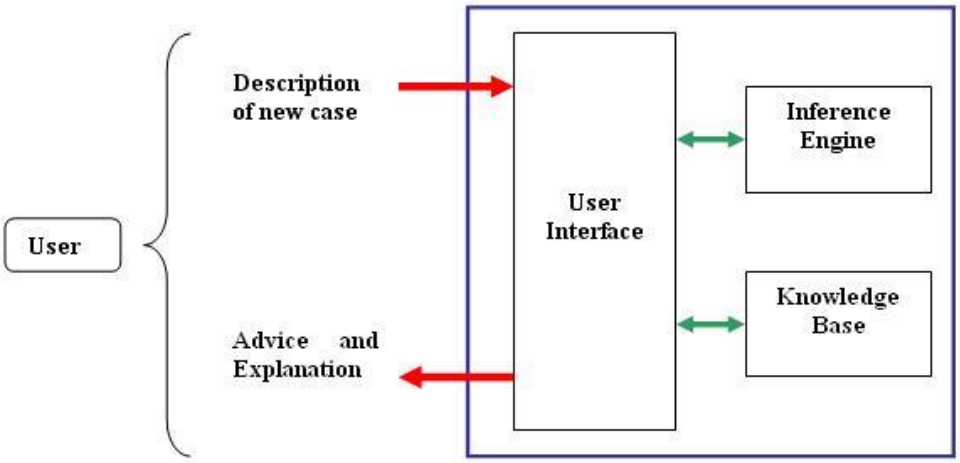


Figure 4: Context Diagram

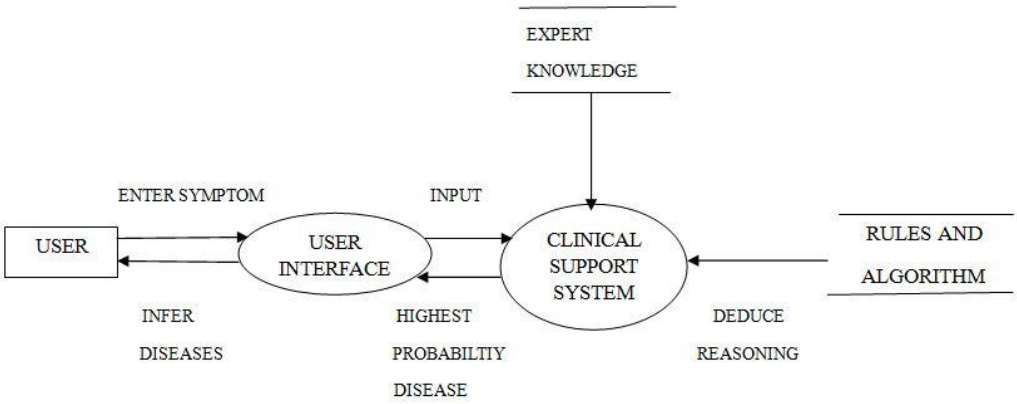


Figure 5: DATA FLOW DIAGRAM

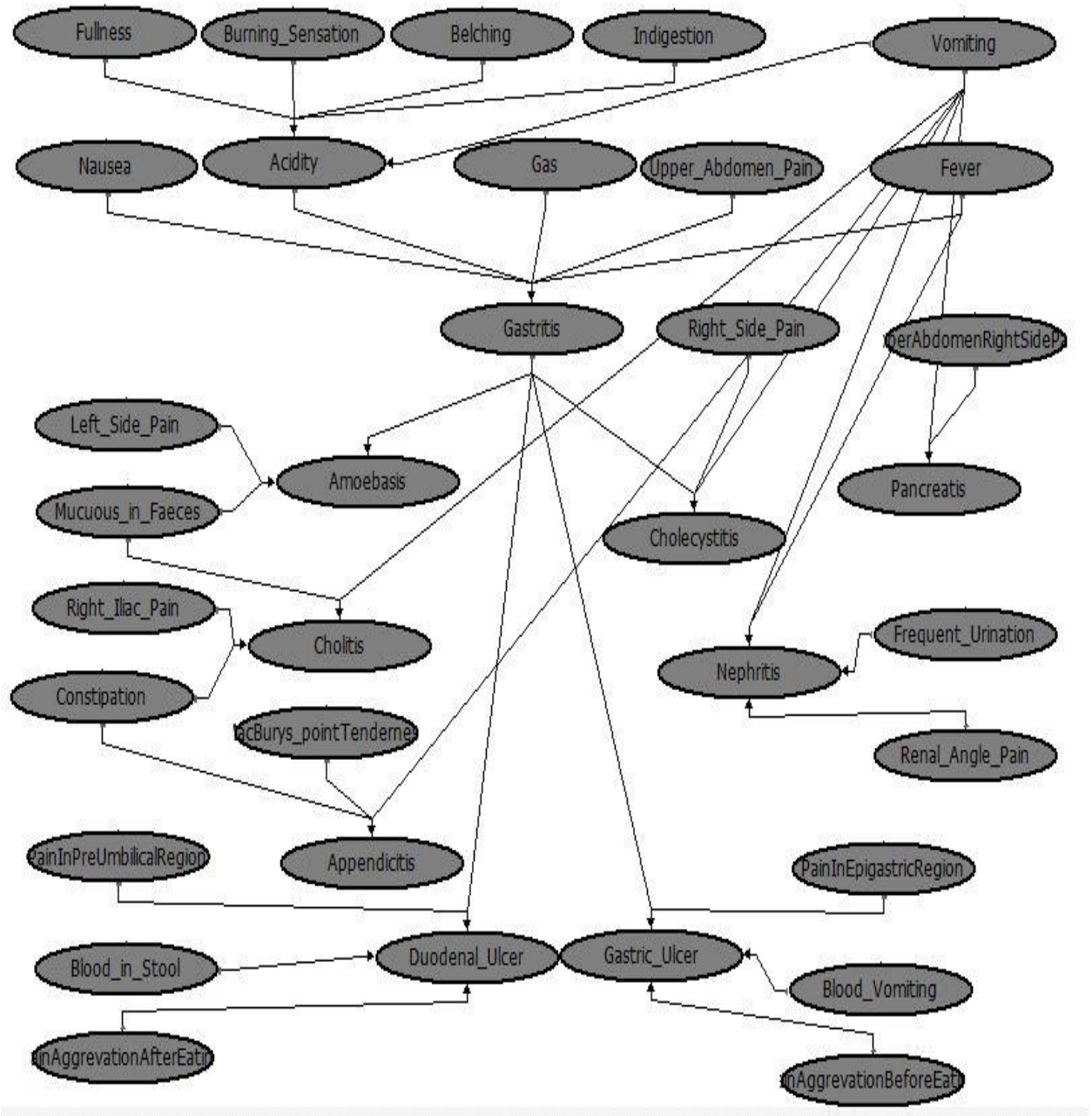
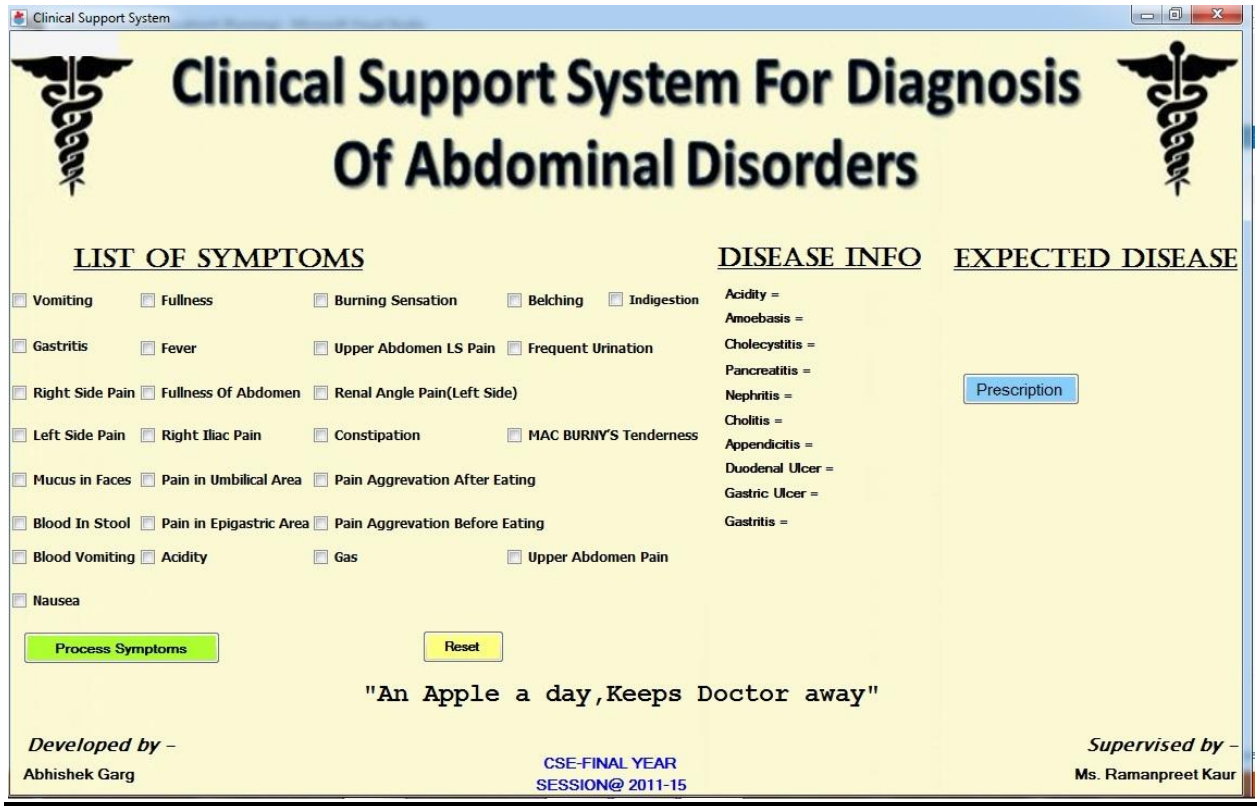


Figure 6: Entire Bayesian Network

SNAPSHOTS OF THE PROJECT



Snapshot 1: System Interface

TOOL TO DESIGN THE NETWORK

MSBNX: MSBNX is a Microsoft Windows software application that supports the creation, manipulation and evaluation of Bayesian probability models. Each model is represented as a graph or diagram. The random variables are shown as ellipses, called *nodes*, and the conditional dependencies are shown as arrows, or directed *arcs*, between variables. At the present time, MSBNX only supports discrete distributions for its model variables. Models are saved to and loaded from disk-based text files in an XML-based format.

MSBNX supports simultaneous viewing and evaluation of multiple models**.** MSBNx (Bayesian Network Editor and Tool Kit) is a component-based Windows application that supports the creation, manipulation and evaluation of Bayesian probability models, created at Microsoft Research. Each model is represented as a graph or diagram. The random variables are shown as ellipses, called nodes, and the conditional dependencies are shown as arrows, or directed arcs, between variables.

At the present time, MSBNx only supports discrete distributions for its model variables. Models are saved to and loaded from disk-based text files in an Extensible Markup Language (XML)-based format. MSBNx supports simultaneous viewing and evaluation of multiple models.

File Formats -- The only document type in MSBNx is the model network document. Each saved document represents one Bayesian belief network and includes its prior probabilities and the properties associated with the model, including all of its variables, or nodes.

MSBNx supports two distinct file formats:

1) The XBN format is XML-based and is the default file format used by MSBNX.
2) The DSC format is a legacy text format used within Microsoft Research.

The XBN format is intended to replace the legacy DSC format.

Both of these file formats are basically text and can be edited or generated by such tools as EMACS, Notepad or WordPad.

Bayesian Networks -- Bayesian Networks are useful for diagnosis and troubleshooting.

Recommendations --

When doing diagnosis and troubleshooting, MSBNx can recommend what evidence to gather next. If you give MSBNx cost information, it does a cost-benefit analysis. If No

cost information is available, MSBNx makes recommendations based on the Value of Information (VOI).

Assessment --

MSBNx tries to make it easy for you to specify your probabilities for a Bayesian Network:

1) With the Standard Assessment Tool, you can specify full and causally independent probability distributions.
2) With the Asymmetric Assessment Tool, you can avoid specifying redundant probabilities.
3) If you have sufficient data and use 'machine learning' tools to create Bayesian Networks, you can use MSBNx to edit and evaluate the results.

Use It in Programs --

MSBNx is fully component based. Its most important component is MSBN3, an ActiveX Dynamic-Link Library (DLL). MSBN3 offers an extensive Component Object Model (COM)-based Application Programming Interface (API) for editing and evaluating Bayesian Networks. You'll find MSBN3 especially easy to use from COM- friendly languages such as Visual Basic and JScript. MSBNx also includes graphical components, for example, both the Standard Assessment and Asymmetric Assessment tools are ActiveX controls and can be used in other applications.
Also, you can extend the editing and evaluation abilities of MSBNx by creating add-ins. For example, MSBNx ships with an (undocumented) add-in for editing and evaluating Hidden Markov Models (HMMs).

6. RESULT AND ANALYSIS

Working:

The specification of a Bayesian network can be described in two parts, a qualitative and a quantitative part.

The qualitative part is the graph structure of the network. The quantitative part consists of specification of the conditional probability tables or distributions.

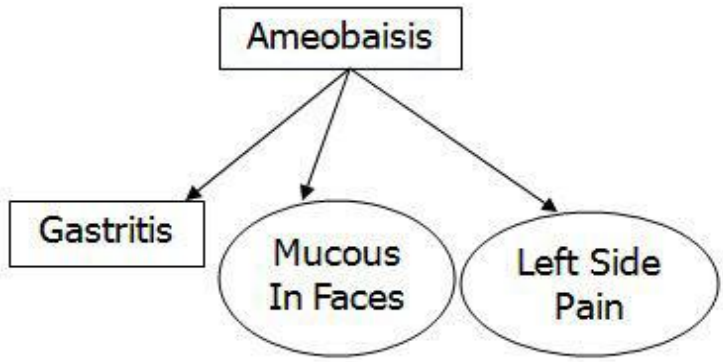
Ideally both specifications are inferred from data, however, data is often insufficient even for the quantitative part of the specification. The alternative is to do the specification of both parts by hand, in collaboration with domain experts.

We create a model and add new nodes to it. Every node represents either a symptom or a disease. Once a new node is added and named, discrete values are assigned to that node.

For example, we consider a particular disease named “Amoebiasis”.

There are three symptoms, namely,

- Mucous in faeces
- Left side pain
- Gastritis



The values for all the three symptoms will be unobserved, yes and no. The arcs are then added between these nodes showing the dependencies and usually flow from top to bottom.

The next step is to add knowledge about what is known based on probabilities. These probabilities will then be used along with the model to infer things about the disease. Given the prior probabilities of the root nodes in a diagram and the conditional probabilities of the other nodes, we can then calculate the other probabilities as well.

After incorporating the three algorithms on the basic Bayes' Probability, the compact mathematical inference drawn can be written as follows:

$$P(A|B)=P(A,B)/P(B)$$

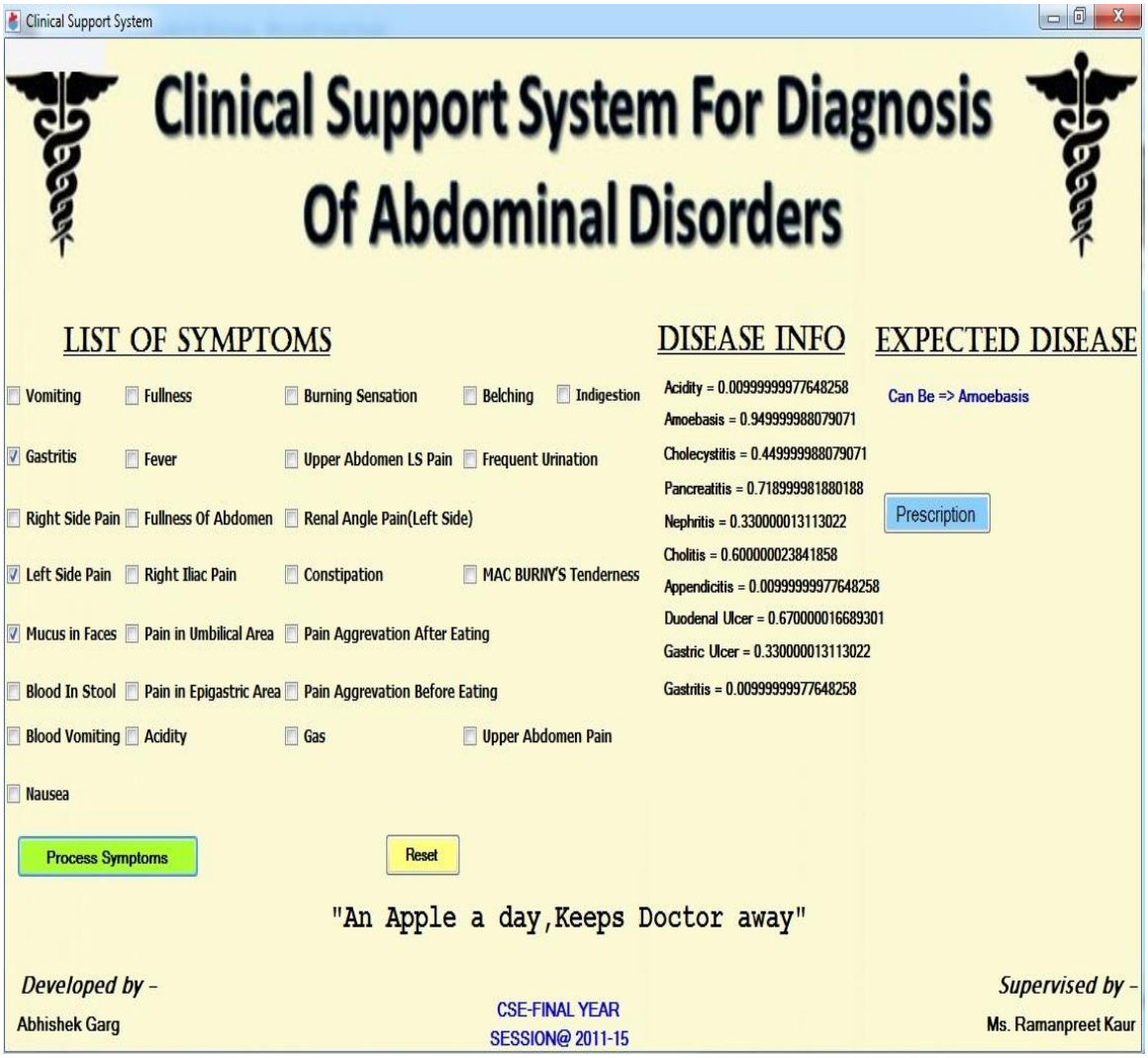
Output:

The output of whether the disease is actually present or not will be a probability. Based on that, how does one decide what to do? The answer is that one can make an educated decision based on decision theory about the presence of the particular disease from that probability.

Depending on the result of the calculation, the system will perform the appropriate action.

There are two levels of outputs shown in the environment. One shows the entire set of probabilities of all the diseases and the other specifies the disease actually present on the basis of highest probability.

The disease with the highest probability is considered to be the resultant disease and thus the best inference out of all. The main inference task in the application is to compute the probabilities of diagnoses given the observed values.



Snapshot 2: Output for Disease Amoebasis

7. Conclusion and Future Work

Conclusion:

Certainty factor based support systems and Bayesian Networks are two popular frameworks to perform uncertain reasoning. Both have their own advantages and disadvantages. Here, we have proposed a system that combines the ability to efficiently elicit expert knowledge and generate user understandable explanation using MSBNx tool with the automated reasoning capabilities of the Bayesian Network.

Future Work:

Currently, the proposed system can handle only 10 diseases. This support system can be made more elaborate by adding many more diseases that comes under abdominal disorders giving it a broader dimension to a doctor's work and reduces time and space complexities.

We have designed the networks using static nodes. The use of dynamic Bayesian Networks should be explored as a way to more accurately model it. This will allow the system to handle more expressively. The network that is presently modelled can be furthermore expanded and detailed so as to incorporate the little shortcomings that have been vaguely left out.

This system can be implemented in form of mobile application for operating systems like ios and android, improve the GUI of the application to provide more appropriate look in relation to medical field and add more illustration.

8. References

1. Jyotirmay Gadewadikar and Ognjen Kuljaca. *Exploring Bayesian networks for medical decision support in cancer detection.* Research Vol. 3(10), pp. 225-231, October 2010
2. Mark L Krieg .*A Tutorial on Bayesian Belief Networks* Surveillance Systems Division, Electronics and Surveillance Research Laboratory. DSTO-TN-0403 December, 2001
3. Muhammad Zubair Asghar, Muhammad Junaid Asghar. *Expert System For Online Diagnosis of Red-Eye Diseases.* International Journal of Computer Science & Emerging Technologies (IJCSET) 35 Volume 1 Issue 2, August 2010
4. Carl M. Kadie, David Hovel and Eric Horvitz. *MSBNx: A Component-Centric Toolkit for Modeling and Inference with Bayesian Networks.*Adaptive Systems and Interaction Microsoft Research MSR-TR-2001-67, 28 July 2001
5. Artificial Intelligence: A modern approach, Stuart Russel and Peter Norvig(Pearson prentice hall) 2nd edition.
6. Azaab S., Abu Naser S., and Sulisel O.,2000. A proposed expert system for selecting exploratory factor analysis procedures, Journal of the college of education, 4(2):9-26.
7. Jeffrey, Richard C. The Logic of Decision, The University of Chicago Press 1983.
8. Jensen, Finn V. An Introduction to Bayesian Networks, Springer-Verlag, NewYork 1996.
9. P. Santosh Kumar Patra, Dipti Prava Sahu, Indrajit Mandal(2010):An Expert System for Diagnosis of Human Diseases.©2010 International Journal of Computer Applications (0975 – 8887) Volume 1 – No. 13.