Project No.6

Bayesian Network Diagnostic System

Presented by

1. Miss Siriwimon Suksukhon 58070503436
2. Mr. Paween Surimittragool 58070503457

Advisor

Asst. Prof. Dr.Santitham Prom-on

"I've read and approved the content of this report"

……………………………………………

(…………………………………………)

Advisor

Bayesian Network Diagnostic System

Miss Siriwimon Suksukhon
Mr. Paween Surimittragool

A Project Submitted in Partial Fulfillment of the Requirements
for the Degree of Bachelor of Engineering
Department of Computer Engineering, Faculty of Engineering
King Mongkut's University of Technology Thonburi
Academic Year 2018

Bayesian Network Diagnostic System

Miss Siriwimon Suksukhon
Mr. Paween Surimittragool

A Project Submitted in Partial Fulfillment of the Requirements
for the Degree of Bachelor of Engineering
Department of Computer Engineering, Faculty of Engineering
King Mongkut's University of Technology Thonburi
Academic Year 2018

Project Committee

…………………………………………………… Committee
(Dr. Peerapon Siripongwutikorn)

…………………………………………………… Committee
(Dr. Sally E. Goldin)

…………………………………………………… Committee
(Dr. Naruemon Wattanapongsakorn)

Project Title	Bayesian Network Diagnostic System
Project Credit	3 credits
Project Participant	Miss Siriwimon Suksukhon
	Mr. Paween Surimittragool
Advisor	Asst. Prof. Dr. Santitham Prom-on
Degree of Study	Bachelor's Degree
Department	Computer Engineering
Academic Year	2018

Abstract

Because the number of patients in Thailand increases rapidly while number of doctors does not. This leads to the problem that people in some area may not have the chance to meet doctors when they get sick. This project aims to solve this problem. It consists of two modules. Diagnostic system, and model improvement system. The diagnostic system uses the theory of Bayesian network to find the feasible diseases based on the probability of given symptoms and behavior. Moreover, we implement the support system in order to improve the model further by feeding some training data.

For this project, we have a specialist providing some advice about our interested diseases. The specialist gave us suggestions about the probability and dependencies among symptoms, behavior, and diseases. After the prototype is developed, we brought this technology to be tested in the hospital. There will be cases where doctors have made diagnosis to patients. We take symptoms, and behavior of patients as input, and let the model predict the output as diseases. Doctors will be testing it to ensure that the model can make basic diagnosis correctly. Until the prediction result is acceptable, the system will be provided to society.

หัวข้อโครงงาน	ระบบวินิจฉัยโรคด้วยเครือข่ายแบบเบย์
หน่วยกิตของโครงงาน	3 หน่วยกิต
จัดทำโดย	นางสาวสิริวิมล สุขสุคนธ์
	นายปวีณ สุริมิตรตระกูล
อาจารย์ที่ปรึกษา	ผศ.ดร.สันติธรรม พรหมอ่อน
ระดับการศึกษา	วิศวกรรมศาสตรบัณฑิต
ภาควิชา	วิศวกรรมคอมพิวเตอร์
ปีการศึกษา	2561

บทคัดย่อ

เนื่องจากจำนวนคนไข้ในประเทศไทยเพิ่มขึ้นอย่างรวดเร็วซึ่งตรงข้ามกับจำนวนแพทย์ที่ไม่ได้เพิ่มขึ้นตามสาเหตุนี้นำไปสู่ปัญหาการขาดแคลนแพทย์คนไข้ในบางพื้นที่ไม่มีโอกาสได้พบแพทย์เวลาเจ็บป่วยโครงการนี้มีวัตถุประสงค์เพื่อแก้ไขปัญหาเหล่านั้น โครงการนี้ประกอบด้วย 2 ระบบย่อยคือระบบวินิจฉัยและระบบเพิ่มประสิทธิภาพการวินิจฉัย ระบบวินิจฉัยถูกสร้างขึ้นจากทฤษฎีเครือข่ายงานแบบเบย์เพื่อหาผลลัพธ์โรคที่เป็นไปได้จากค่าความน่าจะเป็นซึ่งขึ้นอยู่กับอาการและพฤติกรรมของคนไข้ นอกจากนั้นทีมผู้พัฒนายังสร้างระบบเพิ่มประสิทธิภาพการวินิจฉัย เพื่อให้การวินิจฉัยเป็นไปได้อย่างถูกต้องและแม่นยำมากขึ้นโดยการใส่ข้อมูลเข้าไปในระบบวินิจฉัยเพื่อให้ระบบเรียนรู้และปรับปรุงค่าความน่าจะเป็นให้ทันสมัย

ในการพัฒนาโครงการ ทีมผู้พัฒนาได้ติดต่อผู้เชี่ยวชาญทางการแพทย์เพื่อให้คำแนะนำเกี่ยวกับโรคที่ทีมผู้พัฒนาสนใจ ผู้เชี่ยวชาญจะให้คำแนะนำเกี่ยวกับค่าความน่าจะเป็นและความสัมพันธ์ระหว่างอาการ พฤติกรรม และโรคที่เกิดขึ้นหลังจากแบบจำลองระบบวินิจฉัยถูกพัฒนาขึ้น ทีมผู้พัฒนาจะนำไปทดสอบในโรงพยาบาลโดยการรวบรวมข้อมูลการวินิจฉัยของแพทย์ต่อผู้ป่วย อาการและพฤติกรรมจะถูกนำไปเข้าระบบวินิจฉัยเพื่อทำนายโรคที่เป็นไปได้แพทย์จะทำการทดสอบจนกว่าแบบจำลองจะวินิจฉัยได้ถูกต้องในเกณฑ์ที่ยอมรับได้ ระบบจะถูกนำไปใช้จริงโดยการรวมเข้ากับระบบโต้ตอบอัตโนมัติ

Acknowledgement

We would like to express our special thanks of gratitude to Advisor: Asst. Prof. Dr. Santitham Prom-on as our advisor, Dr. Peerapon Siripongwutikorn, Dr. Sally E. Goldin, Dr.Naruemon Wattanapongsakorn as our committees who give us the suggestion to help our project complete in the term of research and report on the Bayesian Network for diagnosis disease. We learn a lot of things from this project. We are thankful to them. One last thing, we would like to say thank you to our family and friends who supported us in doing this project until the end of semester.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Problem Statement and Approach

Currently, Thailand is facing the problem of shortage of personnel with specialized knowledge in medicine. The ratio of doctors to patients is small comparing to many countries in ASEAN or even the same economic level countries. This indicates that a number of doctors in Thailand is not enough to handle people. In Thailand, some district has only a district doctor while our neighboring countries has a sub district doctor each. Although, in some cases, we can substitute the doctor's responsibilities by using the nurse such as for asking the symptoms that the patient has faced because the proportion of nurse is much more than doctors, the ability of nurses might not be good at diagnosis. This results that we can only save a little bit of time of the doctors. However, the number of patients increases rapidly. This leads to the problem of shortage of personnel, and the basic diagnosis responsibility seems to be waste of time for the doctors. Especially, asking about the symptoms the patients had found every time they come to use the service. For example, asking about the symptoms, the number of happening days, congenital disease, or even drug allergy. In some wilderness area, the number of doctors is very low, and it's not enough to serve patients. This makes patients use a lot more time to see doctors at a time. The worst case happens with patients living in far away because if they get sick, they may not be addressed to doctors in time, and die finally.

From those mentioned problems, we are interested to bring technology to solve these problems by implementing the system that is able to give basic diagnosis to patients. Bayesian network would help doing this because it works in the way of dependency among causes and effect. So, we want to apply it into our system so that we can give basic diagnosis as doctors does to reduce the Q&A time between patients and doctors. This benefits to both doctors and patients. Doctors can save time for basic diagnosis, they can partially ensure that they diagnose patients in proper way. Also, patients can reduce time to find the direct doctor about diseases they faced. Our project tends to save time for diagnosis so that the hospital services get better. This also reduces the inequality of healthcare services that is now happening in Thailand. At the same time, nurses would have more time to focus and take care of patients effectively. Patients also get served quicker and friendly. This project belongs to benefiting society category. We mainly focus on bringing technologies to improve healthcare system so that it is easily accessible to people in the country.

1.2 Objectives

1. To create the model that can diagnose and suggest the treatment for helping patients in the hospital.
2. To apply Bayesian network technique to be used in diagnosis system.

1.3 Scope

1. A Bayesian network model predicting associated diseases from given symptoms and behavior, and suggest basic treatment

The model was built applying Bayesian network. We take symptoms, behavior, and diseases as variables. The dependency for each variable will be illustrated using edges. The conditional probability for each variable comes from both statistical data and a specialist. After making diagnosis, we also provide the way to treat that disease basically. Also, patients will be able to give information based on their examination sheet to gain more accuracy. Users also can specify an interested disease alone, and the system will immediately answer causes and treatment.

2. A Bayesian network model improvement system

The system was created to support the main diagnosis system. As time passed, we cannot ensure that the way of our diagnosis would be correct over time. This system takes data of the real diagnosis from doctors and try to update the conditional probability table in the network so that the prediction is still reliable. The system can also display the overall network in case the doctor wants to see the diagnostic process explained by conditional probability table. Moreover, this system provides the way to recreate the network to be used in diagnostic system.

1.4 Tasks and Schedule

Tasks	Assigned to	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	April	May
1) Requirement Analysis											
1.1. Meet advisor	Poo, Jab										
1.2. Write a proposal	Poo, Jab										
1.3. Requirement document	Poo, Jab										
1.4. Research on Bayesian Network	Poo										
1.5. Research on Medical Diagnosis	Jab										
1.6. Meet specialized	Poo, Jab										
2) Software Design											
2.1. Design Document	Poo, Jab										
2.2. Find and Study tools	Poo, Jab										
3) Implementation											
3.1. Design a Bayesian network	Jab										
3.2. Define probability distribution	Poo, Jab										
3.3. Implement Diagnostic system	Jab										
3.4. Network Evaluation	Poo, Jab										
3.5. Implement support system	Poo										
4) Integration Testing											
4.1. Integrate with Chatbot	Jab										
4.2. Fix bug after integrating	Jab										

Tasks	**Assigned to**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**	**April**	**May**
5) User Testing											
5.1. User Testing	Poo										
5.2. Fix bug	Poo										
6) Presentation											
6.1. 1st Semester Midterm report	Poo, Jab										
6.2. 1st Semester Midterm presentation	Poo, Jab										
6.3. 1st Semester Final report	Poo, Jab										
6.4. 1st Semester Final presentation	Poo, Jab										
6.5. 2nd Semester Midterm report	Poo, Jab										
6.6. 2nd Semester Midterm presentation	Poo, Jab										
6.7. 2nd Semester Final report	Poo, Jab										
6.8. 2nd Semester Final presentation	Poo, Jab										

Table 1.1 Schedule and time plan of our project

Chapter 2
Background, Theory and Related Research

2.1 Bayesian Network Background

Bayesian network is a specific type of graph models showing dependency between two variables. Bayesian network consists of three things; nodes, edges (also called arc) and conditional probability table (CPT) in each node. Normally, the edges in a graph can be directional or undirected. However, Bayesian network is said to be a directed acyclic graph (DAG) because it is treated as a causal network; a network where some nodes are effects from a cause node. We can call the starting node as parents while the pointed nodes are their children.

Directed acyclic graph (DAG) is a graph where 1) all edges are directed, and 2) there is no cycles. As shown in Figure 2.1, A and C are conditionally independent so, P (A|B, C) = P(A|B), and P (C|A, B) = P(C|B). Therefore, we can conclude that the conditional probability of B and C directly depend on A. Apart from representing dependency, the edges in Bayesian network also encode the joint probability distribution. [1]

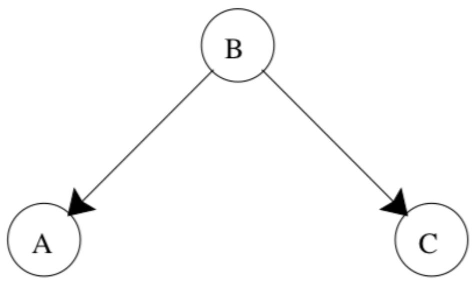


Figure 2.1: An example of a directed acyclic graph

For figure 2.1, A and C are conditionally independent while both of them directly depend on B (probabilities are omitted). The joint probability of this graph can be written as P (A, B, C) = P(A|B) * P(B) * P(C|B). [1]

The equation above illustrates the joint probability of this network. It is the probability that A, B and C occurs concurrently. We can compute this by multiply probabilities of all of them depending on their parents.

The probability of X in a Bayesian network can be described using the following formula.

$$P(\mathbf{X}) = \prod_{i=1}^{n} P(X_i | parents(X_i))$$

X represents a set of nodes in Bayesian network. The capital Pi is the product of nodes in a network. The joint probability of the network can be computed as the multiplication (product) of all nodes given their parents. This is a general formula to calculate the probability of a node from default knowledge. However, if new evidences can be identified, we can calculate probability by making *inference*.

Inference

$$P(X|e) = \frac{P(e|X)P(X)}{P(e)}$$

Inference is the process of calculating a probability distribution of interest e.g. P (X | e=False). In order to make inference on a given problem using Bayesian network, there are four terms to be considered.

1) Prior: the initial belief probability of a variable without any observations, denoted by P(X)
2) Evidence: the probability of evidence; the things we know, the known variable will become a certain variable, denoted by P(e)
3) Likelihood: the probability of how likely it is that X caused the evidence, denoted by P(e|X)
4) Posterior: the probability of variable X, given some evidences, denoted by P(X|e)

Bayes Theorem

Bayes theorem states that:

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

Bayes theorem allows us to update the belief probability in the network if we have new evidence. Bayesian network uses Bayes theorem to make inference. [2] Using Bayes theorem, the joint probability distribution in Figure 1 can be translated to

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

Note that this is because the formula in Figure 2.1 can also be represented as above according to Bayes Theorem. Therefore, Bayesian network as shown in Figure 2.1 can also be rewritten as Figure 2.2.

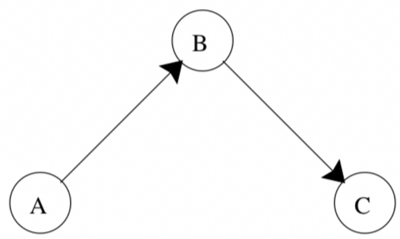


Figure 2.2: A Bayesian network that represents the same joint probability as in Figure 2.1 (probabilities are omitted). [1]

Naive Bayes Classifier

Naive Bayes classifier can be used to solve a prediction task. The model works based on an assumption that all the features are conditionally independent of each other (as shown in Figure 2.3). However, Bayesian network does not have those assumption. Naive Bayes Classifier describe briefly about the relationship between features, but it works well in many cases. However, if we have enough data to derive relationship among features, Bayesian network might be appropriate to use. [3]

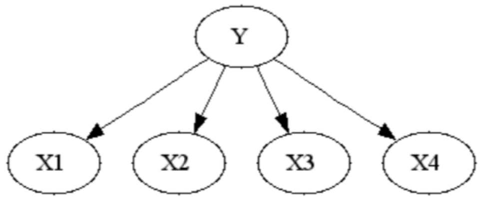


Figure 2.3: Naive Bayes Classifier assumes that all X features are conditionally independent. [11]

"Importantly Bayesian networks handle missing data during inference (and also learning), in a sound probabilistic manner." [4]

2.2 Medical diagnosis

The process for diagnosis of doctor that is used to analyze patient's symptoms, signs, physical examination, medical history. In some case, medical and chemical technology can be used for making diagnosis in order to find cause and treatment guidelines.

2.2.1 Clinical diagnosis

The diagnosis that can analyze diseases from asking about patient's symptoms, medical history and physical examination. Mostly, diseases in this case will be general diseases that people always sick. For example, fever, sore throat, vomit, stuffy nose, stomach ache. Those diseases are not complex and severe.

2.2.2 Medical investigation

In case where the clinical diagnosis can't analyze in general, the doctor will call it 'Investigation' method. The medical investigation is used when the doctor needs more information that can't know from asking patient. Mostly, it will use for analysis the disease that very complex and severe. In some case they use for check symptoms of the disease that has low chance to happen.

2.2.2.1 Laboratory diagnosis

This type of diagnosis uses science knowledge for diagnosis disease. Normally, it uses lab technology for analysis such as blood test, microscopic examination or urinalysis.

2.2.2.2 Biochemical diagnosis

This uses chemical knowledge for diagnosis disease by using chemical reaction for diagnosis. For example, biochemistry tests with the blood.

2.2.2.3 Immunology diagnosis

Analysis immune system in the body to search for the cause of disease. Mostly, this diagnosis uses vaccine.

2.2.2.4 Radiological diagnosis

This uses imaging medicine for captures the imaging in the body. This diagnosis works with radiology. For example, x-ray imaging, ultra-sonogram or magnetic resonance imaging (MRI)

2.2.2.5 Histologic diagnosis

Diagnosis disease with biopsy by cut some part of lump rather than analyze it directly.

2.2.2.6 Cytologic diagnosis

Cutting cell, tissue and organ for diagnosis disease such as using cell wall or cell membrane for analysis the cause of disease

2.2.3 Final diagnosis

It's the process where doctor collect all information such as asking for symptoms, signs, physical examination, medical history and/or investigation until the doctor can diagnosis the disease. In general, it's called "final diagnosis" for this step.

2.2.4 Differential diagnosis

In some case that the doctor can't diagnosis or confused about the disease. The doctor may give a treatment follow the probability of disease or cause. So if the patient get well the doctor call this "therapeutic diagnosis". [4]

2.3 Related Research

2.3.1 Bayesian network modeling for diagnosis of social anxiety using some cognitive-behavioral factors

Social anxiety can be hard to spot as they have some physical symptoms in common to other mental health problems. Therefore, the Bayesian Network is constructed because it can identify the relationship of symptoms and diseases precisely based on the conditional probability. Two parts of constructing Bayesian Network consist of construct structure, and conditional probability tables.

Some important features are selected based on the relevance. For example, shyness (X_4node) is one of the important factors of social anxiety disorder. Behavioral inhibition is also considered to be related. After having a Bayesian network, the conditional probability tables inside Bayesian network are defined by the data learning and testing model process. The data from was split into ⅔ for training set, and ⅓ for a test set.

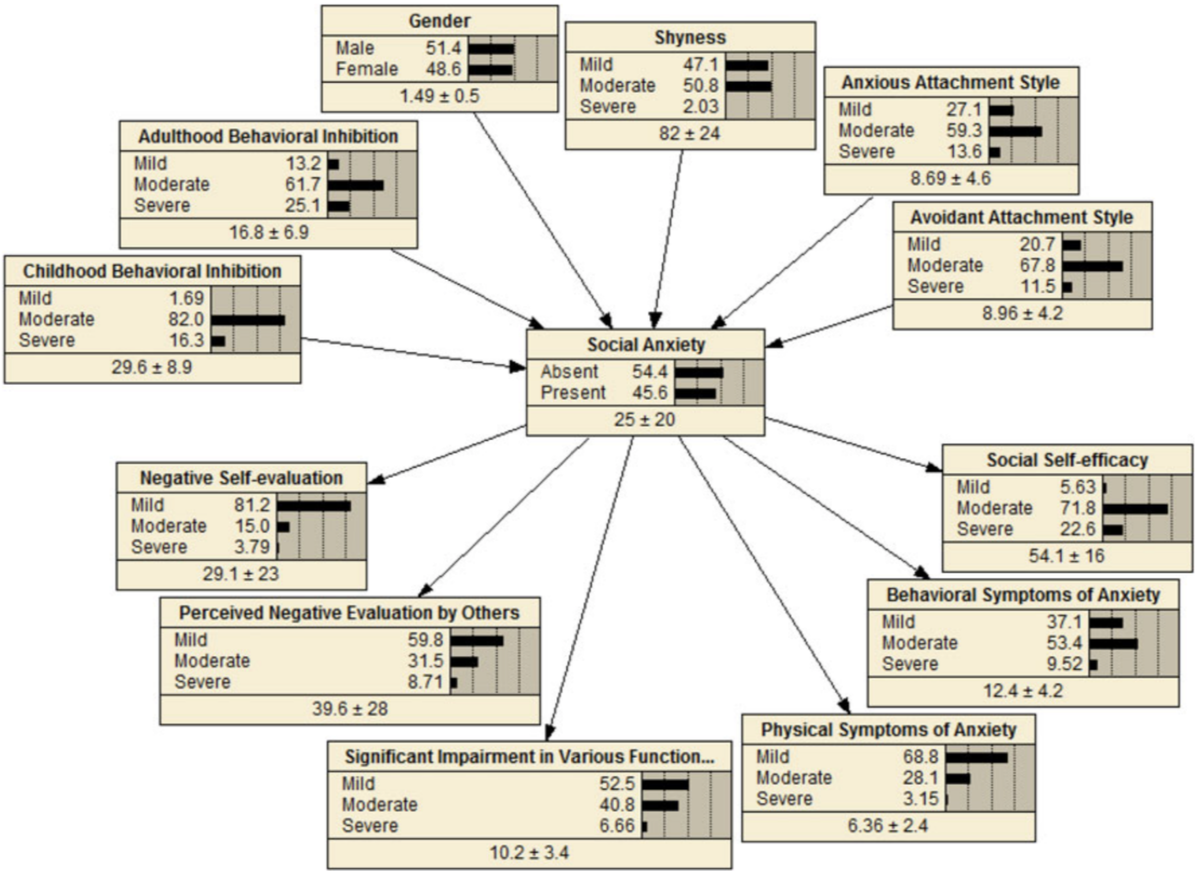


Figure 2.4 The BN model of SAD in Netica after learning [5]

The network was evaluated by treating an interested node as unobserved (which in this case is the social anxiety nodes). The result provides 14.38% of error of the cases. In other words, the model predicted 14.38% wrong out of all test data. Also, it can be seen in Figure 2.5, severe behavioral symptom is likely the variable that is most related to social anxiety. However, the population is only limited to the students of one university so, generalization is needed. [5]

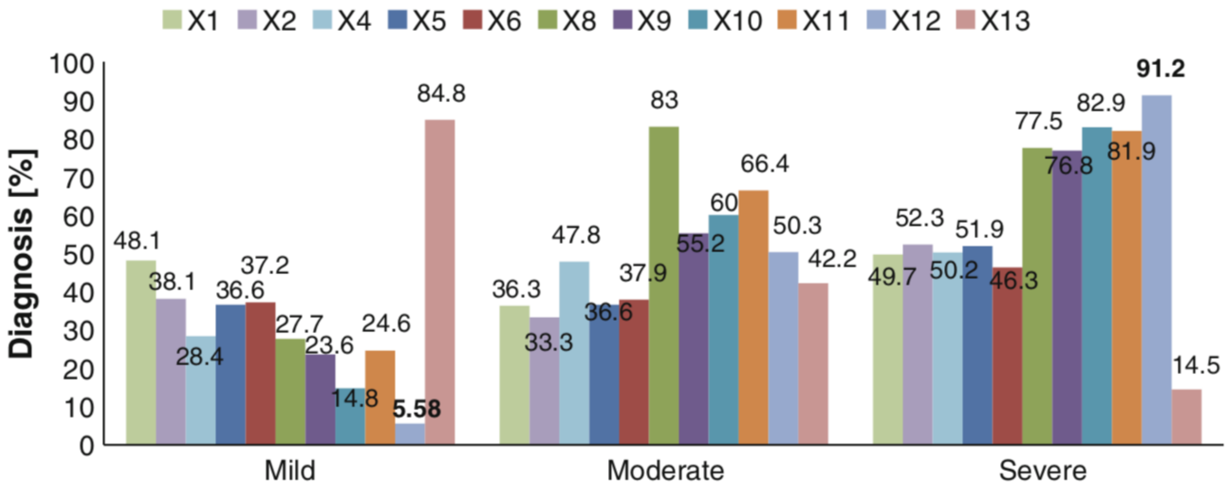


Figure 2.5: The probability of diagnosing a student with social anxiety, taking in consideration only one evidence at time [5]

2.3.2 A Bayesian Network Model for Diagnosis of Liver Disorders

In this research, they present about how to use Bayesian network model to diagnose liver disorders. The inspiration of this project is that they know about the success of Bayesian network with diagnosis of other disease. The data of this project came from the clinic records and use the information to create a model and conditional probability of each feature. The model parameters consist of 93 features and each feature has its own value. For example, blood sugar can be low, normal, high and very high. The outcome is 16 of liver disorders type.

The model use classification accuracy for training and viewed each disease as a separate class that the model predicted based on the values of all the other variables. They train the model 4 times and separate data to 4 group. The result of each training is 34% in the first time and follow by 47%, 56% and 67%. From result the accuracy increasing every time that they are training because every time that the new train happen the model will get more data for training.

After they get accuracy of the model they think this model performance is quite good because they have 2 problem of training are a small number of data and missing data.

In the research they create an application for study about relationship between nodes when set the environment. The feature of application is it can choose nodes and change the variables inside. It will affect to another node and the outcome probability will change follow the environment.

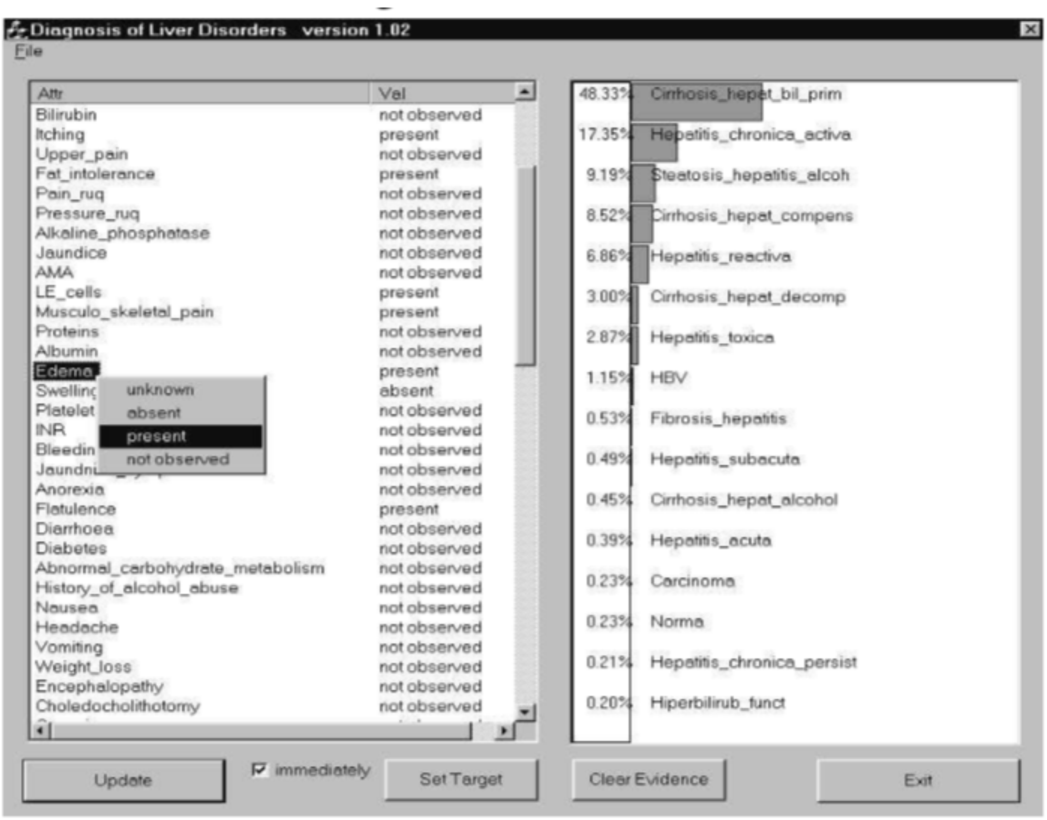


Figure 2.6: The application for study about relationship of probability between the nodes that affect to the output result

This research is good for study about how to design the nodes, feature and variable and adapt the Bayesian network model. They show the detail about the probability change by application and how to evaluate with the lack of data such as small size and missing data. [6]

2.4. Artificial Intelligence in Healthcare

Artificial intelligence (AI) in healthcare is bringing machine learning algorithms to process medical data in order to capture some hidden patterns inside those data. After learning through a huge number of that data, it will have the ability to predict an approximate output. The difference between the traditional software and artificial intelligence is that AI is heuristic, not algorithmic. AI uses statistical data to conduct the model predicting the given output from a set of input. This process will be done repeatedly, and the error rate will be reduced over training time. The goal is to reach the minimum error rate between the real output and the predicted output so that it can make the optimal conclusion. [https://en.wikipedia.org/wiki/Artificial_intelligence_in_healthcare]

Since 1960, there were various types of projects initialized. An example of this was called Dendral which aimed to solve a specific task such as analyzing mass spectra of unknown organic molecules using chemistry knowledge. [16] Dendral was derived to develop another healthcare project called MYCIN. MYCIN was an artificial intelligence to recommend antibiotics with an appropriate amount for patient's body weight by asking a set of yes/no questions. Also, identify bacteria causing severe inflections. After gathering answers from patient, a list of bacteria will be given based on the probability (high to low) along with cause and treatment using principle of certainty factors. Later, the certainty model was proved that it can be transformed to graphical model such as Bayesian networks. [17]

Recently, AI has been brought to be the solution to many fields of healthcare. For example, in radiology, AI has been used to do image processing. Some clinicians may not detect some changes in an image because it is too small, but AI can help this. [20]

2.5. Technologies

2.5.1 Python

Python is programming language that created by Guido van Rossum, 1991. This language is an interpreted high-level programming language. Python has a design that can adapt to use for created programming on both small and large scales. Python is open source code and has a large community-based development model and available on many operating systems.

For the feature Python has many features such as this language has many libraries than support the programming that are multiple paradigms. Moreover, Python is object-oriented for easier to using and about Python system has a dynamic type system and automatic memory management. [7]

We select Python language because this language is very popular for create a model. This language has many libraries which support our project such as aGrUM and pyAgrum that can create a graphical model.

2.5.2 SQL

SQL is Structured Query Language use for managing a data in database in the term of a relational database management system (RDBMS) or in the stream the process will use a relational data stream management system (RDSMS). It is a domain specific language and use for handling structure data.

Moreover, SQL has many advantages, for example, good when working with the large amount of data, using for viewing the data without storing the data in the object, show the object that contain the table more than one table. [8]

In our project has a large amount of training data from this reason we found an advantage of SQL that very compatibility with our project. SQL is very good when handle a large amount of data.

2.5.3 MySQL Workbench

MySQL Workbench is a visual database design for configuring, managing and administering the database. It has 4 main features. First, the program support SQL development user can create and manage connections to database servers. Second, Data modeling or the design part that has many features such as you can create models of your database schema graphically and has the table editing for help to config value in the row and column easier. Third, Server Administration is method for create instant of admin for user it includes a feature about backup and recovery a server. The last feature is Data Migration for allowing user to migrate database between computers. it has many programs that support such as Microsoft SQL Server, Microsoft Access, Sybase ASE, SQLite, SQL Anywhere, PostgreSQL. So anyway, this software should be better when works with MySQL because it fully supports MySQL server. [16]

2.5.4 HTTP Protocol

In 1989 at CERN the one of developers TIM Berners-Lee was create HTTP, from the first this protocol working with the World Wide Web Consortium (W3C) and the Internet Engineering Task Force (IETF) and Requests for Comments (RFCs)

HTTP protocol is a protocol that allow the fetching of resources in the term of HTML documents. It is protocol about exchange data on the web and a client-server protocol. Web browser is requester for a complete document that reconstructed from the different sub-documents fetched.

For the World Wide Web for communication the data on internet it must to use the protocol, so HTTP is one of many protocols that can distributed, collaborative and hypermedia information systems. This protocol creates for help user to access the resources easies by using hyperlink. [9]

2.5.5 RESTful Web Services

Creating web services has a many part that important and an architectural is the one of those. Representational State transfer (REST) is the one of famous architectural that defines a set of constraints to be used for creating Web services. For the web services that use REST style it will provide working together between computer systems on the Internet. REST-compliant use for requesting when user want to access manipulate textual representations of Web resources. The properties of this architecture include scalability that can handle many components and interactions, uniform interface is very simple, visibility of communication between components by service agents, component easy to modify. [10]

2.5.6 aGrUM and pyAgrum

aGrUM is a library that design for create a graph-based model such as Bayesian network, influence diagrams, decision trees, GAI network or Markov decision processes. This library base on C++ that easily for create application with graphical models.

pyAgrum is one of library in aGrUM it is high-level interface that be a part of aGrUM it means pyAgrum base on C++ like aGrUm. It can create, handle and make computations into Bayesian Network.

Those are important library that we use for creating a Bayesian network model in our project and study about Bayesian network theory. [12]

2.5.7 Flask

Flask is a microframework that was created by the one of developers in pocoo, his name is Armin Ronacher in 2004. A microframework is a framework providing some features for web application development, but some are not included such as handling authorization. [18] Flask doesn't need tools or libraries. It doesn't have form validation, database abstraction layer, or any components that need a pre-existing third-party library provide normal functions. Flask can include an extension such as object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools.

Since our project requires two important features including serve as web services and database management, using Flask is the appropriate one. Flask provides features for RESTful request dispatching. [17]

2.5.8 .BIF file structure

.BIF file is used to define Bayesian network structure. It contains three types of block (network, variable, probability). Components in each block include block type, block name, attribute name, attribute value. Below is the format of a block in .bif.

```
block-type block-name {
        attribute-name  attribute-value;
        attribute-name  attribute-value;
        attribute-name  attribute-value;
}
```

Since Bayesian network contains nodes, edges, and conditional probability tables, this information will be explained via network blocks, variable blocks, and probability blocks in .BIF file.

Network blocks explains about network's name and properties. An example of a network block is shown below.

```
network Robot-Planning {
        property version 1.1;
        property author Nobody;
}
```

Variable blocks contain information of variables in a network. An example of a variable block is shown below.

```
variable Leg {
        type discrete[2] { long, short };
        property temporary yes;
}
```

Probability blocks define all possible values of a node depending on their parents, and conditional probability tables as illustrated below. [19]

```
probability ( Leg | Arm ) {
        table 0.1 0.9 0.9 0.1;
}
```

2.5.9 Node.js

Node.js is the open source JavaScript tool used for server-side. We bring Node.js as our project has some part that needs to handle file uploading from html page. [20]

Chapter 3
Design and Methodology

3.1. General Requirement

3.1.1. This system will be provided as a web service.
3.1.2. The system can show relevant diseases based on the given symptoms.
3.1.3. The system can find causes and treatment of the disease.
3.1.4. User can provide information from the medical examination (if any).
3.1.5. User can specify an interested disease to request general description.
3.1.6. User can give symptoms and behavior and get associated diseases as a response.
3.1.7. Administrators can import a network as a text file.
3.1.8. Administrators can export a network as a text file.
3.1.9. Administrators can view the overall of the network.
3.1.10. Administrators can edit the structure of the network including nodes, edges, and conditional probability tables.
3.1.11. Administrators can feed training data to update conditional probability table.

3.2. Use Case Analysis

3.2.1. Use Case Diagram

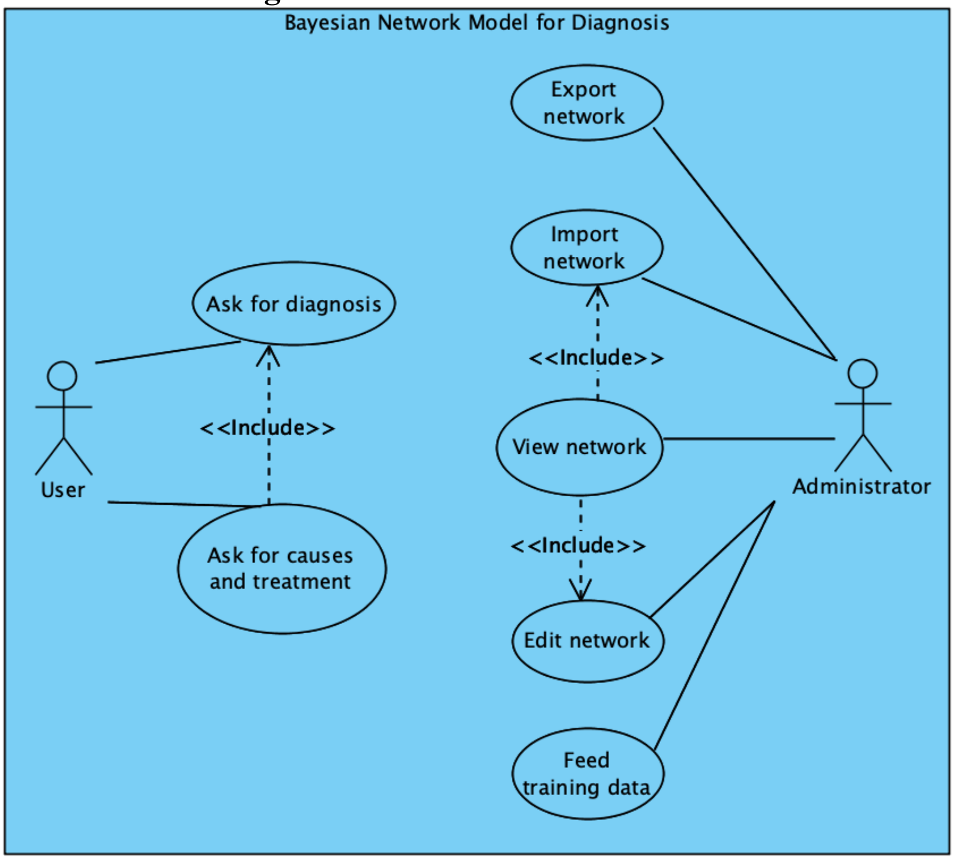


Figure 3.1: Use Case Diagram of Bayesian Network Diagnostic System

There are two roles of users in the system; users and administrators. Users will be able to ask for diagnosis and ask for cause and treatment about the disease. Administrators will be able to import, export, view, edit and feed data to users.

3.2.2. Use Case Narratives

Scenario 1: Ask for diagnosis
Actor: User
Goal: To get associated diseases based on given symptoms, behaviors, and medical examination sheet result.
Precondition: None
Main Success scenario:
1. User selects existing symptoms.
2. User selects behavior.
3. User provides medical examination sheet result.
4. System responses associated diseases as a result.
5. User selects an interested disease.
6. System shows causes and treatment of the disease.

Extension scenario (a):
3a. User doesn't have any examination sheet result.
4a. User clicks no examination result button.
Extension scenario (b):
4b. System cannot find associated diseases.
5b. System shows a message saying that associated diseases not found.
6b. User selects symptoms, behavior, and gives medical sheet result again.
Extension scenario(c):
5c. User doesn't select an interested disease.
6c. User presses back button.
7c. User selects symptoms, behavior, and gives medical sheet result again.
Post condition: None

Scenario 2: Ask for causes and treatment
Actor: User
Goal: To get causes and treatment of an interested disease.
Precondition: None
Main Success scenario:
1. User fills disease name.
2. System shows causes and treatment of the disease.

Extension scenario (a):
1a. User fills incorrect format.
2a. User re-enter disease name.
Extension scenario (b):
1b. User fills non-existing disease name.
2b. User re-enter disease name.
Post condition: None
Scenario 3: Import network
Actor: Administrator

Goal: To import network file (.bif) in order to construct Bayesian network model.
Precondition: None
Main Success scenario:
1. Administrator selects existing network file.
2. System shows Bayesian network model.
Extension scenario (a):
1a. User selects non-existing file.
2a. User re-selects file.
Extension scenario (b):
2b. System cannot read a file.
3b. System shows a message saying that incorrect type of file.
4b. User re-selects file.
Post condition: Bayesian network model inside the diagnostic system is re-initialized.

Scenario 4: Export network
Actor: Administrator
Goal: To export network file (.bif) in order to bring out Bayesian network model.
Precondition: Bayesian network model does exist.
Main Success scenario:
1. User selects export network as .bif file.
2. System gives .bif file.
Post condition: None

Scenario 5: View network
Actor: Administrator
Goal: To view Bayesian network inside the diagnostic system.
Precondition: Bayesian network model does exist.
Main Success scenario:
1. User selects view network button.
2. System shows Bayesian network inside the diagnostic system.
Post condition: None

Scenario 6: Edit network
Actor: Administrator
Goal: To edit Bayesian network inside the diagnostic system.
Precondition: Bayesian network model does exist.

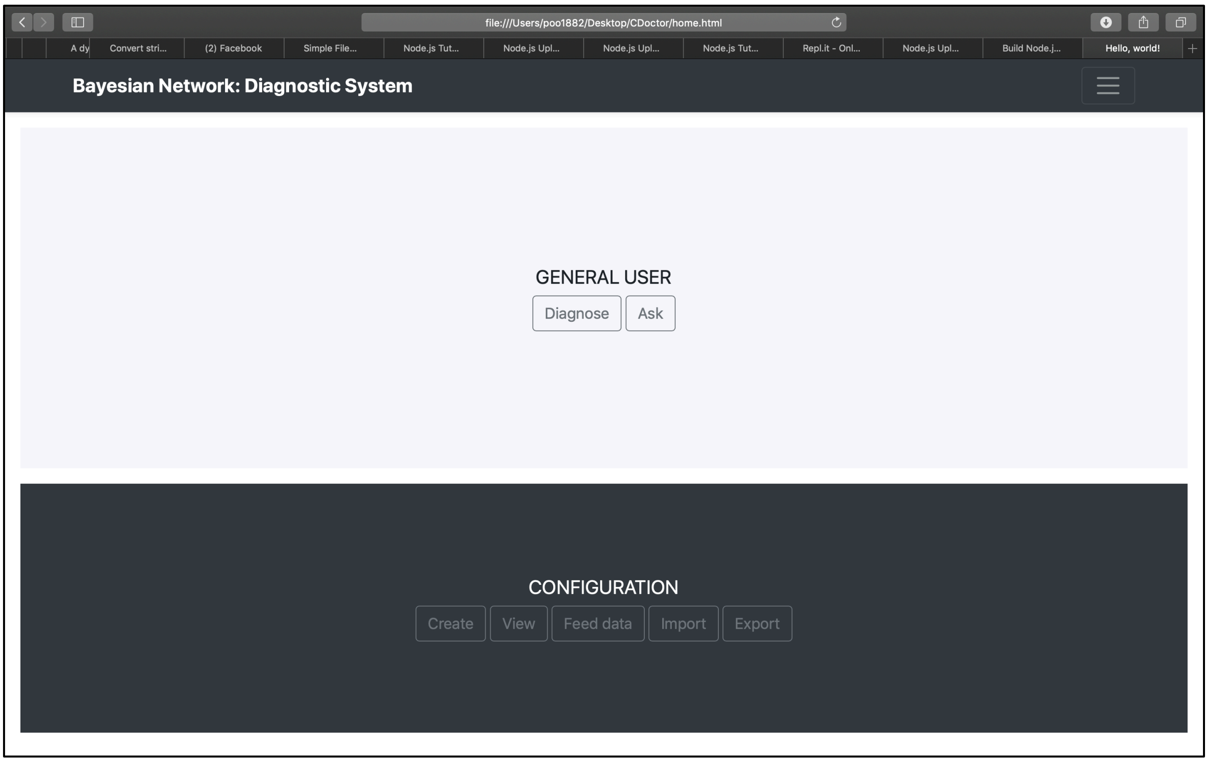
Main Success scenario:
1. User selects what to edit (nodes, edges, conditional probability tables).
2. User fills in what to be added/removed/edited.
3. User clicks save the network.

4. System shows the network that has been edited.
Extension scenario (a):
2a. User fills in incorrect format of input.
3a. User re-enters the input.
Post condition: None

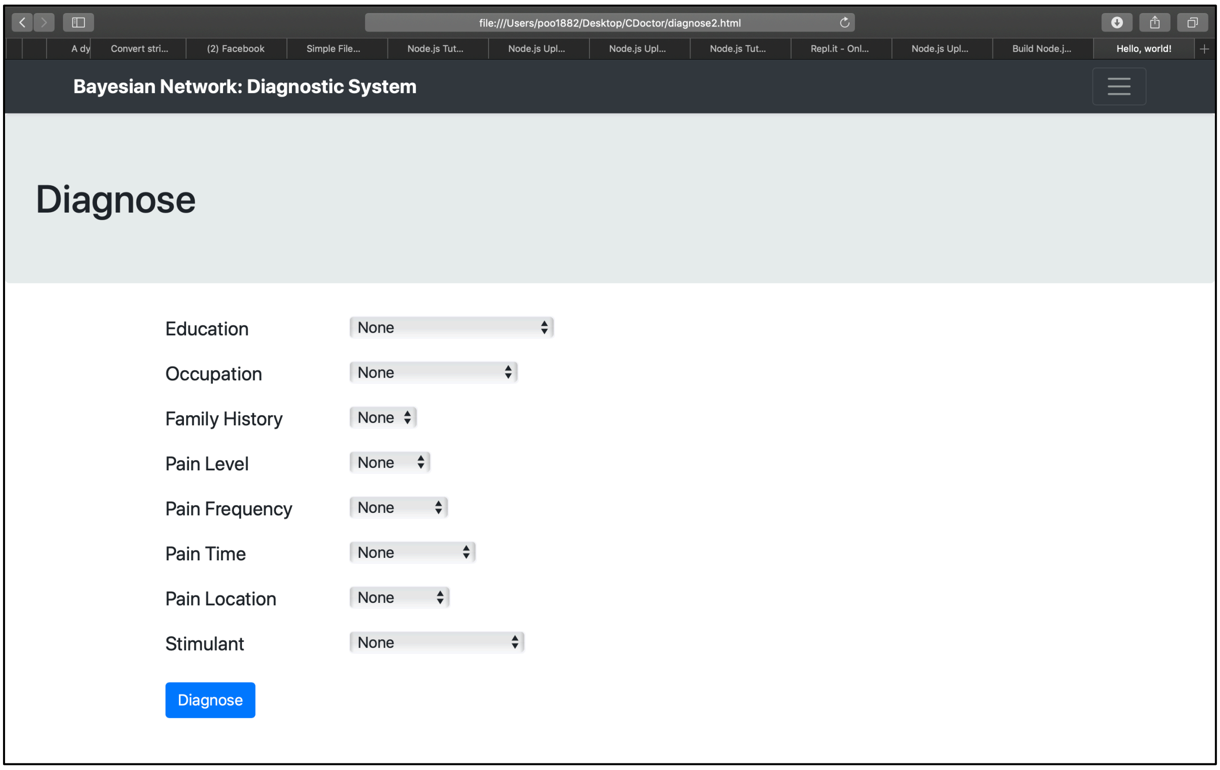
Scenario 7: Feed training data
Actor: Administrator
Goal: To update conditional probability table.
Precondition: Bayesian network model does exist.
Main Success scenario:
1. User selects file to feed in.
2. User selects button send data to a model.
3. System reads a file.
3. System shows success message.
Extension scenario (a):
1a. User selects non-existing file.
2a. System shows error message.
3a. User re-selects a file.
Extension scenario (b):
3b. System cannot read the file.
4b. System shows error message.
5b. User re-selects a file.
Post condition: None

3.3. Screen Layout and Navigation

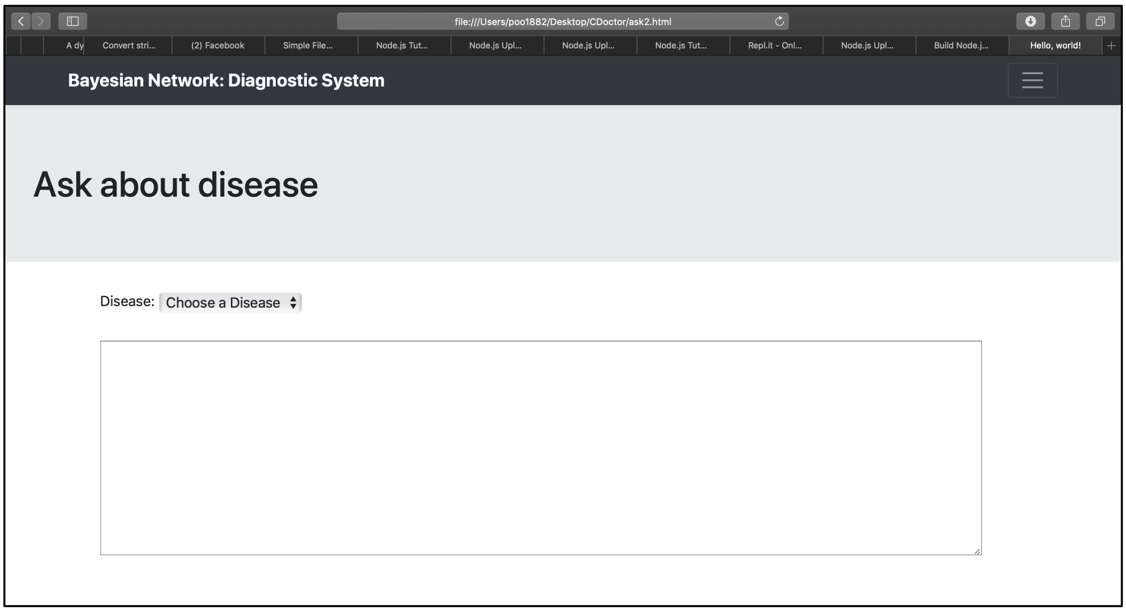
3.3.1. Home



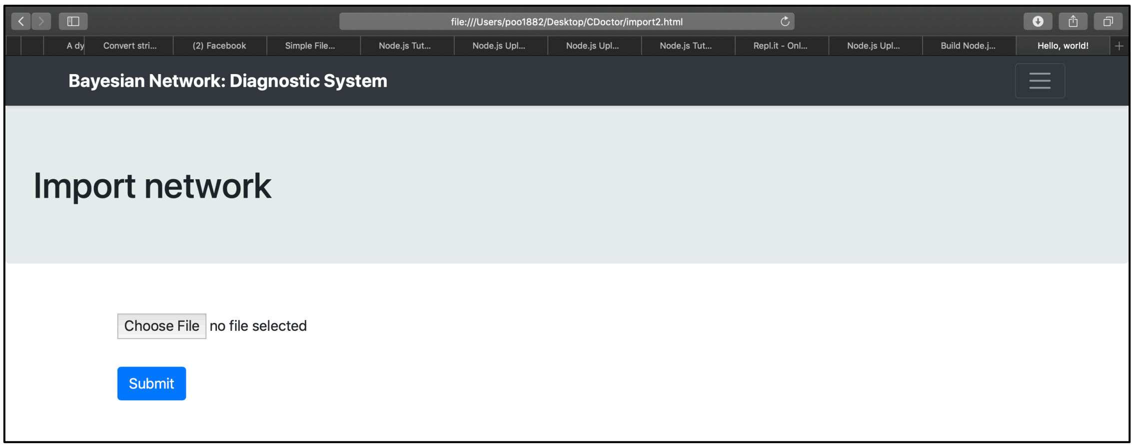
3.3.2. Ask for Diagnosis

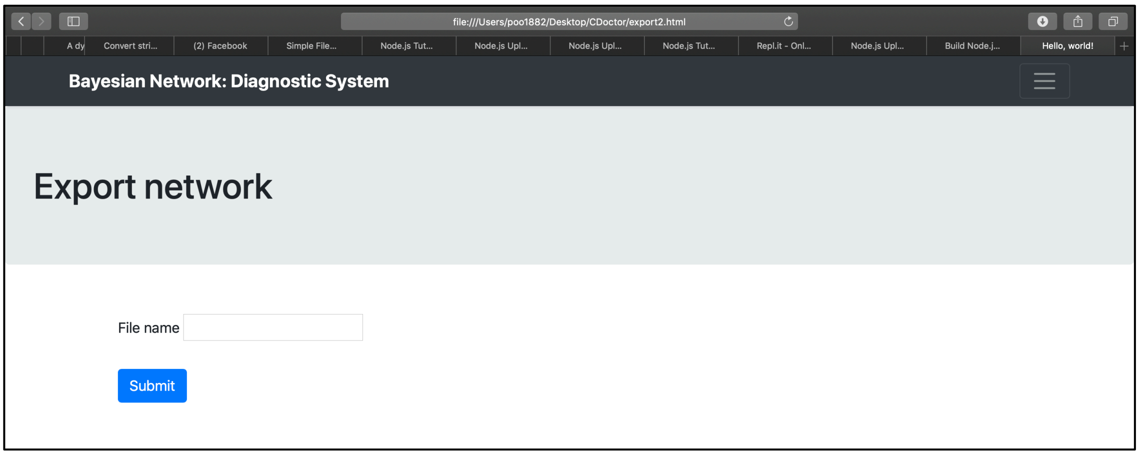


3.3.3. Ask Cause and Treatment

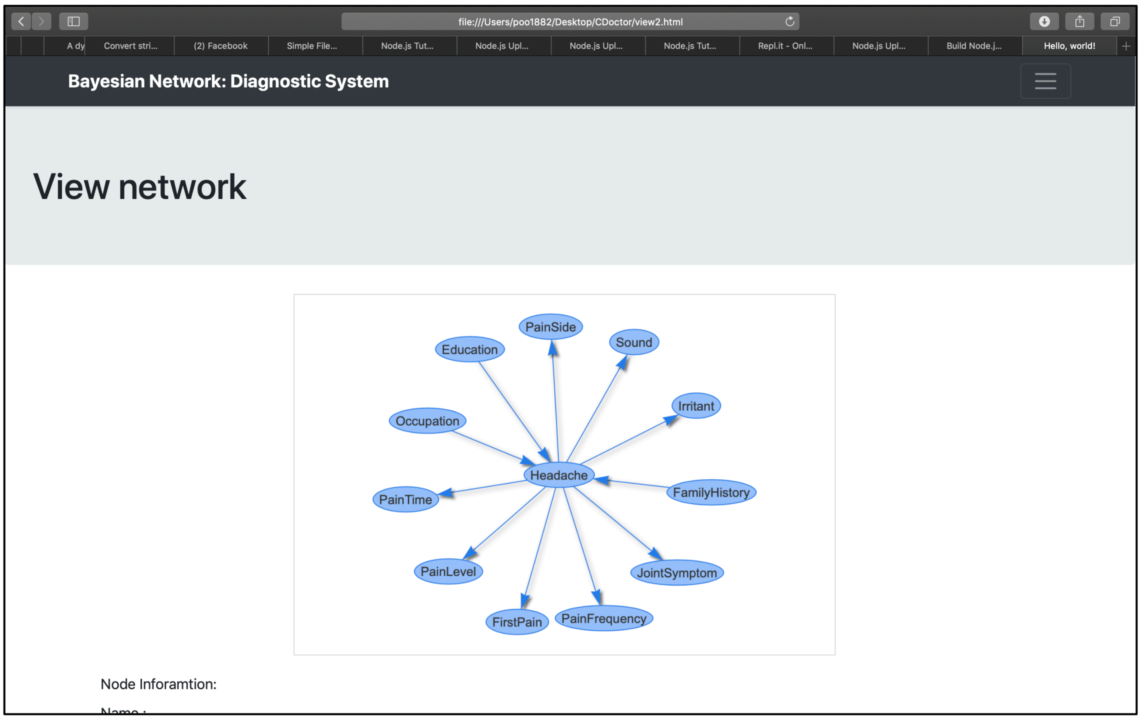


3.3.4 Export/Import

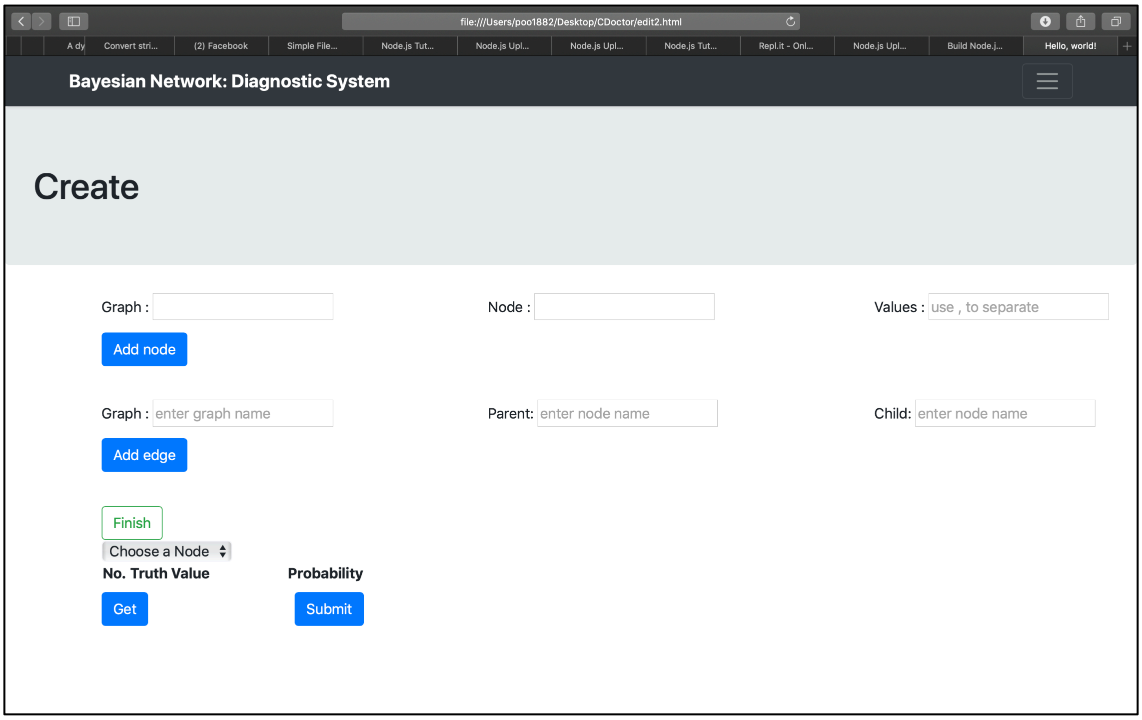




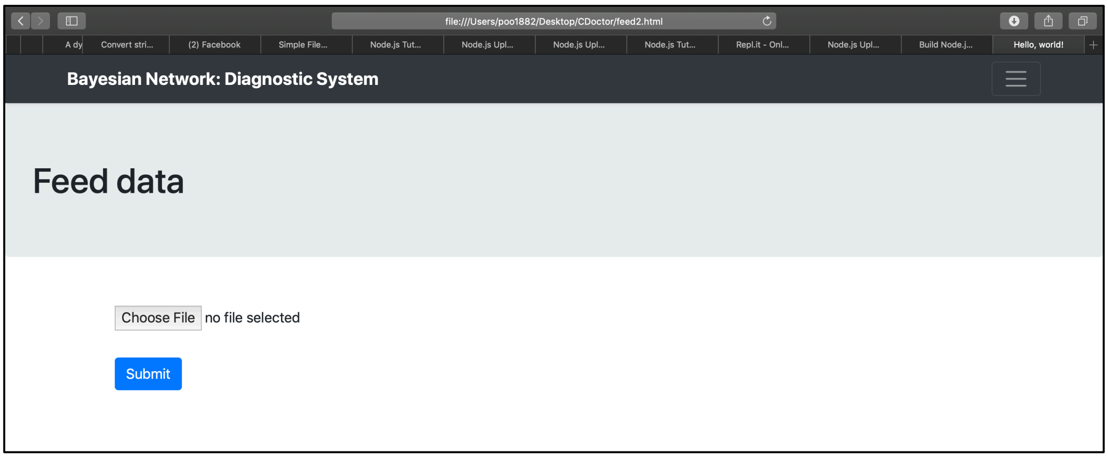
3.3.5. View Network



3.3.6. Edit Network



3.3.7. Feed Training Data



3.4. Discussion of Architecture

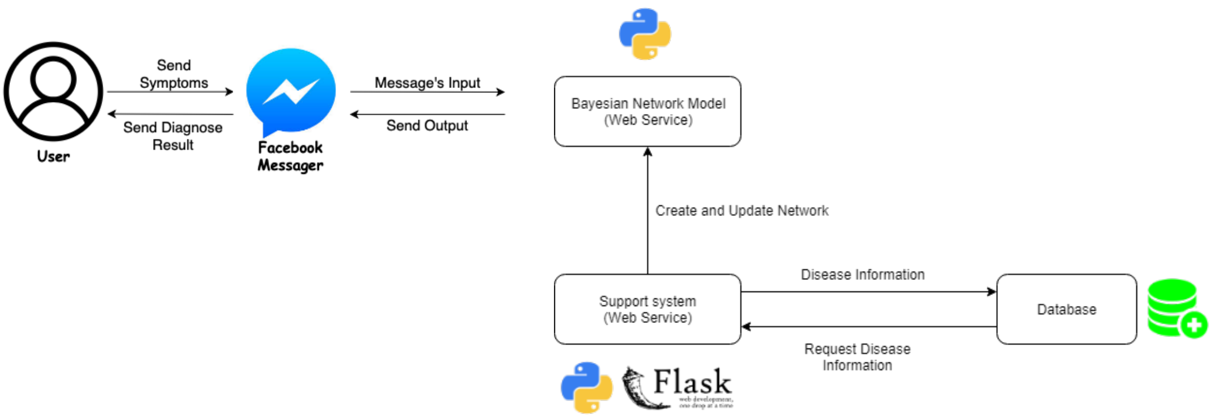


Figure 3.2: An Architecture of Bayesian Network Diagnostic System

Our web service will be served on cloud server to be called by Chatbot project. This is done using Heroku cloud server. The support system will handle for editing the database about the diagnostic system and diseases information.

3.4.1. Diagnostic System

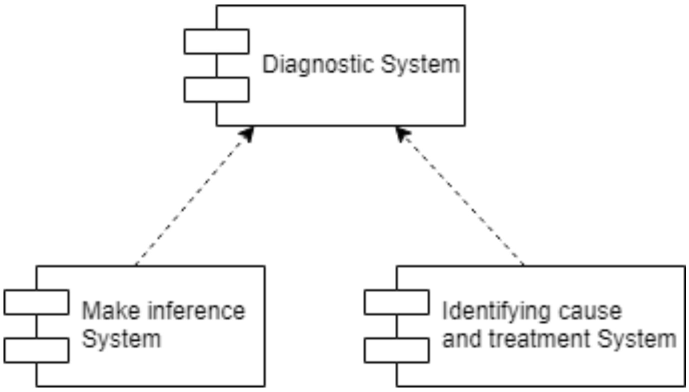


Figure 3.3: An Overview of Diagnostic System

The main diagnostic system contains two modules. Firstly, it is a making inference system - the system that tries to find associated diseases. The input will be symptoms, behavior and examination sheet results. After taking all input in, Bayesian network model will be calculating the probability for each disease (this is called *inference*), and printing out all associated diseases (diseases which reach high probability) as an output.

The second module is called identifying cause and treatment system. This system takes the name of the disease as an input from users and gives causes and treatment of that disease as the output.

3.4.2. Support System

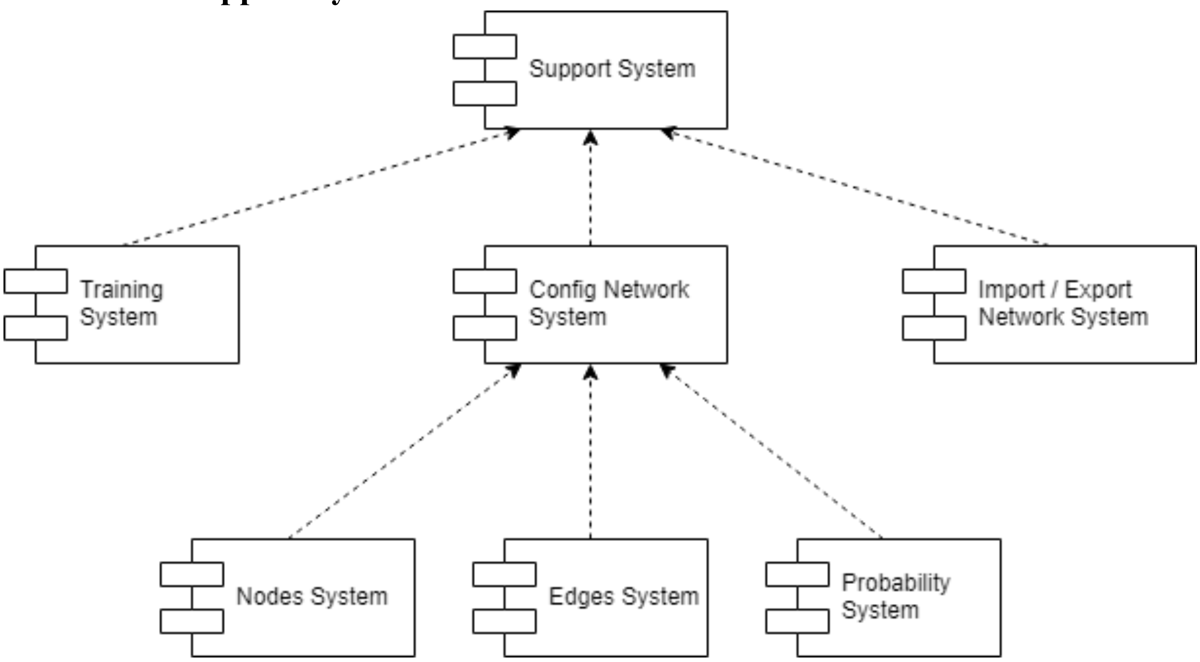


Figure 3.4: An overview of Support System

The support system consists of three modules. The first one is the training system. This will be used to update the conditional probability table inside each node in Bayesian network. It takes record data as input. Each column in a record represents value of each node in Bayesian network. The conditional probability in each node will be updated based on values exist in the record data. For example, the record data says cloudy and raining always occurs concurrently (both of them are true). Therefore, the probability of raining given that the weather is cloudy will be increased.

The second module is the configuration system. This can be separated into three sub modules. As a Bayesian network contains three main components (nodes, edges, conditional probability tables), there will be three sub modules responsible for configuration each of them.

The third module is the import/export network system. This module deals with the .bif text file. It is the file type used to construct a Bayesian network. The system provides both import and export task for users. The import task takes .bif file as an input and will perform a Bayesian network locating inside diagnostic system. The export task will be providing the .bif file as the output in case users want to bring out the model to be used later.

3.5. Database Schemas

3.5.1. Database Design

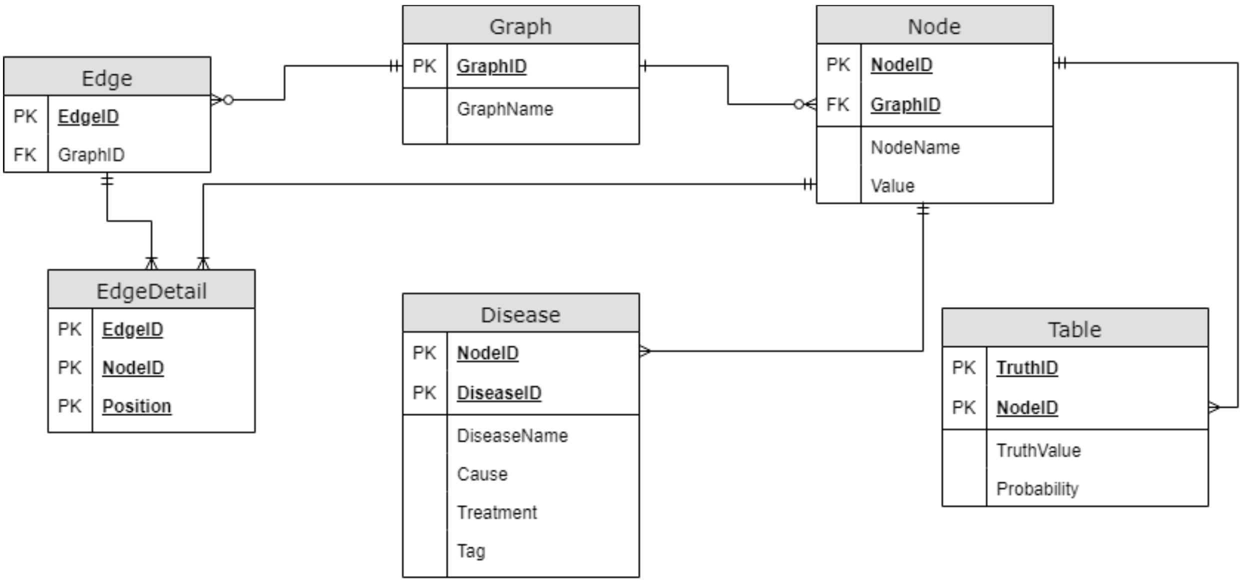


Figure 3.5: An ER diagram describes tables in the system

3.5.2. Data Dictionary

3.5.2.1. Graph

Field Name	Data Type	Data Format	Field Size	Description	Example
GraphID	Integer	NNNN	4	Unique number ID for all Graph	1234
GraphName	text		20	Name of Graph	

3.5.2.2. Node

Field Name	Data Type	Data Format	Field Size	Description	Example
NodeID	Integer	NNNN	4	Unique number ID for all Node	1234
GraphID	Integer	NNNN	4	Unique number ID for all Graph	1234
NodeName	text		20	Name of Node	Tension
Value	text	NNNN, NNNN, NNNN,..	30	Set of possible value inside the node	True,False

3.5.2.3. Tables

Field Name	Data Type	Data Format	Field Size	Description	Example
TruthID	Integer	NNNN	4	Unique number ID for all Tables	1234
TruthValue	text	NNNN NNNN NNN	10		TTT
NodeID	Integer	NNNN	4	Unique number ID for all Node	1234
Probability	double	NNNN	3	Probability Value	99.99

3.5.2.4. Edge

Field Name	Data Type	Data Format	Field Size	Description	Example
EdgeID	Integer	NNNN	4	Unique number ID for all Edge	1234
GraphID	text	NNNN	20	Unique number ID for all Graph	1234

3.5.2.5. EdgeDetail

Field Name	Data Type	Data Format	Field Size	Description	Example
EdgeID	Integer	NNNN	4	Unique number ID for all Edge	1234
NodeID	text	NNNN	4	Unique number ID for all Node	1234
Position	text		8	Position of Node(Parent or Child)	Parent

3.5.2.6. Disease

Field Name	Data Type	Data Format	Field Size	Description	Example
DiseaseID	Integer	NNNN	4	Unique number ID for all Disease	1234
NodeID	Integer	NNNN	4	Unique number ID for all Node	1234
Cause	text	“NNN”	500	Cause of Disease	“Cause of headache”
Treatment	text	“NNN”	500	Treatment of Disease	“Treatment of headache”
Tag	text	NNN,NNN,NNN	50	Tag of disease	head,arm,body

3.6. Bayesian Network Design

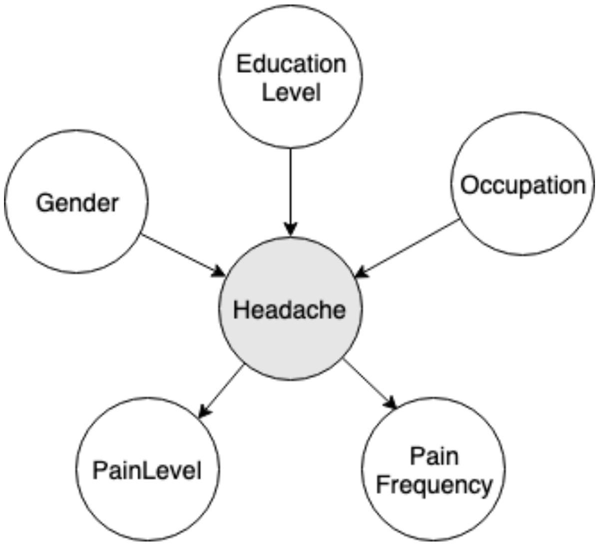


Figure 3.6: Tentative Bayesian Network (Probabilities are omitted.)

The tentative Bayesian network is shown above. We tend to target two types of headache which includes Tension and Migraine (highlighted in grey color). From this network, we can say that Gender, Education Level, Occupation have effect to both of the diseases. Tension and Migraine also provides pain level and pain frequency differently. That means pain level and migraine are conditionally depended on Tension and Migraine.

3.7. UML Diagrams

3.7.1. MVC Class Diagram

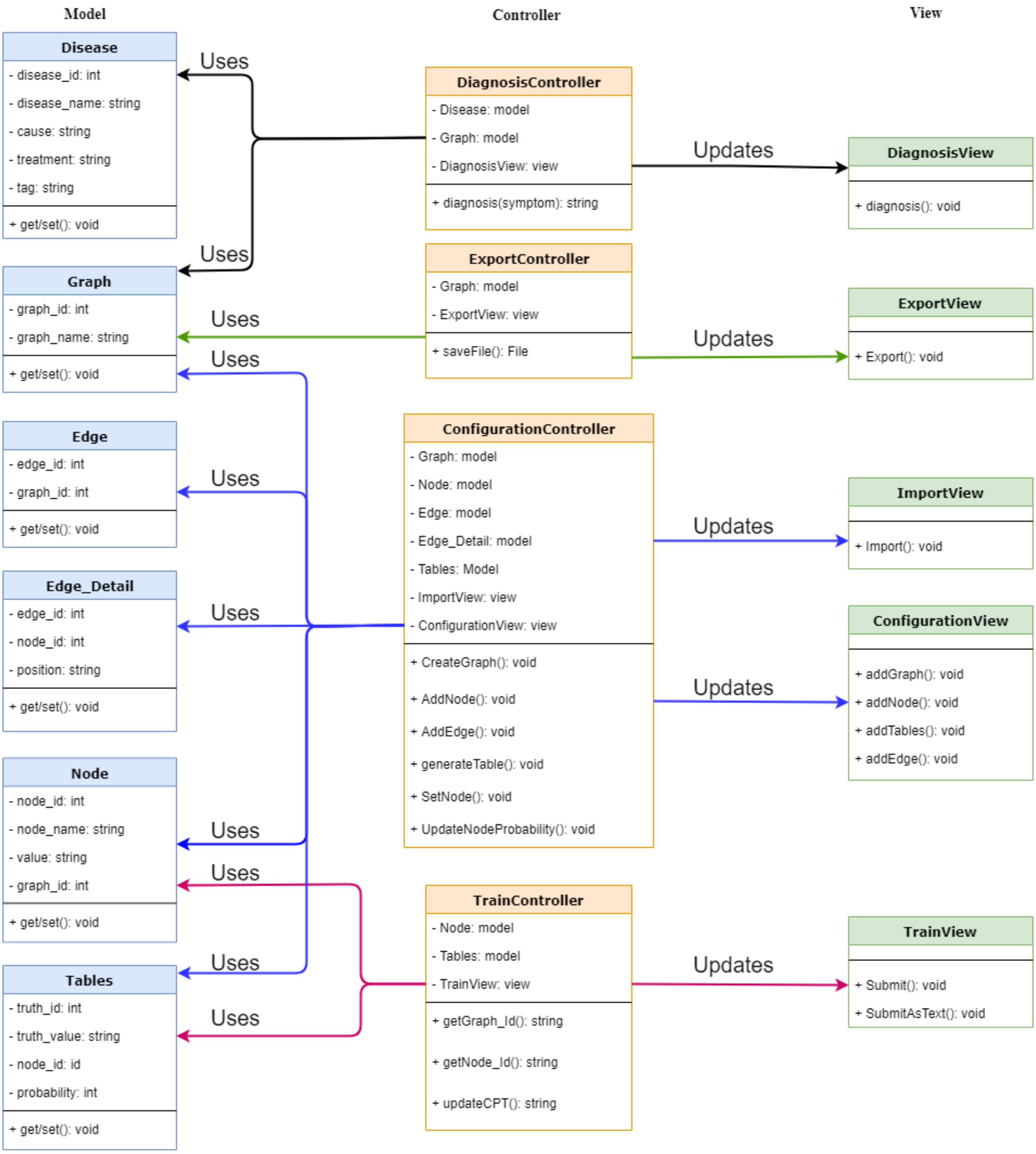


Figure 3.7: An MVC Class Diagram of Bayesian Network Diagnostic System

The MVC class diagram represents the way of our system works. There will be controllers handling for API routing. Those APIs will be called by the user interface. If there are changes made, the controller will update the database.

3.7.2. Sequence Diagram

3.7.2.1. Diagnose

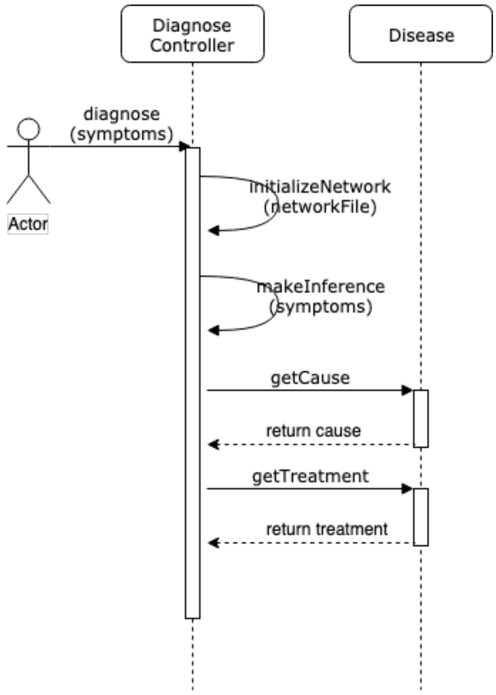


Figure 3.8: A Sequence Diagram of Diagnosis

3.7.2.2. Ask for Cause and Treatment

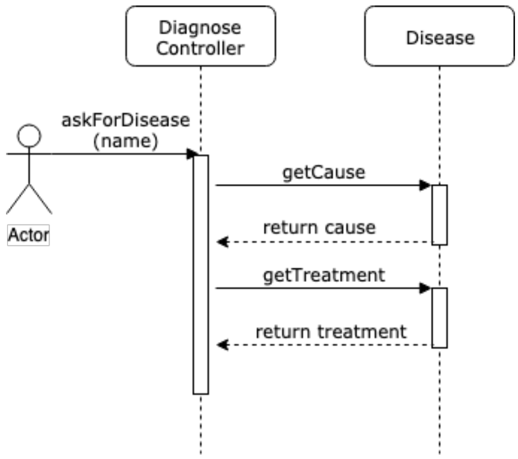


Figure 3.9: A Sequence Diagram of Asking for Cause and Treatment

3.7.2.3. Import Network

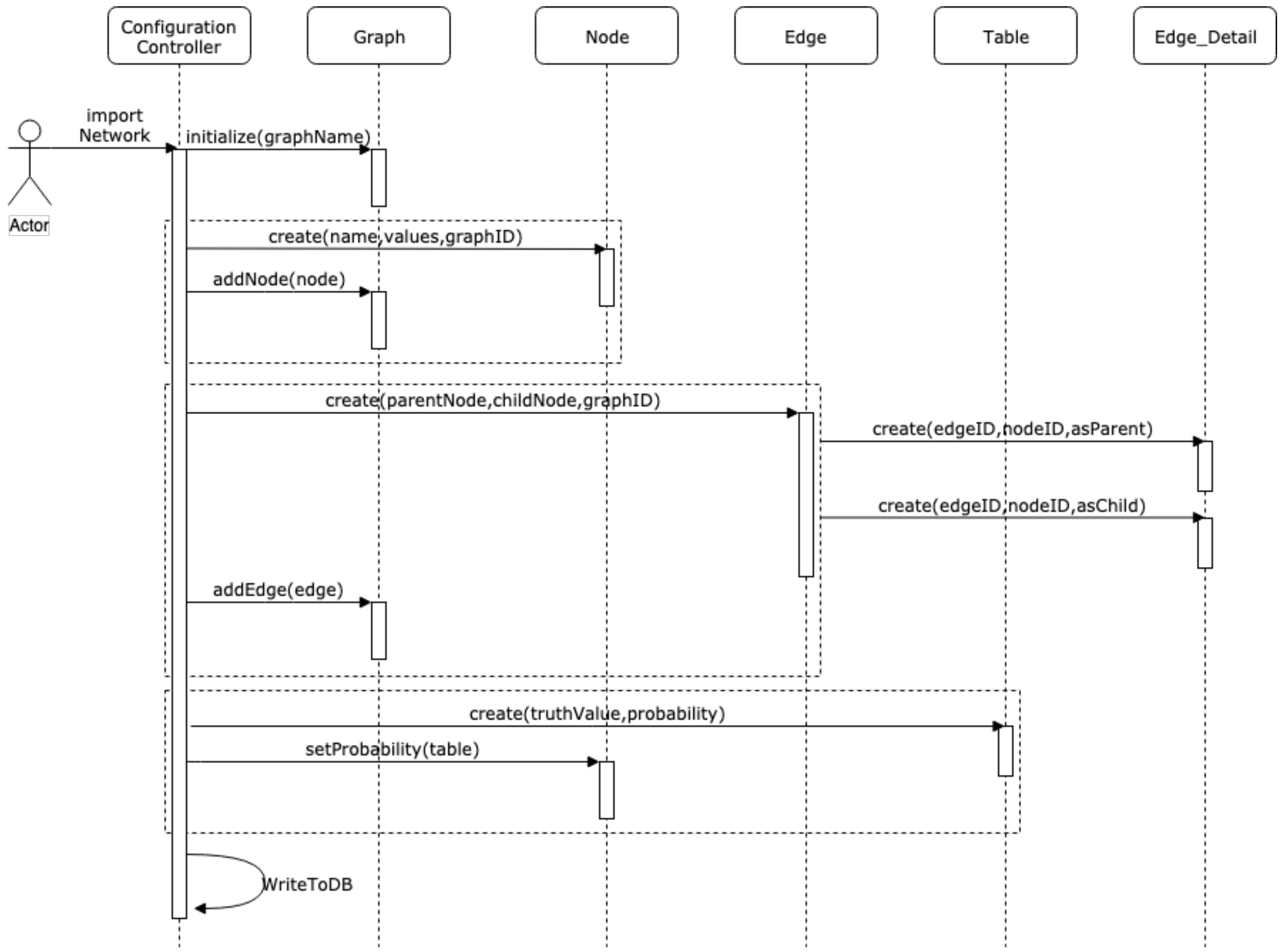


Figure 3.10: A Sequence Diagram of Importing Bayesian Network from Text File

3.7.2.4. Export Network

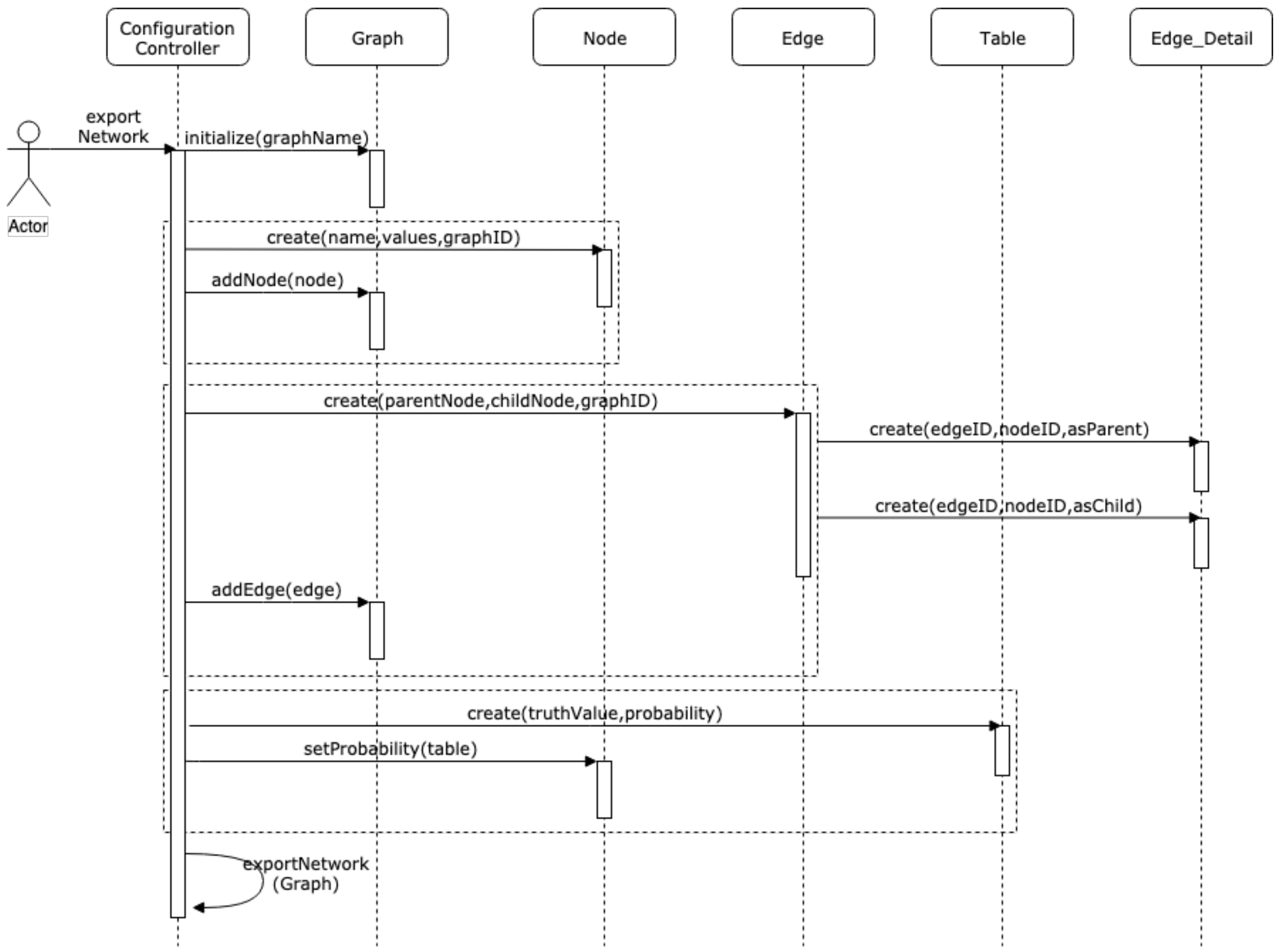


Figure 3.11: A Sequence Diagram of Exporting Bayesian Network as Text File

3.7.2.5. Edit Network

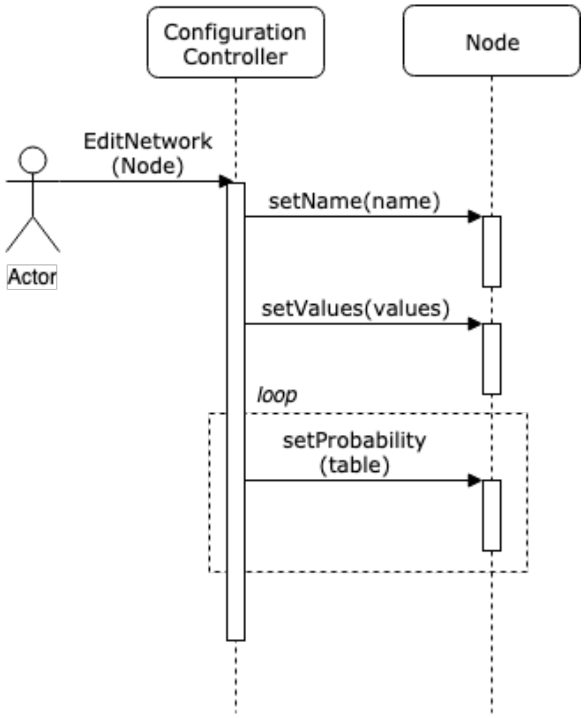


Figure 3.12: A Sequence Diagram of Editing the Components in Bayesian Network

3.7.2.6. Feed Training Data

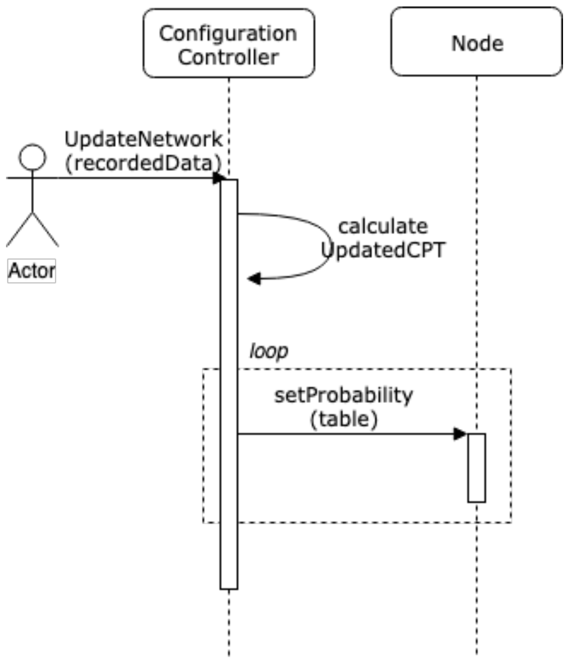


Figure 3.13: A Sequence Diagram of Feeding the Training Data to Update Probability in the Network

Chapter 4
Result

Section 1: Diagnostic System

1.1) Inference Module

Overview

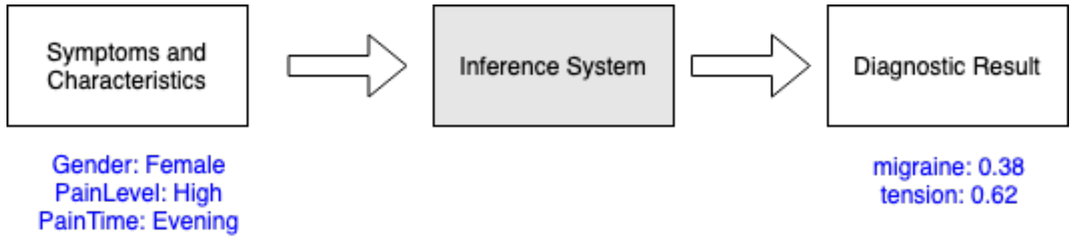


Figure 4.1: A diagram describing input and output of the diagnostic system

As shown above, the system is responsible for diagnosis. By taking symptoms and characteristics of patient as the input, the inference inside Bayesian network model will be made in order to calculate the probability of an interested node which is in this case referred to Headache node (migraine and tension).

Design and Working Process

We picked drop down as a choice since the set of possible values at each node is fixed according to Bayesian network model. User can choose one value for each node, or choose nothing. If nothing is selected, the output will be predicted based on the statistical data inside Bayesian network.

This service has been deployed to Heroku(a cloud server). It can be called via a http request (GET method) using the URL as follow(wait for a few minutes to get response): https://stark-oasis-75400.herokuapp.com/diagnose

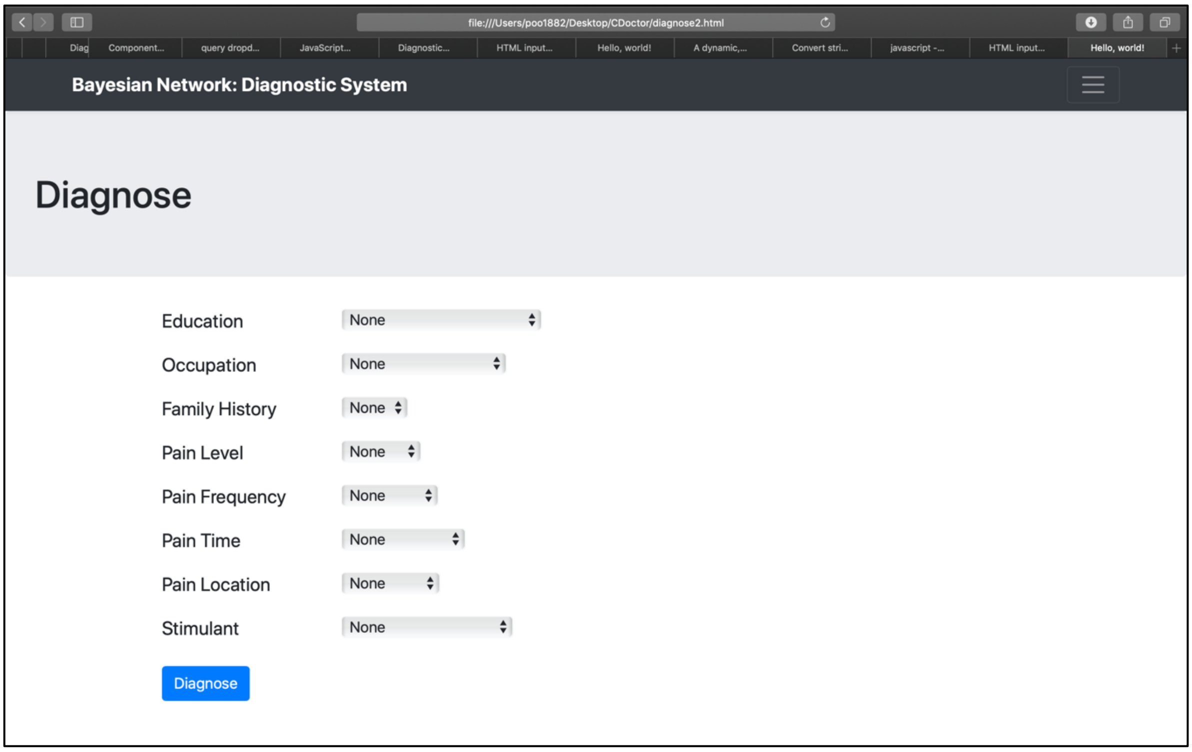
After value at each node is selected, the GET request will be performed. The whole set of values selected will be sent along with the request to a web service located on Heroku. After making inference, the output of getting each disease will be sent back as a probability.

<u>Test Cases</u>

Clinical Features	Tension Headache		Migraine Headache	
	Count	%	Count	%
PainLevel				
Low	25	36.2	4	6.7
Normal	38	55.1	47	78.3
High	6	8.7	9	15
PainTime				
EarlyMorning	7	10.1	15	25
LateMorning	2	2.9	4	6.7
Afternoon	8	11.8	8	13.3
Evening	10	14.5	4	6.7
Night	10	14.5	13	21.7

Figure 4.2: A picture describing the ratio of people getting Migraine and Tension classified by clinical features, retrieved from Chiangmai Neurological Hospital's website

Case 1: A http request with no parameters passing



Input -

Output Migraine 0.4803

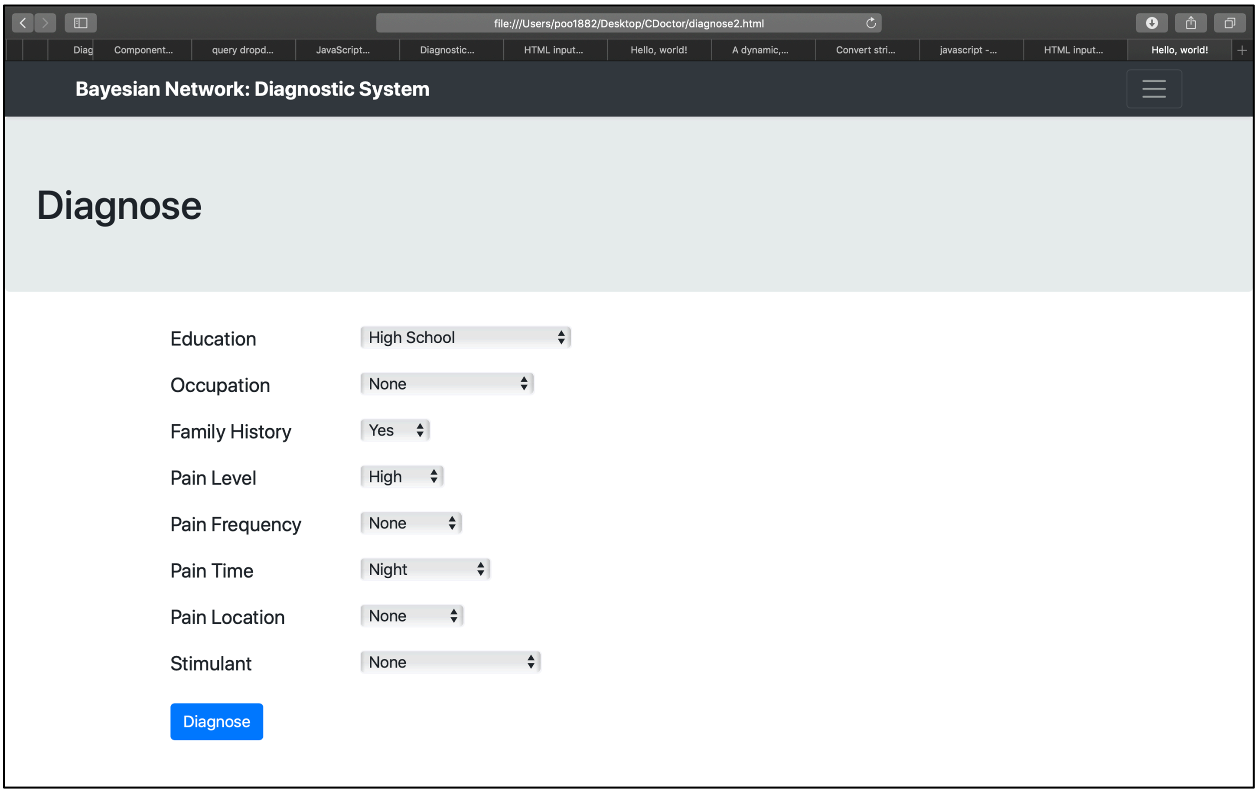
Tension 0.5197

Result

The value at each node is set to be ‘none’. This implies that there is no evidence given to Bayesian network model so, the request will be as follow: https://stark-oasis-75400.herokuapp.com/diagnose

After the request is sent, Bayesian network model will make inference to calculate the probability of getting migraine and tension from given evidence (in this case nothing is given) so, the prediction result will be provided from default knowledge inside Bayesian network model. The response will be sent back as a json containing probability of getting migraine and tension, along with basic cause and treatment.

Case 2: A http request with some information about patient



Input	PainLevel as *High*, PainTime as *Evening*
Output	Migraine 0.45
	Tension 0.55

Result

For this case, the value at node PainLevel, and PainTime are known. The request will be as
http://stark-oasis-75400.herokuapp.com/diagnose?PainLevel=2&PainTime=3

As seeing from the statistical data retrieved from Chiangmai Neurological Hospital’s website above, this patient is more likely to get Tension rather than Migraine. Therefore, the percentage of getting Migraine is lower so, the cause and treatment of Tension are attached as a response.

1.2) Identifying Cause and Treatment Module

Overview

This system is responsible for identifying cause and treatment of an interested disease. User needs to select the disease from the dropdown list, and the result will contain basic cause and treatment of that disease.

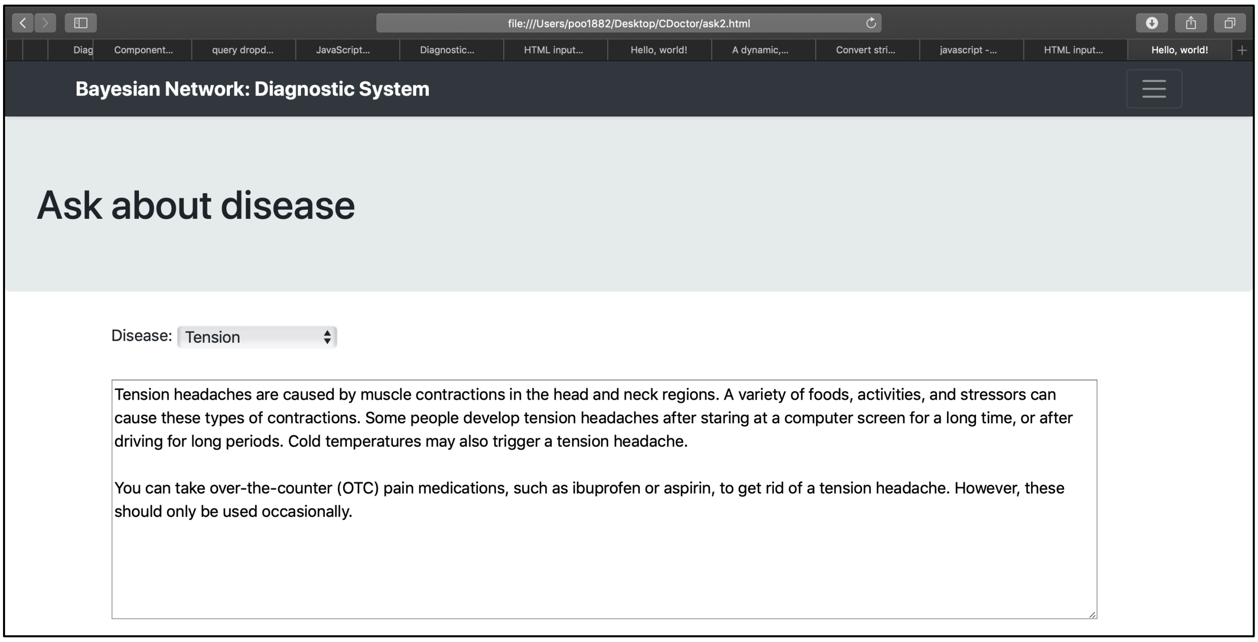
Design and Working Process

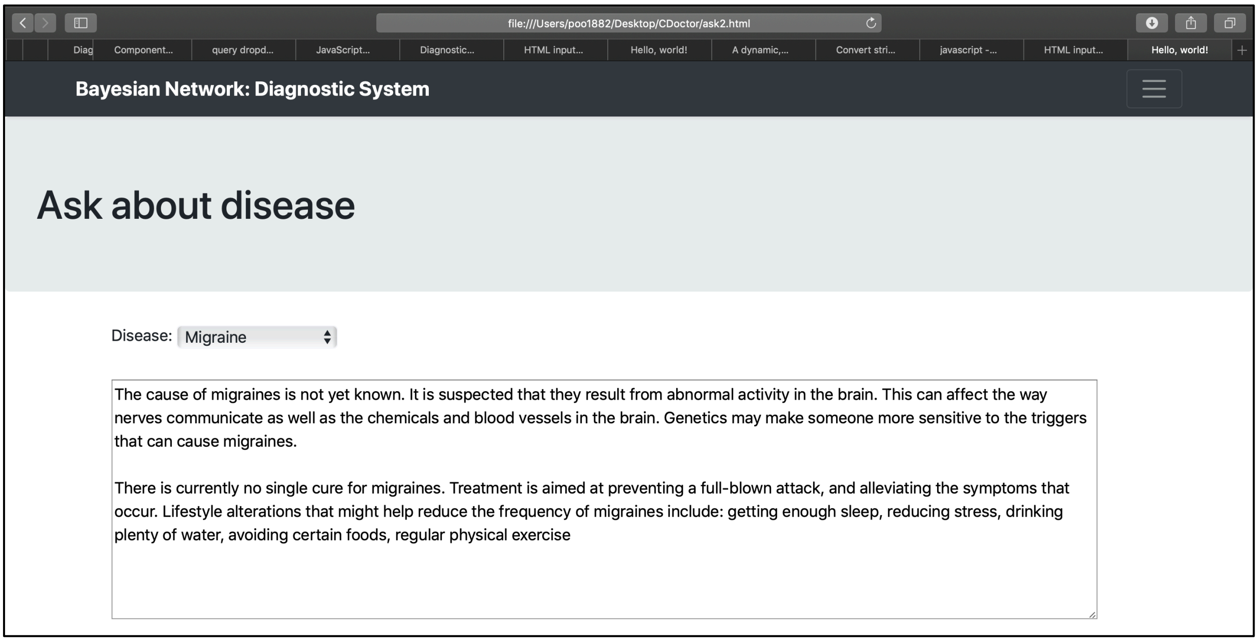
This page uses dropdown list to display diseases retrieved from database because it is easy to make a choice. Also, textarea is good for displaying the description of a selected disease.

When the page is opened, the description of all disease will be requested from the database. After user selects an interested disease, the onchange event will be triggered, and the disease's description will be automatically displayed on the textarea.

Test Case

Case 1: An interested disease is specified from the dropdown list.





Input A disease's name

Output A description containing basic cause and treatment

Result

From the dropdown list, user needs to pick one of them. After selecting the disease, the onchange event will occur. The value of a selected disease from dropdown list will be used to find cause and treatment. After finding, it will be displayed on the textarea.

Section 2: Support System

2.1) Configuration Module

<u>Overview</u>

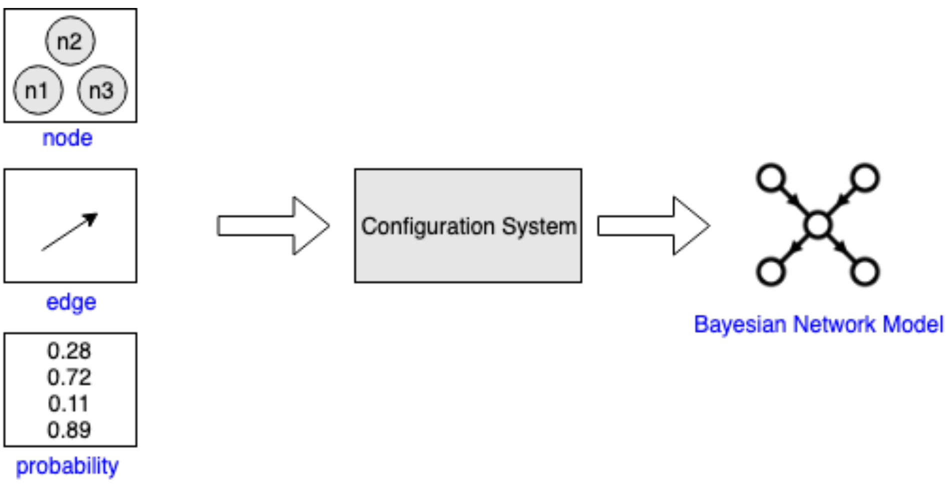


Figure 4.3: A diagram describing input and output of the configuration system

Configuration system allows user to create the new network, or modify the existing network. This consists of 3 sub modules including node, edge, and probability. After filling in all nodes, edges, and conditional probability table, the system will be responsible for creating the network as a Bayesian network.

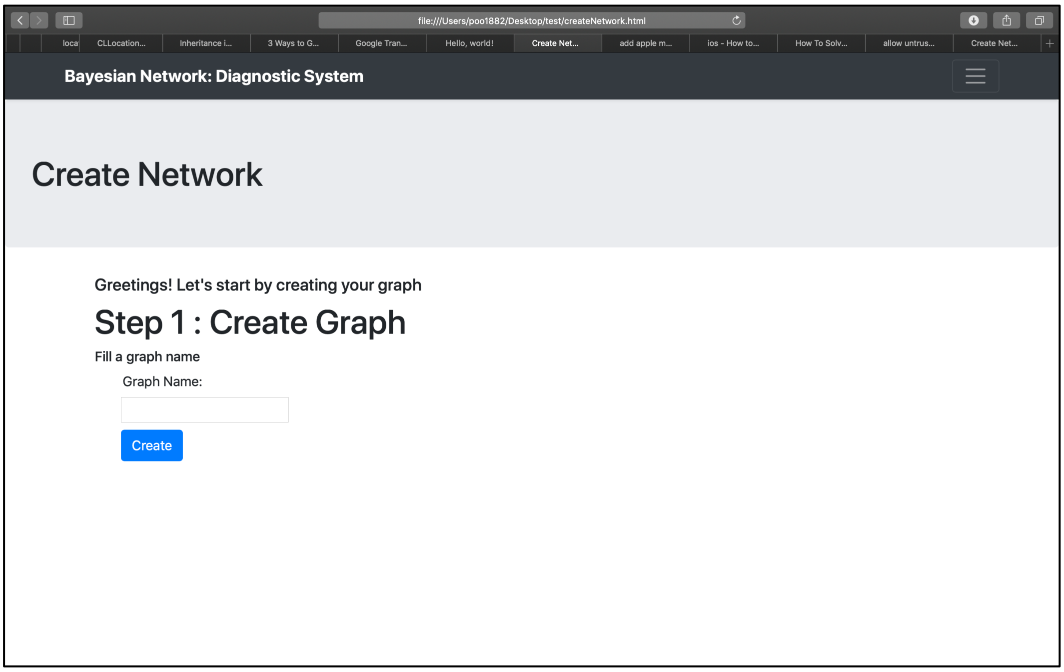
Design and Working Process

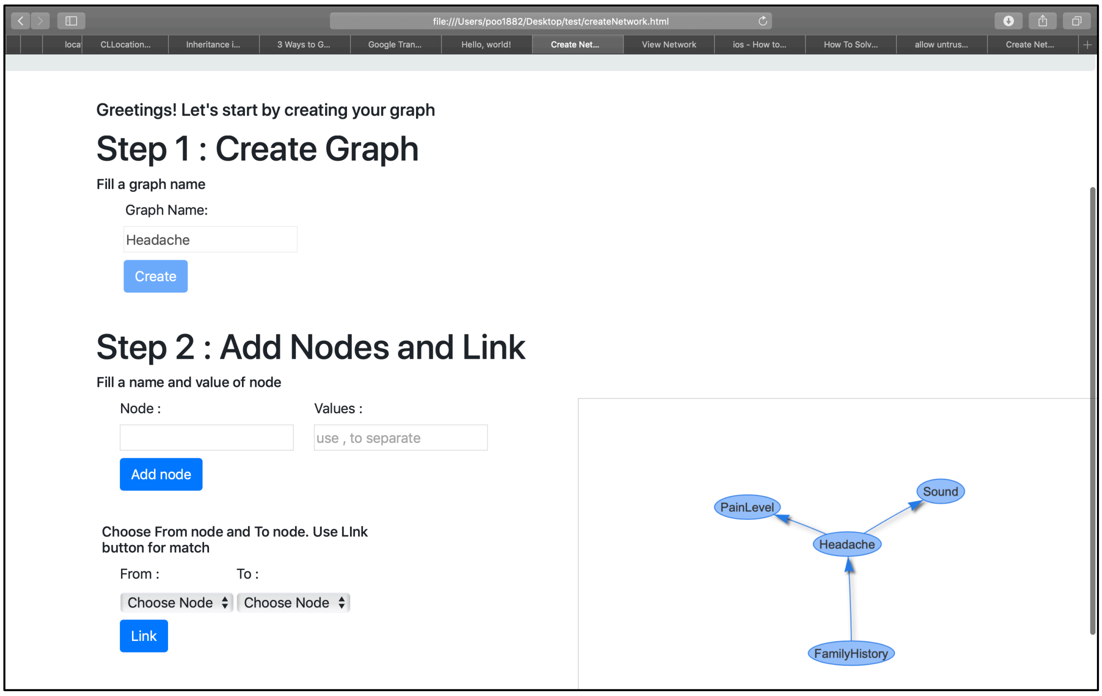
We decided to use textfield to store information about the node, edge, and probability. However, to set the conditional probability at each node, user needs to finish adding nodes and edges first. After that, they will be able to select the node they want to set. The textfield will be showing NULL value if it is the first time of setting. If user aims to modify the existing probability value, the textfield will show the current value instead.

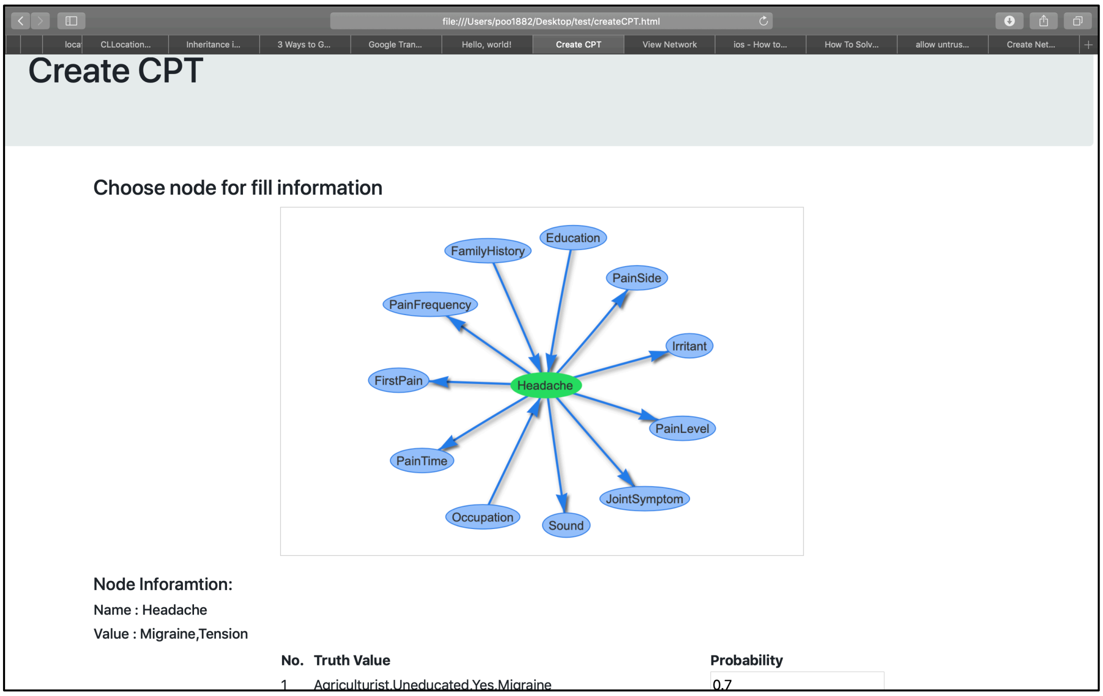
When pressing add button, the information about node, or edge will be stored in the database. For setting probability, user needs to press Get button to request the current value first.

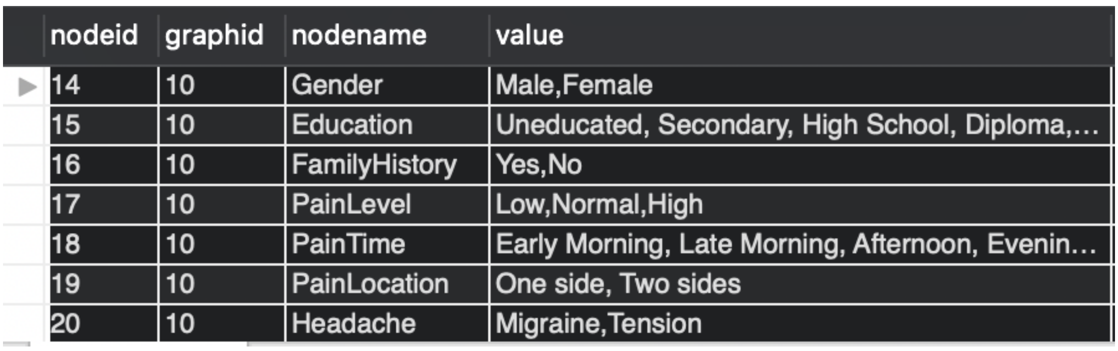
Test Cases

Case 1: Construct the new Bayesian network









nodeid	graphid	nodename	value
14	10	Gender	Male,Female
15	10	Education	Uneducated, Secondary, High School, Diploma,…
16	10	FamilyHistory	Yes,No
17	10	PainLevel	Low,Normal,High
18	10	PainTime	Early Morning, Late Morning, Afternoon, Evenin…
19	10	PainLocation	One side, Two sides
20	10	Headache	Migraine,Tension

A picture showing rows of node in the Nodes table

edgeid	graphid
11	10
12	10
13	10
14	10
15	10
16	10

A picture showing rows of edge in the Edges table

edgeid	nodeid	position
11	14	parent
12	15	parent
17	15	parent
13	16	parent
14	17	child
15	18	child
16	19	child
11	20	child
12	20	child
13	20	child
14	20	parent
15	20	parent

A picture showing rows of detail of each edge in the EdgeLines table

tableid	truthvalue	nodeid	probability
9735	Male	14	0.3
9736	Female	14	0.7
9737	Uneducated	15	0.1
9738	Secondary	15	0.2
9739	HighSchool	15	0.15
9740	Diploma	15	0.4

A picture showing rows of conditional probability table in the Tables table

Input Information about the network including name, nodes, edges, probability

Output The constructed Bayesian network will be displayed.

Result

After the node information are filled in, they will be sent to the database. User will be able to identify parent and child, and they will be stored as a relationship between two nodes in the database. The conditional probability table of a selected node will be displayed as an editable textfield. User will be able to enter the new value if they want to. After adding nodes, edges, and probability, user can export the network as a .bif file as described in *Export System* below.

2.2) Export Module

Overview

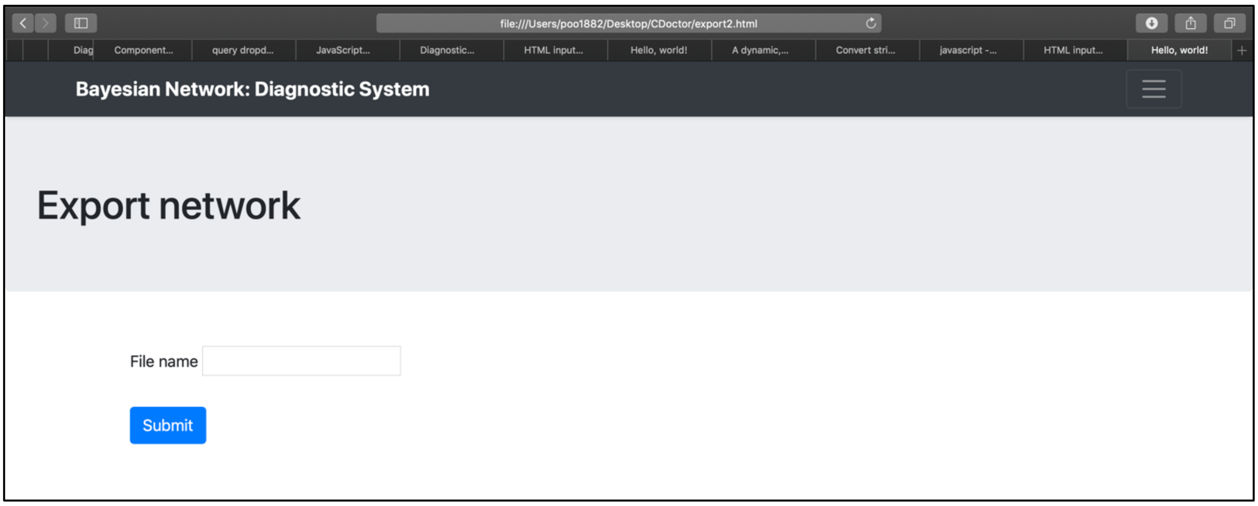
Export system aims to download a network as a .bif file to be used on another computer. This system will release a file to user and it allows user to determine the name of an output file.

Design and Working Process

Export system has only 1 input field for receiving the file name from user and submit button to confirm the user request. After the user click submit button, the file will be released along with the specified name.

Test Case

Case 1: Export Network



```
network "PredictHeadache" {
// written by aGrUM 0.13.6
}

variable Gender {
   type discrete[2] {Male, Female};
}

variable Education {
   type discrete[5] {Uneducated, Secondary, High School, Diploma, Undergraduate or Higher};
}

variable FamilyHistory {
   type discrete[2] {Yes, No};
}

variable PainLevel {
   type discrete[3] {Low, Normal, High};
}

variable PainTime {
   type discrete[5] {Early Morning, Late Morning, Afternoon, Evening, Night};
}

variable PainLocation {
   type discrete[2] {One side, Two sides};
}

variable Headache {
   type discrete[2] {Migraine, Tension};
}

probability (Gender) {
   default 0.3 0.7;
}
probability (Education) {
   default 0.1 0.2 0.15 0.4 0.15;
}
probability (FamilyHistory) {
   default 0.7 0.3;
}
```

A picture showing the network file generated by Configuration System (as .bif)

Input The name of an output file

Output The file is released with the specified name.

Result

After user fills in the name of an input file and click submit, the system will release a downloaded file to user as .bif format.

2.3) Import Module

<u>Overview</u>

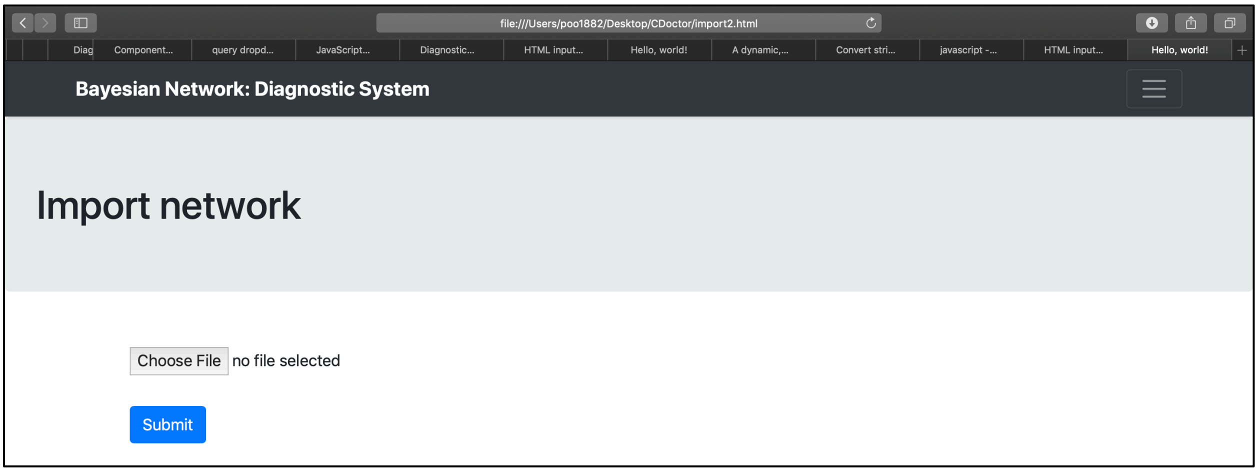
Import system aims to send a network as a .bif file to store on database. This system will move the upload file to same directory with modal after that read the file for store to database.

<u>Design and Working Process</u>

Import system has input field for receiving the file from user and submit button to confirm the user request. After the user click submit button, the file will move to model directory and store in database.

<u>Test Case</u>

Case 1: Import Network



Input The network file
Output The network was stored in database
Result

After user uploads a file and click submit, the system will move a file to model directory and store in the database.

2.4) Display Network Module

<u>Overview</u>

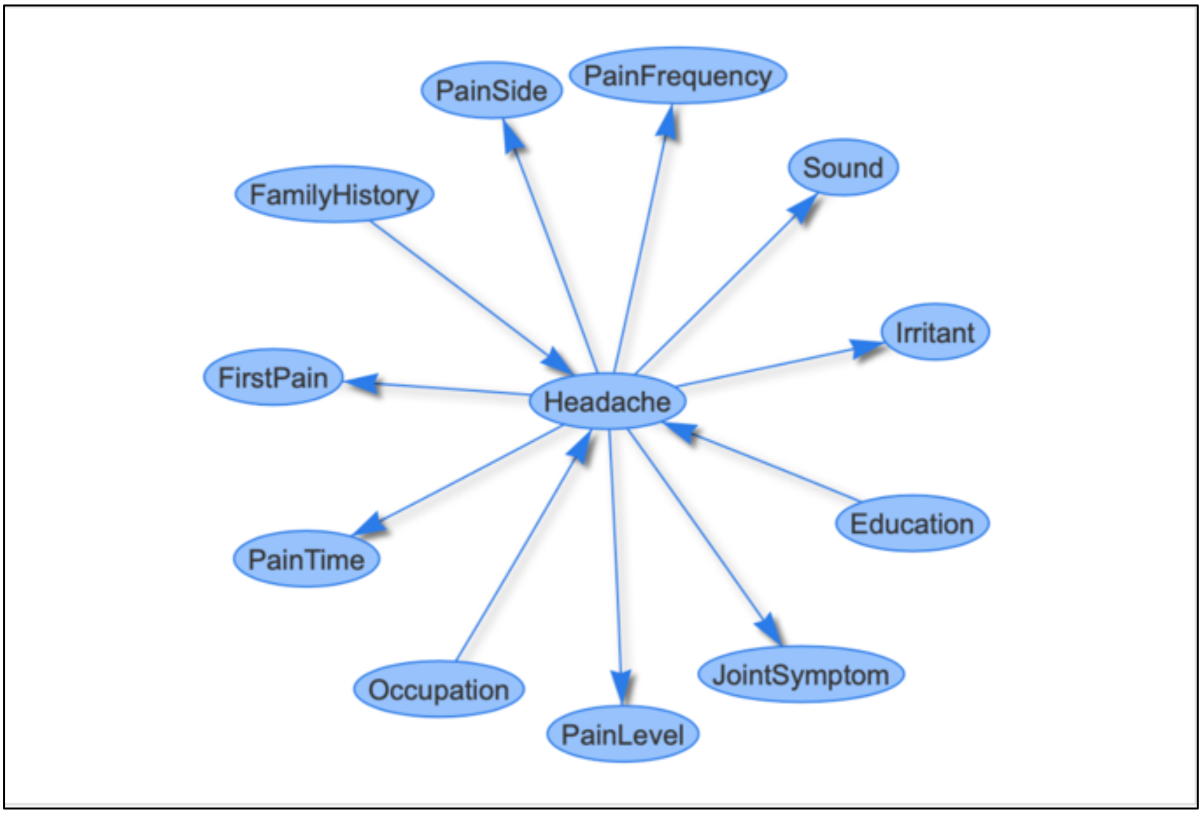
The display network module aims to show overall picture of current Bayesian network. It allows user to select a node in the network. After the node is selected, the information about that node will be shown.

<u>Design and Working Process</u>

The middle of the page shows picture of current network. The bottom part is responsible for showing a node's information. This is done by letting user click an interested node. If there is a click event happens on any node, that node will be colored as green. Also, this module allows user to move the network in the box.

<u>Test Case</u>

Case 1: Show Network



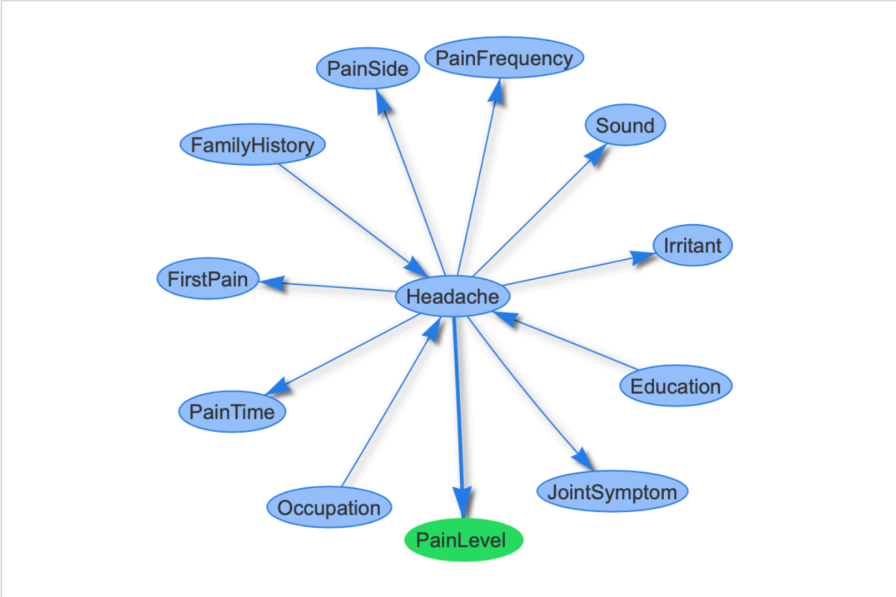
Input -

Output Overall picture of the network is shown.

Result

For the first time when open the page, it will show only the network.

Case 2: Select a node



Node Inforamtion:

Name : PainLevel

Value : Low,Normal,High

No.	Truth Value	Probability
1	Migraine,Low	0.75
2	Migraine,Normal	0.21
3	Migraine,High	0.04
4	Tension,Low	0.75
5	Tension,Normal	0.21

Input Click on a node

Output The node's detail is shown.

Result

When click on the node, the color of selected node will change to be green and the information will be shown at the bottom of the graph.

2.5) Feed Data Module

Overview

This allows user to generate conditional probability tables at each node from a given .json data file.

Design and Working Process

User will be able to select a data file (as .json) to feed into the network so that the conditional probability tables inside the network will be automatically generated based on the given data. The .json data file needs to contains columns existing in the network. Also, it has to start with an object of properties, following by the dataset.

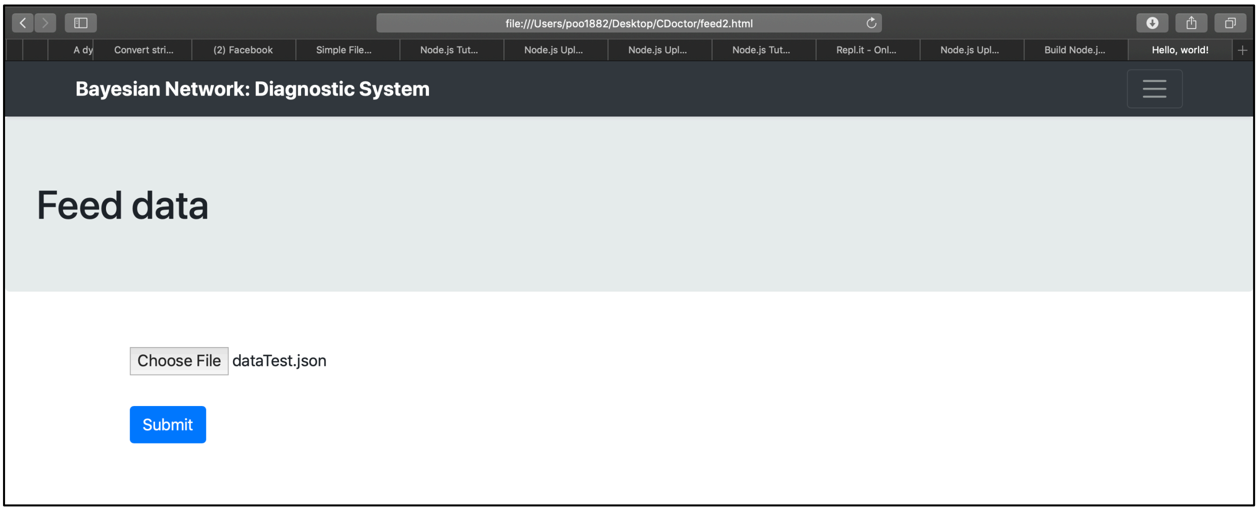
<u>Test Case</u>

Case 1: Input .json data into the system


```json
{
 "properties":{
    "Occupation": "",
    "Education": "",
    "FamilyHistory": "",
    "Sound": "",
    "PainLevel": "",
    "PainTime": "",
    "PainSide": "",
    "PainFrequency": "",
    "FirstPain": "",
    "Irritant": "",
    "Headache": "",
    "JointSymptom": ""
    },

"data":{
    "1":{
        "Occupation": "Agriculturist",
        "Education": "Primary",
        "FamilyHistory": "No",
        "Sound": "JeedJeed",
        "PainLevel": "Normal",
        "PainTime": "LateMorning",
        "PainSide": "TwoSides",
        "PainFrequency": "OneToThree",
        "FirstPain": "Unknown",
        "Irritant": "NotEnoughSleeping",
        "Headache": "Migraine",
        "JointSymptom": "Squeamish"
        },
    "2":{
        "Occupation": "GovernmentOfficer",
```


An example of .json data file



tableid	truthvalue	probability	nodeid
220	Tension,Temples	0.01	9
221	Tension,Eyes	0.09	9
222	Tension,MiddleHead	0.33	9
223	Migraine,Stress	0.07	10
224	Migraine,NotEnoughSleeping	0.04	10
225	Migraine,BadSmell	0.04	10
226	Migraine,Traveling	0.12	10
227	Migraine,HotWeather	0.02	10
228	Migraine,Food	0.19	10
229	Migraine,Alcohol	0.53	10

Input dataTest.json

Output The conditional probability tables in the network is generated.

Result

The data at each row will be calculated as a conditional probability table. Some set of data occurring more frequently will gain more ratio and become more probability also. The conditional probability tables will be stored inside the database. User will be able see it through web interface, or even export it as a.bif file format.

Chapter 5
Conclusion

5.1) Summarization

Our project goal is to implement a diagnostic system by applying Bayesian network theorem. The scope of work includes develop a Bayesian network model predicting associated diseases from given symptoms and characteristics of users, and suggest for basic treatment. Also, we aim to create a system where users are able to create or manage Bayesian network on their own for further development if they want the network to cover more diseases.

5.2) Completion Status

Deliverable Type	**Deliverable**	**Status**
Diagnostic System	Make inference and serve as a web service	Completed
	Identifying cause and treatment	Completed
Support System	Feed training data to automatically construct conditional probability tables	Completed
	Create the new network and managing nodes, edges, probability tables of the existing network	Completed
	Import the network from .bif file, and store in database	Completed
	Export the network as .bif file from the database	Completed
Web Application	Front-end development	Partially Completed
	Back-end development	Completed

- We have created a diagnostic system where users are divided into 2 types: general users and administrators. General users will be able to ask for diagnosis from the existing network, and also ask for general description about a disease. Administrators will be able to modify the network. This includes import network from .bif file, create a network via user interface, view the structure of the current network, edit node, edges, and conditional probabilities, export the network to be used on another computer, and also feed the data to construct the conditional probability tables.

5.3) Issues Encountered and Solved

5.3.1) Since a statistical data of symptoms, characteristics, and associated diseases explained as a probability is not provided publicly, our system covers only two types of headache so far. However, user can add more diseases to the system via support system.

5.3.2) Bayesian network provided on cloud server as a web service is not updated real time when changes happen to local Bayesian network. We still have to manually push the new version of the network on cloud server to apply changes.

5.3.3) There is still problem with the user experience since the user interface is not intuitive enough. There are some parts where user needs explanation what to do. So, we tried to redesign our website and test with new users and ask for suggestion, and it becomes the latest version so far.

5.4) What We Had Learned

By working on Bayesian network diagnostic system, we have learned several things, and also develop ourselves in both hard and soft skills.

Starting with hard skills. Since this project requires both statistical and mathematics knowledge, we have got to study about the theory of Bayesian network, and how it can be used to make inference. As Bayesian network contains nodes, edges, and conditional probability, we have to define all of them according to our system. We need to specify relationship of factors that we got from Chiangmai Neurological Hospital. In computer engineering aspect, we need to design the database schema for handling all features in the system. Especially, the integration part, because our project needs to serve as a web service for the chatbot. We have learned the way of building API and deploy it to a cloud server so that it is easier for the chatbot to test.

The next skills we also acquired are soft skills. Since we do this project in pair, a lot of appointments need to be made. We both have to manage time to find the suitable time slot of each other. We also have to discuss several things to finalize the final design of implementation so that both of us will understand the same thing. This helps us to improve communication skill since there were some points that we understand differently. Lastly, we have improved presentation skill because we have to prepare ourselves for the project presentation.

5.5) Future Extending Suggestion

5.5.1) The diagnosis result should cover more diseases so that it would be more useful to users.

5.5.2) The diagnosis result should provide an urgency level from given symptoms and characteristics of user.

5.5.3) In case there are many diseases getting similar probability, the system should provide the list of those diseases, and let user chooses the disease they are interested in.

References

1. An Introduction to Bayesian Network Theory And Usage [Online]. Available: http://publications.idiap.ch/downloads/reports/2000/rr00-03.pdf Retrieved 2018, September 10
2. Bayesian networks - an introduction [Online]. Available: https://www.bayesserver.com/docs/introduction/bayesian-networks Retrieved 2018, September 10
3. bayesian network vs bayes classifier [Online]. Available: https://stackoverflow.com/questions/12298150/bayesian-network-vs-bayes-classifier Retrieved 2018, September 10
4. Medical diagnosis [Online]. Available: https://en.wikipedia.org/wiki/Medical_diagnosis Retrieved 2018, September 10
5. Zakiyeh Shojaei Estabragh • Mohammad Mansour Riahi Kashani •Farnaz Jeddi Moghaddam • Simin Sari • Zahra Taherifar • Shima Moradi Moosavy • Koosha Sadeghi Oskooyee (2013) Bayesian network modeling for diagnosis of social anxiety using some cognitive-behavioral factors. Islamic Azad University
6. Agnieszka Onisko • Marek J. Druzdzel •Hanna Wasylyk (1999) A Bayesian Network Model for Diagnosis of Liver Disorders. Center for Biomedical Informatics, University of Pittsburgh
7. Python (programming language) [Online]. Available: https://en.wikipedia.org/wiki/Python_(programming_language) Retrieved 2018, September 10
8. SQL [Online]. Available: https://en.wikipedia.org/wiki/SQL Retrieved 2018, September 10
9. Hypertext Transfer Protocol [Online]. Available: https://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol Retrieved 2018, September 10
10. Representational state transfer [Online]. Available: https://en.wikipedia.org/wiki/Representational_state_transfer Retrieved 2018, September 10
11. Bayesian network representation of the naive Bayes classifier [Online]. Available: http://rspa.royalsocietypublishing.org/content/465/2109/2927 Retrieved 2018, September 10
12. aGrUM/pyAgrum[Online]. Available: http://agrum.gitlab.io/ Retrieved 2018, November 11
13. Dendral [Online]. Available: https://en.wikipedia.org/wiki/Dendral Retrieved 2018, November 20
14. Mycin [Online]. Available: https://en.wikipedia.org/wiki/Mycin Retrieved 2018, November 22
15. Artificial intelligence in healthcare [Online]. Available: https://en.wikipedia.org/wiki/Artificial_intelligence_in_healthcare Retrieved 2018, December 1

16. Chapter 1 General Information [Online]. Available: https://dev.mysql.com/doc/workbench/en/wb-intro.html Retrieved 2019, March 14
17. Flask (web framework) [Online]. Available: https://en.wikipedia.org/wiki/Flask_(web_framework) Retrieved 2019, March 14
18. Microframework [Online]. Available: https://en.wikipedia.org/wiki/Microframework Retrieved 2019, May 5
19. The Interchange Format for Bayesian Networks [Online]. Available: http://www.cs.washington.edu/dm/vfml/appendixes/bif.htm Retrieved 2019, May 5
20. Node.js [Online]. Available: https://en.wikipedia.org/wiki/Node.js Retrieved 2019, May 5