Assignment 1: Application of Bayesian Networks

1. Introduction

Tuberculosis (TB) is an infectious disease caused by the Mycobacterium tuberculosis bacteria. TB can either be latent where no symptoms are shown or active where they are prominent. It primarily affects the lungs and spreads airborne through oral emissions from people infected with active Tuberculosis. TB is the second leading cause of death from infectious disease after COVID-19 with prevention methods involving early detection and screening (WHO, 2023).

Our network aims to determine the probability of having Tuberculosis infection or disease based on factors such as gender, age and Tuberculosis Prevention Treatment (TBPT) status. This information can encourage individuals to get tested or vaccinated and, therefore, increase the rate of early detection of TB and prevent the prevalence of the disease. Furthermore, a utility function, that quantifies the desirability of different outcomes, is used to evaluate the action of getting treatment based on the possibility of having TB disease

South Africa (SA) is one of the 30 high burden tuberculosis countries contributing 87% of the estimated incident TB cases worldwide, on its own SA accounts for 3% of cases globally (van Der Walt & Moyo, 2018, p.3) The user community will thus be South Africans with a focus on the rural communities where TB is known to be common.

2. Problem analysis

POTENTIAL FACTORS CONSIDERED

When looking at the factors influencing the possibility of having a TB infection there are 2 main factor categories: physical and location. We have decided to focus on physical aspects as the delineation of location can get overly complex for the scope/size of the network we want to construct.

We divided age into discrete sections from ages 0 - 14, 15 - 44, 45 - 64 and 65 +. Male and Female with no delineation for non-binary were chosen as the dataset we used did only include the first two represented by 0 (female) and 1 (male). This TB infection result dataset used was based on a UK research article, but we used the distribution of ages and gender to estimate the TB infection distribution for South Africa.

Furthermore there exists a TB preventive treatment, which can be a highly effective way to reduce the risk of developing active TB disease. However, it does not fully protect against the disease, and this is a factor that needs to be considered. Given that TB has a latent and active stage, the relationship between these two stages need to be treated. In addition, as the symptoms of TB disease include many, we have selected a range of them including both common and rare symptoms, and have operated under the assumptions that an individual can only experience one of these symptoms.

There exist multiple tests for TB, including a blood test, skin test and a X-ray test that distinguish the presence of only the infection and the disease. However, to avoid a high degree of complexity, we have decided to create a generalized node for TB testing. Furthermore, the rates of false-positives and false-negatives are determined by a combination of obtained statistics about testing and sensible assumptions by the group.

There are many other factors that affect the probability of having TB which are not taken into account in this model. For example, an weakened immunity system through HIV or diabetes can highly increase the probability of getting the TB infection and developing the disease (CDC, 2018). However, the inclusion of all factors makes it difficult to create a manageable network.

DATA SETS AND EXPERT KNOWLEDGE

- We assume gender is 50/50

Gender	
0	1
0.5000	0.5000

- According to WHO, 57% of the population have done Tuberculosis Prevention Treatment before (WHO, 2023)

TBPT	
0	1
0.4300	0.5700

- We got the population distribution from the Statista (Statista, 2022)

age			
age_0_To_14	age_15_To_44	age_45_To_64	age_65_Plus
0.2800	0.4900	0.1700	0.0600

- Factors such as age, gender, and TB prevention treatment (TBPT) can impact the likelihood of an individual contracting a latent TB infection. Generally, people aged between 14 to 45 have a higher risk of getting infected with TB (Schaaf, 2010). Men also tend to have a higher chance of contracting TB compared to women. Additionally, those who have never undergone TBPT are more susceptible to TB infection (WHO, 2023). It is worth noting that, on average, around 80% of South Africans have experienced a TB infection in their lifetime (TBFacts.org, n.d.). With the above facts, we formulated the probability table.

			latent	
TBPT	Gender	age	0	1
0	0	age_0_To_14	0.2000	0.8000
		age_15_To_44	0.0500	0.9500
		age_45_To_64	0.1500	0.8500
		age_65_Plus	0.1200	0.8800
	1	age_0_To_14	0.2500	0.7500
		age_15_To_44	0.0200	0.9800
		age_45_To_64	0.1000	0.9000
		age_65_Plus	0.1000	0.9000
1	0	age_0_To_14	0.5500	0.4500
		age_15_To_44	0.1000	0.9000
		age_45_To_64	0.3000	0.7000
		age_65_Plus	0.2700	0.7300
	1	age_0_To_14	0.5500	0.4500
		age_15_To_44	0.0700	0.9300
		age_45_To_64	0.3000	0.7000
		age_65_Plus	0.2000	0.8000

- We learnt that only around 10% of latent TB infections will become active TB disease from the CDC (CDC, 2018)

	TB	
latent	0	1
0	1.0000	0.0000
1	0.9000	0.1000

- We learnt that there is about 0.1% to 2.3% rate of getting false positive/false negative results. Therefore, to avoid complexity, we decided to deduct a probability of 0.01 from the probability of 1 and add 0.01 to the probability of 0, which is why we don't have a probability of 1 if someone has TB and gets diagnosed with TB. (Farhat et al, 2006). If a person did not get tested, then we say that they are not diagnosed with TB therefore have a probability of 1.

			Diagnosed	
testing	TB	latent	0	1
0	0	0	1.0000	0.0000
		1	1.0000	0.0000
	1	0	1.0000	0.0000
		1	1.0000	0.0000
1	0	0	0.9900	0.0100
		1	0.0100	0.9900
	1	0	0.0100	0.9900
		1	0.0100	0.9900

- Common symptoms of a tuberculosis (TB) infection include cough, chest pain, fever, and night sweats. In individuals without TB, the probabilities of experiencing no symptoms, cough, chest pain, fever, and night sweats are 0.28, 0.03, 0.20, 0.08, and 0.41, respectively (HHP. 2020, Summer. 2023, CDC. 2018, Ford. 2006). Conversely, for those with TB, the probabilities of experiencing no symptoms, cough, chest pain, fever, and night sweats are 0.00, 0.35, 0.15, 0.23, and 0.27, respectively (Menberu, 2016. Miller, 2000. Lee, 2019).

	Symptom				
TB	None	Cough	Chest_Pain	Fever	Sweating_at_night
0	0.2800	0.0300	0.2000	0.0800	0.4100
1	0.0000	0.3500	0.1500	0.2300	0.2700

- The optimal utility of 70 occurs when an individual has no TB disease and receives no treatment. Conversely, the worst-case scenario involves contracting TB without receiving treatment, resulting in a utility of -70. If an individual receives treatment for TB and is TB active, the utility is 30. However, if treatment is administered without the presence of TB infection, the utility is -10, as there may be adverse side effects from the treatment.

		utility
treatment	TB	0
0	0	50.0000
	1	-70.0000
1	0	-10.0000
	1	30.0000

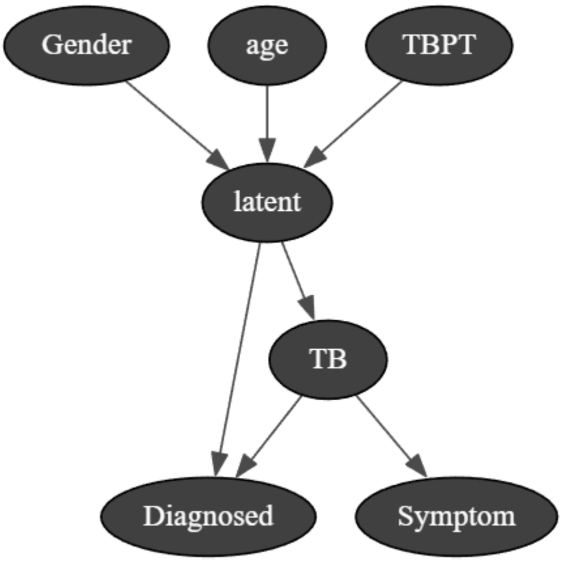
3. Decision Network model

The incidence of tuberculosis (TB) infection can differ among various groups, such as gender, age, and previous administration of Tuberculosis Prevention Treatment (TBPT). Latent TB infection has the potential to progress into active TB disease, which will have some common symptoms irrespective of the individual. However, to be diagnosed with TB (disease or infection), the person will have to perform some TB testing (an action person will have a make). Both active TB disease and latent TB infection will return positive in the diagnostic test. Only individuals who receive a positive for TB diagnosis are eligible for TB treatment, however treatment would also be an action by the person because it is their choice to do it or not . The utility of treatment is measured by considering whether the person has TB or not, and whether they took the treatment or not.

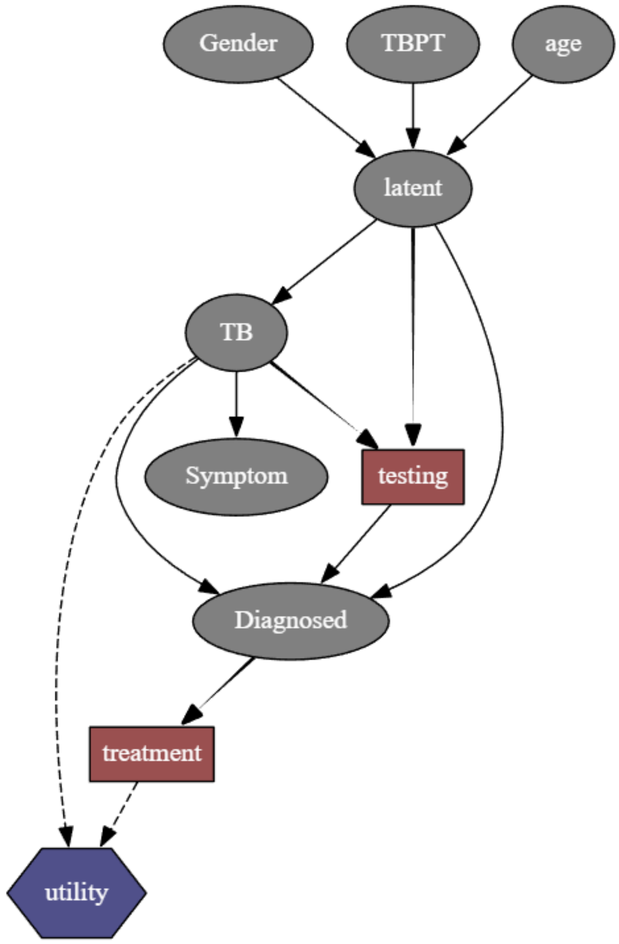
Our program generates an inference diagram from our decision diagram, allowing us to create scenarios and set some designated observed variables, the inference diagram generated will contain the unobserved variables and their associated probabilities. Additionally, the program calculates the expected utility of each scenario. The expected utility calculation formula is as below:

$$\mathrm{EU}(a,e) = \mathrm{E}_{P(S|e)}[U(s,a)] = \sum_{s} P(s|e)U(s,a)$$

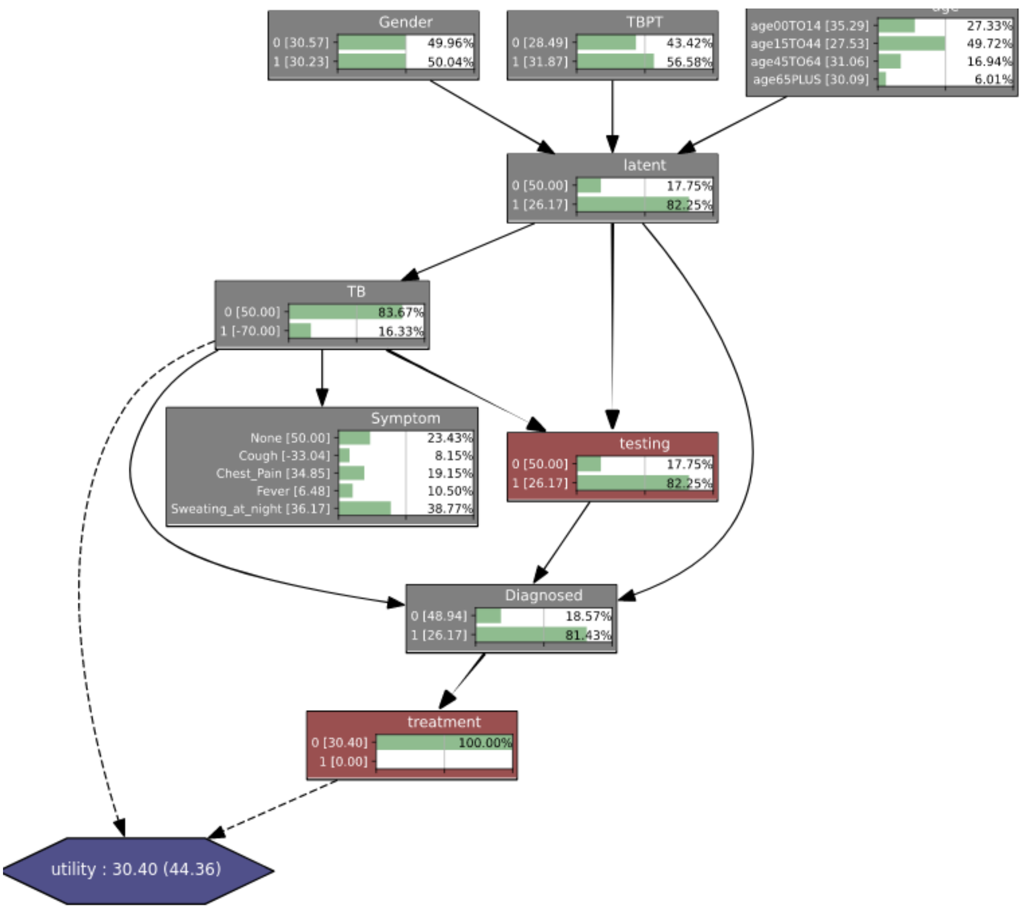
The diagram below shows our Bayesian Network:



The diagram below shows our decision network

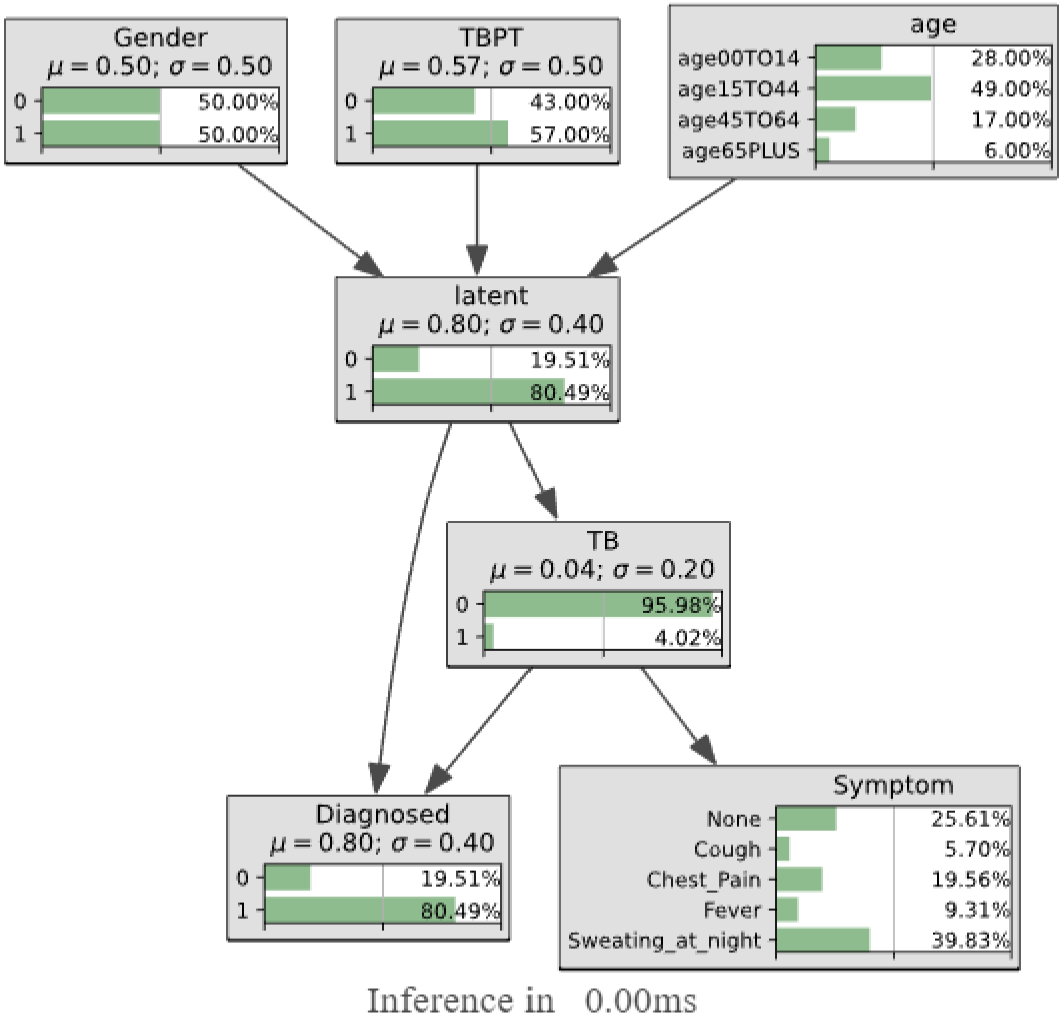


The model below shows the inference tables of our decision network:



MEU 30.40 (stdev=44.36)
Inference in 0.00ms

The model below shows the inference tables of our bayesian network:



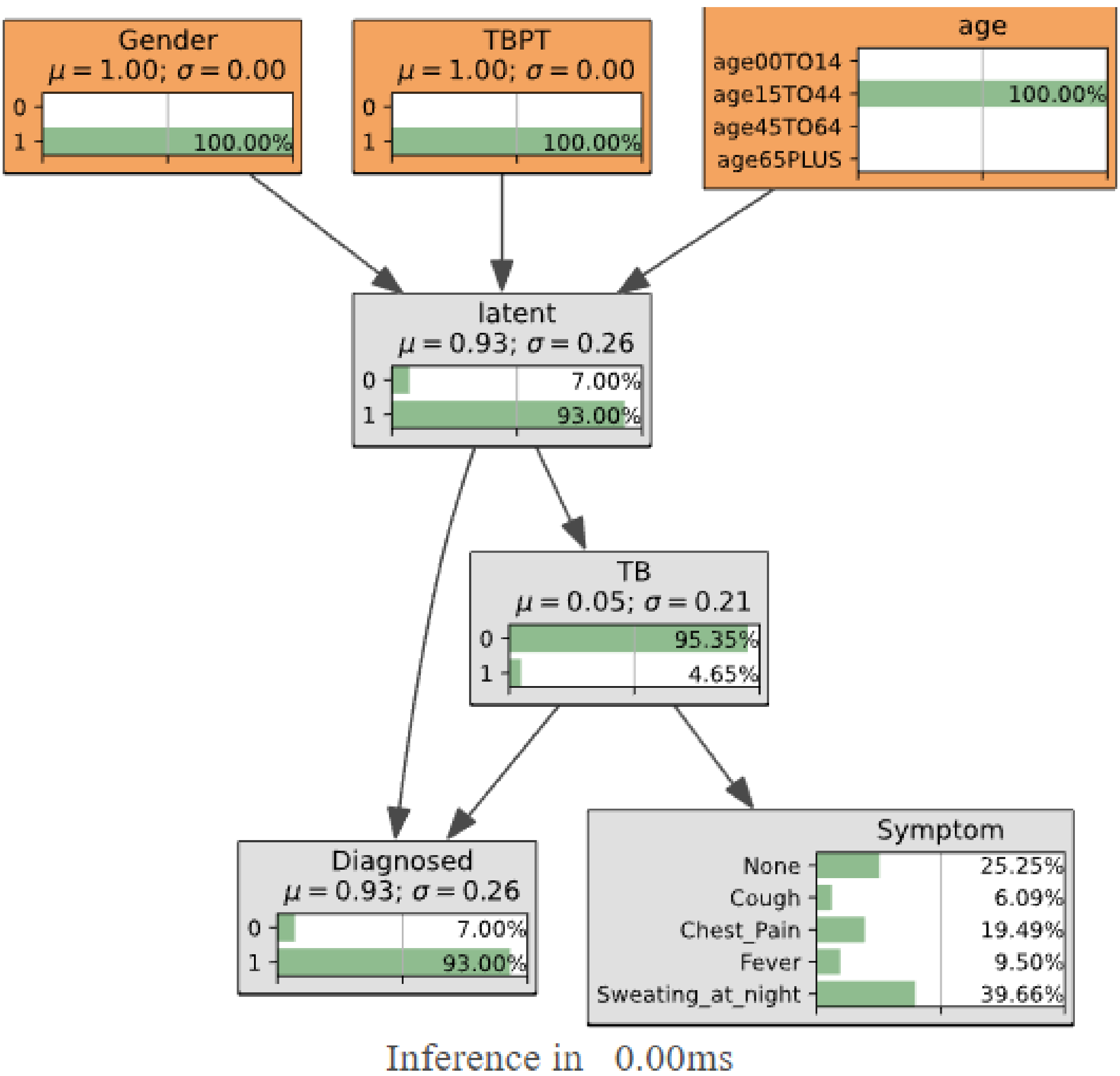
4. Model testing and evaluation

To test our Bayesian and Inference diagram, we will set various observed variables. We have four user personas, that represent individuals in the chosen user community. Each of the cases give insight into how the network could prove useful.

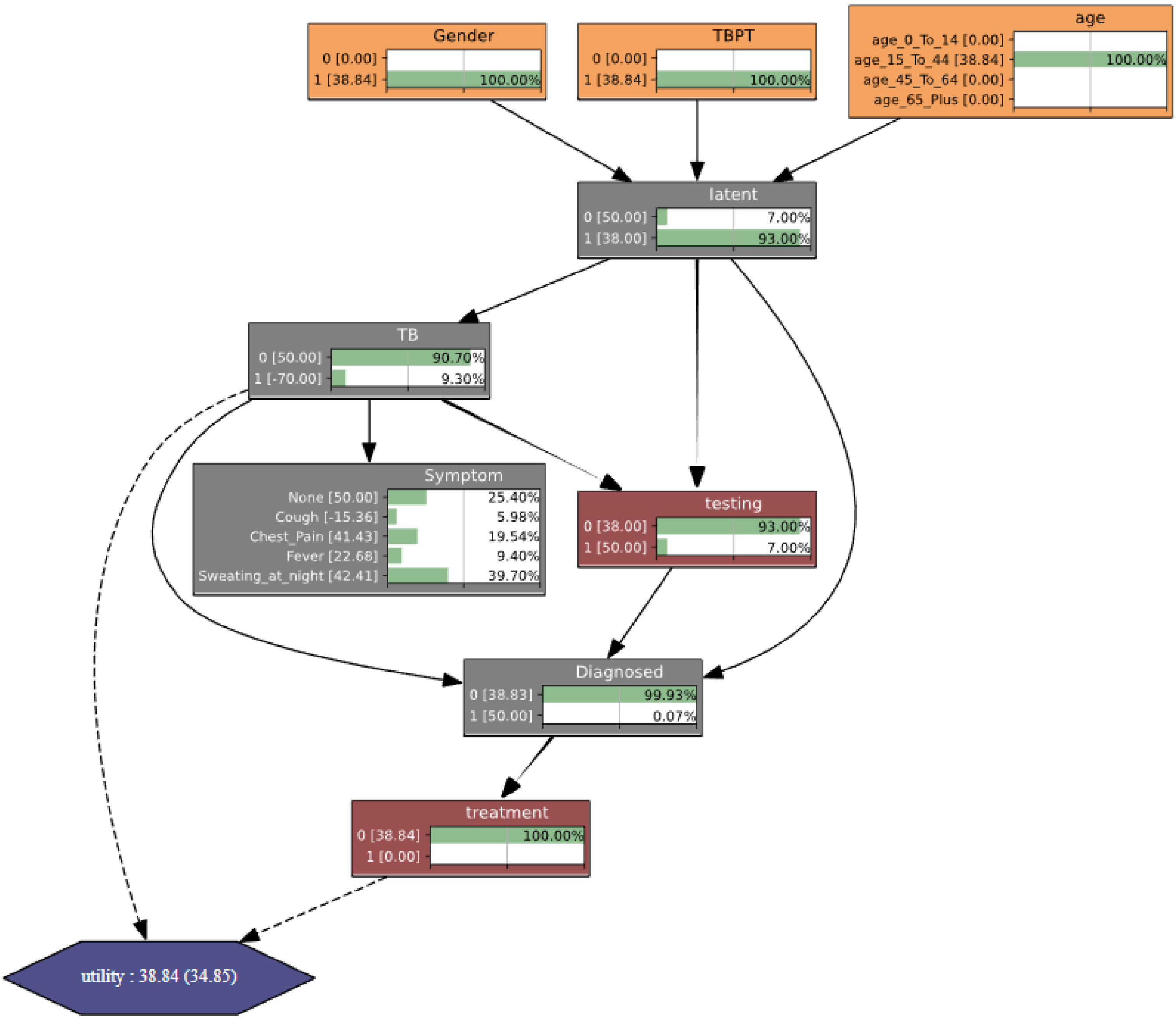
Scenario 1: Jillian, Male, 23

Jilian, a 23 year old male, was worried about his chances of getting TB. He has received TBPT before and wants to know the probability of getting TB and what actions he should take.

Bayesian Network Inference Diagram



Decision Network Inference Diagram

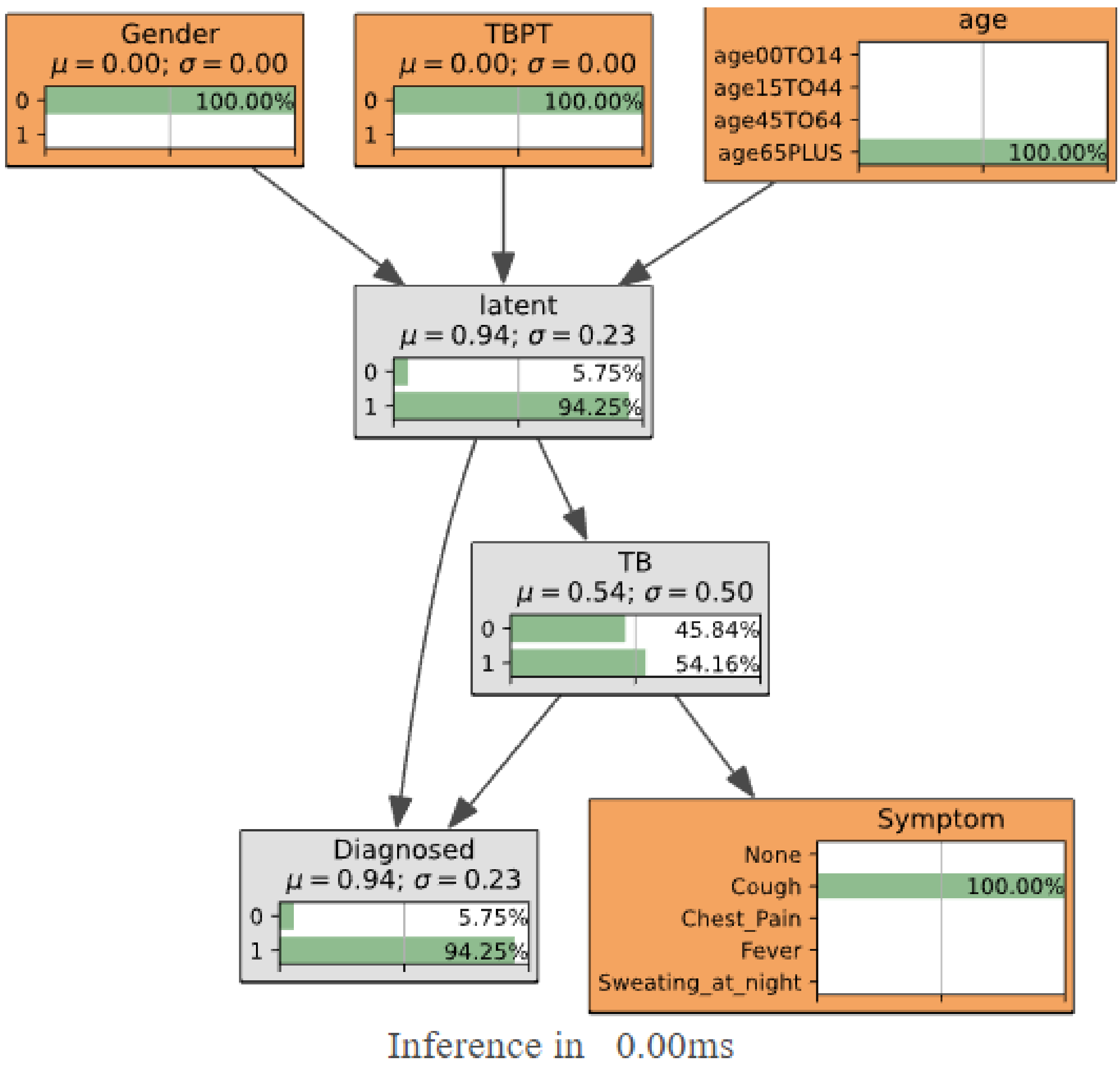


MEU 38.84 (stdev=34.85)
Inference in 1.00ms

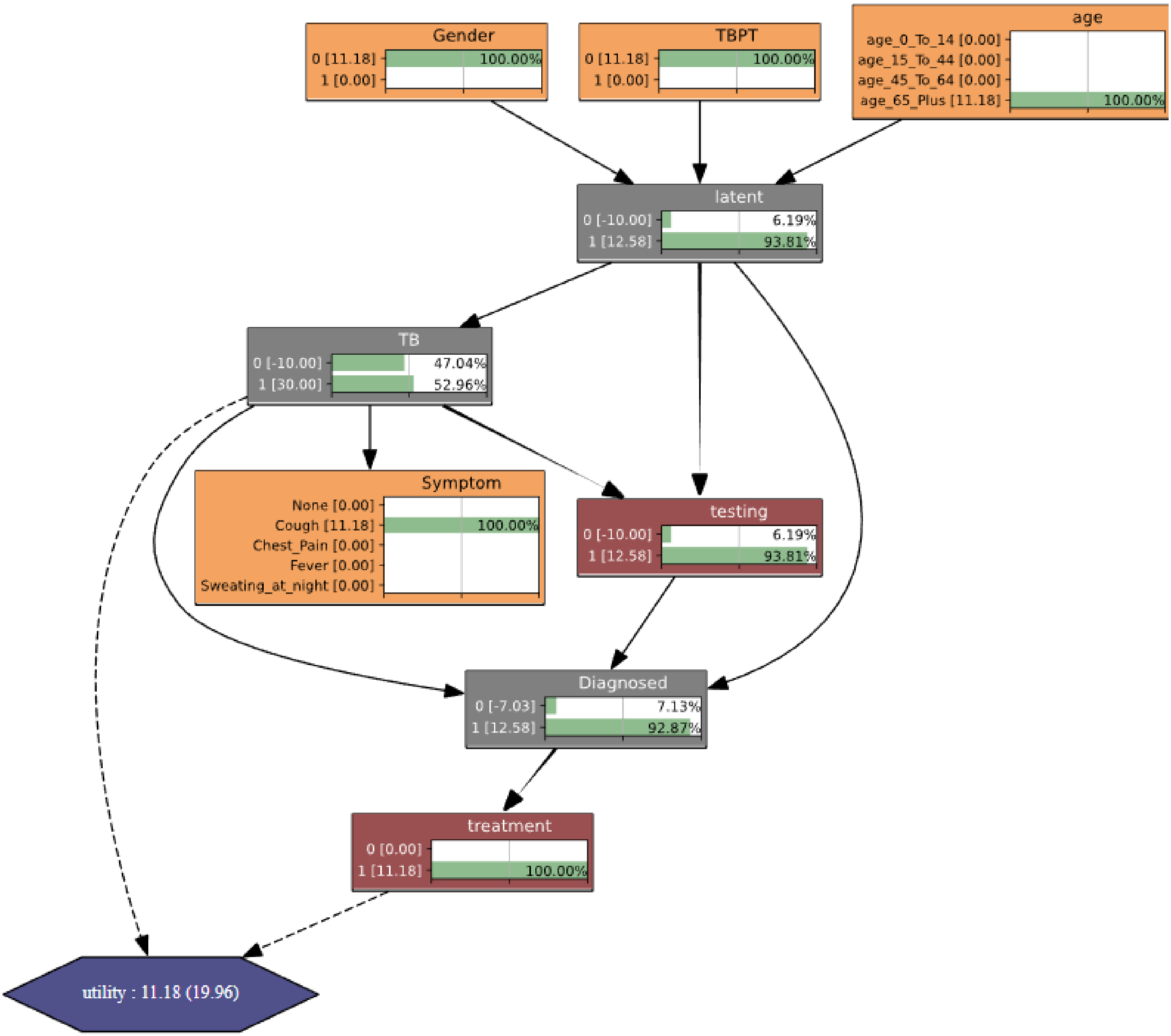
Scenario 2: Elizabeth, Female, 89

Elizabeth, an 89 year old lady, recently developed TB-like symptoms like coughing. She has not been through an TBPT and is worried about whether this TB-like symptom is TB. The network to calculate her probability of this coughing is TB or if she is just sick, we will also calculate her expected utility in this scenario.

Bayesian Network Inference Diagram



Decision Network Inference Diagram

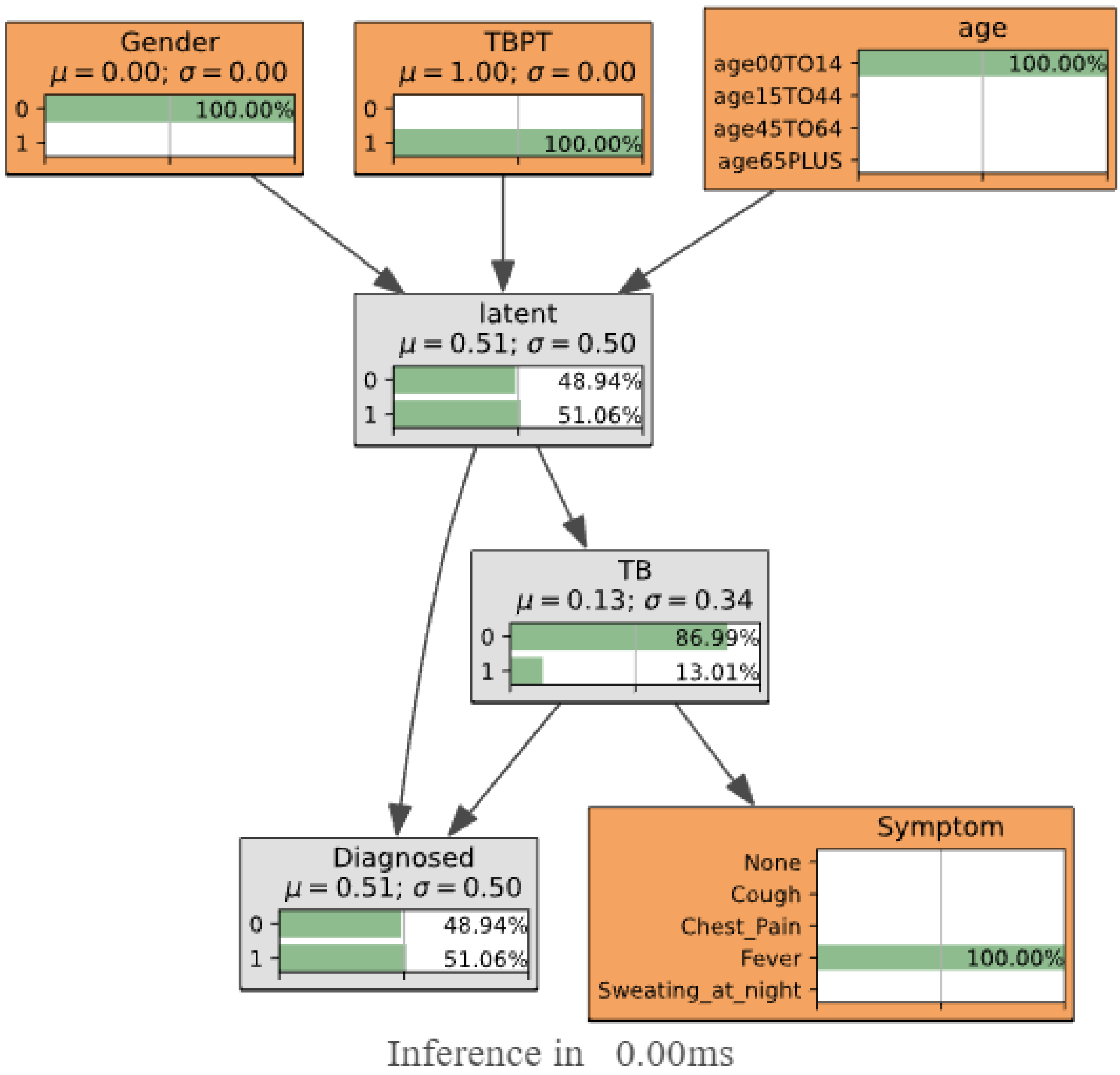


MEU 11.18 (stdev=19.96)
Inference in 1.00ms

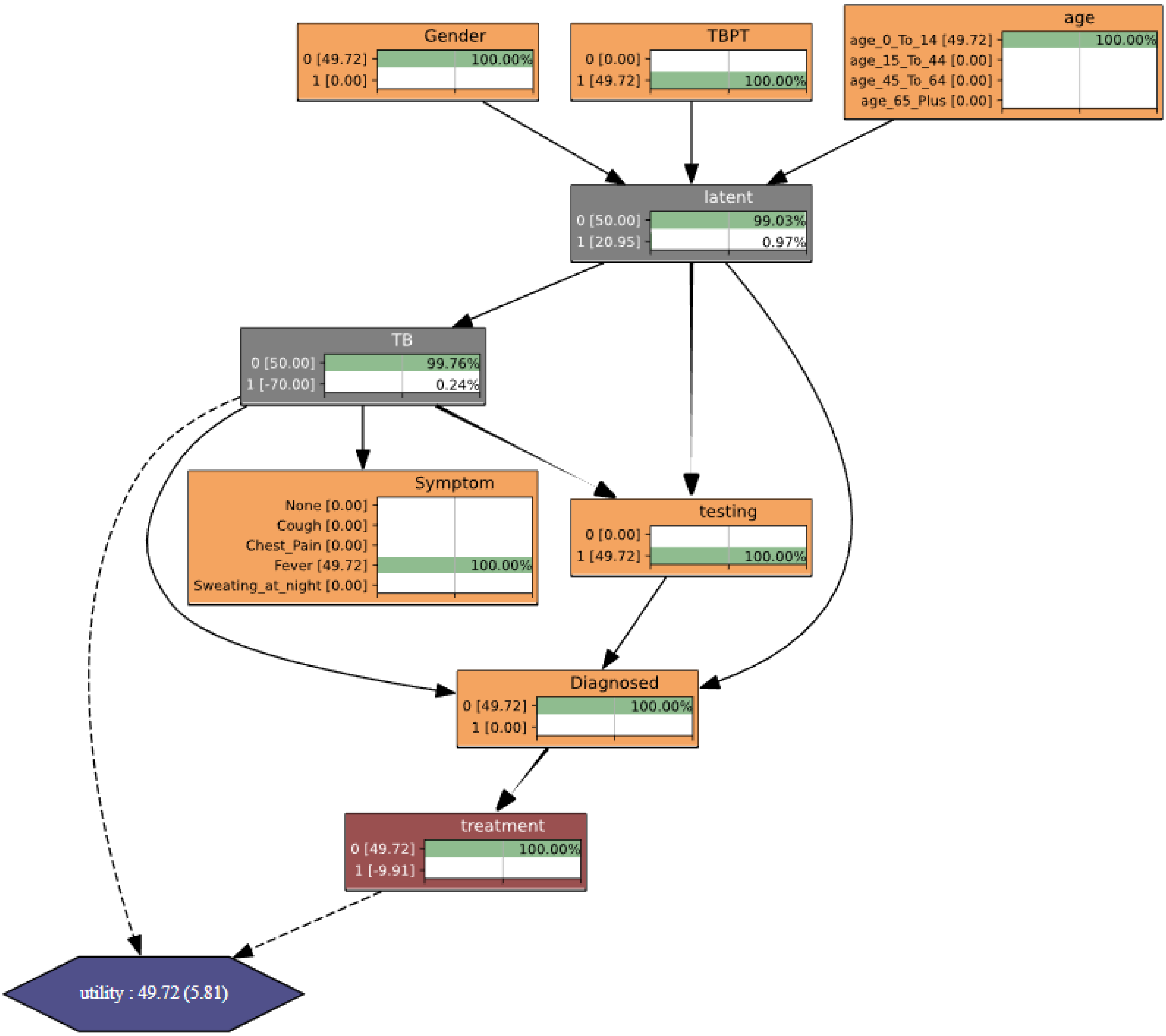
Scenario 3: Jenna, Female, 12

Jenna, is a 12 years old primary school girl, she has been through TBPT. However, she got a really bad fever recently and was hospitalized. She was tested for TB and received a negative result. However, her parents are wondering if they should get her TB treatment anyway and want to know if that is a good decision.

Bayesian Network Inference Diagram



Decision Network Inference Diagram

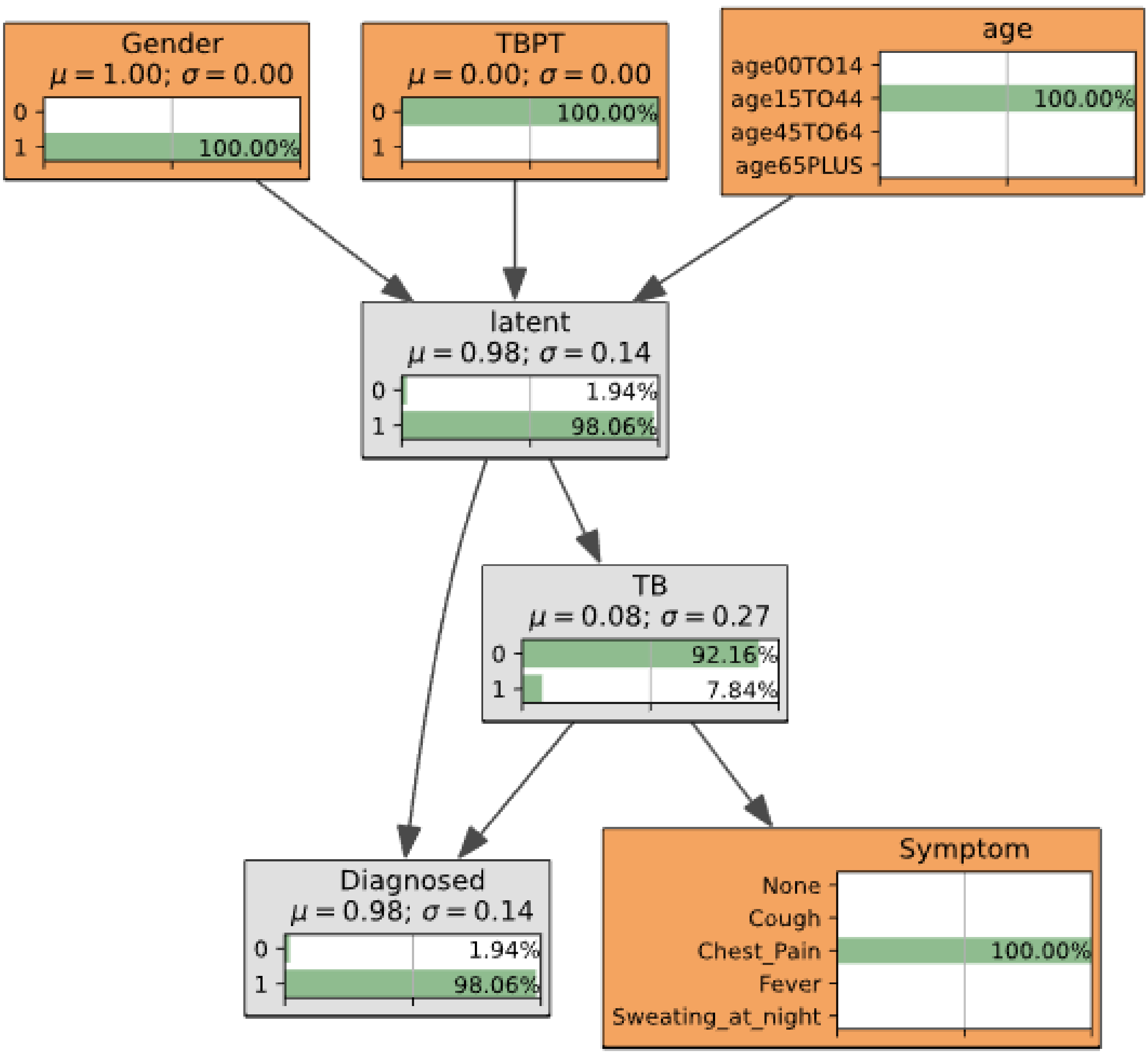


MEU 49.72 (stdev=5.81)
Inference in 1.00ms

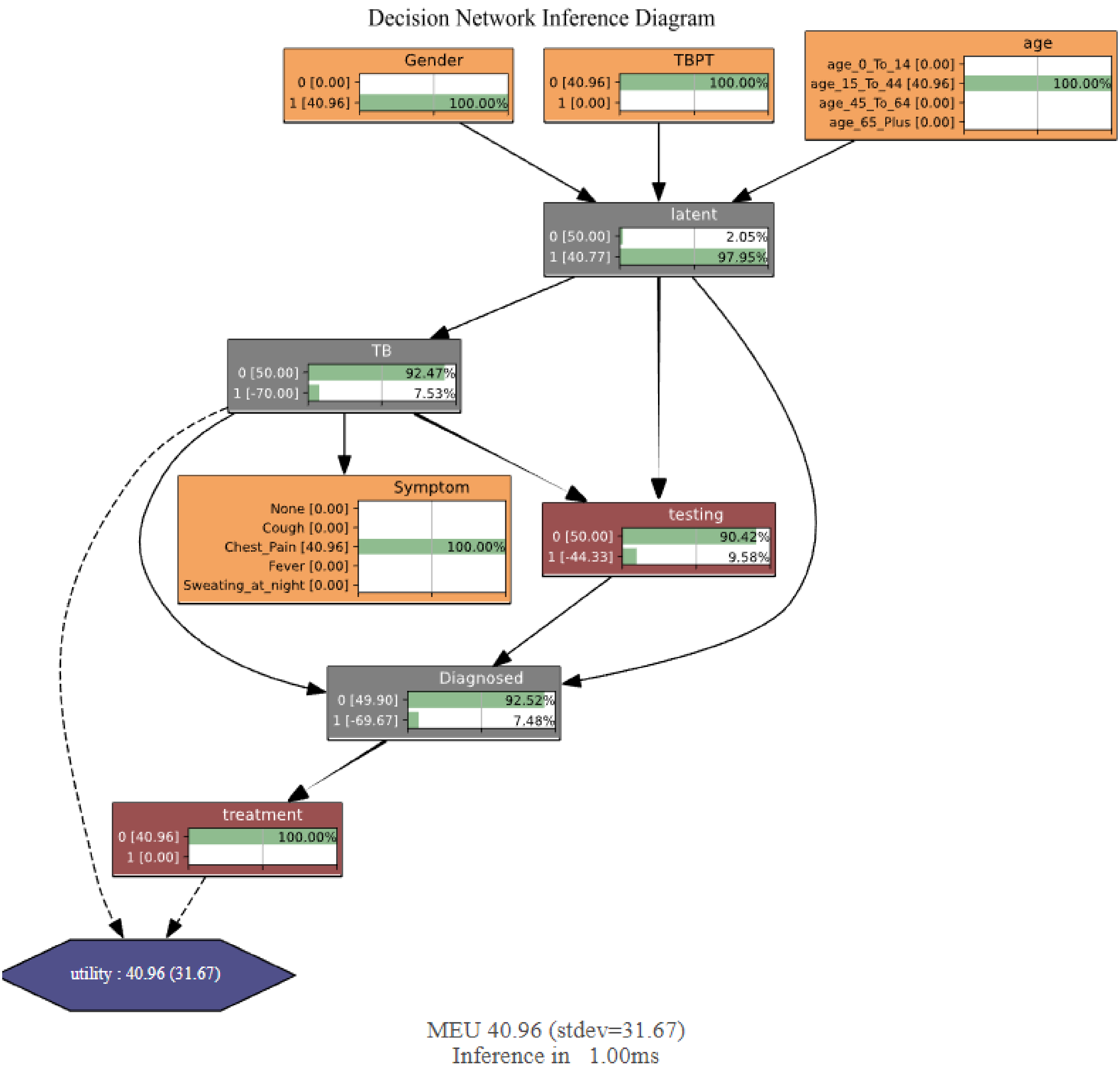
Scenario 4: Chad, Male, 21

Chad, a 21 years old male student studying his final year in UCT and never had TBPT before. He thinks he might have TB because of a really bad chest pain he got after attending a weekend gathering. Because of his young age and general good health, he is wondering if this concern should be taken seriously.

Bayesian Network Inference Diagram



Inference in 0.00ms



Decision Network Inference Diagram
Gender
0 [0.00]
1 [40.96]
100.00%
TBPT
0 [40.96]
1 [0.00]
100.00%
age
age_0_To_14 [0.00]
age_15_To_44 [40.96]
age_45_To_64 [0.00]
age_65_Plus [0.00]
100.00%
latent
0 [50.00]
1 [40.77]
2.05%
97.95%
TB
0 [50.00]
1 [-70.00]
92.47%
7.53%
Symptom
None [0.00]
Cough [0.00]
Chest_Pain [40.96]
Fever [0.00]
Sweating_at_night [0.00]
100.00%
testing
0 [50.00]
1 [-44.33]
90.42%
9.58%
Diagnosed
0 [49.90]
1 [-69.67]
92.52%
7.48%
treatment
0 [40.96]
1 [0.00]
100.00%
utility : 40.96 (31.67)
MEU 40.96 (stdev=31.67)
Inference in 1.00ms

In the first scenario, we can see the probability of having a latent infection for a male in his 20s that went through TBPT is 93%. However, the chance of having the disease is 9,30%. In the second scenario, there is a minimal increase from the first scenario of having the latent infection (93,81%). However, the network shows that the probability of having the disease increases to 52,96% when you are an elderly woman that has not been through TBPT and shows symptoms of coughing. In the third scenario, the probability of having the latent infection is close to 0,97% when you are a young girl that went through TBPT and got a negative test. The probability of having TB disease is approximately 0,24%. In the fourth scenario, we have a similar case as in the first one since we are considering a man in his 20s. The difference is that TBPT status was negative and that he showed the symptom of coughing. Therefore, The probability of having the disease is 7,53% and the infection is 97,95%.

Through these scenario tests, we have demonstrated how the model can be used to estimate the probability of getting DB based on the observed variables (gender, age, TBPT, symptoms). Individuals can evaluate the effect of their decision-making regarding TB preventive treatment. Additionally, individuals can decide whether to undergo TB testing and potentially receive treatment based on the current probability of having the infection/disease. With this knowledge from the network, the user community can choose to undergo actions such as getting TBPT or testing, which can ultimately increase their expected utility.

5. Conclusion

Our Bayesian network aims to estimate the probability of having Tuberculosis disease based on factors such as age, gender and if they have received Tuberculosis Prevention Treatment. The model we have built is simplistic but has a wide range of use. Based on our research and the probability tables we have constructed, the model appears to provide sufficiently reliable results to guide someone to take the correct actions if they have an inclination that they might have come into contact with TB or are infected with the active TB disease. However, our data does not take into account every factor such as a weak immune system due to HIV or diabetes, but it offers a valuable starting point for raising awareness about the importance of early detection. We believe that with the small amount of information the range of people that

this could help is impressive, but the model is very simplistic and with no real world testing the actual data that we obtain does not have enough validity attached to it. Coupled with the fact that the subset of weights we use is heavily derived and simplified to fit within our own expertise and time limit, the model is far from an accurate tool for diagnosing TB.
The model can be expanded and refined with additional factors and variables that may influence the risk of TB infection. We believe that with the help of future development, research and funding artificial intelligent networks could have a wide application to help and assist with diagnosis for early signs of TB infection.
In summary, continuous development and adaptation of the model can contribute significantly to the global effort to reduce the burden of TB and improve public outcomes.

6. References

1. World Health Organization (WHO), 2023. Tuberculosis profile: South Africa. Available: https://worldhealthorg.shinyapps.io/tb_profiles
2. Statista. 2022. Total population of South Africa in 2022, by age group. Available: https://www.statista.com/statistics/1116077/total-population-of-south-africa-by-age-group/
3. Centers For Disease Control and Prevention (CDC). 2018. Deciding When to Treat Latent TB Infection. Available: https://www.cdc.gov/tb/topic/treatment/decideltbi.htm
4. Farhat, M, Greenaway, C, Pai, M, Menzies, D. 2006. False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria?. The International Journal of Tuberculosis and Lung Disease. 10 (11), 1192–1204.
5. Harvard Health Publishing (HHP). 2020. Chest pain: A heart attack or something else? Available: https://www.health.harvard.edu/heart-health/chest-pain-a-heart-attack-or-something-else
6. Jay Summer. 2023. What Causes Night Sweats in Women?. Available: https://www.sleepfoundation.org/night-sweats/women
7. Centers For Disease Control and Prevention (CSC). 2022. Key Facts About Influenza (Flu). Available: https://www.cdc.gov/flu/about/keyfacts.htm
8. Ford, A. C. et al. 2006. Cough in the community: a cross sectional survey and the relationship to gastrointestinal symptoms. Thorax. [Online] 61 (11), 975–979.
9. Lee, S. et al. 2019. The Presence of Cough and Tuberculosis: Active Case Finding Outcomes in the Philippines. Tuberculosis research and treatment. [Online] 20191–9.

10. Miller, L. G. et al. 2000. A Population-Based Survey of Tuberculosis Symptoms: How Atypical Are Atypical Presentations? Clinical infectious diseases. [Online] 30 (2), 293–299.
11. Menberu, M. A. 2016. Performance of the WHO 2011 TB Symptom Screening Algorithm for Pulmonary TB Diagnosis among HIV-Infected Patients in Gondar University Referral Hospital, Ethiopia. International journal of microbiology. [Online] 20169058109–5.
12. Schaaf, H. S. et al. 2010 Tuberculosis at extremes of age. Respirology (Carlton, Vic.). [Online] 15 (5), 747–763.
13. TBFacts.org. 2022. Latent TB - What is latent TB, symptoms of latent TB, latent TB treatment. Available: https://tbfacts.org/latent-tb/
14. Van der Walt, M., Moyo, S. 2018. The first National Prevalence Survey. Available: https://www.nicd.ac.za/wp-content/uploads/2021/02/TB-Prevalence-survey-report_A4_SA_TPS-Short_Feb-2021.pdf