CSC4025Z (2023): Artificial Intelligence
Assignment 1: Application of Bayesian Networks*

Fardoza Tohab
THBFAR002

Kristen Basson
BSSKRI003

Mosamat Sabiha Shaikh
SHKMOS004

1 INTRODUCTION

1.1 Background and Objective

Worldwide, 415 million individuals have diabetes, and it's thought that 193 million more do not have a diagnosis. More than 90% of diabetic patients have type 2 diabetes, which causes microvascular and macrovascular problems that are extremely distressing for patients and caregivers alike and place a significant strain on healthcare systems. The incidence and prevalence of type 2 diabetes are rising internationally despite a better understanding of the risk factors for the illness and evidence of effective preventative programs [1]. By avoiding or delaying problems, early diagnosis through screening programs and the accessibility of safe and effective medicines lowers morbidity and mortality. Increased knowledge of certain diabetes phenotypes and genotypes may lead to more individualized treatment plans for individuals with type 2 diabetes.

The risk of T2D developing in persons whose blood glucose levels are higher than normal but who do not meet the criteria for a diagnosis of T2D (prediabetes) is quite high [2]. T2D is a preventable disorder. Similar to those who have diabetes, those with prediabetes also have an increased risk of cardiovascular disease and death [3]. For this group, the yearly rate of progression to T2D is between 5 and 10% [4][5]. According to the IDF Diabetes Atlas, 4.63 billion people globally, or 6.0% of the population, are thought to have prediabetes [6]. Preventing or delaying T2D may be possible if action is taken while prediabetes is present. To assess the intricate association between various risk factors and the development of T2D, further research and data analysis frameworks are required. In this regard, a Bayesian network (BN)-based model for T2D is taken into account.

The objective of the network that detects type 2 diabetes will be used to accurately determine the presence or absence of the disease in individuals based on a set of observable symptoms or risk factors. The Bayesian network might assist people in taking preventive steps to lower their likelihood of acquiring type 2 diabetes or help them manage their illness more efficiently if they currently have it by identifying the problem early.

1.2 Potential User Community

The network may be primarily used by medical professionals including doctors, nurse practitioners, and diabetes educators. They may use the model to determine whether individuals have a high risk of getting type 2 diabetes and then offer the right treatments to help manage or avoid the condition. Community health workers can also use this network in low-resource communities to make accurate diagnoses.

Another group of users would be the patients themselves. The model might help patients better understand their risk of acquiring type 2 diabetes and help them decide whether to adjust their lifestyle or engage in other methods to lower that risk.
The model may also be used by health insurers to identify people who are most likely to acquire type 2 diabetes and to distribute funds for both disease care and preventative measures. This might aid insurers in better controlling the potentially high cost of treating problems associated with diabetes.

Officials in charge of public health and decision-makers might also be potential users of the network. They might utilize the model to create and implement focused preventive and intervention programs and to better understand the incidence and risk factors for type 2 diabetes in various groups.

2 PROBLEM ANALYSIS

2.1 Potential Factors

Multiple factors can contribute to the development of type 2 diabetes. Some important ones to consider are:

1. Gender: Different genders societally tend to have different lifestyles regarding many factors of type 2 diabetes determiners such as smoking, diet, and physical activities. Additionally, differences in fat distribution in the body and pregnancy contribute.
2. Smoking: Smoking is a contributor to multiple health problems including type 2 diabetes. Some of the side effects of smoking include increased blood pressure, unhealthy diets, and weight gain. These are risk factors for type 2 diabetes as well.
3. Age: Aging leads to insulin resistance and increases the risk of type 2 diabetes which is why it needs to be considered. Additionally, factors like diet and physical activity are commonly known to be different for different age groups.
4. Physical Activity: Regular activity is commonly known to improve insulin and glucose levels whose control is essential in controlling/preventing type 2 diabetes.

5. BMI: BMI can be used as an overall health indicator for many health issues including type 2 diabetes due to associations between fat and insulin resistance.
6. Diet: A diet consisting of high levels of unhealthy fats, sugar and carbohydrates affects insulin levels and hence a factor to consider.
7. Blood Pressure: Diabetics are twice as likely to have hypertension. Moreover, hypertension tends to result in insulin resistance making it a risk factor to consider [20].
8. Hba1c levels: Nondiabetics' level usually ranges between 4.0%-5.6%, prediabetes between 5.7%-6.4% and finally those with diabetes usually have 6.4% or higher levels [21].

As a result, we used these factors to construct our network due to their direct and indirect relationship to the risk and development of type 2 diabetes.

2.2 Datasets and Expert Knowledge

To generate the network and determine the conditional probabilities for each variable in the network, we made use of a variety of published scientific papers. The datasets used included information on various risk factors for Type 2 diabetes, such as gender, age, smoking, physical activity, BMI and diet. Other factors such as blood pressure and hba1c levels (glycosylated haemoglobin test) were also included in the data set, as they are important indicators of type 2 diabetes risk.

2.2.1 Identifying Risk Factors

A study conducted by Fuster-Parra *et al.* consisted of analyzing a cohort of adults with initial prediabetes to find the risk variables for type 2 diabetes. This study sought to identify the factors that have a significant impact on the development of Type 2 diabetes once prediabetes has been established and can be used to prevent Type 2 diabetes. Furthermore, the Markov blanket was used to identify the set of characteristics that have the greatest effects on T2D [7]. The collection of nodes made up of node A's parents, children, and other parents (i.e., spouses) is known as the node's Markov blanket. All the nodes that would make node A independent of all other nodes if we knew their states, i.e., if we had concrete evidence for these nodes, are contained in the Markov blanket of node A. This indicates that the only information required to forecast the posterior probability distribution of node A is the Markov blanket for that node [8].

To build a BN model for Type 2 diabetes, a dataset made up of 12 important aspects of the T2D domain was used. In a multivariate setting, dependencies and conditional independencies were identified from empirical data. Using the R language's bnlearn package, the structure and parameters were learnt while injecting past information. The Markov blanket was taken into consideration to identify the characteristics (variables) that raise the risk of T2D [7]. The study discovered that prediabetes raises a person's likelihood of having T2D, which is linked to other illnesses including cardiovascular disease and is a major public health problem worldwide. According to the study, age, physical activity (PA), fatty liver index (FLI), glycosylated haemoglobin (HbA1c), body mass index (BMI), and age all have significant roles in the development of T2D. By inference, it was possible to arrive at a high estimated likelihood value of T2D when the BMI was instantiated to the obese value, the HbA1c to the more than 6 value, the FLI to the more than 60 value, the physical activity (PA) to the no state, and the age to the 48–62 state. Rankings were given to the warning indicators that contributed to T2D in particular states [7].

2.2.2 Influence of Age and Physical Activity on Special Diets

The ages considered for the Age variable were 18 – 32, 33 – 47, and 48 – 62 years old.

To find the percentage of people on a diet, key findings by the National Centre for Health Statistics were consulted. These studies made use of data from the National Health and Nutrition Examination Surveys (NHANES) for the years 2007–2008, 2009–2010, 2011,–2012, 2013–2014, 2015–2016, and 2017–2018. According to the National Health and Nutrition Examination Survey, from 2015 to 2018, 17% of American adults were following a special diet of some kind, with more women than males doing so. Weight reduction or low-calorie diets were the most typical kind of special diet. From 2007–2008 to 2017–2018, the proportion of adults following any special diet, weight reduction or low–calorie diets, and low–carbohydrate diets grew, whereas the proportion of people following low–fat or low–cholesterol diets fell. With higher weight categories and educational levels came a rise in the number of individuals on special diets. Compared to non-Hispanic blacks and non-Hispanic Asians, non-Hispanic whites were more likely to follow a special diet. Statistic values from these findings were utilized to determine the conditional probabilities between age and being on a special diet [9].

To see how physical activity and diet were related, the Weight-loss Statistics 2023 by Sam Phoenix was consulted. According to the survey, diets and exercise are the two major methods used by people to reduce weight. With 51.4 million Americans reporting they cycle at least once a year, biking is the most popular form of exercise. In the meanwhile, about 40% of respondents claim that dieting is done "to lose weight". 10% of people who are dieting utilize calorie counting, according to the International Food Information Council. Overall, it can be said that nutrition and exercise are both essential for losing weight [10].

Another study conducted by Fuster-Parra *et al.* [7] showed the influence of diet and physical activity on type 2 diabetes. The study showed that Diet has an impact, as shown by the predicted chance of acquiring T2D being 0.0618 (6.18% stated in percentage) or 0.3310 (33.10% expressed in percentage) once Diet has been instantiated to the No or the Yes state, respectively. Body Mass Index (BMI) in the state of obesity has the biggest effect on the No diet group, whereas physical activity in the No state and BMI in the state of obesity have the highest influences on the Yes diet group. When BMI is instantiated into the obese condition, the risk of acquiring T2D rises in both groups, with the estimated likelihood of developing T2D reaching 0.7130 (71.30% expressed in percentage) and 0.4660 (46.60% stated in percentage) in each case, exhibiting the largest effect. The chance of acquiring T2D is projected to be 0.8430 (84.30% stated in percentage) and 0.5910 (59.10% expressed in percentage) at step 5, respectively. Other characteristics that raise the risk of developing T2D include smoking in the former state, age in the 48-62 state, and gender in the male state.

2.2.3 Influence of Diet and BMI on Blood Pressure

To investigate the relationship between BMI and Blood pressure, the study conducted by Landi *et al.* [11] investigated the relationship between different levels of body mass index (BMI) and hypertension using an unselected sample of 7907 community-living adults assessed during the Longevity Check-up 7+ project in Italy was consulted. According to the study, there is a gradient between rising blood pressure and increased BMI levels, with hypertension prevalence ranging from 45% in people with a normal BMI to 87% in people who are classified as obese class III. The average systolic and diastolic blood pressure increased considerably and linearly across BMI levels, the study also revealed. The study finds that, independently of other clinical risk factors, BMI may have a direct impact on blood pressure.

The study found a link between a healthy diet and a decreased prevalence of hypertension. A healthy diet is defined as eating at least three servings of fruit and/or vegetables per day. According to the study, people with healthy diets had a lower prevalence of hypertension (72%) than those without healthy diets (68%) [11].

2.2.4 Influence of Gender on Smoking

The dataset used for this section came from a cross-sectional study done on smoking and physical activity in adults [14]. A total of 2,602 healthy adults in the age range of 30-60 years participated. Data was collected regarding their demographic characteristics, smoking status, physical activity, weight, and height. 55.4% of the participants were male while 44.6% were female. Data collection found that in both male and female, the percentage of non-smokers were more (62.2 and 63% respectively). While almost similar, the rate of male smokers was more by 0.08%.

2.2.5 Influence of Age and Smoking on Physical Activity

Different age groups tend to exhibit different behaviours relating to smoking and physical activity. We investigated the trends of smoking and physical activity in 3 age groups namely young adults, middle-aged and elderly. The goal was to attain the probability of doing physical activity given their age and smoking status.

A cross-sectional study done by the National Institution of Tuberculosis and Lung Disease in Iran investigated the relationship between smoking and physical activity in adults [14]. It was part of a larger study relating to the prevalence of non-communicable disease risk factors conducted by the Research Centre for Tobacco Prevention and Control and the National Research Institute of Tuberculosis and Lung Diseases. It notes a direct relationship between smoking and various health issues. To be identified as a smoker, the participants had to have had at least 100 cigarettes and had no plan of quitting. Physical activity was defined by the time spent doing various activities such as sleeping, lying down, walking, and running. The age group in the study was about 50-50 for the later 2 groups we were investigating. Logical regression was used to investigate the effect of cigarette smoking and an appropriate level of daily activity. It found that more non-smokers (71.0%) had an appropriate level of physical activity compared to non-smokers (28.9%).

Another cross-sectional study carried out in India, sought to examine the relationship between tobacco use and physical activity among college-bound young people between the ages of 18 and 24. According to data collected from 500 students, physically active individuals had much lower probabilities of smoking than those who were not. 20.2% of people reported using tobacco regularly. Younger age groups, females, and non-professionals had greater probabilities of smoking, and females were less active than males. The results point to the necessity for frequent physical exercise and tobacco-discouragement sessions integrated into the academic curriculum, as well as the promotion of physical activity among the young to reduce tobacco use [13].

2.2.6 Influence of Physical Activity and Smoking on BMI

A dataset of 1,246 participants representing an average urban population in Denmark was used to determine the relationship between physical activity and smoking on the BMI rate[15]. The data collected was used in the study to determine the role of age, gender, BMI, physical activity, smoking habits, and diet on appetite rating. 55.1% of the participants were female and 44.9% were male all of them in the age range 20-60. The study categorized the participants as either being above or below 25 with above 25 being the overweight group.

An assumption was made to divide the under-25 category into normal and underweight with equal proportions. Examining the dataset found that 59.88% of the population who did physical and smoked were overweight. A similar trend was found in participants who smoked and did not do physical activity where 63.12% of them were categorized as overweight. Conversely, in non-smokers who did or did not do physical activity, a majority of the participants were below the 25 mark. 39.74% of non-smokers who did physical activity were above 25 and 29.46% of non-smokers who did not smoke were above 25.

2.2.7 Influence of Smoking on HbA1c Levels

An HbA1c test, also known as haemoglobin A1c, is a blood test that measures an average blood sugar level over 3 months [18]. Who defines 6.5% as the recommended cut-point for the diagnosis of diabetes [17]. It has been used as a gold standard for decades [16]. However, it should be noted that one can have a level below 6.5% and still be diabetic. A study in South Africa found that 6.1% was the optimal cutoff point for South Africans and hence was used to define the HbA1c levels. Having a level above 6.1 is considered high otherwise normal.

A study was conducted by the National Diabetes Register in Sweden to examine smoking trends in diabetic patients and the relationship between smoking and glycemic control. Data for this study was obtained from hospitals and primary health care and it was found that smoking is associated with an increase in HbA1c levels. Around 50% of smokers had high levels of HbA1c compared to 40% of non-smokers.

2.2.8 Influence of Blood Pressure and HbA1c Levels on Type 2 Diabetes

A South African study consisting of 946 participants of which 23.7% were male and 76.3% were female [16]. The study aimed to test the cut-off point of 6.5% in a South African population. The participants were screened for diabetes to test the 6.5% cut-off point. 80.3% of the participants had a high HbA1c level and had diabetes whereas 16.7% of those with low HbA1c levels had diabetes.

To test the relationship between blood pressure and type 2 diabetes, 2 separate studies were consulted. The first study was related to the coexistence of hypertension in type 2 diabetes [19]. 556 participated in the study. 38.2% were male and 61.8% were female. The study divided the participants into those that had hypertension as well as diabetes and those that had 'diabetes alone'. It was found that about 50% of the 'diabetes alone' were unaware of their coexistent hypertension. Due to the conclusion of the coexistence of hypertension in diabetics going often undiagnosed, we decided to make an independent assumption relating to hypertension (high blood pressure). A second study showed that the prevalence of Type 2 diabetes mellitus (T2DM) has increased significantly over the past few decades, according to statistics from the National Health and Nutritional Examination Survey (NHANES) and the Centers for Disease Control and Prevention (CDC). According to estimates, 34.2 million Americans, or 10.5% of the population, have diabetes. Hypertension affects 73.6% of diabetics who are 18 years or older. It is not accidental that a substantial percentage of patients have both hypertension and diabetes [12].

These studies were used to find the relationship between type 2 diabetes conditioned on blood pressure and HbA1c. The values obtained from both tables were combined and normalized to form the new data.

3 DECISION NETWORK MODEL

3.1 Diagram

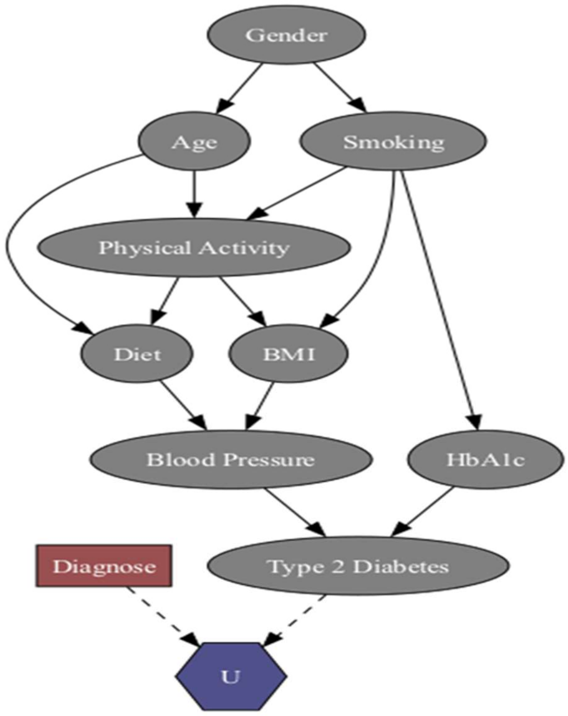

Diagram 1: Final Network

Table 1: Utilities set for the network

Type 2 Diabetes	Diagnose	Utility
Yes	Yes	100
No	Yes	-100
Yes	No	-100
No	No	100

Table 2: Nodes and their values from the Network

Node	Values
Gender	Male, Female
Age	18-32, 33-47, 48-62
Smoking	No, Yes
Physical Activity	No, Yes
Diet	No, Yes
BMI	Underweight, Normal Weight, Overweight
Blood Pressure	Normal, High
HbA1c	Normal, High
Type 2 Diabetes	No, Yes

3.2 Structure and Weights

Based on the research we did it can be assumed that factors such as physical activity and smoking are risk factors in developing diabetes. Since it is common knowledge that males and females depict different characteristics relating to the previously mentioned traits, we chose to start with gender. Currently, the male and female gender ratio is approximately equal and hence the 50-50 split. This is followed by the age and the smoking node being conditionally dependent on gender due to these characteristics differing in different genders. Physical activity, diet and BMI can be conditioned on the previous nodes and amongst each other. We started with common knowledge relationships such as BMI range conditioned on physical activity then based on our research came to this conclusion. Finally, after consulting with 2 doctors, we found that blood pressure and HbA1c are common determiners for diabetes and hence our final two nodes before our objective node. The dependencies are backed by formal knowledge done in the previous section as well as consultations with the doctors.

The utility represents the accuracy of the diagnosis and the ability of the network to correctly identify individuals with diabetes. We chose the golden number 100 for correct diagnosis since it is a significant feat and the goal of our objective. Conversely, misdiagnosis would be an undesirable outcome and hence it is the negative complementary of correct diagnosis.
*Research done in section 2 was used to justify the probabilities used in the nodes.

3.3 Probability Tables

Age | Gender

	Age		
Gender	18-32	33-47	48-62
Male	0.0681	0.4301	0.5018
Female	0.1147	0.5301	0.3552

Smoking | Gender

	Smoking	
Gender	No	Yes
Male	0.6220	0.3780
Female	0.6300	0.3700

Physical Activity | Smoking, Age

		Physical Activity	
Smoking	Age	No	Yes
No	18-32	0.3298	0.6702
Yes	18-32	0.9050	0.0950
No	33-47	0.2900	0.7100
Yes	33-47	0.7110	0.2890
No	48-62	0.2900	0.7100
Yes	48-62	0.7110	0.2890

BMI | Smoking, Physical Activity

		BMI		
Smoking	Physical Activity	Underweight	Normal weight	Overweight
No	No	0.3527	0.3527	0.2946
Yes	No	0.1844	0.1844	0.6312
No	Yes	0.3013	0.3013	0.3974
Yes	Yes	0.2006	0.2006	0.5988

Diet | Age, Physical Activity

		Diet	
Age	Physical Activity	No	Yes
18-32	No	0.6260	0.3740
33-47	No	0.5920	0.4080
48-62	No	0.5550	0.4450
18-32	Yes	0.6520	0.3480
33-47	Yes	0.6260	0.3740
48-62	Yes	0.6310	0.3690

Blood Pressure | BMI, Diet

		Blood Pressure	
BMI	Diet	Normal	High
Underweight	No	0.8620	0.1380
Normal Weight	No	0.5260	0.4740
Overweight	No	0.4030	0.5970
Underweight	Yes	0.6530	0.3470
Normal Weight	Yes	0.3330	0.6670
Overweight	Yes	0.1980	0.8020

HbA1c | Smoking

	HbA1c	
Smoking	Normal	High
No	0.5600	0.4400
Yes	0.5050	0.4950

Type 2 Diabetes | HbA1c, Blood Pressure

		Type 2 Diabetes	
HbA1c	Blood Pressure	No	Yes
Normal	Normal	0.7020	0.2980
High	Normal	0.5940	0.4060
Normal	High	0.5940	0.4060
High	High	0.2980	0.7020

4 MODEL TESTING AND EVALUATION

4.1 Bayesian Network

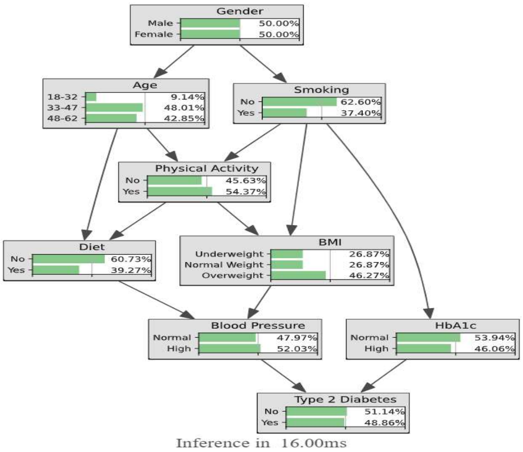

Figure 1: The diagram represents the Bayesian network with no evidence provided. As expected, the probability of having diabetes is lower. Without any evidence, we cannot diagnose someone with diabetes without any prior knowledge. However, due to the prevalence of type 2 diabetes, there is still a chance of having diabetes. Without any evidence, the diagnosis can go either way, but the conclusion would be no diabetes.

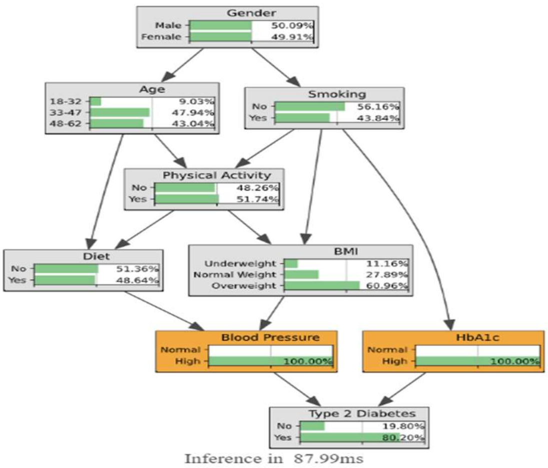

Figure 2: The above diagram shows an inference when the Blood pressure and the HbA1c level are the only observed variables and they are both high, then the patient is at a high risk of being diagnosed with type 2 diabetes. This is expected because hypertension and a high HbA1c are common determiners for diabetes and hence the probability of having it is increased to a more likely probability.

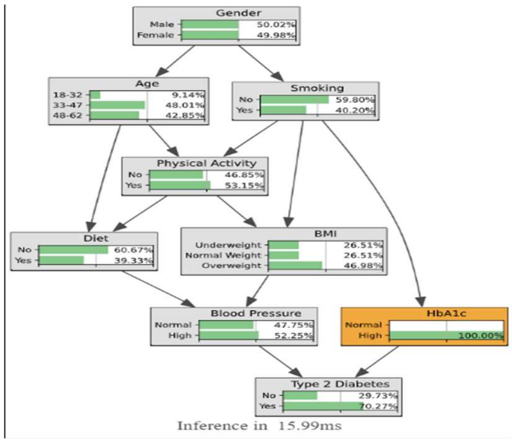

Figure 3: The diagram shows an inference when the HbA1c level is the only observed variable and it is high, the probability of having Type 2 Diabetes is high. This is as expected. With the only evidence being high HbA1c, there is a higher probability of having the disease but it is not as high as in Figure 2 where both hypertension and high HbA1c were observed as the likely determiner.

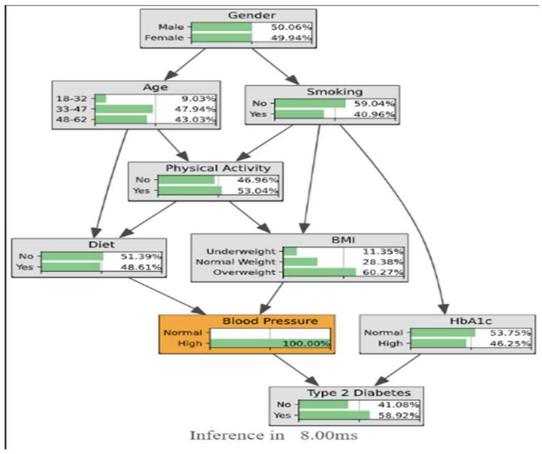

Figure 4: The diagram shows an inference with blood pressure being the only evidence set and it is set as high. While the liklihood of having type 2 diabetes is higher, it is much lower than in the case of Figure 2 and 3. High HbA1c and hypertension make a stronger case for having type 2 diabetes based on the research done in section 2. Between either HbA1c and hypertension, while both play a role in determining diabetes, HbA1c should play a stronger role. This is visible in the probability comparison.

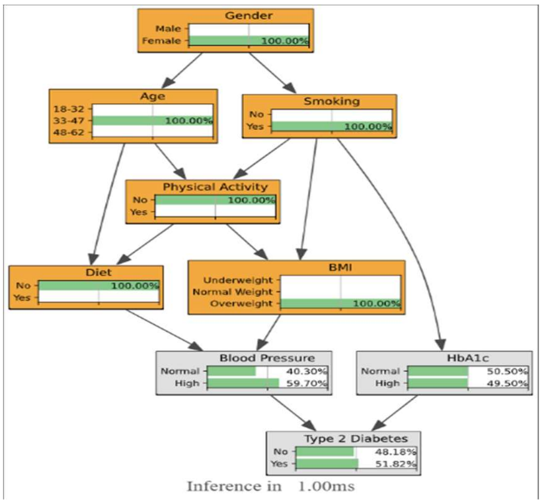

Figure 5: The diagram depicts an inference of someone who has an unhealthy lifestyle. No evidence was set for blood pressure and HbA1c. Based on the given unhealthy lifestyle, the probability of having diabetes is higher but almost equal to not having it. This is because blood pressure and HbA1c levels are still crucial in diabetes determining.

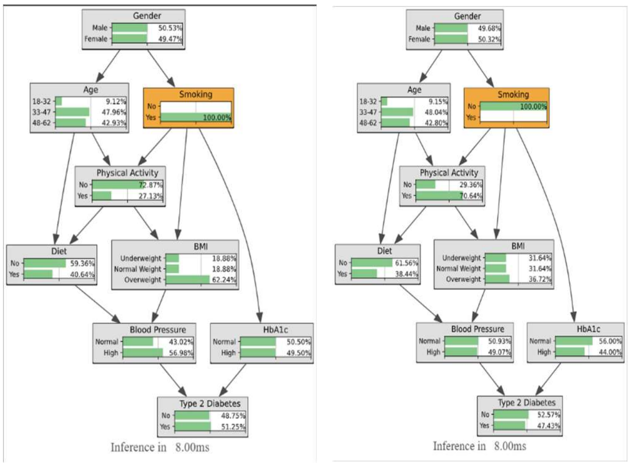

Figure 6 and 7: The diagrams only has smoking as evidence. As expected, a smoker is more likely to have diabetes than a non-smoker. But the probabilities are not much higher than the chance of not having diabetes because it is not the only deciding factor.

4.2 Decision Network

For this section, we consulted with one of the doctors we consulted for developing the diagrams. We presented him with the following scenarios and based on the diagnosis opinion, we matched them with the utilities we obtained. This was used to validate our diagram.

Note: The decision was based on given evidence for the network

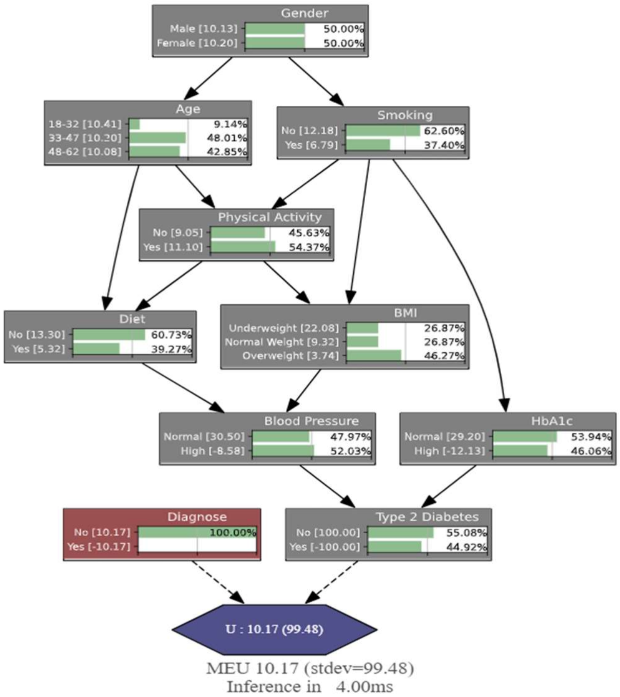

Figure 8: The above diagram represents the utility inference with no evidence. As expected the network made the decision not to diagnose the patient because without evidence (without any risk factors) it is the optimal decision not to diagnose.

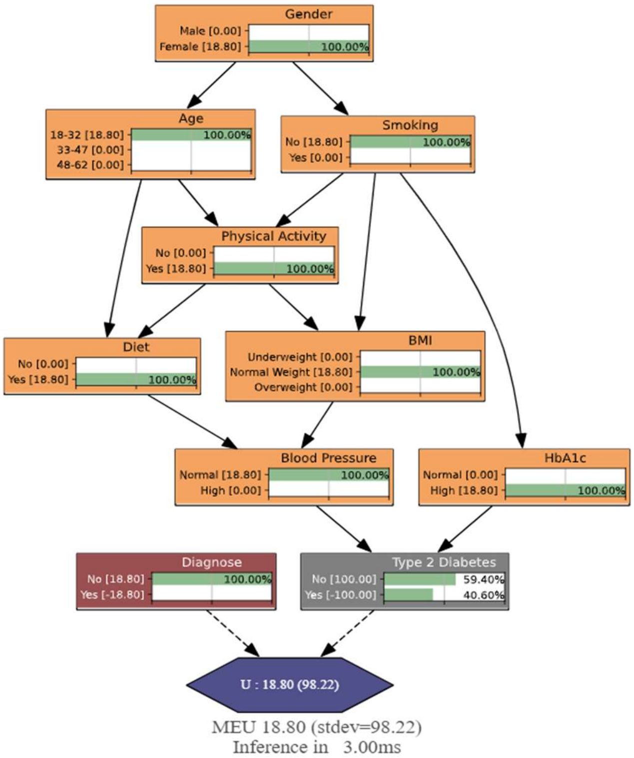

Figure 9: The diagram above represents the utility inference with evidence. We provided the network with evidence that the patient is a female between the ages of 18 - 32 years old, does not smoke, exercises, is on a diet, and has normal weight and normal blood pressure. However, the patient has a high HbA1c level. Considering all these factors the patient would be considered not having type 2 diabetes since the HbA1c levels alone are not a determiner of diabetes. It is only one of the factors. As expected the network made the decision not to diagnose the patient and hence, it is the optimal decision.

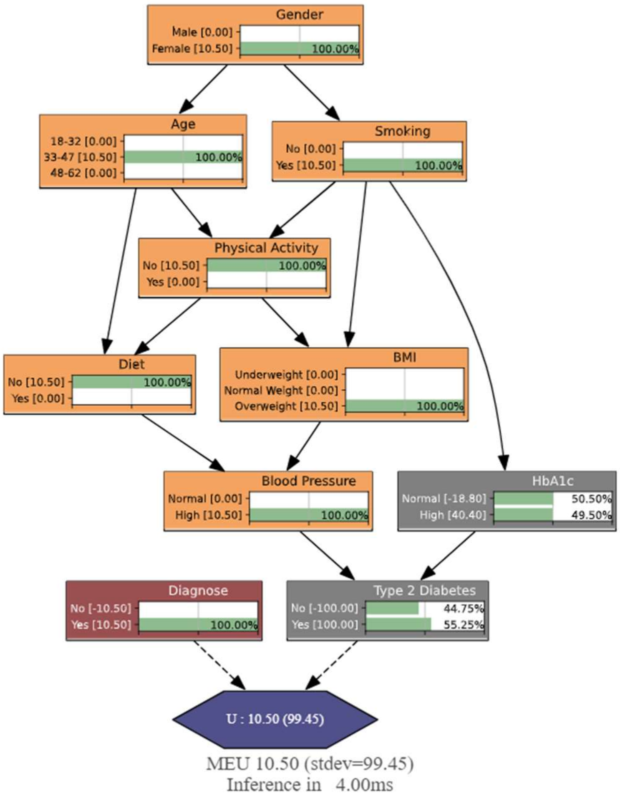

Figure 10: The diagram above represents the utility inference with evidence. We provided the network with evidence that the patient is a female between the ages of 33 - 47 years old, is a smoker, does not exercise, is not on some special diet, is overweight and has high blood pressure. There was no evidence provided for the patient's HbA1c levels. Considering all these factors the patient would be considered to be diagnosed with type 2 diabetes since all but one of the risk factors are present. As expected, the network decided to diagnose the patient and hence, it is the optimal decision.

4.3 Use Case Scenarios

4.3.1 Scenario 1: Diagnosis in a Standard Hospital/Clinic

A standard hospital or a clinic will have the equipment and capabilities to perform all forms and tests. A young female patient who lives a healthy lifestyle comes into the hospital for her regular checkup. She wishes to get checked for diabetes. The doctor takes the following notes after asking standard check-up questions:

- Diet = Good diet
- Smoking = Never
- Physical Activity = Acceptable level

The doctor takes the patient for some regular tests, and these are the results that game back:

- BMI = Normal Weight
- Blood Pressure = Normal
- HbA1c = High

These notes and results were plugged into the Bayesian and the decision can be seen in Figure 9. As seen, the patient will not be diagnosed with type 2 diabetes and that is the optimal decision. The doctor can continue with another testing

4.3.2 Scenario 2: Diagnosis in a low-Resource Area

A rural clinic has only a few standard pieces of equipment such as a stethoscope as well as a blood pressure machine. A middle-aged female patient who has been feeling unwell comes in for a checkup. A nurse tends to her by asking a few questions. The following notes are taken:

- Smoking = Current smoker
- Physical Activity = Limited
- Diet = Not good enough

The nurse proceeds to take her weight on a weighing scale and her blood pressure using a bp machine. These are the outcomes:

- BMI = Overweight
- Blood pressure = High

The nurse places these figures into the decision network. Figure 10 depicts that the optimal decision would be to diagnose the patient with type 2 diabetes. The patient can now be given available medicine and a pamphlet/information on diabetes control.

5 CONCLUSION

Type 2 diabetes is a prevalent and treatable/controllable disease but due to a lack of education and resources, it often goes undiagnosed and/or untreated. The Bayesian network we developed aims to provide a quick, efficient, and accurate diagnosis of type 2 diabetes. It should be noted that it is not comprehensive as certain risk factors such as genetics have been excluded. However, it provides a good approximation of the utility of the decision to take. The decision network not only aids in speeding up the diagnosis process in a standard hospital setting but also aids in giving accurate diagnoses in low-resource areas.

REFERENCES

[1] Sudesna Chatterjee, Kamlesh Khunti, and Melanie J Davies. 2017. Type 2 diabetes. The Lancet 389, 10085 (June 2017), 2239–2251. DOI:https://doi.org/10.1016/s0140-6736(17)30058-2

[2] Tabak AG, Herder C, Rathmann W, Brunner EJ, Kivimaki M. Prediabetes: a high-risk state for diabetes development. Lancet. (2012) 379:2279–290. doi: 10.1016/S0140-6736(12)60283-9

[3] Huang Y, Cai X, Mai W, Li M, Hu Y. Association between prediabetes and risk of cardiovascular disease and all cause mortality: systematic review and meta-analysis. BMJ. (2016) 355:i5953. doi: 10.1136/bmj.i5953

[4] 7. Ezquerra-Lazaro I, Cea-Soriano L, Giraldez-Garcia C, Ruiz A, FranchNadal J, Diez-Espino J, et al. Lifestyle factors do not explain the difference on diabetes progression according to type of prediabetes: results from a Spanish prospective cohort of prediabetic patients. Diabetes Res Clin Pract. (2019)153:66–75. doi: 10.1016/j.diabres.2019.05.033

[5] Bennasar-Veny M, Fresneda S, López-González A, Busquets-Cortés C, Aguiló A, Yañez AM. Lifestyle and progression to Type 2 diabetes in a cohort of workers with prediabetes. Nutrients. (2020) 12:1538. doi: 10.3390/nu12051538

[6] Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalance estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. Diabetes Res Clin Pract. (2019) 157:107–843. doi: 10.1016/j.diabres.2019.107843

[7] Pilar Fuster-Parra, Aina M. Yañez, Arturo López-González, A. Aguiló, and Miquel Bennasar-Veny. 2023. Identifying risk factors of developing type 2 diabetesfrom an adult population with initial prediabetes using a Bayesian network. Frontiers in Public Health 10, (January 2023). DOI:https://doi.org/10.3389/fpubh.2022.1035025

[8] 2020. Articles. Bayesia.com. Retrieved May 6, 2023 from https://www.bayesia.com/articles/#!bayesialab-knowledge-hub/key-concepts-markov-blanket

[9] 2023. Products - Data Briefs - Number 389 - November 2020. Retrieved May 6, 2023 from https://www.cdc.gov/nchs/products/databriefs/db389.htm

[10] Sam Phoenix. 2023. Weight-Loss Statistics 2023 | Surprising Facts & Data - Great Green Wall. Great Green Wall. Retrieved May 6, 2023 from https://www.greatgreenwall.org/supplements/weight-loss-statistics/#:~:text=62.9%25%20of%20people%20tried%20to,weight%20by%20drinking%20more%20water.

[11] Francesco Landi, Riccardo Calvani, Anna Picca, Matteo Tosato, Anna Maria Martone, Elena Ortolani, Alex Sisto, Emanuela D'Angelo, Elisabetta Serafini, Giovambattista Desideri, Maria Tecla Fuga, and Emanuele Marzetti. 2018. Body Mass Index is Strongly Associated with Hypertension: Results from the Longevity Check-up 7+ Study. Nutrients 10, 12 (December 2018), 1976. DOI:https://doi.org/10.3390/nu10121976

[12] Sowjanya Naha, Michael J Gardner, Darshan Khangura, L. Romayne Kurukulasuriya, and James R Sowers. 2021. Hypertension in Diabetes. Nih.gov. Retrieved May 7, 2023 from https://www.ncbi.nlm.nih.gov/books/NBK279027/#:~:text=It%20is%20estimated%20that%20diabetes,years%20or%20more%20have%20hypertension.

[13] Sonu Goel and Divya Monga. 2018. Does level of physical activity influence tobacco use behaviour among youth? A cross-sectional survey from India. Tobacco Induced Diseases 16, 1 (March 2018). DOI:https://doi.org/10.18332/tid/83808

[14] Heydari, G., Hosseini, M., Yousefifard, M., Asady, H., Baikpour, M., & Barat, A.2015. Smoking and Physical Activity in Healthy Adults: A Cross-Sectional Study in Tehran. *Tanaffos*, *14*(4), 238–245.

[15] NikolajT. Gregersen, BenteK. Møller, Anne Raben, SørenT. Kristensen, Lotte Holm, Anne Flint, and Arne Astrup. 2011. Determinants of appetite ratings: the role of age, gender, BMI, physical activity, smoking habits, and diet/weight concern. *Food & Nutrition Research* 55, 1 (2011), 7028. DOI: https://doi.org/10.3402/fnr.v55i0.7028

[16] Annalise E. Zemlin, Tandi E. Matsha, Mogamat S. Hassan, and Rajiv T. Erasmus. 2011. HbA1c of 6.5% to Diagnose Diabetes Mellitus—Does It Work for Us?—The Bellville South Africa Study. *PLoS ONE* 6, 8 (2011). DOI: https://doi.org/10.1371/journal.pone.0022558

[17] 2011. Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus. *Diabetes Research and Clinical Practice* 93, 3 (2011), 299-309. DOI: https://doi.org/10.1016/j.diabres.2011.03.012

[18] All About Your A1C. *CDC*. Retrieved May 8, 2023 from https://www.cdc.gov/diabetes/managing/managing-blood-sugar/a1c.html

[19] Uthman, M., Qaisar, A. M., & Anwar, A. (2015). Coexistence of Hypertension in Type 2 Diabetes Mellitus Patients (Co-tension-D2 study). *Age (years)*, *35*(44), 124.

[20] John R. Petrie, Tomasz J. Guzik, and Rhian M. Touyz. 2018. Diabetes, Hypertension, and Cardiovascular Disease: Clinical Insights and Vascular Mechanisms. *Canadian Journal of Cardiology* 34, 5 (2018), 575-584. DOI: https://doi.org/10.1016/j.cjca.2017.12.005

[21] Shariq I. Sherwani, Haseeb A. Khan, Aishah Ekhzaimy, Afshan Masood, and Meena K. Sakharkar. 2016. Significance of HbA1c Test in Diagnosis and Prognosis of Diabetic Patients. *Biomarker Insights* 11, (2016). DOI: https://doi.org/10.4137/bmi.s38440