Breast cancer risk assessment using Bayesian Networks

Joachim Verschelde

May 2024

1 Problem domain

Megan and her colleague Denise have scheduled an preventive screening examination in their local hospital to check if they have an increased chance of developing breast cancer. Because the hospital is short staffed and the waiting list for a mammography or a biopsy is long, they have decided to use a Bayesian Network to help them determine which patients are at an increased risk of developing breast cancer. The hospital will ask questions or perform tests to determine if the patients have an increased risk, if so a mammography or biopsy is necessary to confirm the presence of cancer. To build this Bayesian network, the hospital has a database of patients that have been diagnosed with breast cancer in the past, and they used this data to build the Bayesian Network. The question for this project is if Megan and Denise are at risk of developing breast cancer.
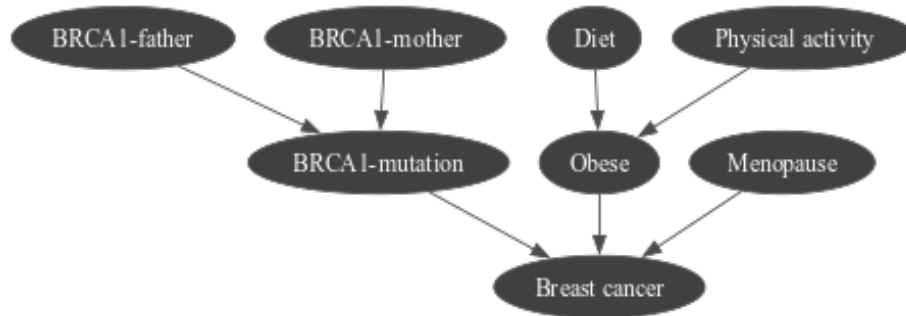
2 Network topology

Figure 1: Bayesian Network for Breast Cancer Risk Assessment

Table 1: Variable Description

Name	Type	Domain
Breast cancer	Categoric	{Yes, No}
Menopause	Categoric	{Before, After}
Obese	Categoric	{No, Moderate, Severe}
Diet	Categoric	{Unhealthy, Normal, Healthy}
Physical activity	Categoric	{None, Little, Regular}
BRCA1-mutation	Categoric	{Yes, No}
BRCA1-father	Categoric	{Yes, No}
BRCA1-mother	Categoric	{Yes, No}

The selection of variables for the Bayesian Network was based on their known or potential impact on breast cancer risk [5] [8]. The primary variables include genetic factors, lifestyle choices, and menopause. Furthermore we discretized the variables to make the network easier to interpret and to reduce the complexity of the network. For example in stead of the variable "Age" we used the variable "Menopause" which is a categoric variable with only 2 values "Before" and "After" indicating if the patient has already reached menopause.

- **Breast cancer**: The target variable representing the presence or absence of breast cancer.

- **BRCA1-mutation**: A mutation in the BRCA1 gene that is strongly associated with an increased risk of breast cancer.

- **BRCA1-father** and **BRCA1-mother**: indicates whether a parent carries the BRCA1 mutation, affecting the likelihood of the offspring having the mutation.

- **Obese**: Obesity is a well-documented risk factor for breast cancer [8].

- **Physical activity**: Regular physical activity burns calories and therefor reduces the chance of becoming obese.

- **Diet**: People who eat a healthy diet are less likely to become obese.

- **Menopause**: The menopause affects hormone levels, which in turn influence breast cancer risk. [7]

2.1 Causal and associative relations

2.1.1 Breast Cancer 1 (BRCA1) gene mutation

There is a causal relation between the BRCA1-father/mother variables and the BRCA1-mutation variable, as you inherit the genes of you parents, and if one of you're parents has a BRCA1 gene mutation the odds are 50% that you will also inherit this mutation [5].

2.1.2   The influence of diet and physical activity on obesity

There is also a causal relation between the physical activity and the obese variable. Regular physical activity burns calories and helps to maintain a healthy weight and reduce obesity, whereas a lack of physical activity, can lead to weight gain and obesity. Therefore, changes in physical activity levels directly affect the likelihood of becoming obese. Furthermore there is also a causal relation between the diet and the obese variable because people that are on a healthy diet eat less calorie-dense foods or are often in a caloric deficit which again directly controls the likelihood of becoming obese.

2.1.3   What causes breast cancer?

There is an associative relation between the menopause and presence of breast cancer, this is because the menopause does not directly increase the risk of breast cancer, but aging factors and the hormonal changes that occur during the menopause do. A mutation in the BRCA1 gene however could cause breast cancer, as the BRCA1 gene helps repair damaged DNA. When the BRCA1 gene is mutated, DNA damage may not be repaired properly and this could lead to breast cancer [9]. It is difficult to tell if obesity causes breast cancer, but there is definitely an association between obesity and breast cancer, especially when the patient has reached menopause [1].

3   Implementation using Pyagrum

I used the fastBN function to create the Bayesian Network, because it is a fast and natural way to declare the variables, their possible values and the relations between them.

```
network_definition = 'Physical activity{None|Little|Regular}→Obese{No|Moderate|Severe}←Diet{Unhealthy|Normal|Healthy};' \
                     'Obese→Breast cancer{Yes|No}←BRCA1-mutation{Yes|No}←BRCA1-mother{Yes|No};' \
                     'BRCA1-father{Yes|No}→BRCA1-mutation; Menopause{Before|After}→Breast cancer'

network = gum.fastBN(network_definition)
```

Figure 2: network definition using pyAgrum

The prior probabilities for the root nodes are defined using the cpt function. However there are multiple ways to use the cpt function, in my opinion the dictionary method is the most readable. I tried to underpin these prior probabilities with data found online. (BRCA1-father, BRCA1-mother [5], Physical activity [2], Diet [3], Menopause [6])

```
network.cpt("BRCA1-father").fillWith([0.0025,0.9975])
network.cpt("BRCA1-mother").fillWith([0.0025,0.9975])
network.cpt("Physical activity").fillWith([0.45,0.17,0.38])
network.cpt("Diet").fillWith([0.329,0.546,0.125])
network.cpt("Menopause").fillWith([0.6,0.4])
```

Figure 3: Declaration of the properly probabilities using pyAgrum

To define the conditional probability tables, I again used the dictionary method as this again was the most readable in my opinion.

```
network.cpt("BRCA1-mutation")[{'BRCA1-mother': 'No' , 'BRCA1-father': 'No'}] = [0.0001, 0.9999]
network.cpt("BRCA1-mutation")[{'BRCA1-mother': 'Yes' , 'BRCA1-father': 'No'}] = [0.5, 0.5]
network.cpt("BRCA1-mutation")[{'BRCA1-mother': 'No' , 'BRCA1-father': 'Yes'}] = [0.5, 0.5]
network.cpt("BRCA1-mutation")[{'BRCA1-mother': 'Yes' , 'BRCA1-father': 'Yes'}] = [0.9999, 0.0001]
```

Figure 4: Example of a conditional probability table declaration using pyAgrum

4   Results

The prior probabilities at the start of the preventive screening examination are computed using the pyAgrum.lib.notebook.showInference function.
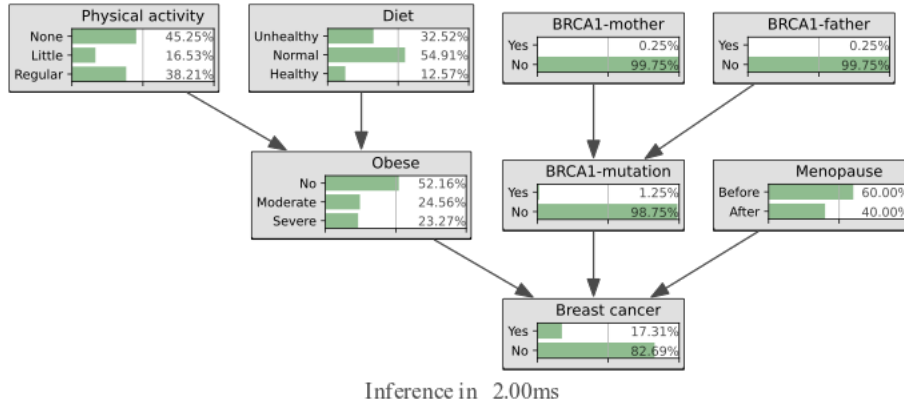
Figure 5: Prior probabilities

As can be seen in figure 5 the prior probability of having breast cancer is 17.31%.

4.1 Family history

The medical files are checked for a family history of breast cancer, and it is found that the father of Megan has a BRCA1 gene mutation, The parents of Denise do not have a mutation in the BRCA1 gene. We update the network with this information and compute the posterior probabilities. Therefor we use the exact inference method LazyPropagation.

```
ie_denise = gum.LazyPropagation(network)
ie_denise.setEvidence({'BRCA1-father': 'No', 'BRCA1-mother': 'No'})
ie_denise.makeInference()
gnb.showProba(ie_denise.posterior("Breast cancer"))

ie_megan = gum.LazyPropagation(network)
ie_megan.setEvidence({'BRCA1-father': 'Yes', 'BRCA1-mother': 'No'})
ie_megan.makeInference()
gnb.showProba(ie_megan.posterior("Breast cancer"))
```
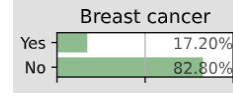
Figure 6: Using the LazyPropagation to compute the posterior probabilities

If we set the evidence for Megan and Denise, we see that given the fact that Megan's father has a BRCA1 gene mutation, the probability of Megan having breast cancer increases, while the probability of Denise having breast cancer decreases.
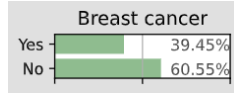
(a) The posterior probabilities for Megan (b) The posterior probabilities for Denise

4.2 Obesity check using BMI

The next step in the screening procedure is a BMI check. The Body Mass Index (BMI) of Megan and Denise is measured and the menopausal status is taken into account. This is done because in the postmenopausal state obesity increases the risk of breast cancer while in premenopausal state, obesity decreases the risk of breast cancer [4]. It is found that Megan has a BMI of 40 and Denise has a BMI of 21. Therefor Denise is moderate obese and Megan is severe obese. Furthermore Denise is 22 years old and in premenopausal state, whilst Megan is 55 years old and is in postmenopausal state. We update the network with this information and again compute the posterior probabilities. By setting the obese variable, the ancestors of the obese variable become independent of the breast cancer variable.

```python
ie_denise.setEvidence({'BRCA1-father': 'No', 'BRCA1-mother': 'No', 'Obese': 'Moderate', 'Menopause': 'Before'})
ie_denise.makeInference()
gnb.showProba(ie_denise.posterior("Breast cancer"))

ie_megan.setEvidence({'BRCA1-father': 'Yes', 'BRCA1-mother': 'No', 'Obese': 'Severe', 'Menopause': 'After'})
ie_megan.makeInference()
gnb.showProba(ie_megan.posterior("Breast cancer"))
```

Figure 8: Using the LazyPropagation to compute the posterior probabilities

(a) The posterior probabilities for Megan (b) The posterior probabilities for Denise

5 Screening result and recommendations

Given the evidence from the family history and the BMI check, Denise has a 2.7% chance of developing breast cancer and therefor can go home immediately. Megan on the other hand has a 78.11% chance of developing breast cancer and therefor needs further examination. Fortunately the result of Megan's mammography is negative, and therefor Megan can also go home. However the doctors have some recommendations for Megan. Using the Bayesian network to compute what the probability of breast cancer would be if Megan would exercise more regularly and eat healthy, the probability of developing breast cancer could decrease to 41.15%.

```
causal_megan = csl.CausalModel(network)

cslnb.showCausalImpact(causal_megan,
                "Breast cancer",
                doing={"Physical activity", "Diet"},
                knowing={'BRCA1-father','BRCA1-mother','Menopause'},
                values={'Physical activity':'Regular', 'Diet':'Healthy','BRCA1-father': 'Yes', 'BRCA1-mother': 'No', 'Menopause': 'After'})
```

Figure 10: Using the causal model to compute impact of diet and physical activity on breast cancer risk

6 Discussion and conclusions

7 References

References

[1] Tanya Agurs-Collins, Sharon A Ross, and Barbara K Dunn. The many faces of obesity and its influence on breast cancer risk. *Frontiers in oncology*, 9:765, 2019.

[2] European Commission. Press release: European commission. `https://ec.europa.eu/commission/presscorner/detail/en/ip_22_5573`. Accessed: [19/06/2024].

[3] Eurostat. Nutritional habits statistics. `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Nutritional_habits_statistics#Consumption_of_fruit_and_vegetables`. Accessed: [19/06/2024].

[4] Laura García-Estévez, Javier Cortés, Silvia Pérez, Isabel Calvo, Isabel Gallegos, and Gema Moreno-Bueno. Obesity and breast cancer: a paradoxical and controversial relationship influenced by menopausal status. *Frontiers in oncology*, 11:705911, 2021.

[5] National Cancer Institute. Brca mutations: Cancer risk and genetic testing. `https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet#:~:text=While%20the%20prevalence%20in%20the,specific%20variants%2C%20called%20founder%20mutations`. Accessed: [19/06/2024].

[6] Our World in Data. Female population by age group. `https://ourworldindata.org/grapher/female-population-by-age-group`. Accessed: [23/06/2024].

[7] Aruna Surakasula, Govardhana Chary Nagarjunapu, and KV Raghavaiah. A comparative study of pre-and post-menopausal breast cancer: Risk factors, presentation, characteristics and management. *Journal of research in pharmacy practice*, 3(1):12–18, 2014.

[8] Susan G. Komen. Body weight and breast cancer risk. `https://www.komen.org/breast-cancer/risk-factor/weight/#:~:text=Body%20weight%20and%20breast%20cancer%20risk%20after%20menopause,weight%20throughout%20life%20%5B27%5D`. Accessed: [19/06/2024].

[9] Wikipedia. Brca1. `https://en.wikipedia.org/wiki/BRCA1`. Accessed: [21/06/2024].