

Domain description

Introduction

The domain I have selected for the third project of Applied Artificial Intelligence is Parkinson's disease. I have modeled this disease as a Bayesian network of causes and symptoms. Parkinson's disease (PD) is a neurodegenerative disease and has a lot of important research ongoing to find its causes and use that to find a cure. It is a really interesting topic because after so much research the causes for this disease have not been identified by the medical community.

I have referred to medical information available online to find the various risk factors and symptoms of the disease and have modeled the network based on the same. The nodes and the corresponding Conditional Probability Tables (CPTs) are based on this information. In some places I have used my own probability values because there is not information I could find.

Based on the information available, the disease's probable causes are:

- Exposure to chemicals (listed as nodes) at occupation
- Age
- Genetic mutation of any of the three genes (Park2, LRRK2, GBA)

Even though there is no concrete evidence in the real world about how much these factors play into causing the disease itself, this network works on probabilities and predicts the possibility of having the disease based on various factors.

Explanation of nodes

There are 24 nodes in total

- The node "Occupational exposure" and the chemical nodes connecting to it
 - These are the different chemicals and their degree of exposure to the person and how they affect the overall exposure level of a person who might go to doctor with this information to check if he has the disease
 - The occupational exposure node is a deterministic node – because it has a single value depending on the values of its parents which can be either "high", "low" or "medium"
 - The chemical nodes each have two states – Exposed and NotExposed

- The chemical nodes states together in the table when NotExposed is ≤ 3 , then the occupational exposure is considered “Low”, when equal to 4, it is “medium” and greater than 4 is “high”
- The node “Age”
 - This node has two states – young age from 35 to 55 and old age above 55 to 90
 - This node is to model the two types of parkinson’s disease that can happen – early onset & normal
- The node “Genetic mutation” and the other nodes connecting to it directly and indirectly
 - The PARK2, LRRK2 & GBA are genes whose mutation can cause the disease. These genes being mutated also depend if the person is from the general population or if he/she belongs to specific ethnic group like “North African Arab” or “Ashkenazi Jewish”. According to medical information, the general population is more likely to have mutations in the GBA gene and the mentioned specific ethnic groups in the LRRK2 gene. The CPTs of both the genes reflect this information as probability of the mutation occurring.
 - The Genetic mutation node is an aggregate of the genetic reasons that influence the presence of the disease. This again is a deterministic node with its result being how probable the mutation is present based on the parent nodes. It is completely absent only when all the parent gene nodes have values absent
- The node “Parkinson’s” and the nodes connected to it as observables and the incoming nodes
 - The presence of the disease itself is influenced by the “Age”, “genetic mutation” and “Occupational exposure”. Based on the values of these three nodes, the disease node shows the probability of the disease being present for the person. The combinations of presence of disease based on the three parents is done by taking into account early age and presence of other factors and then older age.
 - The nodes “Pseudobulbar effect”, “Melanoma”, “Osteopathic hypotension” are effects of having Parkinson’s disease. The first and the last are influenced by having PD whereas Melanoma is also influenced by the mutation in the LRRK2 gene. This suggests a commonality between the two diseases.

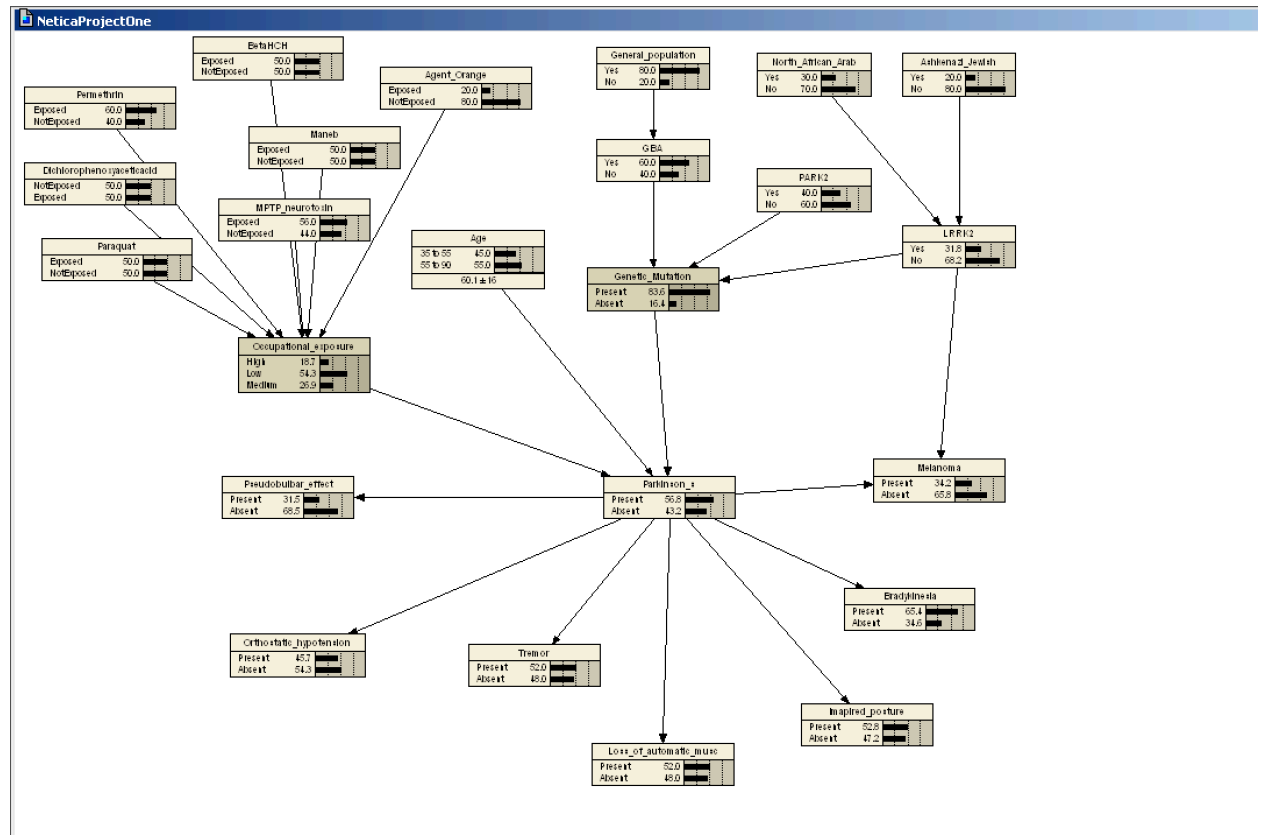
- The nodes “Tremor”, “Loss of automatic muscles”, “Impaired posture” & “Bradykinesia” are direct observables of the disease and are symptoms of the presence of PD.

Evaluating the Bayes network

I have supplied the .neta file that contains this Parkinson’s Bayesian network. You can open it in Netica and play around with it. You can also check the different CPTs by right clicking a node and going to the “Table”.

- Open the file in Netica. File > Open
- Go to Network > Compile
- Now the net is ready and you can play around with it by changing the probability values of the different nodes to check how the causes influence the presence of the disease. And also by setting values for the symptoms you can see how probable it is to get the disease for the non-causal direction

When you compile the Bayes net you should see a network like this picture here



If you encounter a problem compiling or generating the net, kindly let the Professor know before you evaluate so I can let you know how any issue you face could be solved. Thank you in advance.

For the nodes explanation, kindly check the corresponding CPTs to see how the values are achieved.