Differential Diagnosis of Covid-19 and Influenza type diseases with a Bayesian network

1 Introduction

1.1 Project description

Bayesian networks are powerful artificial intelligence tools that can be used to update disease probabilities based on symptom observations.

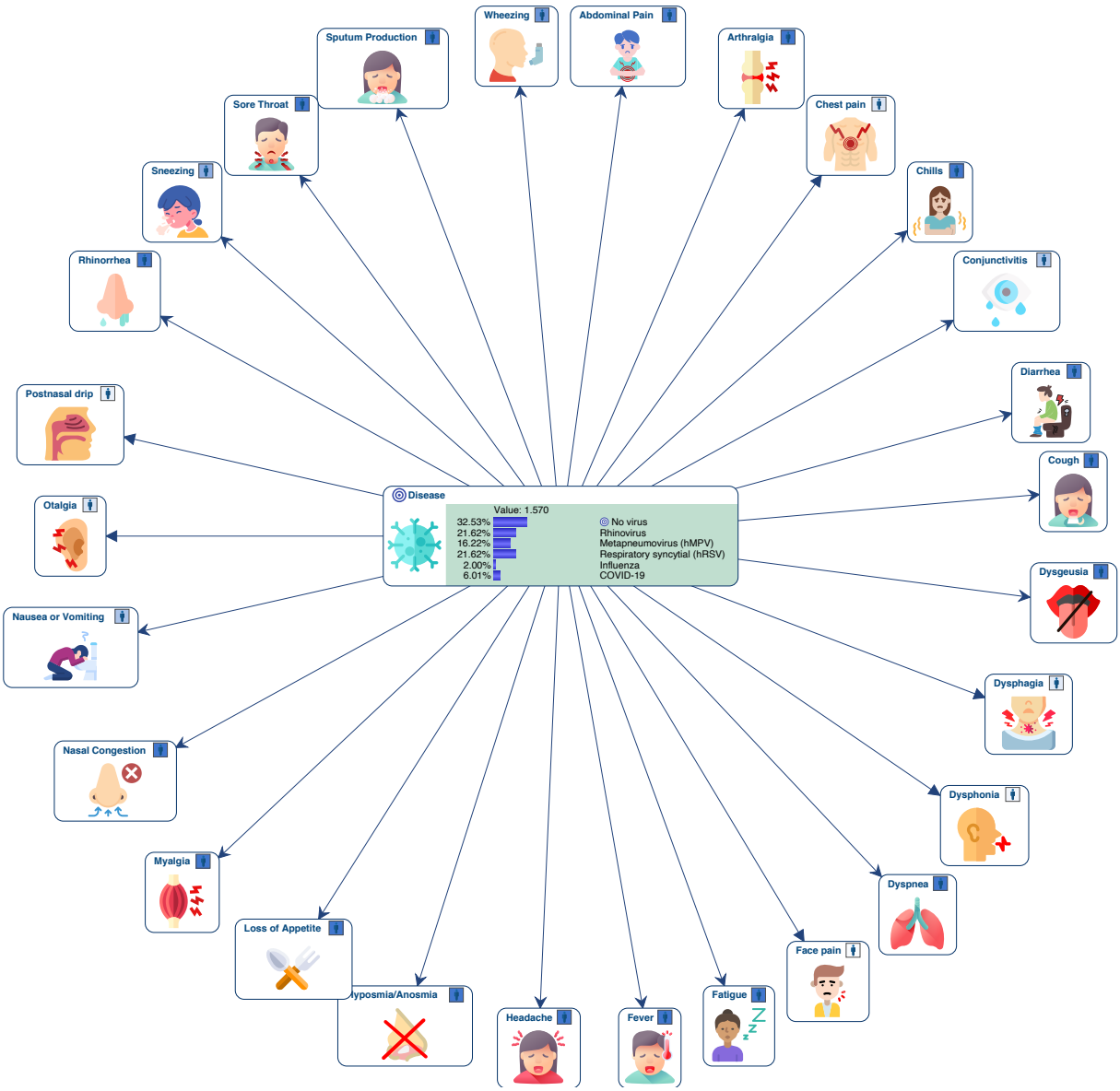
This is typically the situation you are facing in this pandemic period, with the observation of symptoms shared by both Covid-19 and other Influenza-Like Illnesses or allergies. You have then to take into account the likelihoods that these diseases are causing the observed symptoms and calculate, in a rigorous way, the probabilities of each disease knowing these symptoms. Many studies have shown that this task is not at all adapted to our way of reasoning. It is therefore necessary to rely on probabilistic calculation tools, the most efficient one to date being a Bayesian network.

Bayesia is currently the world leader in the field of Bayesian networks. Bayesia counts among its customers the FDA, EPA, NASA, US Air Force, Procter&Gamble, Airbus, ... not to mention some 300 universities (UCLA, Johns Hopkins, CMU ...).

The huge advantage of Bayesian networks is that they can be learned automatically by using machine learning algorithms (like neural networks), but they can also be built without large databases, or even without data set at all, only based on expert knowledge. This is precisely the framework of this project, the exploitation of your knowledge to build the best possible model, while waiting for "patient" data sets.

2 The Model

We have created a first version of a differential diagnosis model (see below) using data from bibliographic research, and expert knowledge coming from ICO (France), the University of Mons (Belgium), and the University of Sydney (Australia).



This model is obviously not perfect (no direct relationships between symptoms), but it is a first step (the addition of intra-symptom relationships will require the elicitation of many more probabilities).

The quantitative part of this model is mainly based on conditional probabilities, i.e. the likelihoods that the diseases are the causes of the symptoms.

2.1 Prior - Prevalence

The probability distribution of **Disease** corresponds to the prevalence of each disease, normalized on this subset of diseases. This distribution needs to be adapted over time, as influenza becomes more rare and as Covid-19 spreads. It may also be necessary to have a specific prevalence

distribution by geographic area. A default distribution has been set in the model. However, it can be modified manually before starting the diagnosis session.

2.2 Likelihoods

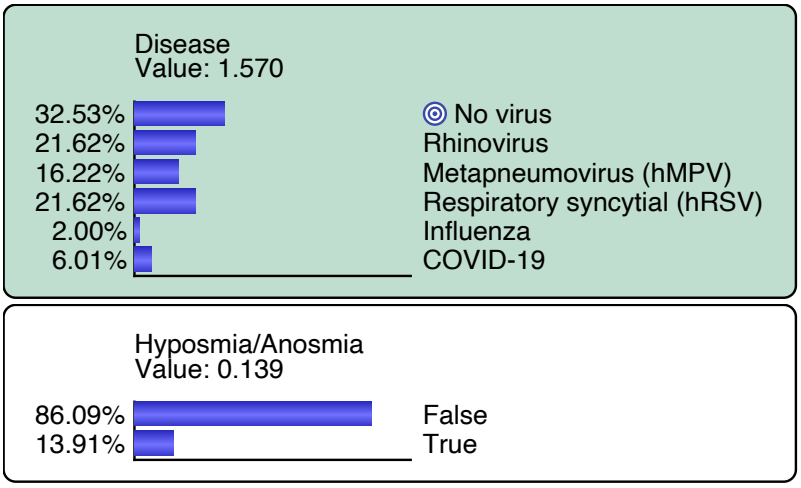
Let's take the example of **Hyposmia/Anosmia**. We need to know the likelihood that each disease is the cause of this symptom.

Disease	False	True
No virus	93.773	6.227
Rhinovirus	86.085	13.915
Metapneu...	89.470	10.530
Respirator...	89.735	10.265
Influenza	91.873	8.127
COVID-19	20.348	79.652

We specify in this table that 73.65% of the patients infected by Covid-19 suffer from Hyspomia/Anosmia, compared to 8.1% for those infected by influenza ...

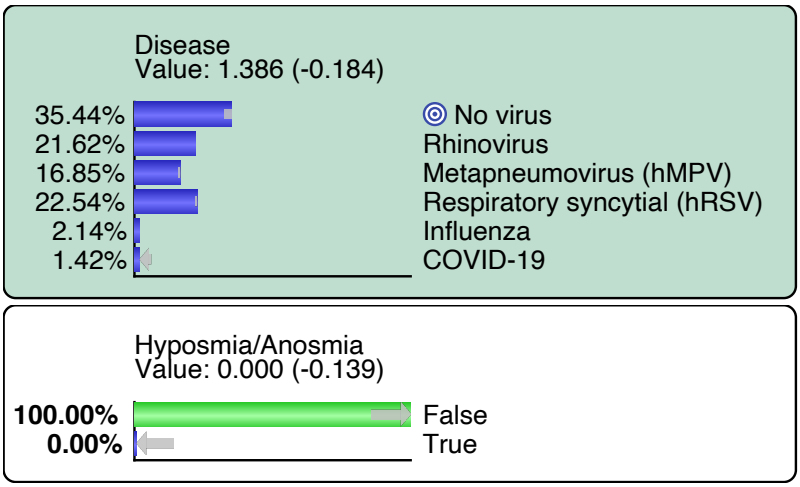
2.3 Probabilistic Inference

Let's suppose we just focus on the relationship **Disease - Sore Throat**. We show below the results of three probabilistic inferences:



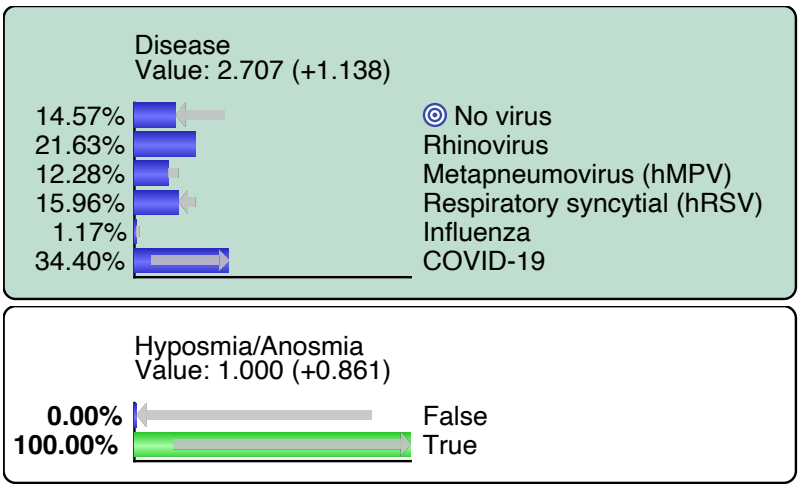
Prior Distributions

The probability of having Covid-19 without any information on symptoms is 6%, compared to a 2% risk of having the flu. The probability to have a patient with Hyposmia/Anosmia is 14%.



Posterior distribution, given no Hyposmia/Anosmia

Without Hyposmia/Anosmia, the probability of Covid-19 decreases to 1.42%, whereas the probability associated with Influenza increases to 2.14%.

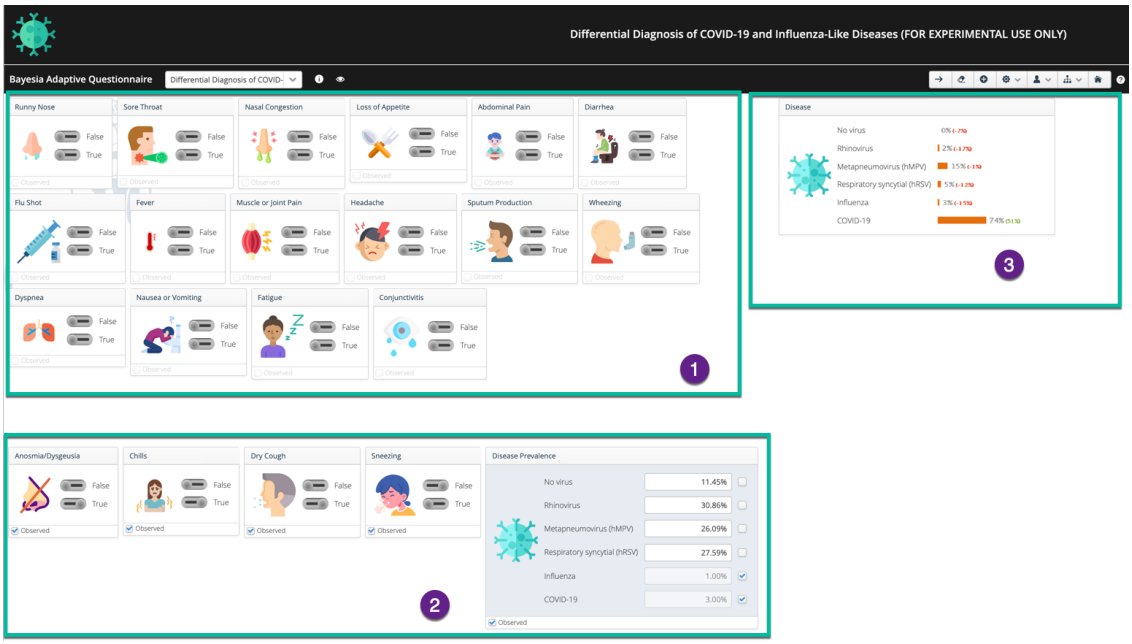


Posterior distribution, given Hyposmia/Anosmia

However, if Hyposmia/Anosmia is observed, the probability of Covid-19 increases to 34.4%, compared to 1.17% for Influenza.

2.4 Web-Simulator

The current version of the model is already available on Internet : https://covid19.bayesialab.com/#!questionnaire/101412724590



It's an Adaptive Questionnaire:

1. List of symptoms sorted by interest to reduce uncertainty about disease. This list is dynamically sorted after each observation
2. List of observed symptoms
3. Probability of the diseases given the observed symptoms

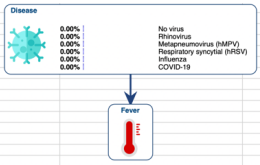
3 Probability Elicitation

The current network requires your intervention! We need your expertise to obtain the most realistic likelihoods for each disease/symptom.

You can simply share your knowledge by filling out the table that is in the "YOUR EXPERTISE" sheet of the attached Excel file.

The screen shot below give you an example of the likelihoods associated with Fever.

Disease	Fever	
	False	True
No virus	95.000	5.000
Rhinovirus	90.000	10.000
Metapneumovirus (hMPV)	80.000	20.000
Respiratory syncytial (hRSV)	75.000	25.000
Influenza	65.000	35.000
COVID-19	30.000	70.000



This relationship between "Disease" and "Fever" is described in the table below.
We need your expert knowledge to populate this table (see Sheet "Your Expertise")

	Diseases					
Symtoms	**No virus**	**Rhinovirus**	**Metapneumovirus (hMPV)**	**Respiratory syncytial (hRSV)**	**Influenza**	**COVID-19**
Abdominal Pain						
Anosmia/Dysgeusia						
Chills						
Conjunctivitis						
Diarrhea						
Dry Cough						
Dyspnea						
Fatigue						
Fever	**5%**	**10%**	**20%**	**25%**	**35%**	**70%**
Headache						
Loss of Appetite						
Muscle or Joint Pain						
Nasal Congestion						
Nausea or Vomiting						
Runny Nose						
Sneezing						
Sore Throat						
Sputum Production						
Wheezing						

4 Data

If you have access to patient data sets, this data has to anonymized.

We need a row per patient. Each row describes the symptoms and the associated diagnosed disease for the patient, without any information that could be used for identifying the patient.

5 Questions?

Please feel free to reach out for any questions, either by email (jouffe@bayesia.com), or by phone at +33 6 07 25 70 05.

Thanks in advance for your help!

Dr. Lionel Jouffe
CEO Bayesia