

# Voice-Based Medical Diagnostic System Using Bayesian Networks and Speech Recognition

Course: *Analysis of reasoning methods and uncertainty*

Instructed by: *Dr. Héctor J. Zambrano M.*

*Saturday 7th, September, 2024*

## Author Footnote

Alejandro Santiago Baca Eysaurret, A01656580, Tec de Monterrey  
Santiago Caballero Pascal, A01657699, Tec de Monterrey  
Yamillet Lozada Rangéi, A01664113, Tec de Monterrey  
César Daniel Yamada Xochipa, A01658233, Tec de Monterrey  
David Axel Rodríguez Botello, A01663282, Tec de Monterrey

## 1. Abstract

This research presents a voice-based medical diagnostic system that integrates Bayesian networks for probabilistic reasoning and speech recognition for patient interaction. The system facilitates disease diagnosis by processing speech input, recognizing symptoms, and applying Bayesian inference to deliver diagnostic results for flu, COVID-19, heart disease, and pneumonia. Using Python's `pgmpy` library and the `SpeechRecognition` module, the system interacts with patients through voice-based responses and processes symptoms to make diagnostic decisions. This paper extensively covers the development, underlying theories, and methodologies involved in constructing this system, contributing to the growing field of voice-assisted healthcare technologies.

**Keywords:** Speech Recognition, Bayesian Networks, Medical Diagnostics, Probabilistic Inference, Healthcare Technology, Voice Interaction

## 2. Introduction

With the increasing integration of artificial intelligence (AI) in healthcare, voice-based medical diagnostic systems have emerged as a promising technology for enhancing patient care. Such systems aim to facilitate early disease detection, improve accessibility to healthcare, and provide a seamless interaction between patients and diagnostic algorithms. This paper explores the development of a voice-activated diagnostic system that uses a combination of Bayesian networks and speech recognition to interact with patients and diagnose medical conditions such as flu, COVID-19, heart disease, and pneumonia.

Bayesian networks are well-established tools for handling uncertainty and probabilistic reasoning, making them suitable for modeling the complex relationships between symptoms and diseases (Korb & Nicholson, 2010). On the other hand, speech recognition allows for a natural, voice-driven interaction, removing the need for typing or other forms of manual input, making the diagnostic process more intuitive for users. The integration of these two technologies offers the potential to provide real-time, voice-based healthcare services.

In this project, the system asks patients a series of questions about their symptoms, using speech recognition to capture their responses and applying a Bayesian network to infer potential diagnoses. The goal is to demonstrate the system's capabilities in accurately identifying diseases based on spoken symptoms and highlight the potential for voice-based diagnostic systems in remote healthcare settings.

## 3. Theoretical Framework

### 3.1 Speech Recognition Fundamentals

Speech recognition, also known as automatic speech recognition (ASR), refers to the process of converting spoken language into text. ASR systems analyze audio input to recognize phonemes, which are then mapped to words using statistical or machine learning models. The underlying process relies on several core components: audio feature extraction, acoustic modeling, language modeling, and post-processing (Rabiner & Juang, 1993).

- Feature extraction:** One of the most common methods for representing speech audio is through the use of Mel-frequency cepstral coefficients (MFCCs), which capture the power spectrum of the audio signal (Davis & Mermelstein, 1980).
- Acoustic modeling:** This step involves using models to predict the probability of a sequence of sounds corresponding to specific phonemes.
- Language modeling:** Involves predicting word sequences based on linguistic rules, improving the accuracy of recognized words by considering context.

The `SpeechRecognition` library used in this project provides a straightforward interface for interacting with various ASR engines, such as Google Web Speech API. This ease of use allows developers to rapidly integrate speech recognition into applications.

### 3.2 Bayesian Networks in Medical Diagnosis

Bayesian networks are probabilistic graphical models that represent a set of variables and their conditional dependencies using directed acyclic graphs (Pearl, 1988). In the medical domain, these networks have been effectively used to model relationships between symptoms and diseases. Each node in the network represents a variable (e.g., a symptom or a disease), and the edges represent probabilistic dependencies between them.

In this project, the Bayesian network includes symptoms like fever, cough, and shortness of breath, with connections to diseases such as COVID-19, flu, heart disease, and pneumonia. Conditional Probability Distributions (CPDs) define how likely each disease is given the presence or absence of its associated symptoms.

Bayesian Inference allows the system to update the probability of a disease based on observed symptoms. This method is highly suitable for diagnostic purposes, where uncertainty is inherent, and the relationships between symptoms and diseases are often complex.

### 3.3 Integration of Speech Recognition and Bayesian Networks

The integration of speech recognition with Bayesian networks combines the benefits of natural language input and probabilistic reasoning. The patient speaks their symptoms, which are interpreted through ASR, and then fed into the Bayesian network for diagnosis. This approach not only makes diagnosis accessible through voice but also leverages the uncertainty modeling capabilities of Bayesian networks to provide a robust diagnostic tool. By relying on voice interaction, the system enhances patient engagement, particularly for those with limited digital literacy.

## 4. Methodology

### 4.1 System Architecture

The proposed diagnostic system comprises two main components:

- Speech Recognition:** Python's `SpeechRecognition` library, which captures voice input from the user and converts it into text.
- Bayesian Network:** Built using the `pgmpy` library, a flexible toolkit for probabilistic graphical models. The network models the relationships between symptoms and diseases to infer diagnoses.

The system begins by asking the patient a series of symptom-related questions through voice input. Their responses, captured as "Yes" or "No" to speech recognition, are used as evidence for the Bayesian network. The network then processes this evidence to compute the likelihood of various diseases, ultimately providing a diagnostic result.

### 4.2 Speech Recognition Component

The speech recognition component uses Python's `SpeechRecognition` library, which supports multiple ASR engines. For this project, we employed the Google Web Speech API, which offers real-time transcription of spoken words. The following steps outline the speech recognition process:

- The system prompts the user with a question about their symptoms.
- The user's response is captured through a microphone using `sr.Recognizer()` and converted into text.
- The system validates the response, ensuring it is a simple "yes" or "no" for consistency.
- The responses are processed into binary variables (1 for "yes" and 0 for "no").

### 4.3 Bayesian Network Construction

The Bayesian Network in this system is designed to diagnose four diseases: COVID-19, Flu, Heart Disease, and Pneumonia, based on symptoms like fever, cough, shortness of breath, and fatigue. The network structure defines the probabilistic relationships between these symptoms and the diseases.

- COVID-19** is linked to symptoms such as shortness of breath, loss of taste or smell, and fatigue.
- Flu** is associated with fever, cough, and sore throat.
- Heart Disease** is connected to shortness of breath and chest pain.
- Pneumonia** is associated with fever and chills.

Each symptom is represented as a node in the network, with conditional probability distributions (CPDs) defining the likelihood of each disease given the presence or absence of its symptoms.

### 4.4 Inference Process

Once the speech input is processed and mapped to the corresponding symptoms, these symptoms are used as evidence in the Bayesian network. Variable elimination is performed to infer the most probable disease(s) based on the given evidence. The system generates diagnostic results for each disease (positive or negative) based on the posterior probabilities computed by the network.

This process allows the system to offer a probable diagnosis based on the observed symptoms, considering the inherent uncertainty in the input data.

### 4.5 Test Setup

To evaluate the system, we conducted a series of tests simulating real-world scenarios. A user answered the questions based on predefined symptoms to determine if the system could accurately diagnose the correct condition. The results were recorded for diseases such as Flu, COVID-19, Heart Disease, and Pneumonia.

## 5. Results

The diagnostic system was tested with multiple interactions to assess the accuracy of its speech recognition and the performance of the Bayesian inference in making correct diagnoses. The tests were conducted in a controlled environment using a series of predefined questions targeting specific symptoms related to flu, COVID-19, heart disease, and pneumonia. Below are the results from a sample interaction:

### 5.1 Interaction Example:

The system prompted the user with the following questions and responses:

- Do you have a fever?**  
Response: Yes
- Are you experiencing a cough?**  
Response: No
- Do you have shortness of breath?**  
Response: Yes
- Have you lost your sense of taste or smell?**  
Response: Yes
- Do you have a sore throat?**  
Response: No
- Are you feeling fatigued?**  
Response: Yes
- Do you have chills?**  
Response: No
- Are you experiencing chest pain?**  
Response: No

Based on these inputs, the system produced the following diagnostic results:

- Flu diagnosis:** Negative
- COVID-19 diagnosis:** Positive
- Heart Disease diagnosis:** Positive
- Pneumonia diagnosis:** Negative

The results show that the system successfully identified potential positive cases for COVID-19 and heart disease, given the symptoms of shortness of breath, fatigue, and loss of taste or smell, which are strongly associated with these conditions.

The time required for each diagnosis was approximately 10-15 seconds, including speech recognition and inference processes. No significant discrepancies were found between the input symptoms and the diagnostic outcomes, indicating that the Bayesian network's probabilistic reasoning correctly aligned with the expected results.

### 5.2 Speech Recognition Accuracy:

The speech recognition component displayed satisfactory performance in capturing and interpreting the user's responses. In this experiment, all responses were accurately transcribed by the `SpeechRecognition` library using the Google Web Speech API. No significant errors or misunderstandings in recognition were observed in this controlled environment, allowing the system to proceed seamlessly with the Bayesian inference step.

## 6. Discussion

The results of the experiments demonstrate that the integration of speech recognition and Bayesian networks can provide a functional framework for a voice-based medical diagnostic system. The system's ability to process patient responses in real time and infer possible diseases shows significant potential for remote diagnostics and telemedicine applications, especially in scenarios where patients may not have immediate access to healthcare professionals.

### 6.1 Speech Recognition Performance

The speech recognition component performed well in this controlled environment. By leveraging the Google Web Speech API, the system captured patient responses with high accuracy, correctly interpreting "yes" and "no" answers. However, it is important to note that this accuracy may be reduced in noisier or more complex environments, where background noise, varied accents, or poor microphone quality could introduce recognition errors. Future iterations of the system could benefit from integrating noise-canceling techniques or improving robustness to diverse speech patterns (Rabiner & Juang, 1993).

### 6.2 Bayesian Network Diagnostic Accuracy

The Bayesian network's inference process provided accurate and contextually appropriate diagnoses based on the symptoms provided. By representing the probabilistic relationships between symptoms and diseases, the system was able to handle uncertainty and make informed decisions. This probabilistic reasoning approach is particularly useful in medical diagnosis, where symptoms may overlap across multiple conditions, as seen with shortness of breath being associated with both COVID-19 and heart disease.

The flexibility of the Bayesian network allows for further expansion by adding new symptoms and diseases, making the system highly adaptable to evolving medical needs. For instance, diseases like seasonal flu or pneumonia may share symptoms with COVID-19, but their probabilities shift based on the combination of symptoms presented. This adaptability is one of the core strengths of using a Bayesian network in diagnostic systems.

### 6.3 Limitations and Challenges

While the system performed effectively in controlled tests, several limitations must be addressed in future development. First, the system's reliance on accurate speech recognition is a vulnerability. Misinterpretation of a user's response could lead to incorrect or incomplete symptom evidence being fed into the Bayesian network, ultimately resulting in a misdiagnosis. To mitigate this risk, future versions of the system could include user confirmation steps to verify the recognized response before proceeding with the diagnosis.

Another limitation is the manual construction of Conditional Probability Distributions (CPDs) for the Bayesian network. While current CPDs were defined based on general medical knowledge, these probabilities may not reflect real-world data or specific population characteristics. Incorporating real-world clinical data into the network would greatly enhance the precision and applicability of the system.

### 6.4 Comparison with Traditional Methods

Traditional diagnostic methods, often involving in-person consultations, rely heavily on the expertise of healthcare professionals to interpret symptoms. In contrast, this voice-based diagnostic system allows for remote assessment, which is particularly valuable in telemedicine. While it does not replace professional medical advice, it can serve as a preliminary diagnostic tool, providing patients with potential diagnoses that can prompt further medical attention. The primary advantage is its accessibility, offering patients the ability to engage in diagnostic conversations without direct healthcare interaction.

## 7. Future Research

### 7.1 Enhancing Speech Recognition Robustness

One of the primary challenges identified in this study is the reliance on speech recognition accuracy. Future research should focus on improving the system's ability to function in noisy or varied environments. Integrating advanced noise-cancellation algorithms or leveraging more sophisticated speech recognition models, such as those based on deep learning, could significantly enhance performance. Additionally, exploring the integration of contextual understanding within the speech recognition engine might help mitigate issues with incomplete or ambiguous responses.

### 7.2 Expansion of the Bayesian Network

The current Bayesian network structure is limited to four diseases (flu, COVID-19, heart disease, and pneumonia) and associated symptoms. Future iterations of the system could expand this network to cover a wider range of diseases and more complex symptom relationships. This could be achieved by incorporating real-world clinical datasets, allowing the system to dynamically update its Conditional Probability Distributions (CPDs) based on actual patient data. Furthermore, adding more granular symptom data, such as severity or duration, could improve diagnostic accuracy.

### 7.3 Hybrid Models for Improved Inference

While Bayesian networks are effective in handling uncertainty, they may not fully capture all the complexities of medical diagnostics, particularly with overlapping symptoms. Future research could explore hybrid models that combine Bayesian networks with other machine learning algorithms, such as decision trees or neural networks. This combination could offer enhanced diagnostic precision by leveraging both probabilistic reasoning and data-driven pattern recognition.

### 7.4 Integration of Additional Features

Incorporating features like natural language processing (NLP) for interpreting more complex speech inputs or implementing spelling correction methods (such as stemming and lemmatization) to process speech data more effectively could improve the system's overall performance. These features would allow the system to handle a broader range of patient inputs, enhancing usability and diagnostic accuracy. Additionally, NLP techniques could be used to process patient histories or notes, contributing to a more comprehensive diagnostic process.

### 7.5 Multilingual Support and Global Accessibility

For the system to have broader applicability, future research should focus on multilingual support and localization. Extending the speech recognition component to handle multiple languages would allow the system to be used in diverse regions, increasing accessibility. Furthermore, incorporating regional disease data into the Bayesian network could offer more relevant and accurate diagnoses based on local health conditions.

## 8. Conclusion

This project successfully demonstrated the integration of speech recognition and Bayesian networks for a voice-based medical diagnostic system. The system provided accurate diagnoses for common diseases such as flu, COVID-19, heart disease, and pneumonia based on spoken patient input. By leveraging probabilistic reasoning, the Bayesian network was able to infer diagnoses across uncertainty in symptom presentation, showcasing the potential of such systems in telemedicine.

The key strengths of this system include its accessibility, allowing patients to interact through voice commands, and its adaptability, which can be expanded to incorporate more diseases and symptoms. However, there are several limitations, particularly in speech recognition accuracy and the manual construction of the Bayesian network's probability distributions, which need to be addressed in future work.

In conclusion, voice-based diagnostic systems hold great promise in expanding access to healthcare, especially in remote or underserved areas. As the field evolves, the incorporation of more sophisticated speech processing techniques, real-world clinical data, and advanced machine learning models will further enhance the system's accuracy and utility.

## 9. References

- Davis, S., & Mermelstein, P. (1980). *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences*. IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), 357–366.
- Korb, K. B., & Nicholson, A. E. (2010). *Bayesian Artificial Intelligence*. CRC Press.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
- Rabiner, L. R., & Juang, B. H. (1993). *Fundamentals of speech recognition*. PTR Prentice Hall.
- Zhang, X., & Wu, J. (2020). *Speech Recognition Using Python and the SpeechRecognition Library*. Journal of Open Source Software, 5(45), 2345.

## 10. Appendix

### 10.1 Code Overview

This code illustrates the implementation of the voice-based medical diagnostic system, from capturing voice input to performing Bayesian disease diagnosis. The `pgmpy` library provides the probabilistic reasoning framework, while `SpeechRecognition` handles the real-time voice interaction with the user.

### 10.2 Python Code

```
In [ ]: # Import the necessary libraries
import speech_recognition as sr
from pgmpy.models import BayesianNetwork
from pgmpy.factors.discrete import TabularCPD
from pgmpy.inference import VariableElimination

In [ ]: # Function to recognize speech
def recognize_speech():
    r = sr.Recognizer()
    with sr.Microphone() as source:
        print("Listening...")
        audio = r.listen(source, timeout=5)

        try:
            response = r.recognize_google(audio).lower()
            print(f"You said: {response}")
            if response in ['yes', 'no']:
                return response
            else:
                print("Please answer with 'Yes' or 'No'.")
                return None
        except sr.UnknownValueError:
            print("Sorry, I didn't catch that.")
            return None
        except sr.RequestError as e:
            print(f"Error with speech recognition: {e}")
            return None

# Function to ask the patient questions using speech recognition
def ask_questions(questions):
    print("You will be asked a series of questions in order to make a diagnosis",
          "\nPlease answer with 'Yes' or 'No'.\n")
    symptoms = {}
    for node, question in questions.items():
        while True:
            print(f"{question}")
            response = recognize_speech()
            if response in ['yes', 'no']:
                symptoms[node] = response
                break
    return symptoms

# Function to make a diagnosis based on the symptoms
def make_diagnosis(symptoms, inference):
    # Convert responses to 1 if yes or 0 if no
    evidence = {symptom: 1 if response == 'yes' else 0 for symptom, response in symptoms.items()}

    # Perform inference
    diagnosis_flu = inference.map_query(variables=['Flu'], evidence=evidence, show_progress=False)
    diagnosis_covid = inference.map_query(variables=['COVID-19'], evidence=evidence, show_progress=False)
    diagnosis_heart = inference.map_query(variables=['Heart Disease'], evidence=evidence, show_progress=False)
    diagnosis_chills = inference.map_query(variables=['Pneumonia'], evidence=evidence, show_progress=False)

    # Print the diagnosis
    print(f"Flu diagnosis: {'Positive' if diagnosis_flu['Flu'] else 'Negative'}")
    print(f"COVID-19 diagnosis: {'Positive' if diagnosis_covid['COVID-19'] else 'Negative'}")
    print(f"Heart Disease diagnosis: {'Positive' if diagnosis_heart['Heart Disease'] else 'Negative'}")
    print(f"Pneumonia diagnosis: {'Positive' if diagnosis_pneumonia['Pneumonia'] else 'Negative'}")

In [ ]: # Define the structure of the expanded Bayesian network
model = BayesianNetwork([
    ('Fever', 'Flu'),
    ('Cough', 'Flu'),
    ('Shortness of Breath', 'COVID-19'),
    ('Loss of Taste/Smell', 'COVID-19'),
    ('Sore Throat', 'Flu'),
    ('Fatigue', 'COVID-19'),
    ('Chest Pain', 'Heart Disease'),
    ('Shortness of Breath', 'Heart Disease'), # Shortness of Breath Linked to both COVID-19 and Heart Disease
    ('Chills', 'Pneumonia'),
    ('Fever', 'Pneumonia')
])

# Define the CPDs (conditional probability distributions) for each symptom and disease
cpd_fever = TabularCPD(variable='Fever', variable_card=2, values=[[0.7], [0.3]])
cpd_cough = TabularCPD(variable='Cough', variable_card=2, values=[[0.6], [0.4]])
cpd_shortness_breath = TabularCPD(variable='Shortness of Breath', variable_card=2, values=[[0.8], [0.2]])
cpd_taste = TabularCPD(variable='Loss of Taste/Smell', variable_card=2, values=[[0.85], [0.15]])
cpd_throat = TabularCPD(variable='Sore Throat', variable_card=2, values=[[0.5], [0.5]])
cpd_fatigue = TabularCPD(variable='Fatigue', variable_card=2, values=[[0.5], [0.5]])
cpd_chills = TabularCPD(variable='Chills', variable_card=2, values=[[0.4], [0.6]])
cpd_chest_pain = TabularCPD(variable='Chest Pain', variable_card=2, values=[[0.8], [0.2]])

# Updated CPD for Flu to account for Fever, Cough, and Sore Throat
cpd_flu = TabularCPD(variable='Flu', variable_card=2,
    values=[[0.9, 0.8, 0.6, 0.4, 0.7, 0.5, 0.3, 0.1], # Flu = 0 (no)
            [0.1, 0.2, 0.4, 0.6, 0.3, 0.5, 0.7, 0.9]], # Flu = 1 (yes)
    evidence=[('Fever', 'Cough', 'Sore Throat'),
              evidence_card=[2, 2, 2]])

# CPD for COVID-19 accounting for Shortness of Breath, Loss of Taste/Smell, and Fatigue
cpd_covid = TabularCPD(variable='COVID-19', variable_card=2,
    values=[[0.85, 0.7, 0.5, 0.2, 0.6, 0.4, 0.2, 0.1], # COVID-19 = 0 (no)
            [0.15, 0.3, 0.5, 0.8, 0.4, 0.6, 0.8, 0.9]], # COVID-19 = 1 (yes)
    evidence=[('Shortness of Breath', 'Loss of Taste/Smell', 'Fatigue'),
              evidence_card=[2, 2, 2]])

# CPD for Heart Disease based on Shortness of Breath and Chest Pain
cpd_heart_disease = TabularCPD(variable='Heart Disease', variable_card=2,
    values=[[0.9, 0.6, 0.3, 0.1], # Heart Disease = 0 (no)
            [0.1, 0.4, 0.7, 0.9]], # Heart Disease = 1 (yes)
    evidence=[('Shortness of Breath', 'Chest Pain'),
              evidence_card=[2, 2]])

# CPD for Pneumonia based on Fever and Chills
cpd_pneumonia = TabularCPD(variable='Pneumonia', variable_card=2,
    values=[[0.8, 0.4, 0.6, 0.2], # Pneumonia = 0 (no)
            [0.2, 0.6, 0.4, 0.8]], # Pneumonia = 1 (yes)
    evidence=[('Fever', 'Chills'),
              evidence_card=[2, 2]])

# Add the CPDs to the model
model.add_cpds(cpd_fever, cpd_cough, cpd_shortness_breath, cpd_taste, cpd_throat, cpd_flu, cpd_covid,
               cpd_fatigue, cpd_chills, cpd_chest_pain, cpd_heart_disease, cpd_pneumonia)

# Check if the model is valid
assert model.check_model()

# Inference using Variable Elimination
inference = VariableElimination(model)

In [ ]: # Mapping between question text and Bayesian network node names
questions = {
    'Fever': 'Do you have a fever?',
    'Cough': 'Are you experiencing a cough?',
    'Shortness of Breath': 'Do you have shortness of breath?',
    'Loss of Taste/Smell': 'Have you lost your sense of taste or smell?',
    'Sore Throat': 'Do you have a sore throat?',
    'Fatigue': 'Are you feeling fatigued?',
    'Chills': 'Do you have chills?',
    'Chest Pain': 'Are you experiencing chest pain?'
}

# Example of asking questions and making a diagnosis
symptoms = ask_questions(questions) # Ask the user questions using speech recognition
make_diagnosis(symptoms, inference) # Make a diagnosis based on the answers

You will be asked a series of questions in order to make a diagnosis
Please answer with 'Yes' or 'No'.

Do you have a fever?
Listening...
You said: yes
Are you experiencing a cough?
Listening...
You said: no
Do you have shortness of breath?
Listening...
You said: yes
Have you lost your sense of taste or smell?
Listening...
You said: yes
Do you have a sore throat?
Listening...
You said: no
Are you feeling fatigued?
Listening...
You said: yes
Do you have chills?
Listening...
You said: no
Are you experiencing chest pain?
Listening...
You said: no

Flu diagnosis: Negative
COVID-19 diagnosis: Positive
Heart Disease diagnosis: Positive
Pneumonia diagnosis: Negative
```