# A System to Diagnose Patients with Cough

Allen Wang[1] and Stylianos Rousoglou[1]

[1]LILY Lab, Department of Computer Science, Yale University, New Haven, CT

LILY Lab

## Introduction

This project involved evaluating different machine-learning and probabilistic techniques to build a diagnosis system for both doctors and patients. The work was split into two phases: determining the appropriate predictive model for diagnosing a principal symptom, and building a functional web-application around this model to demonstrate a proof-of-concept for clinical diagnosis support systems. The predictive model focused on producing diagnoses from a single principal symptom: cough. The web-application handles both doctor and patient inputs and generates a set of diagnoses probabilities as the results.

## Models and Methods

**Relative Likelihoods:**
- For the cough principal, there are 92 symptom-based questions and 23 possible diagnoses.
- For each diagnosis, all of the symptoms have a value relative to 1 that represents the likelihood of this diagnosis if the symptom is present (Table 1).
- Questions, diagnoses, and relative likelihoods compiled from Lajos Pusztai and Frederick Howard, our partners from the school of Medicine.

**Bayesian Network:**
- Lack of training data - could not train based on real patient diagnoses
- Conditional Probabilities - if known, allows for Bayesian networks implementation with small data sets
- Relative Likelihoods - independent probabilities of diagnoses given symptoms
- Unfeasible to generate conditional probabilities of diagnoses based on 92 symptoms
- Maximum Likelihood Estimator - also unfeasible despite requiring less conditional probabilities

**Decision Tree:**
- Lack of training data- generated sample data from self-generated patient inputs
- Small training set led to overtraining
- Classified patient inputs producing labels based on severity

| | Viral infection | Post-viral cough | Influenza | Pertussis | Pneumonia |
|---|---|---|---|---|---|
| < 40 | 1 | 1 | 1 | 1 | 1 |
| 40 - 50 | 1 | 1 | 1 | 1 | 1 |
| 50 - 60 | 1 | 1 | 2.0 | 1 | 2.0 |
| 60 - 80 | 1 | 1 | 5.0 | 1 | 5.0 |
| >80 | 1 | 1 | 10.0 | 1 | 10.0 |
| Gender (if female) | 1 | 1 | 1 | 1 | 1 |
| Ethnicity | 1 | 1 | 1 | 1 | 1 |
| Not influenza season | 1 | 1 | 0.001 | 1 | 1 |
| Sick contact | 2.0 | 2.0 | 1 | 1 | 1 |
| Contact with influenza | 1 | 1 | 5.0 | 1 | 1 |
| Contact with pertussis | 1 | 1 | 1 | 5.0 | 1 |

**Table 1.** Table of Relative Likelihoods for Cough From Doctor Interface

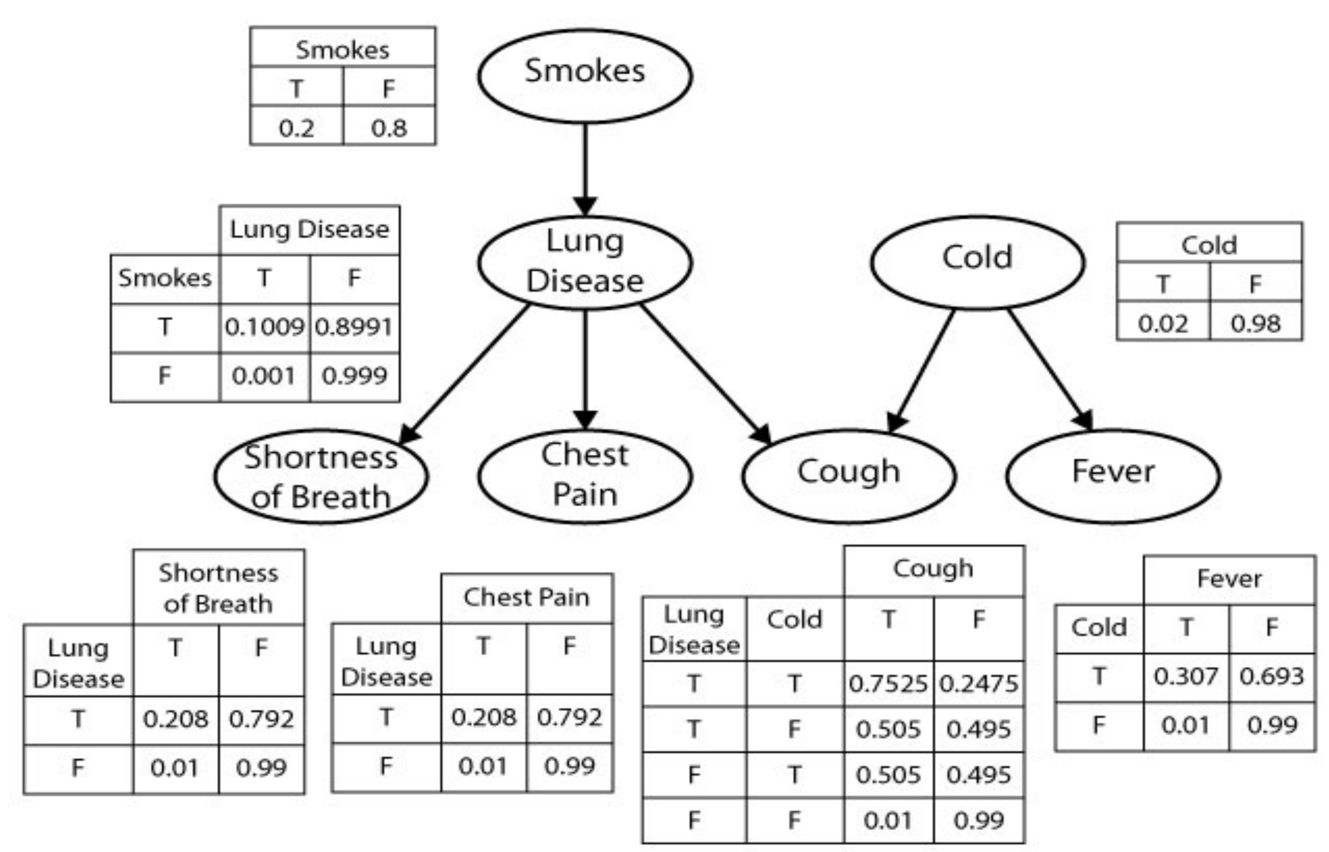


**Figure 1.** Bayesian Network And Conditional Probabilities for a few symptoms

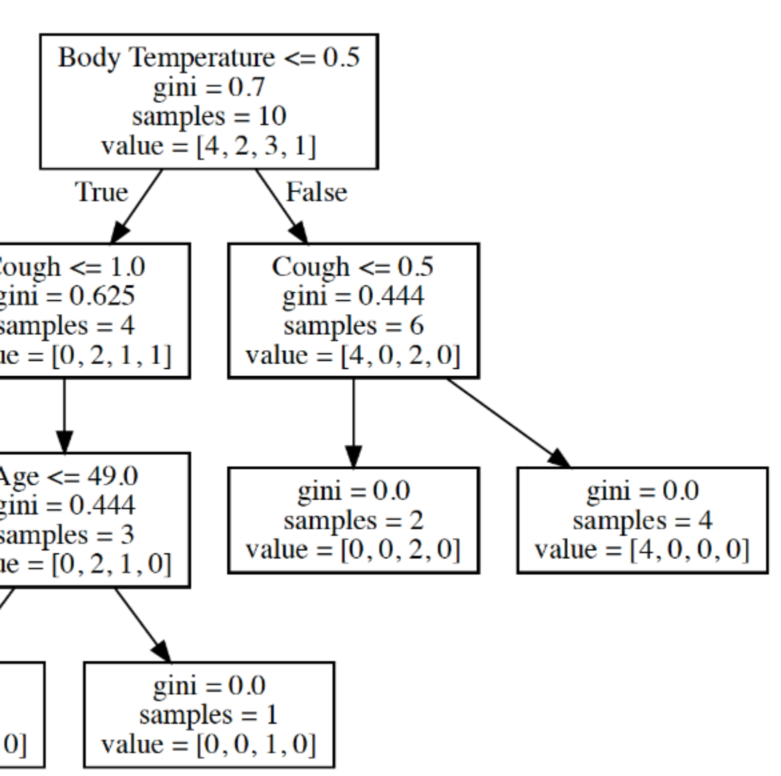


**Figure 2.** Cough Decision Tree

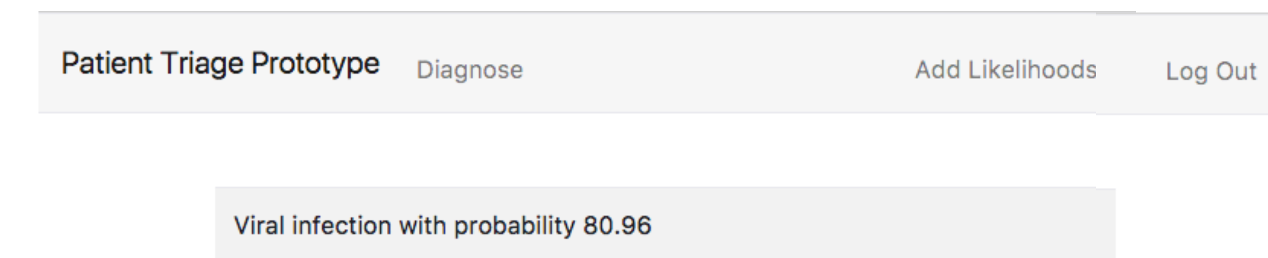


**Figure 3.** Web-application Predictive Model Results

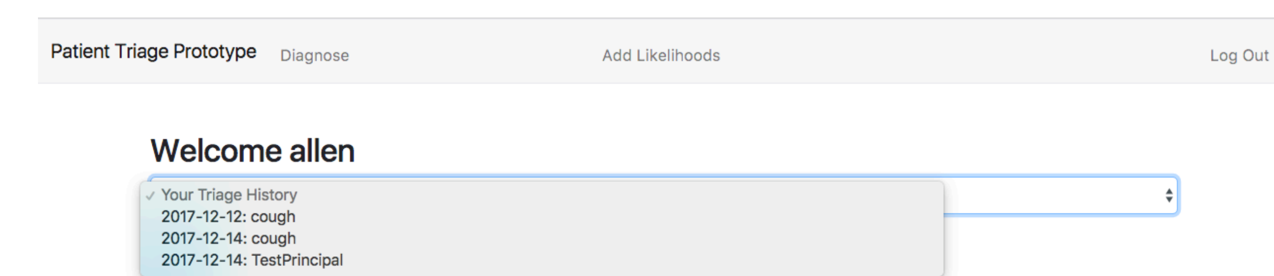


**Figure 4.** Web-application Patient Interface Diagnosis History

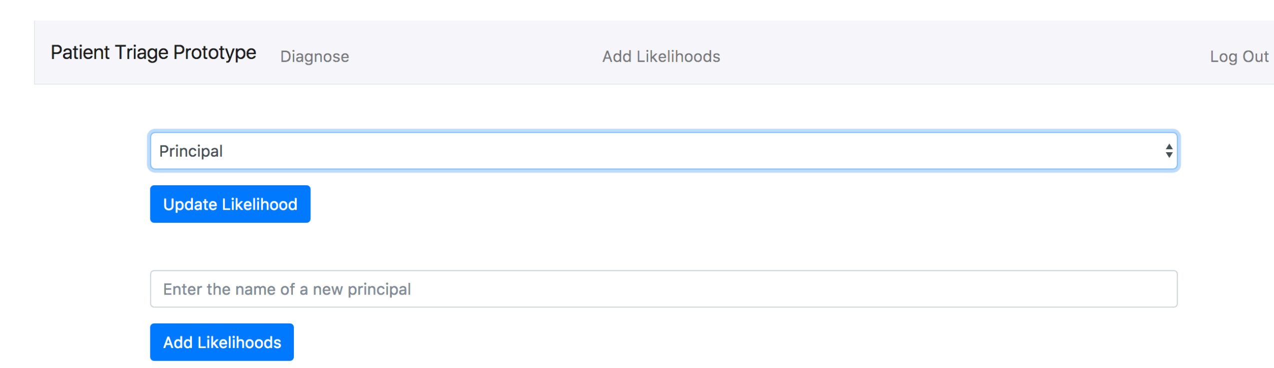


**Figure 5.** Web-application Doctor Interface Update/Add Likelihood

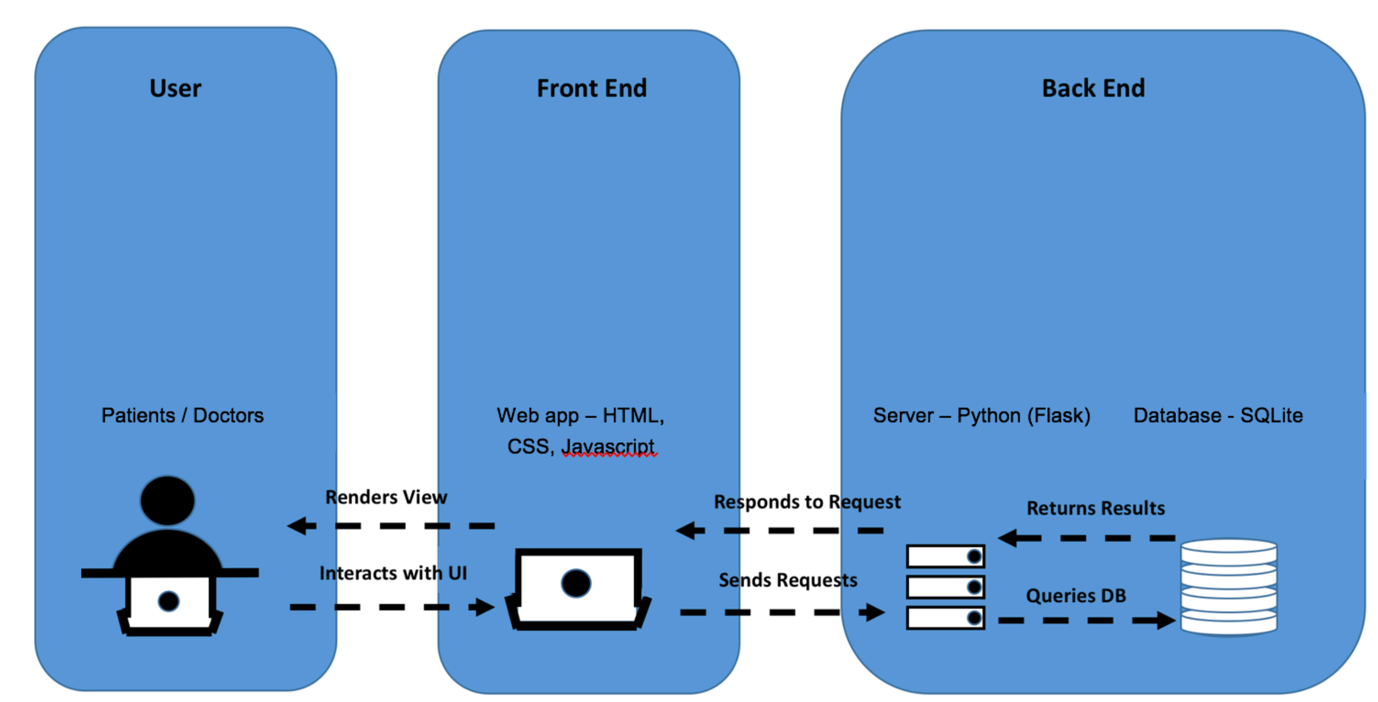


**Figure 6.** Web-application System Architecture

## Web Application Results

**Predictive Model:**
- Relative likelihoods combined with patient input calculations
- Resultant diagnoses likelihoods summed and normalized
- Diagnoses that meet our cutoff (5%) are returned

**Patient Interface:**
- Patients can view a history of their diagnoses
- Start a self-diagnosis by answering several question pages grouped by related symptoms
- Receive a set of likely diagnoses based on patient inputs for the questions and the predictive model

**Doctor Interface:**
- Doctors can modify relative likelihoods and add principal symptoms
- Once a new input grid is filled out, diagnosis results are updated for either existing or new principal symptoms
- Currently stores just the relative likelihoods for cough

## Conclusion and Future Work

This project presented a user-friendly web application that can diagnose cough patients as well as preliminary research for future development. The application's Doctor Interface can incorporate additional principal symptoms. The biggest issue faced was finding a large source of reliable patient diagnosis data. Some possibilities include unstructured free-text data from clinician notes or Electronic Health Records (EHR). NLP techniques could be applied to build models from free-text and with enough data, the machine learning approaches we explored could be implemented. A clinical diagnosis system could draw conclusions from multiple different predicative models as well as update the relative likelihoods based on empirical data or learning parameters.

### Acknowledgement

We would like to thank the LILY Lab and all of our team members that were integral to this project: Frederick Howard, Irene Li, Lajos Pusztai, and Professor Dragomir Radev.