PLAKSHA UNIVERSITY

Identifying Cognitive Biases in Medical Decision Making

Dhruv Menon

AL3001|Machine Learning & Pattern Recognition

Problem Statement

How can we identify and eliminate cognitive biases in medical decision making?

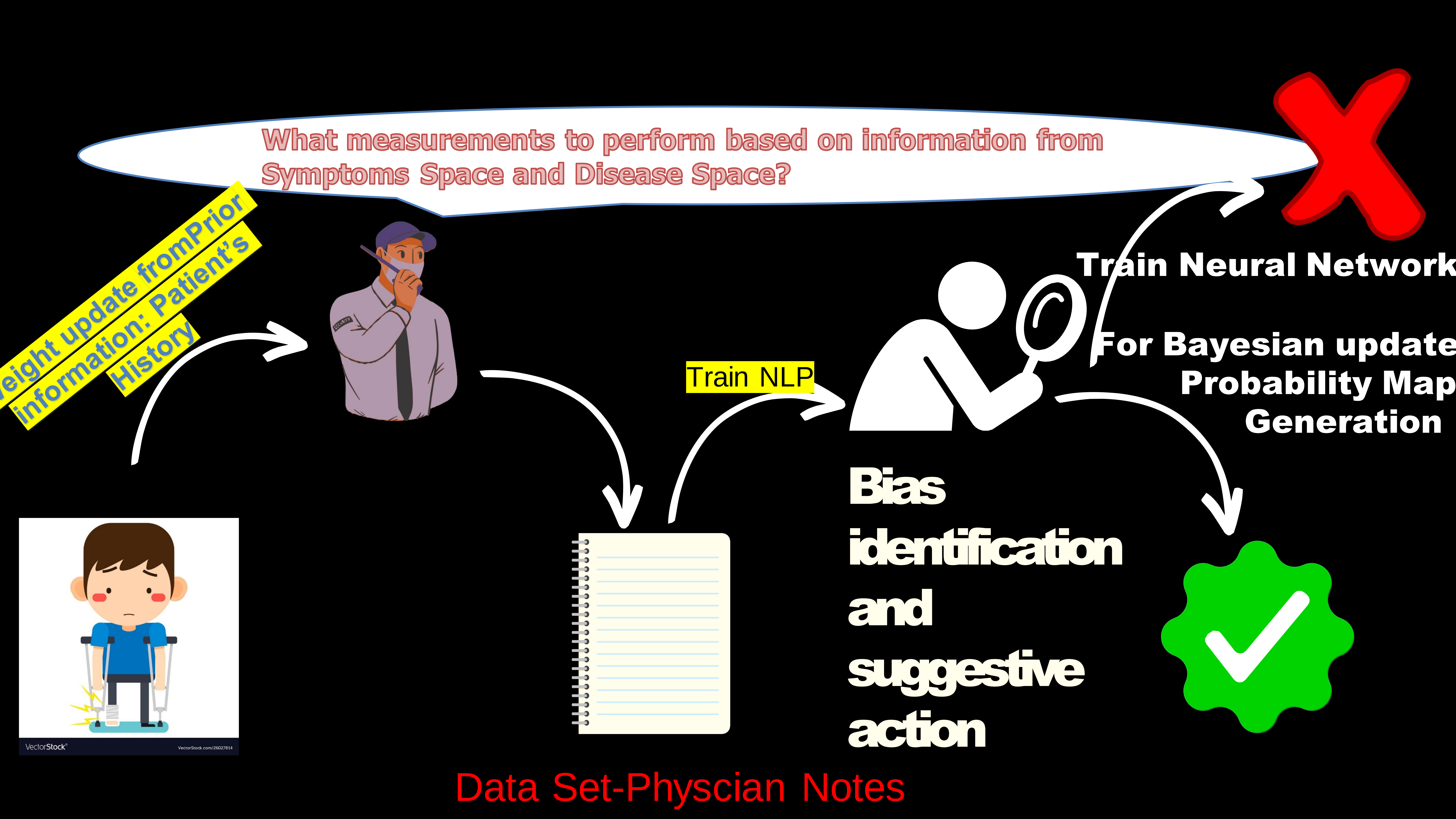
Key concept

Insight: Decision-making equates to resource allocation minus noise.(Jack Welch)

Trade off in Decision-making:

Resource Productivity: Maximizing value from available resources

Capacity Utilization: Minimizing unused resources


What measurements to perform based on information from Symptoms Space and Disease Space?
eight update fromPrior information: Patient's History
Train NLP
Train Neural Network
For Bayesian update Probability Map Generation
Bias identification and suggestive action
VectorStock®
VectorStock.com/26027814
Data Set-Physcian Notes

Literature Review

Errors: 1.7-6.5 % of hospital admissions result in errors, leading to significant mortality and financial costs.

Cost: In 2008 , medical errors cost $19.5 billion in the USA. Incremental cost per error = $4685.

Statistical significant Correlation exists between cognitive biases and occurrence of medical errors

	Research Paper	Cognitive biases identified
1	Saposnik et al. BMC Medical Informatics and Decision Making (2016) 16:138 DOI 10.1186/s12911-016-0377-1	Risk aversion, overconfidence, confirmation bias, anchoring bias
2	Pape, Tom and Kavadias, Stylianos and Sommer, Svenja C., Decision Bias in Project Selection: Experimental Evidence from the Knapsack Problem	Small-Project Bias, Premature Search Termination
3	Schmidgall, S., Harris, C., Essien, I., Olshvang, D., Rahman, T., Kim, J. W., ... & Chellappa, R. (2024). Addressing cognitive bias in medical language models	confirmation, frequeny, cultural, status quo, false-consensus, recen
4	Kostick-Quenet, K.M., Gerke, S. AI in the hands of imperfect users. npj Digit. Med. 5, 197 (2022). https://doi.org/10.1038/s41746-022-00737-z	
5	G. Harris, C. (2020, April). Mitigating cognitive biases in machine learning algorithms for decision making. In Companion Proceedings of the Web Conference 2020 (pp. 775-781).	

Identified Targeted Biases

- **Measurement Bias: biasing judgement based on one experimental measurement like single blood test etc or inappropriate mapping with proxy variable**

- **Confirmation bias is the tendency to search for, interpret, favor, and recall information in a way that confirms one's preexisting beliefs or hypotheses. In clinical settings, this might manifest as a doctor giving more weight to evidence that supports their initial diagnosis.**

- **Automation Bias: Trusting the machine a bit too much or too less**

GPT4 has best performance metric

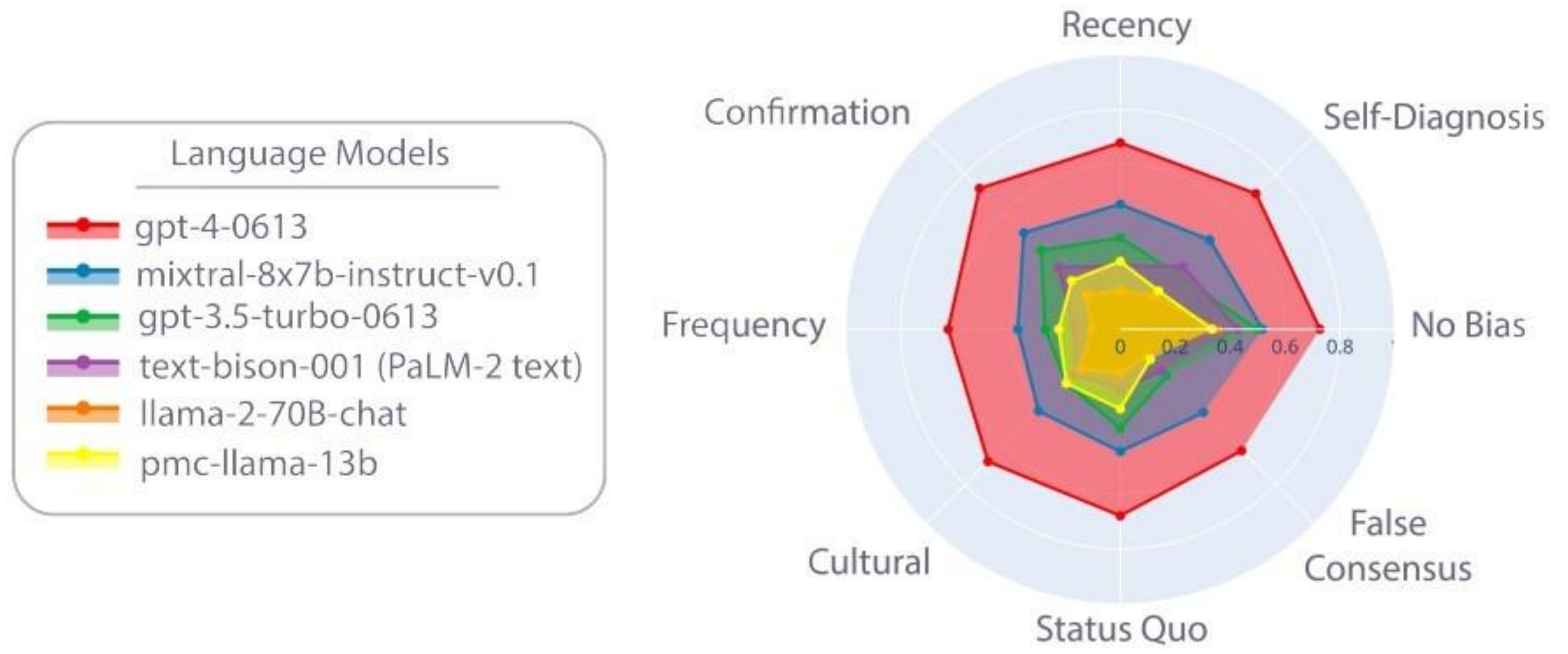


Figure 2. Model comparison following cognitive bias addition. Accuracy is indicated by the distance between each dot and the origin (e.g., a radius of 0.8 corresponds to 80% accuracy). The names of each cognitive bias surround the circle. Table 1 shows the results in tabular format.

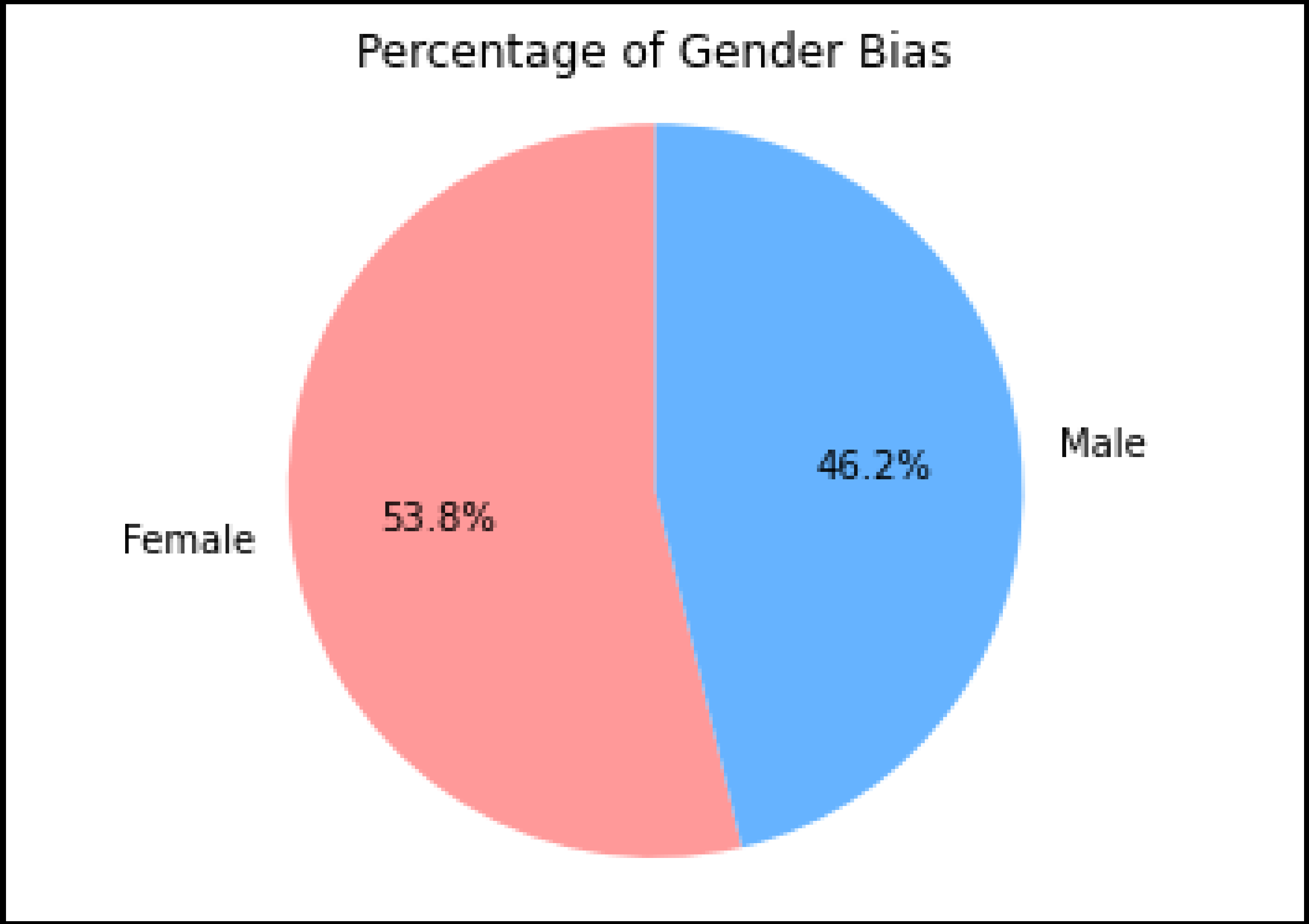
Feature Preprocessing

Handled missing values and removed duplicate

Evaluating fairness metrics to remove gender bias for patients:

Treatment Equality

This fairness measure requires equal false negative rates (FNR) across different groups. It ensures that the model is equally inaccurate for all groups when predicting negative outcomes Minimize Type 2 Errors


Percentage of Gender Bias
Male
46.2%
Female
53.8%

Data Set

MedQA dataset : randomly sampled data from conversations between physcians and patients.

MedQA dataset : randomly sampled data from conversations between physcians and patients.

Obtained ethical clearance

Data was collected by the researchers empirically:

Abacha, Asma Ben, et al. "An empirical study of clinical note generation from doctor-patient encounters." *Proceedings of the 17th*

Features Identified

FAMILY HISTORY/SOCIAL HISTORY (fam/sochx): This includes information about the patient's family medical history, lifestyle, and social context. It helps understand potential genetic risks and environmental factors.

HISTORY OF PRESENT ILLNESS (genhx): This feature captures details about the patient's current health condition, symptoms, and any relevant events leading up to their visit.

PAST MEDICAL HISTORY (pastmedicalhx): Here, we document the patient's previous medical conditions, surgeries, and chronic illnesses. It provides context for their current health status.

CHIEF COMPLAINT (cc): The patient's primary reason for seeking medical attention. It helps focus the assessment and diagnosis.

PAST SURGICAL HISTORY (pastsurgical): Information about any surgical procedures the patient has undergone in the past.

ALLERGY: Details about any known allergies the patient may have, including medications, foods, or environmental triggers.

REVIEW OF SYSTEMS (ros): A comprehensive assessment of various body systems to identify any additional symptoms or issues.

MEDICATIONS: A list of the patient's current medications, including dosage and frequency.

ASSESSMENT: The healthcare provider's evaluation of the patient's overall health and any specific findings.

EXAM: Documentation of the physical examination performed by the provider.

DIAGNOSIS: The identified medical condition or problem based on the assessment and examination.

DISPOSITION: Decisions regarding further treatment, referrals, or hospitalization.

PLAN: The proposed course of action, including treatment options, follow-up, and patient education.

EMERGENCY DEPARTMENT COURSE (edcourse): A summary of the patient's experience during their emergency department visit.

IMMUNIZATIONS: Information about the patient's vaccination history.

GYNECOLOGIC HISTORY (gynhx): Relevant details about the patient's reproductive health and gynecological issues.

PROCEDURES: Any medical procedures or interventions performed on the patient.

OTHER HISTORY (other_history): Additional relevant information not covered by the above features.

LABS: Results from laboratory tests and diagnostic studies.

Sample Data

```
import pandas as pd
import numpy as np

# Load the dataset
file_path = '/mnt/data/cognitive_biases_dataset_detailed (1).xlsx'
df = pd.read_excel(file_path)

# Correct 'Disease' column based on the specified order every 600 entries
diseases = ['Asthma', 'Depression', 'Diabetes', 'Migraine', 'Hypertension']
num_diseases = len(diseases)
for i in range(len(df)):
    df.at[i, 'Disease'] = diseases[(i // 600) % num_diseases]

# Display the first few rows of the dataset
sample_data_output = df.head()
sample_data_output
```

Sample Rule System

Bias Type	Rule	Example in Physician Note
Confirmation Bias	Assuming the cause of symptoms without thorough assessment, and failing to consider alternative diagnoses. Giving too much or too little weight to past medical history without considering current clinical presentation.	If the physician attributes the patient's asthma symptoms solely to inadequate inhaler technique without considering other potential conditions such as chronic obstructive pulmonary disease or congestive heart failure.
Measurement Bias	Overemphasizing clinical parameters and neglecting comprehensive assessment, including psychosocial aspects of care and potential limitations of measurements.	If the physician focuses solely on spirometry results or peak flow measurements without considering factors such as patient effort, technique, or the limitations of relying solely on numerical measurements.
Automation Bias	Overreliance on automated alerts or prompts without critical evaluation of patient presentation and symptoms, potentially leading to misdiagnosis or overlooking comorbid conditions.	If the physician adjusts medication dosage solely based on an automated reminder to increase the dose without considering other factors such as patient-reported symptoms or potential comorbid conditions.

Methodology

Mathematically, the problem is a directed graph translating from the symptom space(space of all possible symptoms associated with the diseases in the disease space) to measurement space(space of all possible measurements that can be performed to identify the diseases associated with all the symptoms in the symptom space) to disease space(diseases in the data set), where the edges represent the probabilities.

The problem can be broken down into 3 major steps:

Step 1: Training NLP model for Bias Identification from annotated data set

Step 2: Disease and bias-specific corrective action

Step 3: Train neural network for bayesian update and probability map generation

Neural Network Output(under training)

```
Epoch 1/50
48/48 [==============================]  - 1s 11ms/step - loss: 0.6213 - accuracy: 0.6966 - val_loss: 0.3562 - val_accuracy: 0.8604
Epoch 2/50
48/48 [==============================]  - 0s 7ms/step - loss: 0.2924 - accuracy: 0.8783 - val_loss: 0.2930 - val_accuracy: 0.8746
Epoch 3/50
48/48 [==============================]  - 0s 7ms/step - loss: 0.1659 - accuracy: 0.9394 - val_loss: 0.2816 - val_accuracy: 0.8896
Epoch 4/50
48/48 [==============================]  - 0s 7ms/step - loss: 0.0899 - accuracy: 0.9732 - val_loss: 0.3089 - val_accuracy: 0.8846
Epoch 5/50
48/48 [==============================]  - 0s 7ms/step - loss: 0.0540 - accuracy: 0.9824 - val_loss: 0.3318 - val_accuracy: 0.8846
Epoch 6/50
48/48 [==============================]  - 0s 7ms/step - loss: 0.0414 - accuracy: 0.9899 - val_loss: 0.3480 - val_accuracy: 0.8996
Epoch 7/50
48/48 [==============================]  - 0s 7ms/step - loss: 0.0267 - accuracy: 0.9924 - val_loss: 0.3763 - val_accuracy: 0.8921
9/9 [==============================]  - 0s 2ms/step - loss: 0.2897 - accuracy: 0.8996
Test Loss: 0.28965526843070984
Test Accuracy: 0.899581015586853
```

Code

```
model.add(Dense(512, input_dim=input_dim, activation='relu'))
  model.add(Dropout(0.5))
  model.add(Dense(256, activation='relu'))
  model.add(Dropout(0.5))
  model.add(Dense(128, activation='relu'))
  model.add(Dense(1, activation='sigmoid'))

  model.compile(optimizer=Adam(learning_rate=0.001),
         loss='binary_crossentropy',
         metrics=['accuracy'])
  return model

# Main function to run the training and evaluation
def main():
  # Load the dataset
  file_path = 'validated_cognitive_biases_dataset.xlsx'
  df = load_data(file_path)

  # Prepare features and labels
  X, y, label_encoder, vectorizer = prepare_features_labels(df)

  # Split the data into training and testing sets
  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

  # Compute class weights to handle class imbalance
  class_weights = compute_class_weight(class_weight='balanced', classes=np.unique(y_train), y=y_train)
  class_weights_dict = dict(enumerate(class_weights))

  # Build the model
  model = build_model(X_train.shape[1])

  # Early stopping to prevent overfitting
  early_stopping = EarlyStopping(monitor='val_loss', patience=5, restore_best_weights=True)

  # Train the model
  history = model.fit(X_train, y_train,
            epochs=50,
            batch_size=32,
```

Performance Metrics

Balanced accuracy = (Sensitivity + Specificity) / 2
where:
Sensitivity: The “true positive rate” – the percentage of positive cases the model is able to detect.
Specificity: The “true negative rate” – the percentage of negative cases the model is able to detect.

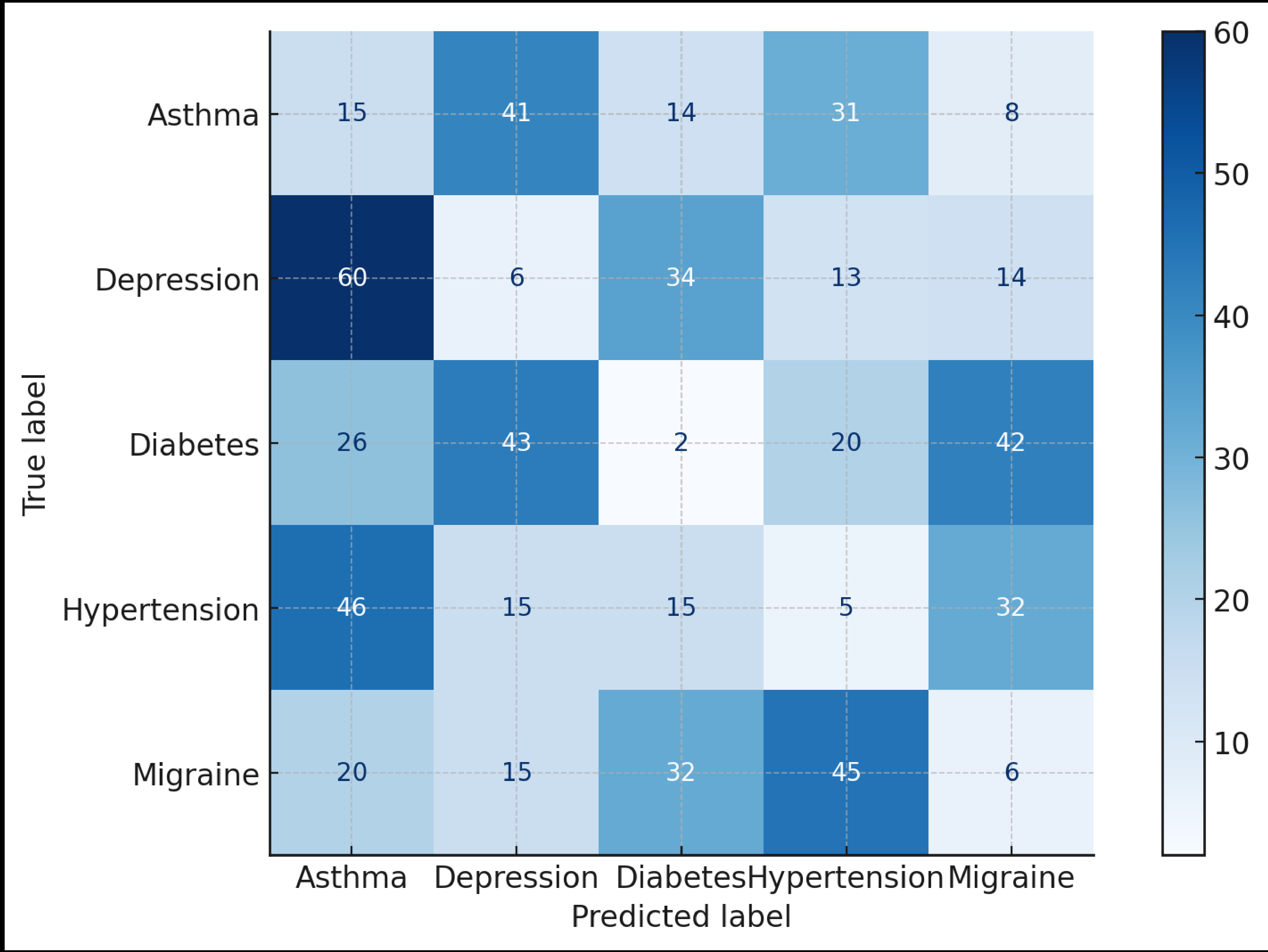
Cognitive biases in decision making are so common according to literature so yes the data set is imbalanced which merits the use of such an approach.

Schmidgall, S., Harris, C., Essien, I., Olshvang, D., Rahman, T., Kim, J. W., ... & Chellappa, R. (2024). Addressing cognitive bias in medical language models. *arXiv preprint arXiv:2402.08113*

Data Set:

precision recall f1-score support Asthma 0.09 0.14 0.11 109 Depression 0.05 0.05 0.05 127 Diabetes 0.02 0.02 0.02 133 Hypertension 0.04 0.04 0.04 113 Migraine 0.06 0.05 0.05 118 accuracy 0.06 600 macro avg 0.05 0.06 0.05 600 weighted avg 0.05 0.06 0.05 600

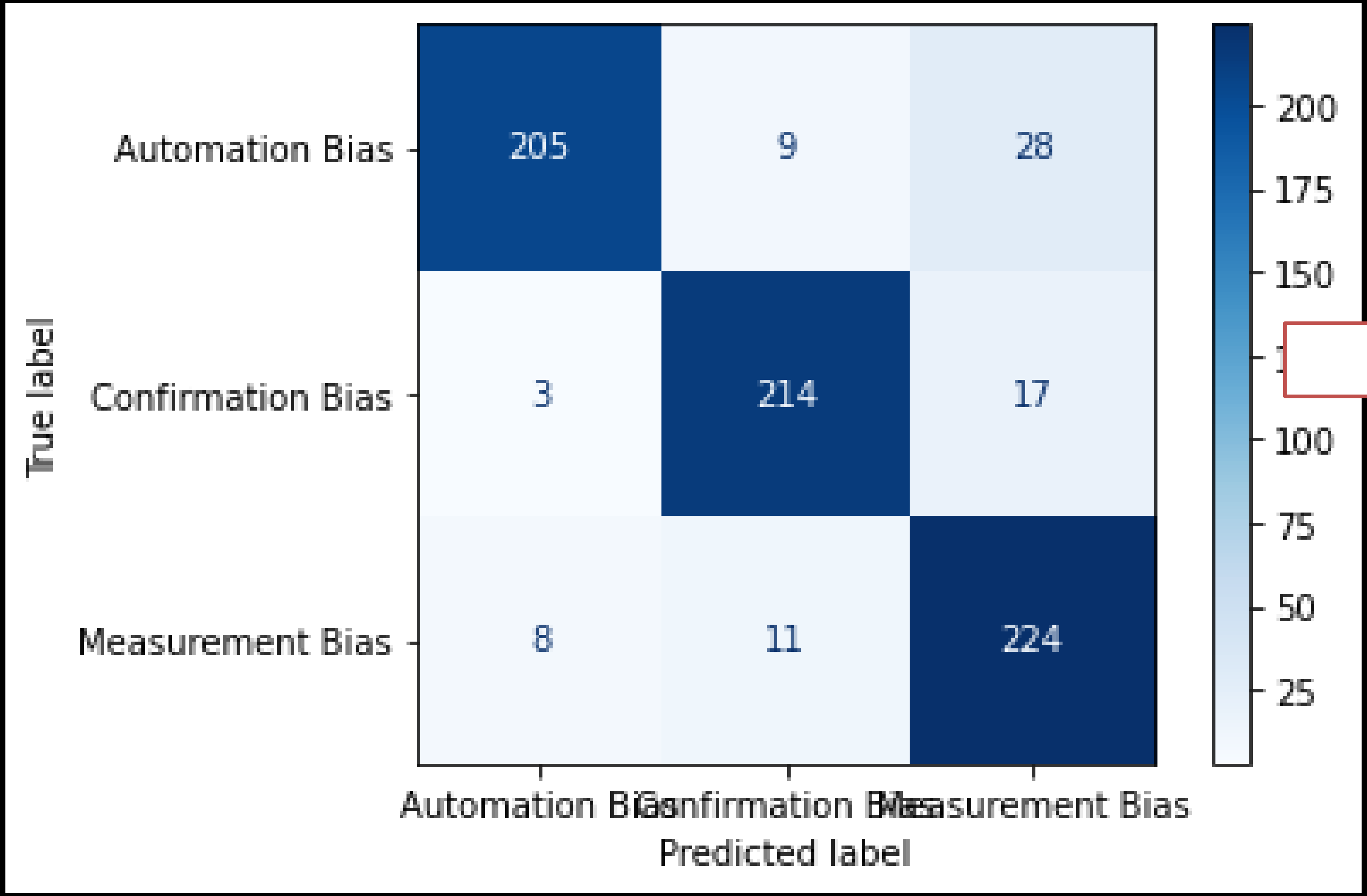
Confusion matrix



Sample Code outputs

Physician Notes	Predicted Disease	Actual Disease	Predicted Bias	Suggestive Action
His brothers had prostate cancer. Father had asthma.	Diabetes	Asthma	Confirmation Bias	Review family history for correct influences
This 19-year-old Caucasian female presents to ...	Asthma	Asthma	None	None
The patient is an 89-year-old lady. She actually...	Asthma	Diabetes	Measurement Bias	Re-evaluate patient's symptoms and diagnostics
None.	Migraine	Migraine	None	None
PUD, ?stroke and memory difficulty in the past...	Hypertension	Hypertension	None	None

Code outputs for biases(Not trained well so showing last)(There was class imbalance during annotating which was handled by assigning weights to each class such that they balance out)



For the observant among you Early Stopping to prevent overfitting

```
72/72 ... 14ms/step - accuracy:
0.9... : 0.8765 - val_loss:
0...
Ep...
72/72 ... s/step - accuracy:
0.9854 - lo... val_accuracy: 0.8800 - val_loss:
0.3293
Epoch 6/50
72/72 ━━━━━━━━━━━━━━━━━━━━ 1s 14ms/step - accuracy:
0.9857 - loss: 0.0383 - val_accuracy: 0.8800 - val_loss:
0.3598
23/23 ━━━━━━━━━━━━━━━━━━━━ 0s 3ms/step - accuracy: 0.8921
- loss: 0.2064
Test Loss: 0.2047780156135559
Test Accuracy: 0.8942976593971252
```

Sensitivity: C1 =0.84, C2=0.91, C3=0.92

```
Text: Physician attributes all symptoms (e.g., wheezing,
coughing,...
Predicted Label: Confirmation Bias
Actual Label: Confirmation Bias

Sample 3:
Text: The patient has a history of multiple medical
problems inclu...
Predicted Label: Measurement Bias
Actual Label: Automation Bias

Sample 4:
Text: Married.  He is retired, being a Pepsi-Cola driver
secondary...
Predicted Label: Automation Bias
Actual Label: Automation Bias
```

Advantage and Limitation

Advantages

Logically speaking, adding additional nodes to the network (include more diseases!) shouldn't affect model generalizability or performance.

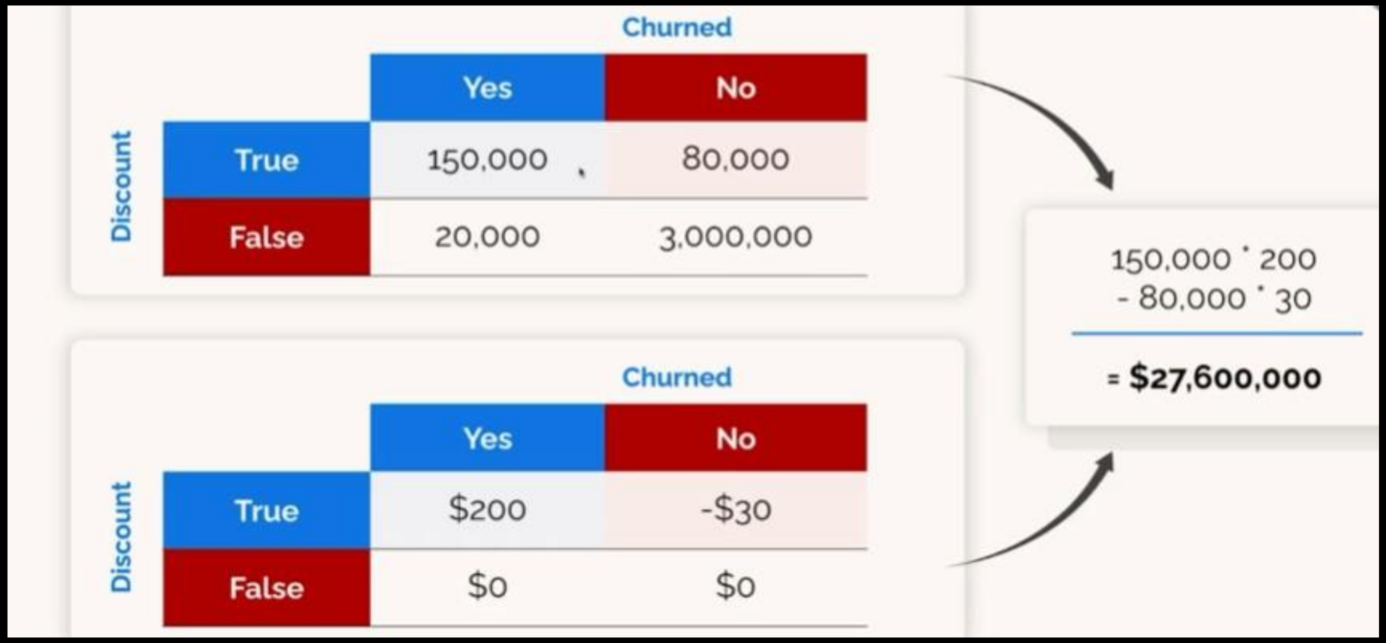
Limitations

Big assumption in our model that diseases are not correlated!!

Depression and Asthma miscorrelation

Next Step(for shark tank pitch)

Build Rumsfeld Matrix
Go deeper into the problem



Discount	Churned: Yes	Churned: No
True	150,000	80,000
False	20,000	3,000,000

Discount	Churned: Yes	Churned: No
True	$200	-$30
False	$0	$0

150,000 * 200
- 80,000 * 30
= **$27,600,000**

Example of Valuation Matrix: Here incremental cost would be $4685

Build Probability Map for Doctor