GPT-4-Powered Insights for Diabetes Risk Prediction

JESUS MINJARES

UT AUSTIN: AI IN HEALTHCARE

Healthcare Issue: Diabetes Risk

PROBLEM: DIABETES AFFECTS MILLIONS GLOBALLY.

CHALLENGE: EARLY RISK DETECTION IS COMPLEX.

GPT-4 ROLE: PREDICTS RISK AND SIMPLIFIES NOTES.

GOAL: ENHANCE EARLY INTERVENTION.

Medical Dataset Used

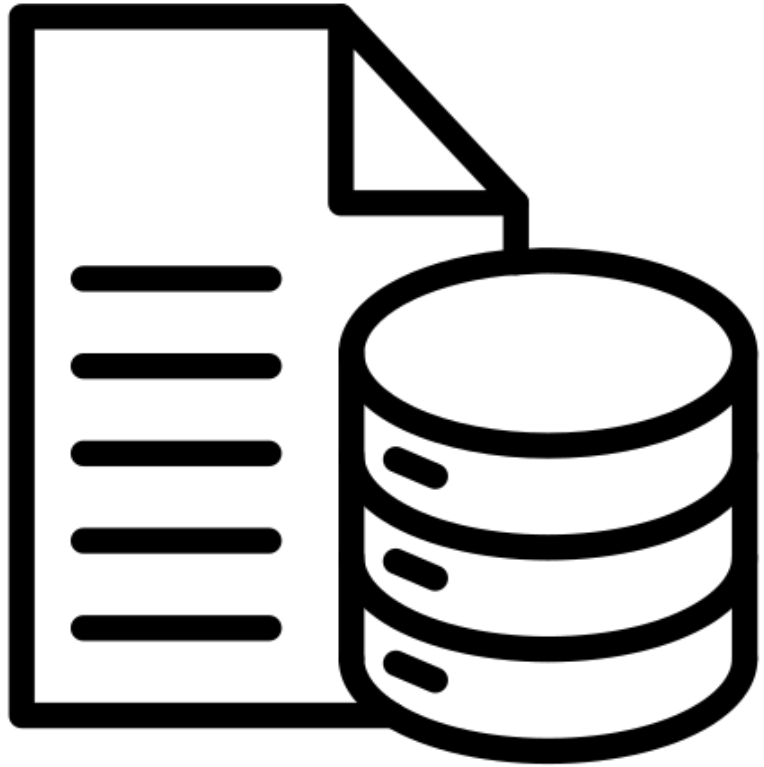
Source: Synthea (synthea.mitre.org).

Size: 7 MB, 100 patients.

Key Files:

conditions.csv: Diagnoses.

observations.csv: Glucose data.



Engineered Prompts for GPT-4

CoT Prompt: Predicts risk (High/Medium/Low).

- *Analyze glucose, obesity, meds*.

```
You are a medical AI predicting diabetes risk. Use
reasoning and respond with 'High', 'Medium', or 'Low'.
1. Conditions (limited sample): {patient_conditions}.
2. Observations (limited sample): {patient_observations}.
3. Medications (limited sample): {patient_medications}.
4. Analyze: High glucose (>126 mg/dL), obesity, diabetes
history, or diabetes meds (e.g., metformin, insulin)
suggest high risk. Normal glucose (<100 mg/dL) suggests
low risk.
5. Conclude: Risk level.
Output only: High, Medium, or Low.
```

ToT Prompt: Diagnoses (e.g., "Diabetes").

- *Check conditions and meds*.

```
You are a medical AI performing differential diagnosis.
Respond with the most likely diagnosis (e.g., 'Diabetes',
'Hypertension', 'Other').
1. Conditions (limited sample): {patient_conditions}.
2. Observations (limited sample): {patient_observations}.
3. Medications (limited sample): {patient_medications}.
4. Check: Diabetes (glucose, meds), Hypertension (blood
pressure), or Other (symptoms).
5. Conclude: Most likely diagnosis.
Output only the diagnosis.
```

Methods Employed

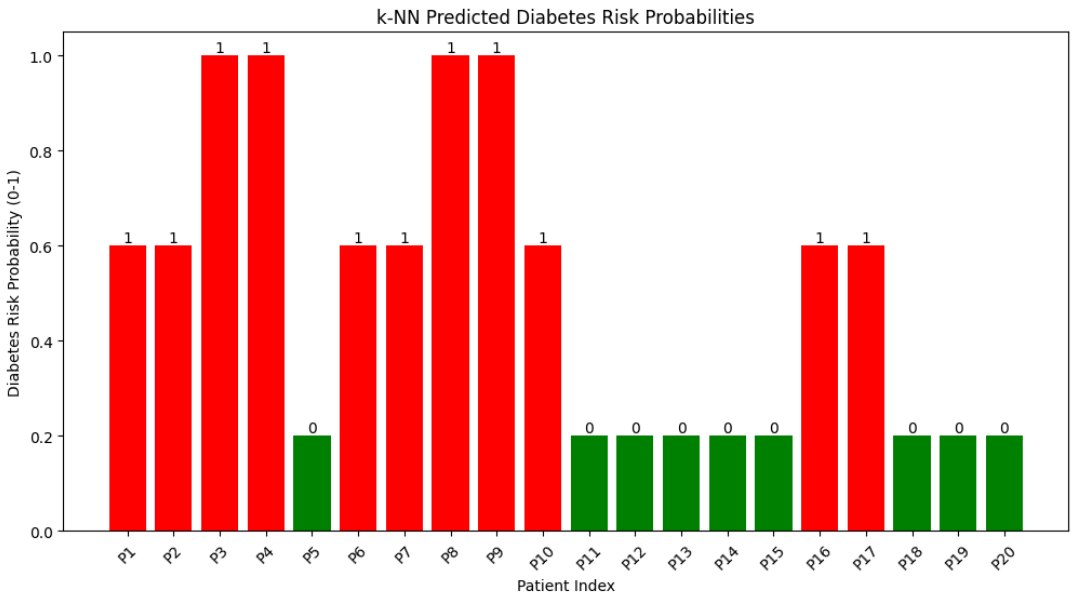
```
Predicting diabetes risk (CoT):  30%|███       | 6/20 [00:16<00:38,  2.77s/it]
Prompt:
You are a medical AI predicting diabetes risk. Us...
Response: Medium

Predicting diabetes risk (CoT):  35%|███       | 7/20 [00:20<00:37,  2.87s/it]
Prompt:
You are a medical AI predicting diabetes risk. Us...
Response: High

Predicting diabetes risk (CoT):  40%|████      | 8/20 [00:23<00:36,  3.01s/it]
Prompt:
You are a medical AI predicting diabetes risk. Us...
Response: Low
```

Preprocessing: Loaded Synthea data.

GPT-4:

- CoT: Risk prediction.
- ToT: Differential diagnosis.

ML: k-NN for prediction/classification.



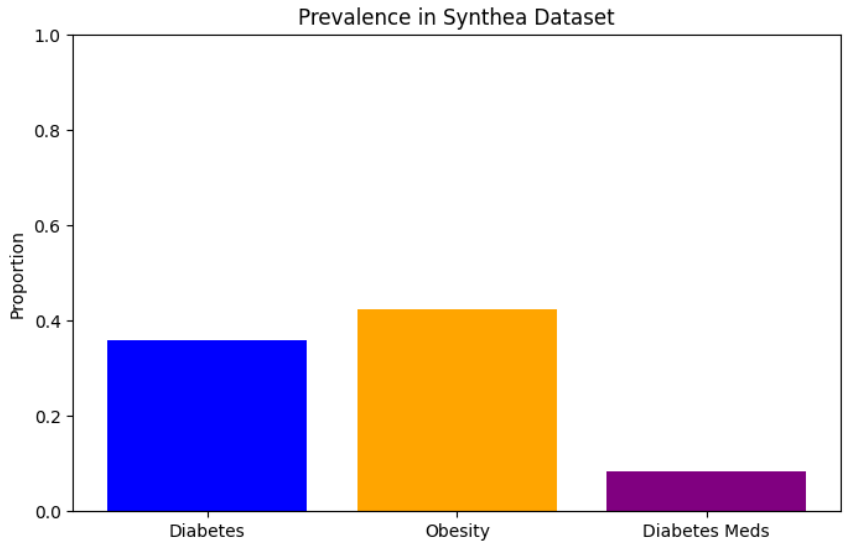
Evaluation of Results

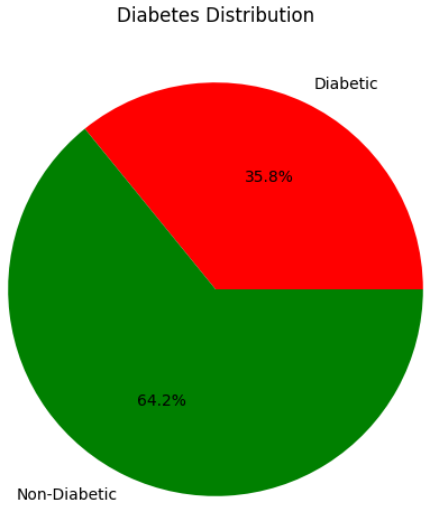
Prevalence:

- Diabetes: 35.85%.
- Obesity: 42.45%.

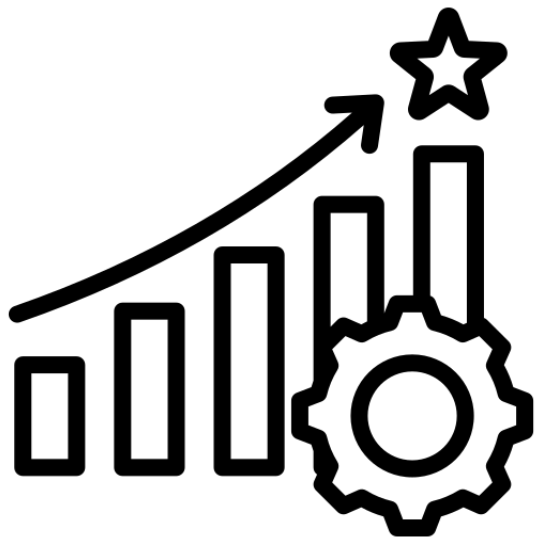
GPT-4 Accuracy:

- CoT & ToT at 0.90.





Ideas for Improvement



- Dataset: Larger Synthea sample.
- GPT-4: Refine prompts with examples.
- ML: Add features, tune k-NN.
- Visualization: Glucose trend plot.