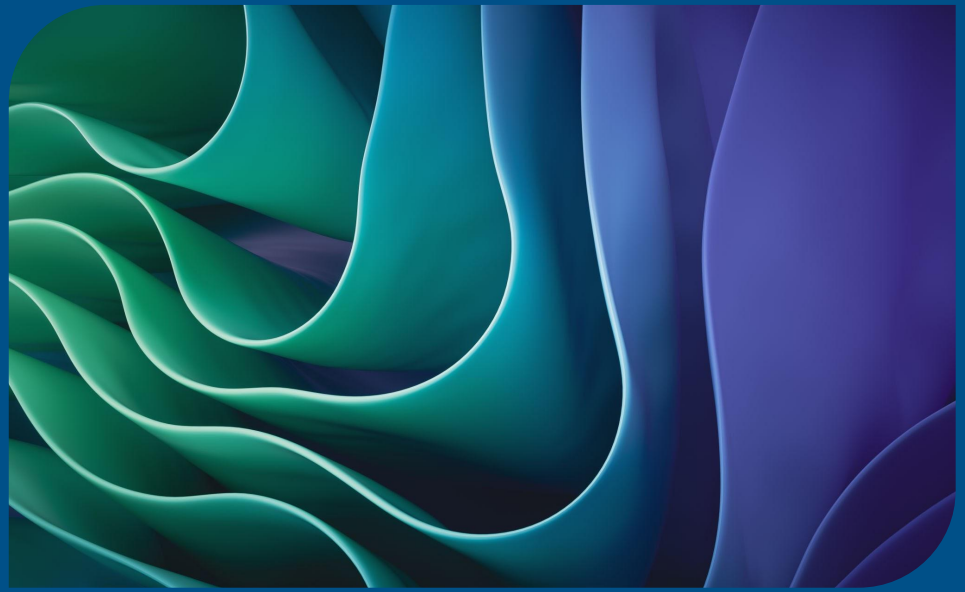
MM.DD.YY

ICD10 Suspecting

Subtitle Goes Here



Lorem Ipsum

1. High level approach
2. Detailed Results
3. Confidence Scores
4. Limitations
5. Improvements for future

Initial design

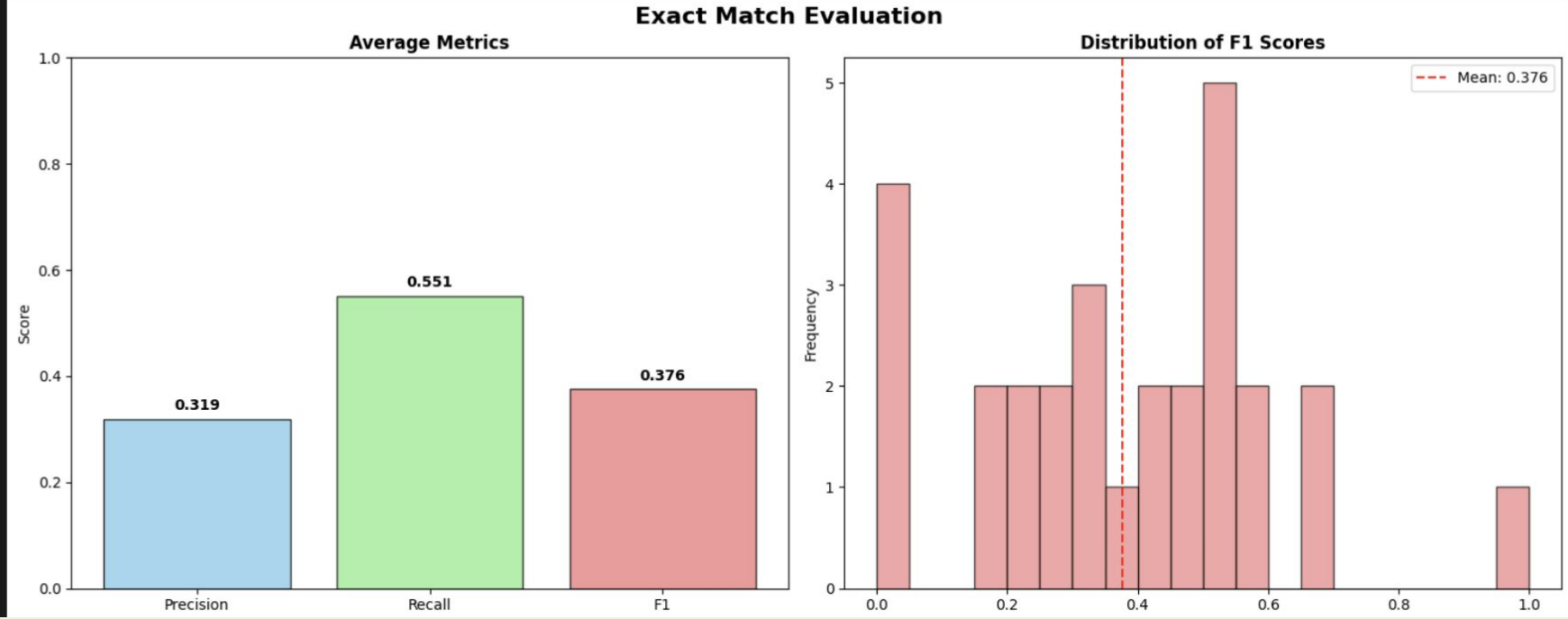
1. Create an agentic implementation using gemini flash 2.5. This has access to three tools:
    a. Lookup ICD-10 alphabetical index - query as an arg
    b. Lookup ICD-10 tabular index - icd code as an arg
    c. Create ICD10 Suspect entry - need {reasoning, code, description}
2. Select a few paragraphs from icd10 guidelines - ICD-10-CM FY25 Guidelines October 1, 2024, these pertain to how to use the indices and how to assign the codes
3. Let the agentic approach just run in a back and forth until it emits a stop token
4. Make use of a preprocessing stage that converts the transcript into medical notes that we can use for iteration. This helps remove fluff
5. For confidence scoring use LLM as a judge with two tags (use gemini pro 2.5):
    a. Confident
    b. requires human review

Confidence scoring is given the transcript and all encoded icds and I asked for a confidence reasoning and an assignment of one of the two tags

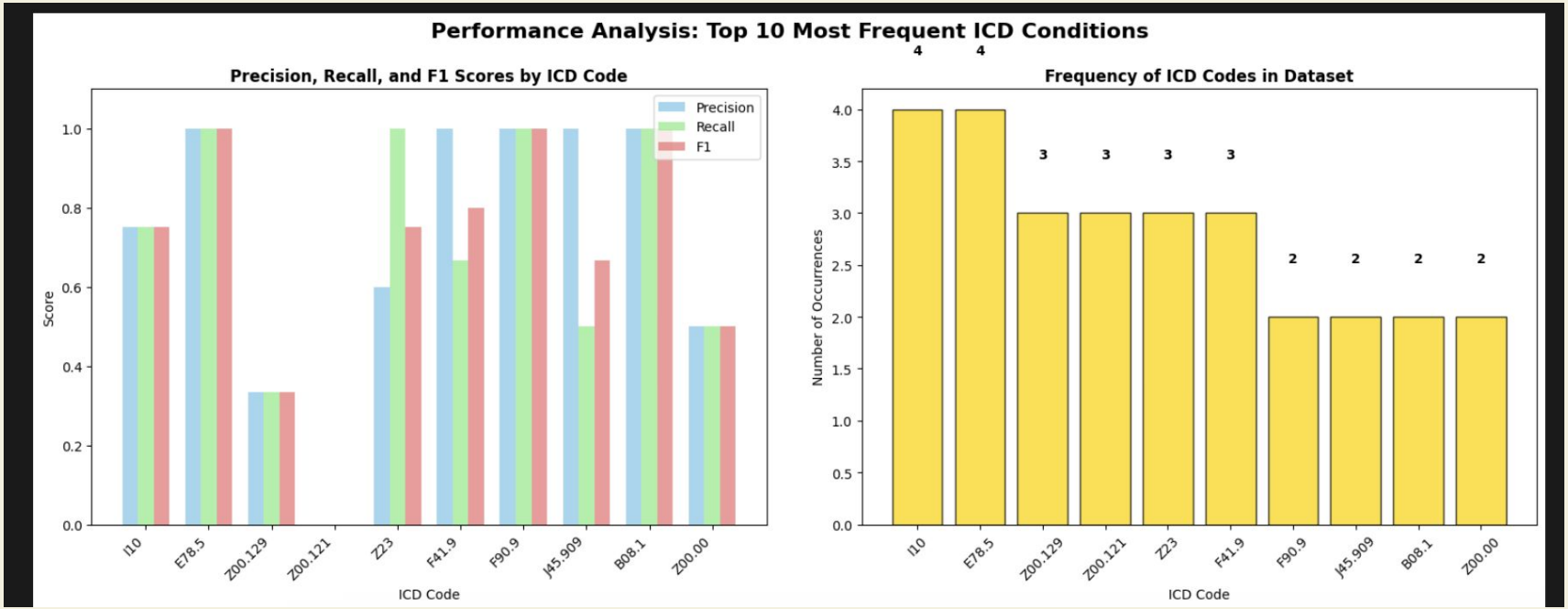
Algorithm I used

1. Pass the entire transcript to preprocessing stage
2. Pass processed notes to the agent along with a system prompt
3. Execute it for 50 turns or stop token whichever is reached first
4. Store the generated icd10_suspect_entries
5. Send the entire chat history and predicted suspects to confidence scoring prompt

Comparison with ground truth - per condition



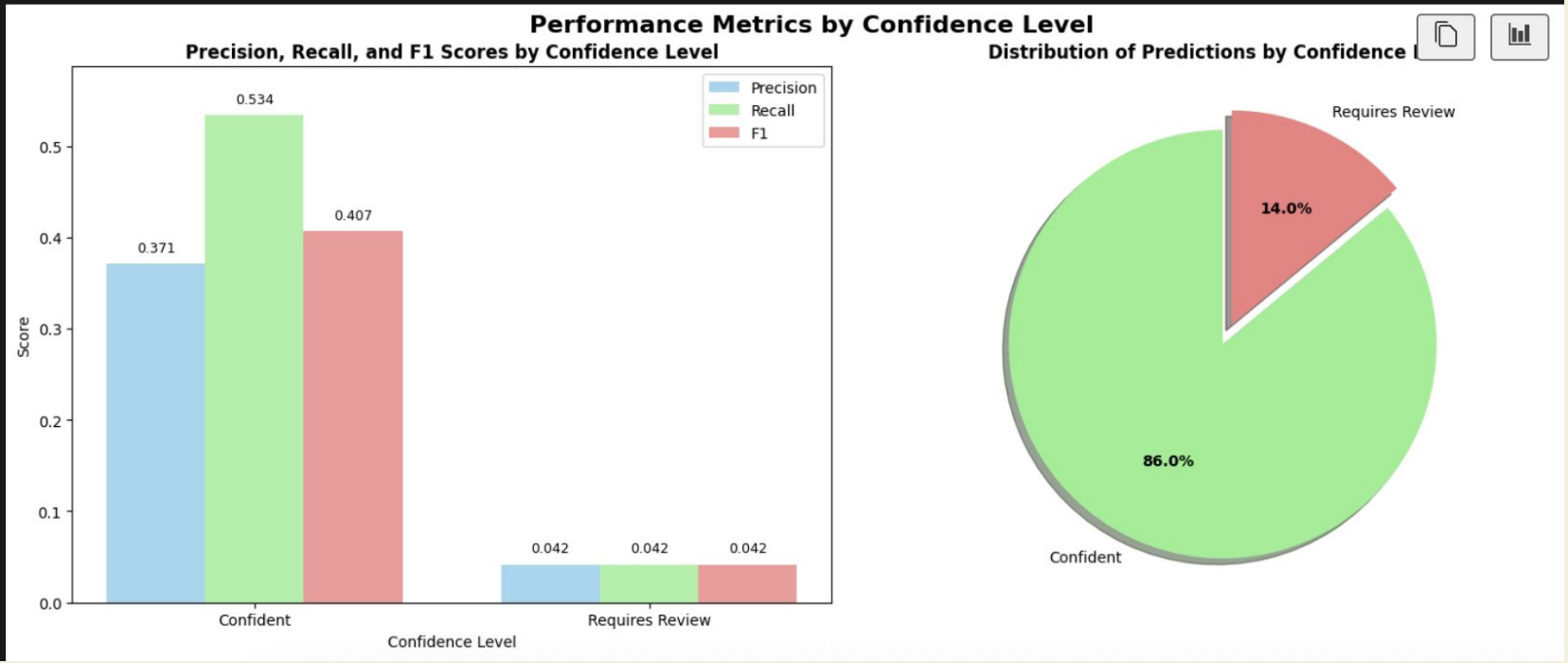
Comparison with ground truth - most frequent codes



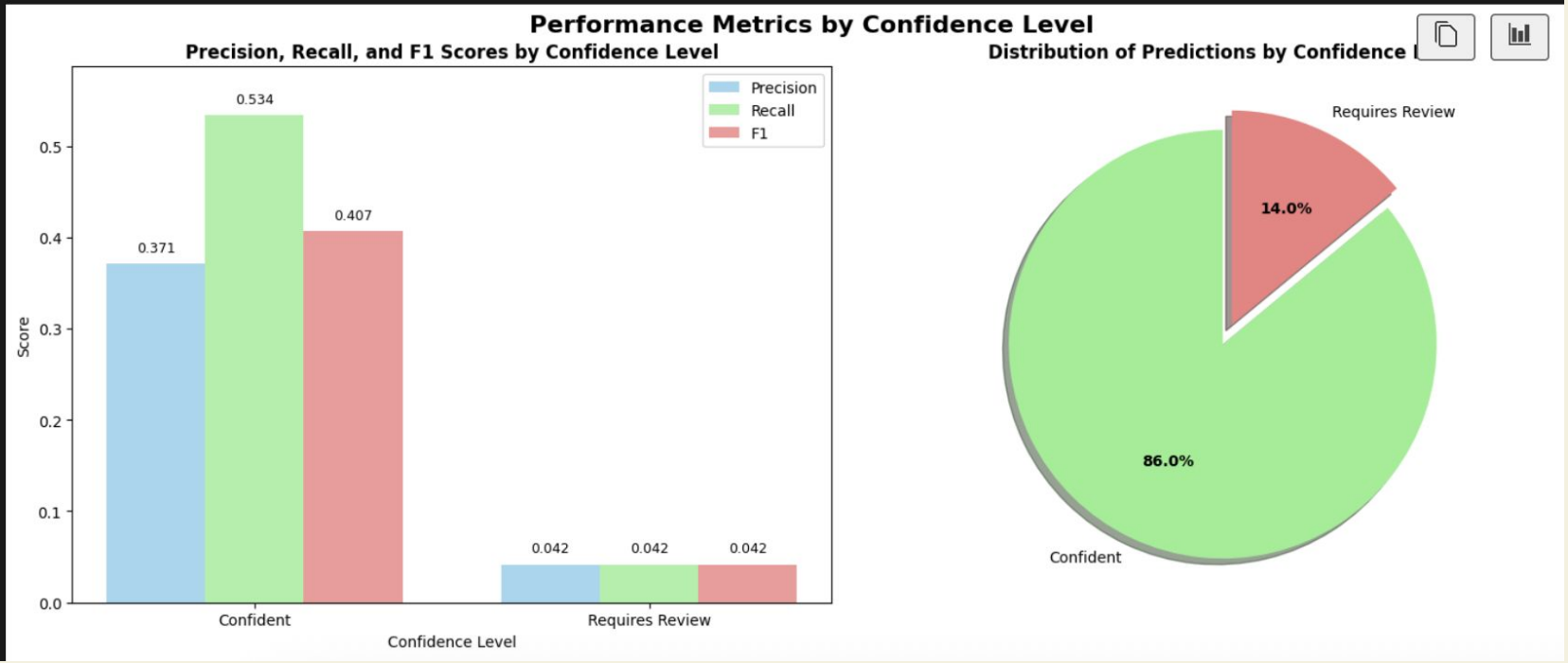
Confidence analysis

```
Performing comprehensive confidence evaluation...
Analyzing confidence patterns in predictions...
Analyzing confidence patterns in predictions...
COMPREHENSIVE CONFIDENCE ANALYSIS
============================================================

PREDICTION-LEVEL ANALYSIS (Individual ICD Code Accuracy)
------------------------------------------------
Total predictions: 164
Confident predictions: 141
Review predictions: 23
Confident accuracy: 0.326
Review accuracy: 0.043
Overall accuracy: 0.287
Calibration gap: 0.283

ENCOUNTER-LEVEL ANALYSIS (Precision/Recall/F1 per Encounter)
------------------------------------------------
Total encounters: 28
Confident encounters: 28
Review encounters: 12
```

Confidence analysis



Confidence analysis



Deep dive example – index 3

****Chief Complaint:****
Persistent cough and exertional dyspnea (difficulty breathing) during swimming, worsening over the past two weeks.

****Current Symptoms:****
* **Chronic cough, especially over the last two weeks.**
* **Congestion, improved with humidifier use and air filter change.**
* **Dyspnea on exertion: Specifically during swim team (freestyle and breaststroke), leading to difficulty breathing and inability to "get air in" after 25 yards; this condition has worsened over the summer.**
* **Wheezing noted on physical examination.**
* **No fever, vomiting, or diarrhea.**
* **Eating well.**
* **Cough intermittently wakes patient at night (once last week), but generally sleeps through the night.**

****Medical History:****
* **Family history of "Reactive Airway" (likely asthma) in both patient and sibling during childhood, managed with inhalers, eventually grew out of it.**

And so on

Deep dive example – index 3 – predictions

Predictions:

J45.901 Unspecified asthma with (acute) exacerbation
R05.3 Chronic cough
Z77.098 Contact with and (suspected) exposure to other hazardous, chiefly nonmedicinal, chemicals

Reference answer:
J98.8 Wheezing-associated respiratory infection (WARI)
Z23 Encounter for immunization
J45.990 Exercise-induced RAD (reactive airway disease)

Deep dive example – index 3 – reasoning

Predictions:

J45.901 Unspecified asthma with (acute) exacerbation
The patient presents with symptoms highly suggestive of an asthma exacerbation (persistent cough, exertional dyspnea, wheezing), with a history of 'Reactive Airway' (likely asthma) and current acute management. J45.901 describes unspecified asthma with acute exacerbation, which fits the lack of a specified asthma type and the acute worsening of symptoms.

R05.3 Chronic cough
The patient reports a 'Chronic cough' that has worsened and intermittently wakes them at night. While cough is a symptom of asthma, the chronic nature and impact warrant a separate code, as per guidelines allowing coding of additional signs/symptoms not routinely associated with a disease process or that are significant enough to be listed.

Z77.098 Contact with and (suspected) exposure to other hazardous, chiefly nonmedicinal, chemicals
The patient is regularly exposed to chlorine on the swim team, which is noted as an irritant. Z77.098, 'Contact with and (suspected) exposure to other hazardous, chiefly nonmedicinal, chemicals,' accurately describes this environmental exposure and its relevance to the patient's respiratory conditio

Deep dive example – compare the two sets

Code	Description	Option #1	Option #2	Better Code	Reason
J45.901	Unspecified asthma with (acute) exacerbation	✅ Present	❌ Absent	✅ **Option #1**	Patient is wheezing, using albuterol and prednisone — classic for asthma exacerbation even if not previously diagnosed.
R05.3	Chronic cough	✅ Present	❌ Absent	✅ **Option #1**	Chronic cough is documented over weeks with history during swim season — clinically accurate.
Z77.098	Exposure to hazardous nonmedicinal chemicals (e.g., chlorine)	✅ Present	❌ Absent	✅ **Option #1**	Chlorine exposure during swimming is clearly discussed as a trigger.
J98.8	Wheezing-associated respiratory infection (WARI)	❌ Absent	✅ Present	❌ Option #1**	WARI is a vague, less commonly accepted code and doesn't match the treatment plan (inhaler, steroids).
Z23	Encounter for immunization	❌ Absent	✅ Present	✅ **Option #2**	Vaccines (flu, COVID) were administered — Z23 **should be coded**, and is correctly included in Option #2.
J45.990	Exercise-induced bronchospasm/RAD	❌ Absent	✅ Present	❌ Option #1**	Technically fits, but J45.901 (asthma with exacerbation) is more appropriate for acute wheezing with treatment needed.

Deep dive example – confidence analysis

J45.901 Unspecified asthma with (acute) exacerbation
confident
The transcript provides strong evidence for an asthma exacerbation, including wheezing on exam, exertional dyspnea, history of reactive airway disease, and acute treatment with Albuterol and Prednisone. The 'unspecified' and 'exacerbation' components of the code are well-supported.

R05.3 Chronic cough
requires_human_review
While 'chronic cough' is explicitly mentioned, it is also a primary symptom of asthma. Per coding guidelines, symptoms integral to a confirmed diagnosis are generally not coded separately. It is debatable whether the 'chronic' nature makes it significant enough to report in addition to the asthma exacerbation. This requires a second opinion.

Z77.098 Contact with and (suspected) exposure to other hazardous, chiefly nonmedicinal, chemicals
confident
The transcript explicitly identifies regular exposure to chlorine as an environmental irritant contributing to the patient's condition. This Z-code accurately captures a factor influencing the patient's health status and is directly supported by the documentation.

Limitations

1. **Data quality issues may exist**
2. **Did not pass the relevant section from icd guidelines for each code, this can further help model come to accurate conclusion**
3. **Issues with gemini api a couple examples did not process**
4. **Costly method? Each analysis costs about 0.5-3 dollars**
5. **Did not use gemini pro for latency/cost**
6. **Did not analyse quality of tool calls**
7. **Could have grouped tool calls together to reduce number of turns**
8. **Uncertainty estimation -**
    a. **Can estimate uncertainty by scoring on each icd code where we have ground truth and using accuracy as the uncertainty table**
    b. **Can also estimate uncertainty by putting replaying the same example multiple times and treating the codes that change as uncertain ones**
9. **Not enough ICD know how was passed to the model, it made mistakes such as multiple icd conditions from same major categories (multiple specifications for diabetes)**