

# Conversational System for Differential Diagnosis of GI Cancer: Track Overview

Manjira Sinha<sup>1</sup>, Rajat Pal and Tirthankar Dasgupta

*Tata Consultancy Services, Research and Innovation, Kolkata, India*

## Abstract

Gastrointestinal (GI) tract cancers, encompassing malignancies in the esophagus, stomach, liver, pancreas, and colon, represent a significant burden on global health, contributing to high mortality rates worldwide. The diagnosis of GI cancers is particularly challenging due to the overlapping symptoms shared by various gastrointestinal conditions and the complex etiologies involved. As a result, accurately differentiating between these cancers remains a formidable task for clinicians, often leading to delays in diagnosis and sub-optimal management. This diagnostic uncertainty has profound consequences. Artificial Intelligent based systems can assist physicians towards faster diagnosis and effective outcome.

## Keywords

Health Analytic, GI Tract Cancer, Diagnostics models,

## 1. Introduction

Gastrointestinal (GI) tract cancers, encompassing malignancies in the esophagus, stomach, liver, pancreas, and colon, represent a significant burden on global health, contributing to high mortality rates worldwide. The diagnosis of GI cancers is particularly challenging due to the overlapping symptoms shared by various gastrointestinal conditions and the complex etiologies involved. As a result, accurately differentiating between these cancers remains a formidable task for clinicians, often leading to delays in diagnosis and sub optimal management. This diagnostic uncertainty has profound consequences, contributing to the alarming statistic that medical errors are the third leading cause of death in the United States, with misdiagnoses playing a central role in this issue.

Additionally, the time constraints placed on healthcare providers—especially given their substantial time spent on administrative tasks—further exacerbate the problem. Physicians often struggle to balance these duties with critical patient care, leaving less time for nuanced diagnosis. This can result in delayed treatment, which negatively impacts prognosis, especially in cancers where early intervention is key to survival. Studies suggest that administrative burdens contribute significantly to clinician burnout and inefficiency, with a consequent reduction in the quality of patient care .

The need for more effective diagnostic support has led to increasing interest in Artificial Intelligence (AI)-driven diagnostic assistants. Recent surveys indicate that majority of physicians believe AI could significantly enhance their diagnostic accuracy and improve treatment decisions, underscoring the demand for AI-based solutions in clinical practice. AI technologies, such as machine learning and deep learning, have the potential to augment human expertise by analyzing large datasets, including medical imaging and histopathology, to provide faster, more accurate assessments. These innovations offer a promising avenue for addressing the diagnostic challenges in GI tract cancers, potentially reducing misdiagnoses, enhancing early detection, and improving patient outcomes.

---

*Forum for Information Retrieval Evaluation, December 12-15, 2024, India*

\*Corresponding author

✉ sinha.manjira@tcs.com (M. Sinha); rajat.pal@tcs.com (R. Pal); dasgupta.tirthankar@tcs.com (T. Dasgupta)

id 0000-0002-2653-5092 (M. Sinha)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## 2. Task Description

Development of a Question Answering based conversational system that can help in the early detection of GI cancer, given information on general symptoms, diagnosis and medical history of a patient.

The task participants were provided with a sample set of 30 questions and corresponding model answers that helped the participants develop the system's capabilities. The participants were encouraged to leverage other open source biomedical data sets and knowledge-bases. They were expected to extract relevant biomedical entities and their relationships and also standardized them from their canonical forms.

<p><b>2. Doctor's Query to Mobile App:</b></p> <p>Patient: Sarah Jones Age: 55 Sex: Female</p> <p>Presenting Symptoms:</p> <ul style="list-style-type: none"><li>Abdominal pain and discomfort</li><li>Early satiety</li><li>Nausea</li><li>Weight loss</li><li>Abdominal mass palpable on examination</li></ul> <p>Vitals:</p> <ul style="list-style-type: none"><li>Temperature: 98.8°F (37.1°C)</li><li>Heart Rate: 82 beats per minute</li><li>Blood Pressure: 118/72 mmHg</li><li>Respiratory Rate: 14 breaths per minute</li></ul> <p>Lab Results:</p> <ul style="list-style-type: none"><li>Complete Blood Count (CBC): Within normal limits</li><li>Tumor Markers: Elevated serum levels of chromogranin A</li><li>Stool Guaiac Test: Negative</li></ul> <p>Imaging Studies:</p> <ul style="list-style-type: none"><li>Abdominal Ultrasound: Large mass identified in the stomach wall.</li><li>Computed Tomography (CT) Scan of Abdomen: Large, well-defined mass in the stomach wall, suggesting a gastrointestinal stromal tumor (GIST). Biopsy obtained and sent for pathology.</li></ul> <p><b>Question:</b> This patient presents with [briefly summarize the main concerns]. The imaging suggests a GIST. Can you provide me with information on the role of KIT and PDGFRA mutations in the pathogenesis of these tumors, their clinical implications, and available treatments?</p>	<p><b>Response :</b> Based on the information you provided, here's a recommended course of action for Sarah Jones:</p> <p>Diagnosis:</p> <ul style="list-style-type: none"><li>The patient's symptoms, imaging findings (abnormal abdominal mass on ultrasound and CT scan), and elevated chromogranin A levels strongly suggest a gastrointestinal stromal tumor (GIST).</li><li>Next Step: Order genetic testing for KIT and PDGFRA mutations on the biopsy obtained during the CT scan. This will confirm the diagnosis and provide valuable information for treatment planning.</li></ul> <p>Treatment:</p> <ul style="list-style-type: none"><li>If a KIT or PDGFRA mutation is identified: Imatinib (a tyrosine kinase inhibitor) is the recommended first-line treatment.</li><li>If the patient doesn't have a KIT or PDGFRA mutation: Alternative treatment options, such as surgery, may be considered.</li><li>If the patient has a KIT or PDGFRA mutation, but the tumor is resistant to imatinib: Other tyrosine kinase inhibitors, such as sunitinib and regorafenib, may be effective.</li></ul> <p>Key Considerations:</p> <ul style="list-style-type: none"><li>Staging: Once the diagnosis is confirmed, it's important to determine the stage of the GIST using imaging studies (CT scan, PET scan). This helps to assess the tumor's size, location, and spread, guiding treatment decisions.</li><li>Patient's Overall Health: Consider the patient's age, general health status, and potential comorbidities when planning treatment.</li><li>Multidisciplinary Approach: Collaborate with a multidisciplinary team, including a medical oncologist, surgical oncologist, and pathologist, to provide the best possible care for the patient.</li></ul>
--	---

Figure 1: Sample Data Provided to the Participants

### 2.1. Evaluation Metrics

For evaluation purposes, as test data, 50 questions were shared with the participants for submitting the corresponding answers as generated by their developed conversational system.

The answers were evaluated against our ground truth answers on the following criteria:

- Concepts/entities/relationships correctly identified
- Linguistics correctness and meaningfulness of the answers
- Consistency in the answers when asked similar question with different paraphrases
- Confidence in the questions when doubted

### 2.2. Participants

Total five teams across various institutions submitted their results for the task.

## 3. Observation

It was interesting to observe that different teams approached the problem very differently. While some teams opted for high performing LLMs such as GPT 3.5, others preferred transformer models, general and domain specific, and also leveraged retrieval augmented generation. Apart from the model diversity, teams also augmented their datasets from sources such as wikipedia, PubMed.

The details of each team's implementation and relative performance are discussed in the individual team working notes. We present here the overall findings:

3. Doctor's Query to Mobile App:

Patient: Michael Brown Age: 50 Sex: Male

Presenting Symptoms:

- Abdominal pain and distention
- Nausea and vomiting
- Jaundice
- Unexplained weight loss

Vitals:

Temperature: 99.0°F (37.2°C)

Heart Rate: 100 beats per minute

Blood Pressure: 140/90 mmHg

Respiratory Rate: 20 breaths per minute

Lab Results:

- Elevated bilirubin levels
- Elevated CA 19-9 levels

Imaging Studies:

- Abdominal Ultrasound: Showed a pancreatic mass
- CT Scan of Abdomen: Confirmed a pancreatic head tumor

**Question:** This patient likely has pancreatic cancer. Can you provide information on the role of BRCA mutations in pancreatic cancer and potential implications for treatment?

**Figure 2:** Data for Evaluation Provided to the Participants

Team Name	Institute
Cancer-Answer	IIT Kharagpur
BITSCSIS	BITS Pilani
Turing	University of Petroleum and Energy Studies
Bug Smashers	University of Petroleum and Energy Studies
SSN_GenAI_AA	Sri Sivasubramaniya Nadar College of Engineering

**Figure 3:** Teams Details

1. GPT-3.5 Turbo based conversational system demonstrates potential for generating accurate and relevant information regarding GI cancers. The system also scored high in entity accuracy and linguistic correctness and meaningfulness.
2. A combined system using query categorization, RoBERTa-based retrieval from a vector database, keyword boosting, and BioGPT-based response generation effectively interprets complex and unstructured user queries related to GI cancers.
3. A BERT-based question-answering system proposed by one of the teams demonstrates the

potential of using alternative to LLMs to provide accurate and relevant information about GI cancers. However, it performs not so good on BLEU and ROGUE scores.

4. A system leveraging Electronic Health Records (EHR) data provides significant performance improvement over the baseline.

## **4. Conclusion**

The shared task on building conversational agents for aiding the diagnosis of GI cancer has yielded valuable insights into the potential and challenges of leveraging large language models (LLMs) and transformer-based architectures for healthcare applications. Teams developed a range of solutions, integrating both general-purpose models and domain-specific models tailored to the nuances of gastrointestinal oncology. One key finding is that domain-specific models, when properly fine-tuned, demonstrated enhanced accuracy in understanding and responding to medical queries, especially those related to complex diagnostic processes and symptom interpretation. However, the general-purpose models, while versatile, often required additional contextualization and retraining to perform optimally in this specialized domain. Another important conclusion is the critical role of data quality and the ethical considerations surrounding the use of sensitive medical data.

The impact of this task is profound, as it highlights the ability of conversational agents to assist healthcare professionals by streamlining the diagnostic process, improving patient outcomes, and offering scalable solutions to support clinicians. This collaborative effort marks a significant step toward more intelligent, efficient healthcare solutions in oncology.

## **Declaration on Generative AI**

During the preparation of this work, the author(s) used open sourced LLMs to process pubmed abstracts for inference, aggregation and generation of sample use-cases . After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.