IJCNLP-AACL 2025

The 14th IJCNLP and 4th AACL Conference

Multilingual Clinical Dialogue Summarization And Information Extraction With Qwen-1.5B LoRA

Kunwar Zaid, Amit Sangroya, Jyotsana Khatri
TCS Research, New Delhi, India
{kunwar.zaid, amit.sangroya, khatri.jyotsana}@tcs.com

MOTIVATION

- Clinical conversations are long, noisy, and multilingual, making automated analysis challenging.
- Healthcare systems require Concise clinical summaries
- Most LLM solution are GPU-intensive
- The shared task restricts model size, requiring efficient adaptation under limited compute.
- Goal: Build an efficient, multilingual system under constrained compute

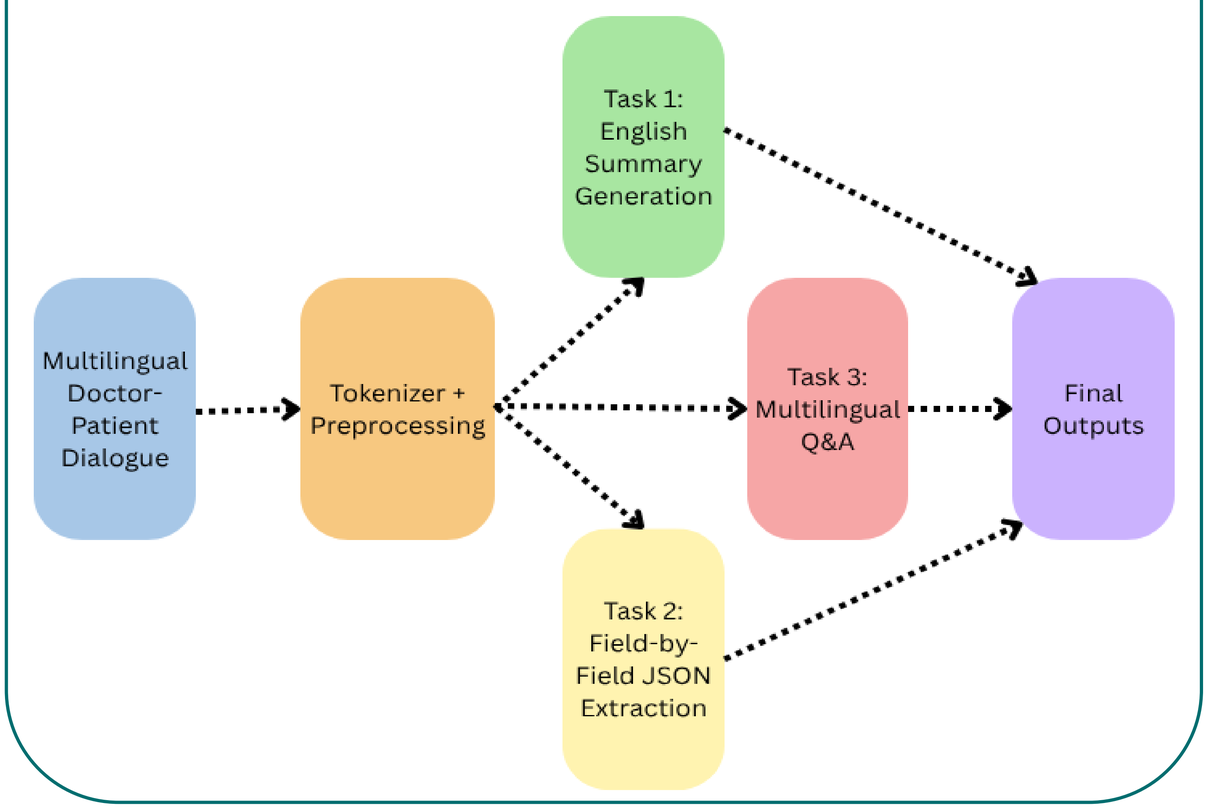
TASK DESCRIPTION

- **Given a doctor-patient dialogue in one of 10 languages, generate**
  - English clinical summary
  - Schema-aligned structured information (JSON)
  - Multilingual question-answer responses
- Languages include:
  - English, Hindi, Gujrati, Tamil, Telugu, Marathi, Kannada, Bangla, Assamese

METHODOLOGY

- **English Summarization**
  - 6-10 sentence clinical summary
  - Focus on diagnosis, symptoms, investigations, and management.
- **Field-wise JSON Extraction**
  - Each schema field queried independently
  - Outputs merged into valid JSON
- **Multilingual Q&A**
  - Answers generated in the dialogue's original language

- **Model & Training**
  - Base model: Qwen-1.5B-Instruct
- **Parameter-efficient fine-tuning:** LoRA
  - Rank r = 8, alpha = 32, dropout = 0.05
- **Quantization:**
  - 4-bit NF4 (BitsAndBytes)
- **Training**
  - Single epoch
  - AdamW optimizer with cosine decay
  - Gradient checkpointing + mixed precision
  - Trained under strict GPU and memory constraints



Experiments And Evaluation

Role-based System-User prompts for each task
- Summarization
- Structured extraction
- Multilingual Q&A

Explicit output constraints
- Fixed length summaries
- "N/A" for missing fields
- Language-controlled Q&A responses

Why Field-wise Json Extraction?
Single shot JSON generation failed
- Invalid JSON syntax
- Empty or null fields
- Schema inconsistency

Our Solution:
- Reformulate each field as a separate Q&A task
- Post processing merges answers into valid JSON

Evaluation Setup

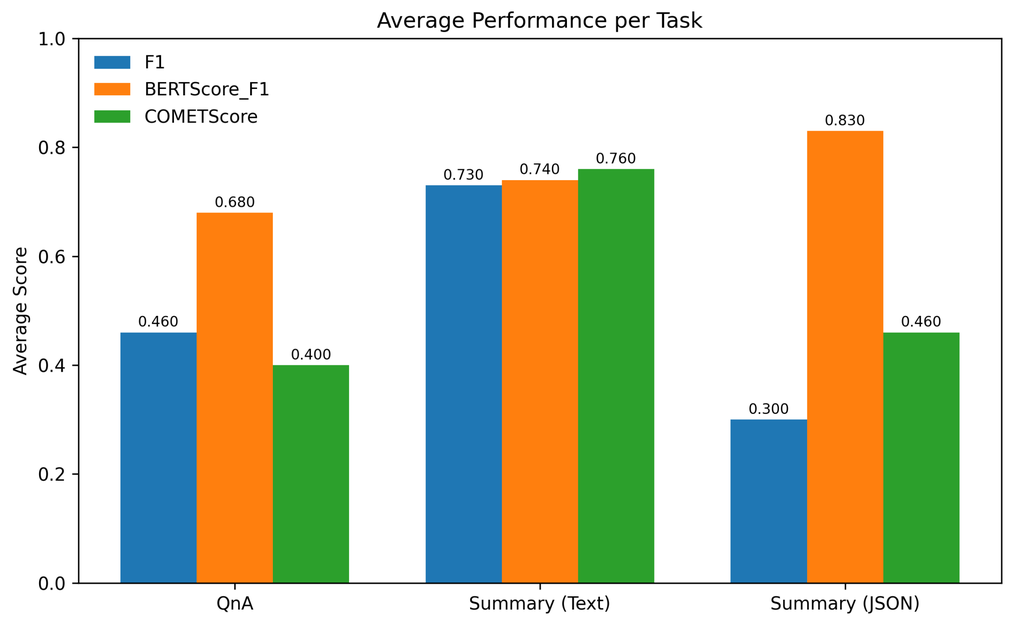
Official NLP-AI4Health 2025 test set
Metrics
- **Q&A:** Macro F1
- **Summarization:** Rouge-L, BERTScore-F1
- **Structured extraction:** Key-Value F1 (KNV)

Result & Observation

Q&A: F1 = 0.46
Summarization
- ROUGE-L = **0.178**
- BERT-F1 = **0.83**

Structured JSON Extraction: KNV F1 = **0.296**

Key Observations
- High semantic alignment despite low lexical overlap
- Strong performance in:
- Moderate degradation in low-resource languages



References

1. Abacha, Asma Ben, et al. "Overview of the MEDIQA-Chat 2023 shared tasks on the summarization & generation of doctor-patient conversations." *Proceedings of the 5th Clinical Natural Language Processing Workshop*. 2023.
2. Dettmers, Tim, et al. "Qlora: Efficient finetuning of quantized llms." *Advances in neural information processing systems* 36 (2023): 10088-10115.
3. Hu, Edward J., et al. "Lora: Low-rank adaptation of large language models." *ICLR* 1.2 (2022): 3.
4. Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries." *Text summarization branches out*. 2004.