

# MedDialogue: Pediatric Malnutrition Assessment

---

## Project Information

---

**Lead/Mentor:** Frederick Gyasi

**Contributors:** Frederick Gyasi

**Institution:** Medical University of South Carolina, Biomedical Informatics Center, Clinical NLP Lab

**Current Funding:** N/A

**Future Funding:** N/A

**IRB #:** N/A

**RMID:** N/A

**SPARCRequest:** N/A

---

## Project Summary

---

MedDialogue is a clinical NLP framework for fine-tuning large language models (LLMs) on medical dialogue tasks, with current implementation focused on pediatric malnutrition assessment using ASPEN, WHO, and CDC guidelines.

The framework enables training domain-specific conversational models that perform multi-turn clinical assessments, supporting evidence-based diagnostic reasoning, temporal pattern analysis, and structured output generation (TEXT, JSON, XML, Markdown).

## Key Research Questions

1. **Can fine-tuned LLMs accurately assess pediatric malnutrition using clinical notes and ASPEN diagnostic criteria?**

2. How effective is multi-turn conversational training for clinical reasoning tasks compared to single-turn approaches?
- 

## Project Experiments

---

### Current Implementation

Component	Description
Task Domain	Pediatric malnutrition assessment
Clinical Guidelines	ASPEN, WHO, CDC
Assessment Fields	11 domains (case presentation, clinical symptoms, growth/anthropometrics, physical exam, nutrition/intake, labs/screening, diagnosis/reasoning, malnutrition status, care plan, social context, clinical insights)
Question Templates	110 total (10 per field)
Output Formats	TEXT (40%), JSON (35%), XML (10%), Markdown (15%)
Supported Models	LLaMA 3.1 8B, Phi-4, Mistral 7B, Qwen 2.5 7B
Training Method	LoRA/QLoRA fine-tuning with Unslot optimization
Quantization	4-bit (bnb-4bit)

### Evaluation Approach

Metric	Description
Classification	Binary malnutrition present/absent
Primary Metrics	Accuracy, Precision, Recall, F1-Score, Specificity
Additional Metrics	MCC, AUC-ROC

Metric	Description
Evaluation Pattern	Multi-turn conversation (6 questions per patient)

## Training Configuration

Parameter	Default Value
Max Sequence Length	16,384 tokens
Batch Size	2
Learning Rate	2e-4
LoRA Rank	16
LoRA Alpha	16
Gradient Accumulation	4 steps
Precision	BF16

## Resources

### Code Repository

**GitHub:** <https://github.com/gyasifred/meddialogue>

### Framework Components

Module	Function
meddialogue/core.py	Main MedDialogue class orchestrating training pipeline
meddialogue/train.py	Training utilities with conversation generation
meddialogue/models.py	Model loading, LoRA application, registry
meddialogue/inference.py	Single/multi-turn inference pipeline

Module	Function
<code>meddialogue/data_prep.py</code>	Data preprocessing and conversation formatting
<code>meddialogue/config.py</code>	Configuration dataclasses (Task, Model, LoRA, Training, Safety, Conversation)
<code>meddialogue/utils.py</code>	Text processing, output formatting, parsing utilities
<code>train_malnutrition.py</code>	Malnutrition-specific training script
<code>evaluate_malnutrition.py</code>	Multi-turn evaluation with classification metrics
<code>gradio_chat_v1.py</code>	Interactive web interface for clinical consultation
<code>malnutrition_system_prompts.py</code>	Domain-specific system prompts

## Datasets

### Input Requirements:

Column	Description
<code>txt</code>	Clinical note text
<code>input_label_value</code>	Binary label (0=absent, 1=present)
<code>case_presentation</code>	Case summary
<code>clinical_symptoms_and_signs</code>	Symptom documentation
<code>growth_and_anthropometrics</code>	Growth measurements
<code>physical_exam</code>	Physical examination findings
<code>nutrition_and_intake</code>	Dietary intake information
<code>diagnosis_and_reasoning</code>	Clinical reasoning
<code>labs_and_screening</code>	Laboratory values
<code>care_plan</code>	Treatment recommendations
<code>clinical_insights</code>	Key insights

Column	Description
social_context (optional)	Social determinants

---

## Acknowledgements

---

**Lead/Mentor:** Frederick Gyasi (gyasi@musc.edu)

**Contributors:** Frederick Gyasi

**Past Contributors:** N/A

**Institution:** Medical University of South Carolina, Biomedical Informatics Center, Clinical NLP Lab

**Current Funding:** N/A

**Future Funding:** N/A

**License:** MIT

---

## Technical Dependencies

---

Package	Purpose
Unsloth	Fast LLM fine-tuning with memory optimization
Transformers	Hugging Face model infrastructure
PEFT	Parameter-efficient fine-tuning (LoRA)
PyTorch	Deep learning framework
Gradio	Interactive web interface
scikit-learn	Evaluation metrics
pandas	Data processing

---

Contact: [gyasi@musc.edu](mailto:gyasi@musc.edu)

