# FLAN-T5 Medical AI

# Confusion Matrix Analysis Report

Generated: 2025-10-19 16:41:52

## Executive Summary

This report presents a comprehensive confusion matrix analysis of the FLAN-T5 medical AI model's diagnostic classification performance. The model was evaluated on 50 test cases across 11 disease categories. **Key Findings:**
• Overall Classification Accuracy: 58.0%
• Correct Predictions: 29/50
• Average Inference Time: 0.068 seconds
• Best Performing Category: Mental_Health
• Lowest Performing Category: Emergency

## Model Information

| Property | Value |
|---|---|
| Model Type | FLAN-T5 |
| Model Path | models\flan_t5_diagnosis_20251018_205708 |
| Total Parameters | 247,577,856 |
| Model Size | 944.4 MB |
| Device | cuda |
| PyTorch Version | 2.5.1+cu121 |

## Classification Performance Summary

| Category | TP | FP | TN | FN | Precision | Recall | F1-Score | Specificity | Support |
|---|---|---|---|---|---|---|---|---|---|
| Cardiovascular | 4 | 4 | 41 | 1 | 0.500 | 0.800 | 0.615 | 0.911 | 5 |
| Emergency | 0 | 0 | 45 | 5 | 0.000 | 0.000 | 0.000 | 1.000 | 5 |
| Endocrine | 4 | 3 | 42 | 1 | 0.571 | 0.800 | 0.667 | 0.933 | 5 |
| Gastrointestinal | 0 | 0 | 45 | 5 | 0.000 | 0.000 | 0.000 | 1.000 | 5 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Infection | 5 | 5 | 40 | 0 | 0.500 | 1.000 | 0.667 | 0.889 | 5 |
| Mental_Health | 5 | 0 | 45 | 0 | 1.000 | 1.000 | 1.000 | 1.000 | 5 |
| Musculoskeletal | 4 | 0 | 45 | 1 | 1.000 | 0.800 | 0.889 | 1.000 | 5 |
| Neurological | 3 | 1 | 44 | 2 | 0.750 | 0.600 | 0.667 | 0.978 | 5 |
| Other | 0 | 6 | 44 | 0 | 0.000 | 0.000 | 0.000 | 0.880 | 0 |
| Preventive | 0 | 0 | 45 | 5 | 0.000 | 0.000 | 0.000 | 1.000 | 5 |
| Respiratory | 4 | 2 | 43 | 1 | 0.667 | 0.800 | 0.727 | 0.956 | 5 |

# F1-Score and Classification Metrics Summary

**Overall Performance Metrics:**
• Overall Accuracy: 0.580
• Macro-averaged F1-Score: 0.476
• Micro-averaged F1-Score: 0.580
• Macro-averaged Precision: 0.453
• Macro-averaged Recall: 0.527
**Aggregate Confusion Matrix Components:**
• True Positives (TP): 29
• False Positives (FP): 21
• True Negatives (TN): 479
• False Negatives (FN): 21
**F1-Score Interpretation:**
• F1-Score is the harmonic mean of Precision and Recall
• F1 = 2 × (Precision × Recall) / (Precision + Recall)
• Range: 0.0 (worst) to 1.0 (perfect)
• Macro F1: Unweighted average across all classes
• Micro F1: Weighted by class frequency

# Detailed Confusion Matrix Analysis

The confusion matrix reveals how well the model distinguishes between different disease categories. Diagonal elements represent correct classifications, while off-diagonal elements show misclassifications. **Key Observations:**
• Best classified category: Infection (100.0%)
• Most challenging category: Preventive (0.0%)
• Most common misclassification: Emergency → Infection (3 cases)

# Sample Predictions Analysis

### ■ *Correct Classifications (Sample):*

**1. Question:** I have fever, cough, and sore throat
**Response:** infection...
**Category:** Infection ■

**2. Question:** What are symptoms of pneumonia?
**Response:** pneumonia...
**Category:** Infection ■

**3. Question:** How to treat viral infection?
**Response:** infection...
**Category:** Infection ■

**4. Question:** Bacterial vs viral infection differences
**Response:** infection...
**Category:** Infection ■

**5. Question:** When to use antibiotics for infection?
**Response:** infection...
**Category:** Infection ■

### ■ *Misclassifications (Sample):*

**1. Question:** I have chest pain and shortness of breath
**Response:** bronchitis...
**Expected:** Cardiovascular | **Predicted:** Respiratory ■

**2. Question:** Breathing difficulty at night
**Response:** hypertension...
**Expected:** Respiratory | **Predicted:** Cardiovascular ■

**3. Question:** Dizziness and balance problems
**Response:** hypertension...
**Expected:** Neurological | **Predicted:** Cardiovascular ■

**4. Question:** Early signs of stroke recognition
**Response:** stroke...
**Expected:** Neurological | **Predicted:** Cardiovascular ■

**5. Question:** Persistent nausea and vomiting
**Response:** headache...
**Expected:** Gastrointestinal | **Predicted:** Neurological ■