

Frac-MAS: Clinician Evaluator Guide

Protocol for Assessing AI Diagnostic Accuracy & Patient Clarity

Overview

Thank you for participating in the validation of **Frac-MAS**. We are building a system designed not only to detect fractures but to explain them in language that patients can understand.

As a clinical expert, your role is twofold:

1. **Verify:** Is the AI medically correct?
2. **Critique:** Would this explanation make sense to a non-medical patient without your intervention?

This document serves as your **grading rubric** for the Google Form evaluation.

1 The Workflow

For each case in the form, you will follow this 3-step process:

Step 1: Blind Diagnosis (The Control)

You will see the **raw X-Ray image** first.

- **Action:** Please enter your diagnosis and recommended treatment in the text fields provided.
- *Note: This establishes the ground truth before you see the AI's opinion.*

Step 2: Review AI Output

The form will then reveal the "**System-Generated Diagnosis**." This includes:

- **Prediction:** The specific fracture type (e.g., "Spiral Fracture").
- **Patient Translation:** A plain-English description (e.g., "A twisting break").
- **Technical Analysis:** A description of visual features and attention regions.

Step 3: Rate the System

You will answer two key questions using a 5-star scale. Please use the rubrics below to ensure consistency.

2 Grading Rubric

Question 1: "How technically accurate is the AI response?"

Focus: Clinical Precision. Does the AI match your expert diagnosis?

Stars	Criteria
5	Perfect. The AI correctly identifies the fracture existence, the specific subtype (e.g., Spiral vs. Oblique), and the anatomical location.
4	Clinically Acceptable. Correct existence and general location, but misses a minor nuance (e.g., identifies "Distal Radius" but misses "Intra-articular extension").
3	Partial Correctness. Correctly identifies "Fracture Present," but misclassifies the type (e.g., calls a Spiral fracture an Oblique fracture).
2	Significant Error. Misses a clear fracture (False Negative) or hallucinates a fracture in a healthy bone (False Positive).
1	Critical Failure. Completely wrong diagnosis (e.g., wrong bone, wrong side, or non-sensical output).

Question 2: "How easy is it for a common person to understand?"

Focus: Patient Communication. Imagine the patient reads this text on their phone before seeing you. Will they understand it, or will they be confused/scared?

Stars	Criteria
5	Empowering & Clear. The AI translates jargon into plain English (e.g., "A twisting break" alongside "Spiral Fracture"). The tone is calm, and the "Recommended Actions" are actionable for a layperson.
4	Good Translation. Uses mostly simple language but leaves 1-2 medical terms undefined (e.g., uses "distal" or "commuted" without context).
3	Textbook Style. The explanation is grammatically correct but reads like a medical report. Too dense for an average patient (e.g., "Disruption of the cortical margin at the diaphysis...").
2	Confusing / Vague. The language is generic (e.g., "Bone issue detected") or creates unnecessary anxiety without explaining why.
1	Incomprehensible. The output looks like raw code, debug logs, or is written in a robotic, unnatural syntax.

Thank you for helping us bridge the gap between AI diagnostics and patient care.