Prompt

You are an assistant to an expert ophthalmologist and have professional level knowledge of a series of eye diseases, specifically Diabetic Retinopathy. You also have extensive experience working with Convolutional Neural Networks, understanding how they work.

A recent Convolutional Neural Network has been developed to diagnose Diabetic Retinopathy from retinal fundus images. The diagnoses are given as one of 4 classes: level 0 – no DR; level 1 – mild DR, level 2 – Moderate & Severe DR and finally level 3 – Proliferative DR.

Your job is to explain the diagnoses provided by the CNN to the patient.

For each diagnosis, you are provided with the original retinal image, the processed image (which is sent to the model for training), a Grad-CAM heatmap which shows the areas of the processed image that contributed to its classification, and the given classification. Given these resources, you write a concise, professional explanation to the patient, detailing how the CNN has come to make its decision and what this means. You write your explanation in under six sentences, maintaining good bedside manner, and writing in a way that is understandable to the patient who has limited technical knowledge of the disease and machine learning techniques. If you are unsure of anything, you pass on this information to the ophthalmologist you assist with rather than writing to the patient.

Your tone is professional and neutral while maintaining empathy for the patient.

Original image:

Processed image:

Heatmap:

Diagnosis:

Your explanation:

Example Response:

Based on the analysis of your retinal images, the system has identified signs consistent with **Moderate to Severe Diabetic Retinopathy (level 2)**. This means that there are noticeable changes in the blood vessels in your retina, such as bleeding, swelling, or abnormal growth, which may affect your vision if not monitored or treated. The heatmap you see highlights areas in the retina that the model focused on—these are regions showing these concerning changes. While this is not the most advanced stage, it does require medical follow-up to help prevent further progression. We recommend scheduling a detailed examination with your ophthalmologist soon, who can confirm the diagnosis and discuss possible treatment options with you. Please know that this finding is manageable with proper care and regular monitoring.

Would you like help understanding what your next steps might involve?

AI Evaluation Result

### **Professionalism score: ⭐⭐⭐☆ (3/5)**

**Critique:**

* While the tone is calm and generally polite, it borders on *too reassuring*, which can come across as minimizing the gravity of the diagnosis.
* It claims "the system has identified..." without clarifying the limitations of AI or the fact that such tools are not diagnostic authorities — this lacks professional transparency.
* It blurs the lines between clinical authority and algorithmic suggestion, which may mislead patients into thinking the AI output is definitive.
* No mention of the fact that this is *not* a substitute for a clinician’s assessment — that’s a critical omission in professional communication.

### **Accuracy score: ⭐⭐☆ (2.5/5)**

**Critique:**

* It states the diagnosis is "Moderate to Severe Diabetic Retinopathy (level 2)" — this is misleading. **“Moderate to severe”** is a vague combo of stages. If using a numeric grade (e.g., level 2), it should explicitly reference a validated scale like the **ICDR** or **ETDRS** classification. Combining terms like “moderate to severe” with “level 2” is confusing and potentially inaccurate.
* The mention of "abnormal growth" (likely referencing neovascularization) is problematic. **Neovascularization is not present in moderate DR**, only in proliferative DR — a much more severe stage. This is a factual error and can significantly misrepresent the diagnosis.
* It attributes clinical signs to image-based AI findings without addressing possible false positives/negatives, variability between models, or limitations of image quality.

### **Understandability score: ⭐⭐⭐⭐ (4/5)**

**Critique:**

* It’s relatively clear in its use of language, and explains findings in simple terms like "bleeding," "swelling," and “abnormal growth.”
* However, it *over-simplifies* the implications without contextualizing uncertainty.
* The mention of the heatmap is a nice touch, but again, it lacks explanation of what the colors mean or how reliable that visualization is — patients might over-trust it.
* There is no clarification that this is **AI-assisted information**, not a doctor’s verdict — which makes it riskier from a patient comprehension standpoint.

### **Suggestion 1: Enforce a Clear Disclaimer About AI Limitations**

**Why:** The prompt fails to require the assistant to *explicitly* convey to the patient that the diagnosis comes from an AI model — which is not a licensed medical authority. Without a disclaimer, it risks misleading patients into viewing the AI output as equivalent to a doctor’s diagnosis.

### **Suggestion 2: Require Clinical Accuracy Checks for Each Stage Description**

**Why:** The prompt allows for potential medical inaccuracy by grouping "Moderate & Severe DR" as level 2 without clarifying which clinical features define each. This could result in misleading or factually incorrect statements about the patient's condition.

### **Suggestion 3: Clarify Expectations Around Visual Data Interpretation**

**Why:** The assistant is told that a heatmap is provided but not instructed on *how* or *whether* to interpret it. This leads to vague or misleading references to the heatmap’s meaning and significance.