

EffloraBot: An AI-Driven Interactive Learning Tool for Structured Dermatological Descriptions

Stephan Blüthgen, Christian Greis

Department of Dermatology, University Hospital Zurich, Switzerland



BACKGROUND

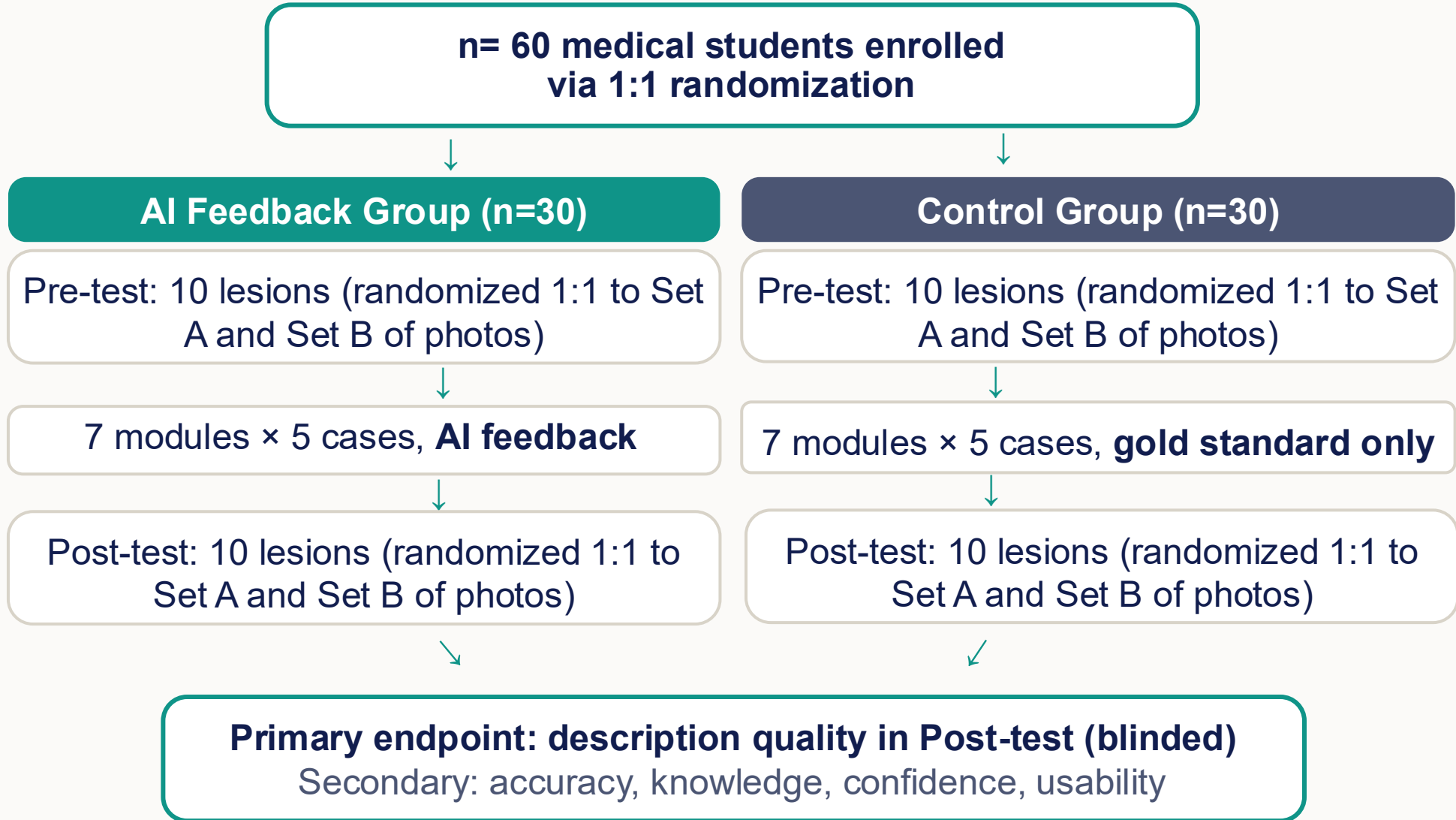
The accurate morphological description of skin lesions is a fundamental clinical skill, yet notoriously challenging for medical students to master. Providing individualized, high-quality feedback is highly resource-intensive for clinical faculty. To close this educational gap, we developed EffloraBot, a web-based application that uses a Large Language Model to give students immediate, structured feedback on their free-text lesion descriptions.

METHODS

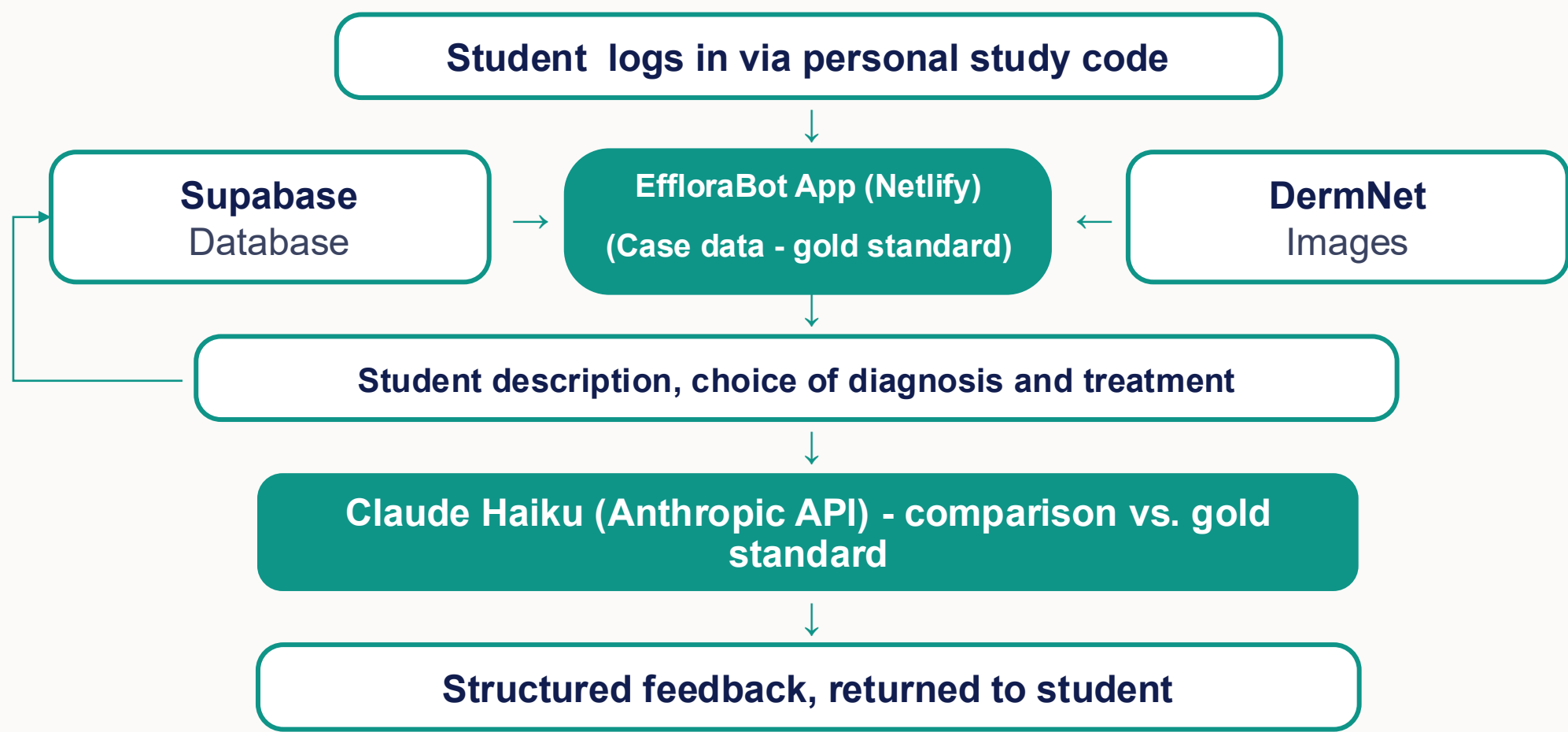
We conducted a prospective, randomized, controlled, two-arm trial with medical students (n = 60). Participants evaluated 55 clinical cases across 7 progressive modules. The intervention group received automated, LLM-generated feedback structured along five standardized axes (lesion type, count, size, color and distribution). The control group received only the gold-standard description, diagnosis, and therapeutic guidelines, without individualized morphological feedback. The primary endpoint was objective description quality, evaluated using automated LLM-based scoring (Claude Sonnet 4.6, Anthropic) compared against our gold standard descriptions. Secondary endpoints assessed diagnostic accuracy, therapeutic knowledge and subjective diagnostic confidence (5-item Likert scale).

DESIGN & ARCHITECTURE

Study Design



System & Data Flow



RESULTS

Sixty medical students were enrolled and randomized (n=30 per group). Both groups improved significantly from pre- to post-test (both p<0.001). Despite higher baseline scores in the control group for description quality and therapeutic knowledge, the AI feedback group showed significantly greater improvement in lesion description quality (+12.4 vs +5.0 points, p<0.001, ANCOVA adjusting for baseline), diagnostic accuracy (+4.4 vs +3.4, p<0.001), therapeutic knowledge (+3.2 vs +1.2, p<0.001), and self-reported confidence (+7.1 vs +5.8, p<0.001). System usability was rated higher in the AI group (18.8 vs 17.5/25, p<0.001).

CONCLUSION

Structured, AI-generated feedback significantly improves medical students' ability to describe skin lesions, with effects extending to diagnostic accuracy, therapeutic knowledge, and self-confidence. These findings support the integration of LLM-based feedback tools into dermatological curricula.

Objective description quality (Pre- vs. Post-Test, maximum score 50 points)

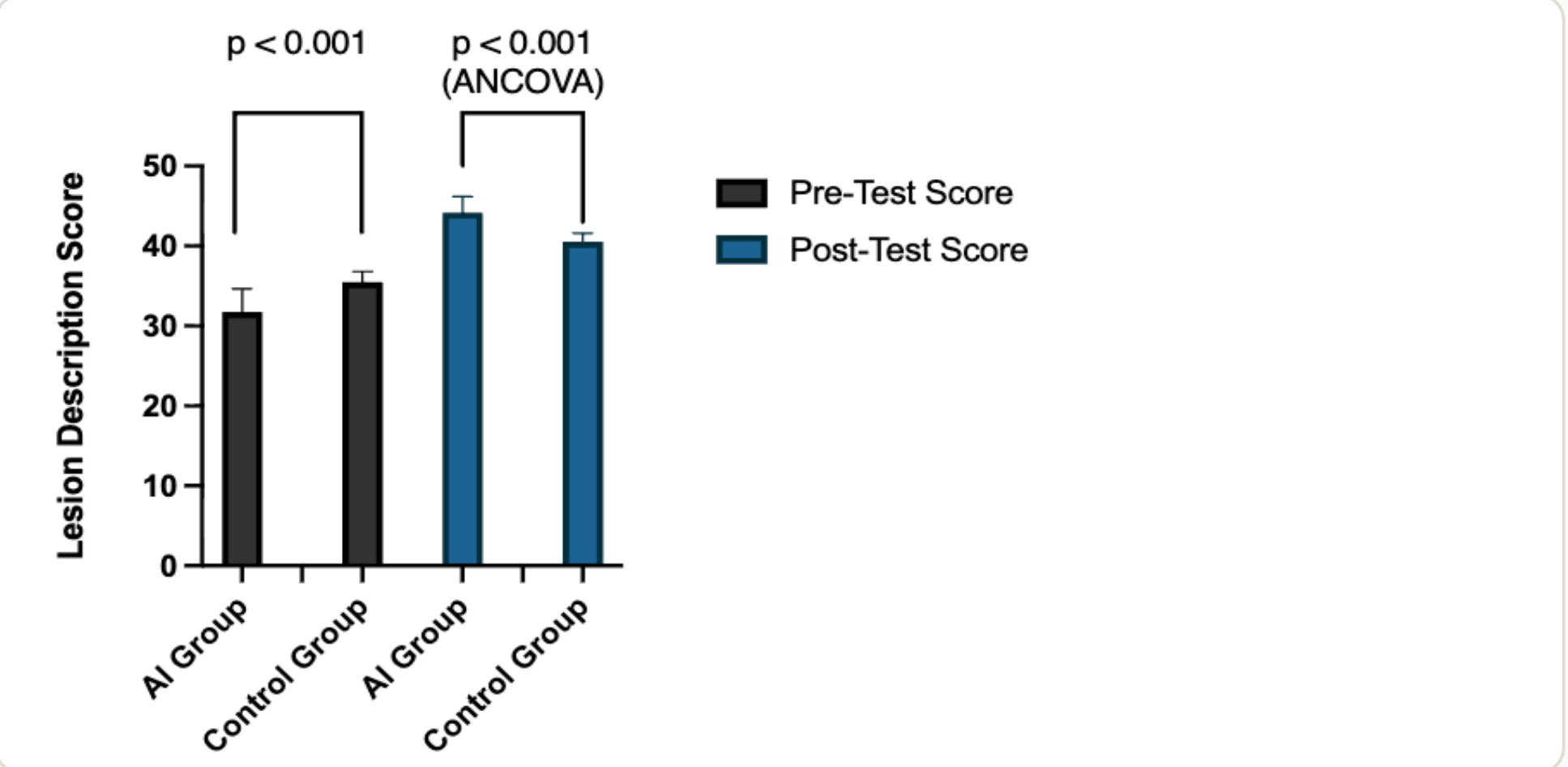


Figure 1. Objective description quality before and after completing all modules (maximum score: 50 points, assessed across 5 categories: primary lesion type, localization, distribution pattern, count, and size). ANCOVA adjusting for baseline, p<0.001.

Improvement across all endpoints (AI vs. Control)

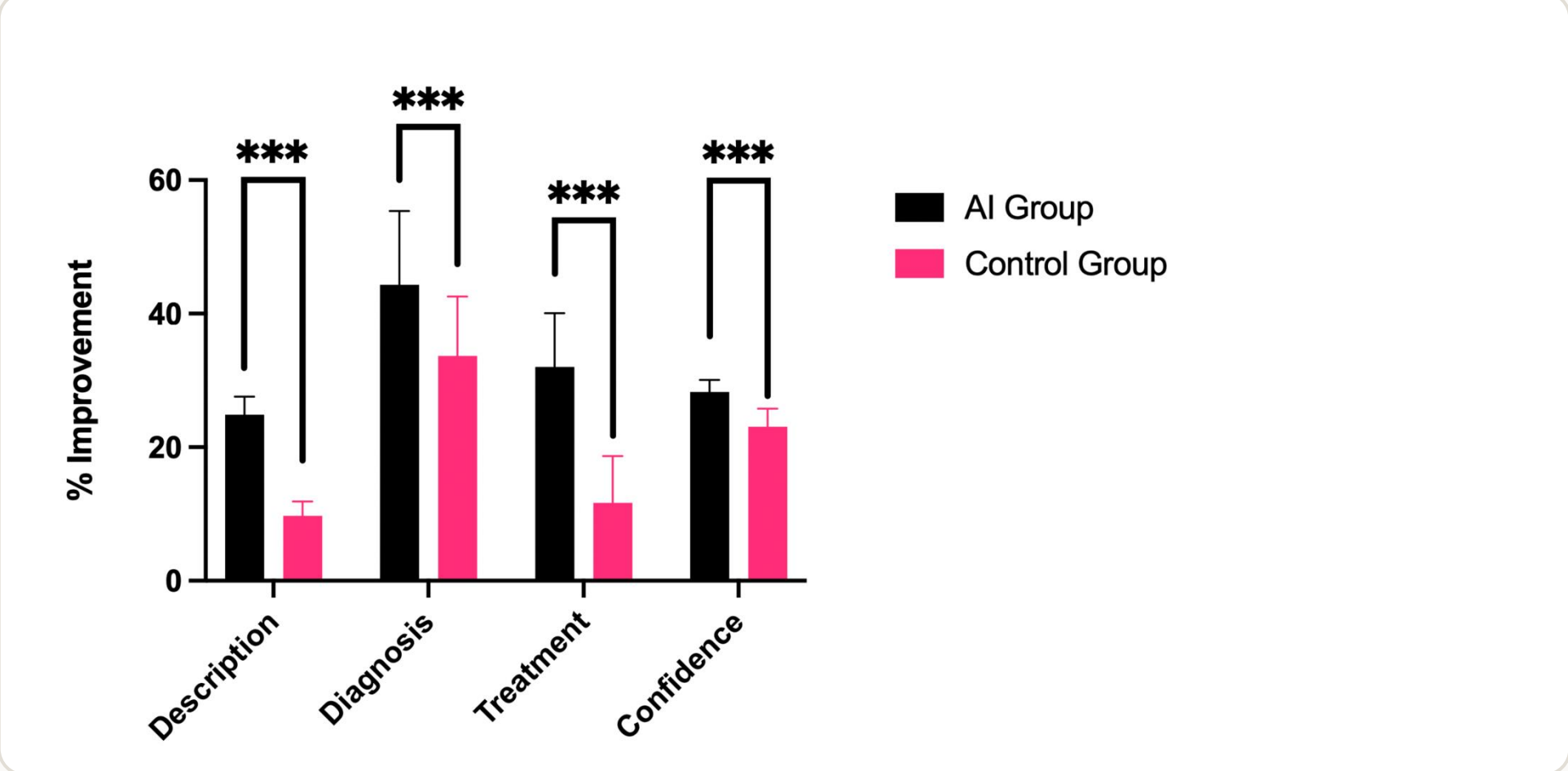


Figure 2. Mean relative improvement across all endpoints. Welch's t-test, *** p < 0.001..

Satisfaction with the app (maximum score 25 points)

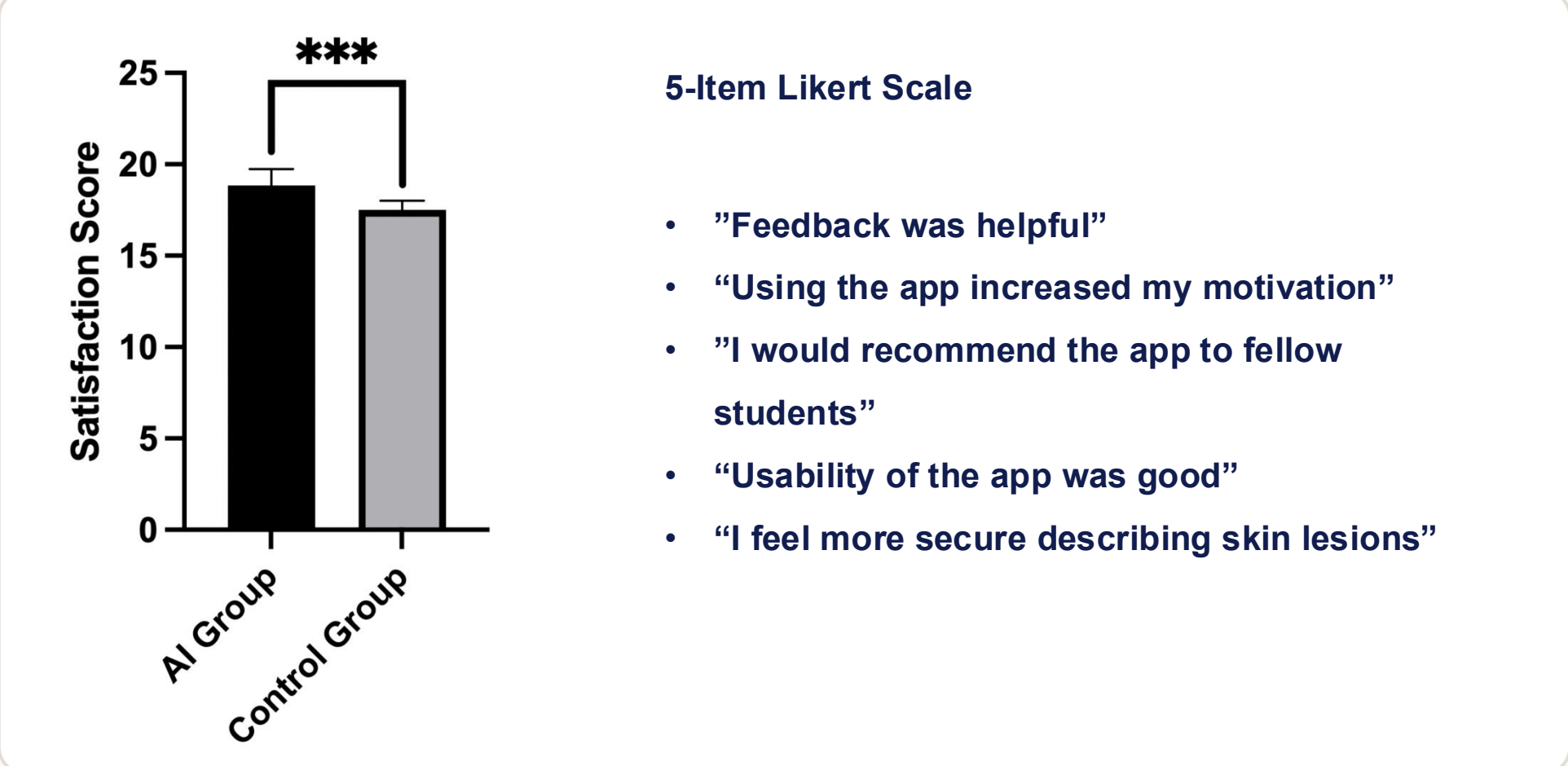


Figure 3. System satisfaction (5-item Likert scale, maximum 25 points). Mann-Whitney U test, *** p < 0.001.

