Potentials and Pitfalls of AI in Wound Care

Navigating the risks of decision support

Dr. med. Dorothee A. Busch[1,2], Florian Kücking M.Sc.[1], Prof. Ursula Hübner[1]

[1]Forschungszentrum Gesundheits- und Sozialinformatik, Hochschule Osnabrück
[2]Universitätsklinikum Erlangen, Hautklinik, Wundzentrum

Objective

AI decision support shows immense promise for chronic wound assessment, but deployment depends on two vulnerable points: the quality of human data *input* and the clinical reception of AI *output*.

Using wound maceration as a blueprint, we evaluate both sides.

Macerated wound

The input side: Are our training data reliable?

The Visual Grading Dilemma

Moderate inter-rater reliability among clinicians on maceration (Fleiss $\kappa = 0.39$).

This reveals high variability and uncertainty in everyday visual grading of features that can't be externally verified.

What Predicts Diagnostic Accuracy?

Works: formal wound care qualifications and high self-confidence.

Doesn't: total years of experience or a general "wound care focus" did not consistently improve performance.

Takeaway for AI Development

If experts don't agree, training and validation data may be inherently unreliable, threatening downstream AI model safety.

The output side: The human element and risk of automation bias

Automation bias is a well-documented cognitive tendency of humans to over-trust automated systems. Without structured safeguards, even experienced clinicians may defer to AI output, increasing risk and impact of incorrect recommendations.

Conclusion

AI can measurably improve wound assessment. However, the same output can harm when it is wrong.
Trustworthy ground truth is essential, but cannot be assumed, even among specialists.
Those who gain most from AI are the most vulnerable to it.
Pertinant professional qualification is the strongest safeguard against this vulnerability.

Who follows wrong AI?

More likely to accept wrong AI output

- Lower baseline diagnostic performance
- No certified wound care training
- Higher perceived benefit from the system

More resistant to wrong AI output

- Higher diagnostic performance
- Certified wound-care training
- Physician role

Govern both ends of the pipeline

INPUT
→ Better ground truth

- Establish standard for labelling
- Ideally multiple qualified raters per image
- Report inter-rater reliability
- Pertinentqualification as a key factor for more reliable labelling

OUTPUT
→ Safe use

- AI-literacy training, esp. non-specialists
- Transparent, explainable recommendations
- AI systems augment, not replace the judgement of clincians

References

Kücking, F., Hübner, U., Przysucha, M., Hannemann, N., Kutza, J. O., Moelleken, M., Erfurt-Berge, C., Dissemond, J., Babitsch, B., & Busch, D. (2024). Automation Bias in AI-Decision Support: Results from an Empirical Study. Studies in health technology and informatics, 317, 298–304. https://doi.org/10.3233/SHTI240871

Kücking, F., Busch, D. A., Przysucha, M., Kutza, J. O., Hannemann, N., Hüsers, J., Babitsch, B., & Hübner, U. (2026). Impact of AI recommendation correctness on diagnostic accuracy in clinical decision-making. International journal of medical informatics, 207, 106223. https://doi.org/10.1016/j.ijmedinf.2025.106223

Kücking, F., Hübner, U. H., & Busch, D. (2025). Diagnostic accuracy differences in detecting wound maceration between humans and artificial intelligence: the role of human expertise revisited. Journal of the American Medical Informatics Association: JAMIA, 32(9), 1425–1433. https://doi.org/10.1093/jamia/ocaf116

HOCHSCHULE OSNABRÜCK
UNIVERSITY OF APPLIED SCIENCES

Uniklinikum Erlangen