

Comparison Of A Novel Deep Learning Model With Healthcare Professionals In Chronic Wound Infection Detection

Bilal M. Chaudhry, MD¹, Palawat Busaranuvong, PhD², Francisco Batiz-Fabella, MD³, Eddy C. Rios, BS³, Emma N. Mastro, BS³, Diane Strong, PhD², Bengisu Tulu, PhD², Emmanuel Agu, PhD², Raymond M. Dunn, MD⁴, Giorgio Giatsidis, M.D Ph.D³.

¹Rutgers University/ RWJ Barnabas Health, Toms River, NJ, USA, ²Worcester Polytechnic Institute, Worcester, MA, USA, ³University of Massachusetts Chan Medical School, Worcester, MA, USA, ⁴University of Massachusetts Chan Medical Center, Worcester, MA, USA.

PURPOSE:

Chronic wounds affect 10.5 million Medicare beneficiaries. With an aging population demand for wound expertise is increasing. SCARWID (Synthetic Caption Augmented Retrieval for Wound Infection Detection), a Deep Learning (DL) model, was created to detect infections in areas without wound expertise. We evaluated SCARWID performance and compared it to healthcare professionals (HCP).

METHODS:

1,208 Diabetic Foot Ulcer (DFU) images with expert assigned labels (infected vs non-infected) were used. SCARWID was trained and evaluated with 5-fold cross-validation. The same images were divided among HCPs—medical students (n=4), plastic surgery residents (n=5), emergency physicians (n=3), and plastic surgeons (n=4). Accuracy, sensitivity (SN), specificity (SP), and positive predictive value (PPV) were calculated. Paired accuracy differences between each HCP group and SCARWID were estimated.

RESULTS:

SCARWID achieved 81% accuracy, 85% SN, 78% SP, and 79% PPV. HCP overall reached 64% accuracy, 52% SN, 74% SP, and 72% PPV. Students averaged 68% accuracy (SN 68%, SP 70%, PPV 70%); residents 63% accuracy (SN 35%, SP 89%, PPV 85%); emergency physicians 57% accuracy (SN 64%, SP 50%, PPV 54%); and plastic surgeons 65% accuracy (SN 48%, SP 80%, PPV 72%). Paired accuracy differences (group – SCARWID) were –0.134 for students (–0.284 to +0.016; p=0.06), –0.188 for residents (–0.276 to –0.101; p=0.003), –0.245 for emergency physicians (–0.412 to –0.078; p=0.02), and –0.167 for plastic surgeons (–0.290 to –0.045; p=0.02).

CONCLUSION:

Overall, SCARWID outperformed every HCP group in accuracy of infection detection. These results support the potential of DL to augment infection triage in a remote setting.