Generative Artificial Intelligence for Recognition of Surgical Site Complications: The PRISCA Project

Isabel Carozzo[1], Elisa Bruzzo[2], Matteo Parodi[3], Luca Giulio Brayda[2], Michele Minuto[4] and Ennio Ottaviani[1]
[1]On Air Srl, Genoa, Italy
[2]Nextage Srl, Genoa, Italy
[3]Flairbit Srl, Genoa, Italy
[4]IRCCS San Martino Polyclinic Hospital, Genoa, Italy

isabel.carozzo@onairweb.com
elisa.bruzzo@nextage-on.com
matteo.parodi@flairbit.io
luca.brayda@nextage-on.com
michele.minuto@unige.it
ennio.ottaviani@onairweb.com

Abstract: The project PRISCA develops digital tools for postoperative monitoring and early detection of wound complications (e.g., infections) through telemedicine solutions. The approach addresses concrete clinical needs, including the follow-up of patients discharged from centers of surgical excellence located far from home, the optimization of hospital access, and the early identification of at-risk situations. The output of the project will be a telemonitoring platform (mobile and web apps), artificial intelligence modules for wound image analysis, and informative content for patients. The technological architecture is designed to support timely intervention in case of wound-related complications, while also reducing unnecessary in-person visits. The project's main innovation is the use of an Artificial Intelligence module. It aims to enable healthcare professionals to perform automated analysis of wound images for early detection of post-surgical complications. The limited availability of public data was tackled by applying a data augmentation method and by integrating a generative AI model. It creates new synthetic images based on real images and textual prompts. All generated images had been validated by clinicians and then included in the final dataset. This approach ensures the model learns from a diverse set of images, increasing its robustness and accuracy. The adopted detection model is YOLOv11, which localizes the wound and performs a pathological/non-pathological classification. Results show good localization and promising classification accuracy. We performed a comparison between the model trained on the original dataset and the version enhanced with synthetic data, in order to assess relative improvements. These comparisons will help refine the model for better performance in real-world scenarios. The first results show an increase in the performance of the model with augmented data but more systematic comparisons are needed. Additional real images from a proprietary dataset currently being collected will also be integrated, further enhancing the AI's ability to identify early complications.

Keywords: Generative AI, Medical image Classification, Telemedicine, Surgical wounds

1. Introduction

Surgical site infections (SSIs) represent one of the most common adverse events among patients undergoing surgical procedures and they can pose serious clinical complications (Haque et al., 2018; Badia et al., 2017). In addition to infections, other wound-related issues may develop over time, such as wound dehiscence or keloid formation. Postoperative follow-up protocols involve at least two face-to-face evaluations: one at a local health center approximately 15 days after surgery, and a second outpatient consultation at the hospital around 30 days later. However, the detection rate of complications during scheduled visits remains low, with many infections ultimately diagnosed only in emergency departments. Telemedicine offers new opportunities for remote patient monitoring, recovery assessment, and timely intervention in postoperative care. A survey conducted among healthcare professionals in Liguria (Italy) indicates that 71% believe an AI-powered service for automatic wound complication detection can effectively optimize the postoperative protocol. Then, the idea of developing digital tools emerges, aiming to transform the current follow-up model through smartphone-based applications. The objective of this study, conducted as part of the PRISCA project, is therefore to implement an automated system for the analysis of wound images, enabling early detection of post-surgical complications.

2. Materials and Methods

2.1 Dataset

The dataset used for the study consists of 583 images of post-operative wounds, collected from various sources. Specifically 189 images, provided by available sources (Fateen, 2017; Oota, Rowtula and Mohammed, 2023) and by Casa di Cura Villa Montallegro (Genoa, Italy), were included. They were labeled by expert clinicians from Ospedale Policlinico San Martino (Genoa, Italy), who classified them as physiological (normal healing phase) or pathological (with complications). Other 394 images were provided by Redscar© (González-Hidalgo et al., 2019a; González-Hidalgo et al., 2019b) with already assigned labels, available online at http://opendemo.uib.es/datasets/Redscar-database/. To ensure the dataset quality, images containing tattoos near the wound or belonging to poorly represented phototypes were eliminated. The final dataset consisted of 554 images, with 434 physiological and 120 pathological wounds.

2.2 Generative AI for Data Augmentation

A data augmentation procedure was adopted to increase dataset diversity. Alongside geometric techniques, an innovative method based on generative AI was introduced, using the open-source Stable Diffusion model to generate synthetic images by combining real images and descriptive text prompts. The choice of Stable Diffusion was motivated by its proven ability to generate high-quality images while preserving fine-grained details, even in domains such as medical imaging. Compared to text-to-image or image-to-image models, Stable Diffusion supports conditioning on both modalities. Previous studies demonstrated its versatility and robustness in generating synthetic datasets for clinical applications (Mou et al., 2023; Kazerouni et al., 2023). A critical aspect concerns balancing conflicting requirements: producing images sufficiently different to avoid overfitting, yet maintaining a high degree of realism with clinically plausible results. To achieve this, the generative process was controlled through two parameters: the level of detachment from the original and the adherence to the textual prompt. By varying them, multiple synthetic versions of each image were generated, altering both skin representation and wound morphology. Most common variations are: presence/ absence of moles, modified hairiness, changed skin pigmentation, and different arrangement of stitches. Sometimes, the wounds appear smoother or more irregular than the original, or have different degrees of secretion and redness.

Figure 1: Pathological wound example, left original, right augmented

A total of 228 synthetic images of physiological and 91 images of pathological wounds were validated as plausible by expert clinicians from Ospedale Policlinico San Martino, and were used to augment the training dataset.

2.3 Model and Metrics

For automatic wound detection and classification, the YOLOv11 model was adopted. The choice was motivated by its ability to achieve high detection accuracy; YOLO models have demonstrated strong generalization performance in medical applications, including dermatological lesion detection. The process involves a localization phase, in which the model identifies the bounding box of the wound, followed by a classification phase, which distinguishes between physiological, pathological, and non-wound regions.

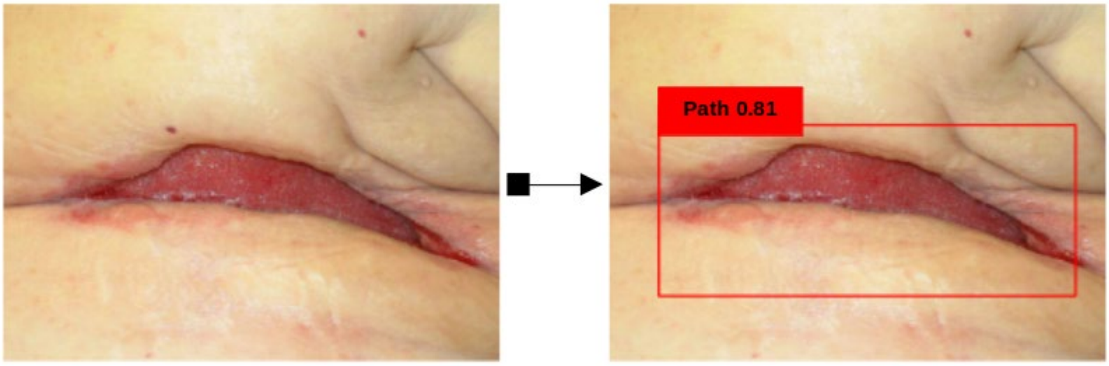

Figure 2: Pathological wound detection with confidence level

The model was trained in two distinct configurations:

- using only the original real images;
- using an extended dataset with synthetic images.

This choice allows to systematically evaluate the contribution of generative AI to improve system performance in terms of accuracy, robustness, and generalization ability.

The following metrics were used to evaluate performance:

- precision: measures the percentage of correct positive predictions out of the total positive predictions;
- recall: measures the percentage of true positive predictions out of the total of positive cases;
- F1-score: it is the harmonic mean between Precision and Recall, useful when the classes are unbalanced.

3. Results

To compare performance between the extended and base models, we adopted the F1-score as the primary evaluation metric. The F1 confidence curve enabled us to assess the balance between precision and sensitivity across varying decision thresholds. The extended model reached a maximum F1-score of 0.90 (Fig.5), outperforming the base model, which peaked at 0.81(Fig.4). The improved behaviour observed for both physiological and pathological classes highlights better generalization and greater sensitivity, particularly in identifying pathological wounds. These findings indicate that the structured inclusion of synthetic images within the data augmentation pipeline can contribute positively and measurably to classification outcomes.

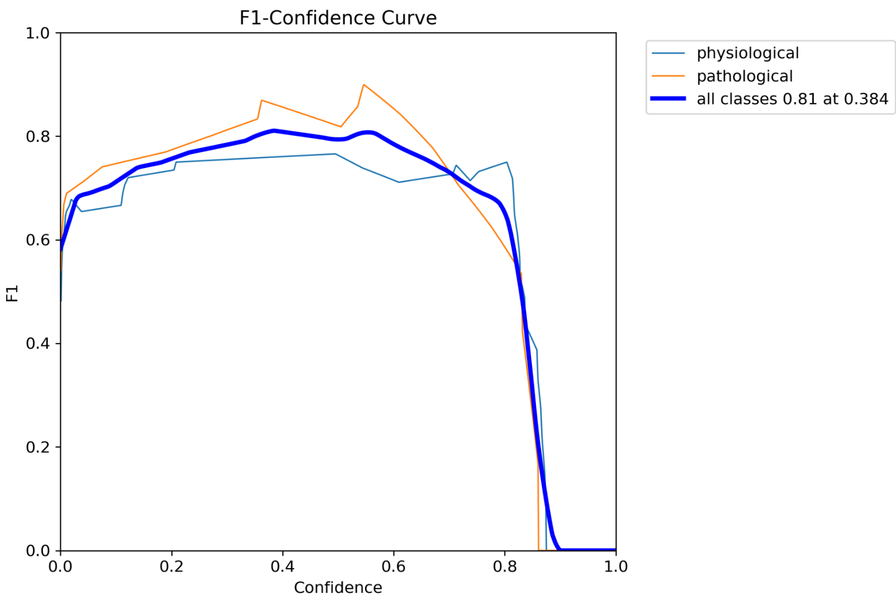

Figure 4: F1-curve for original dataset

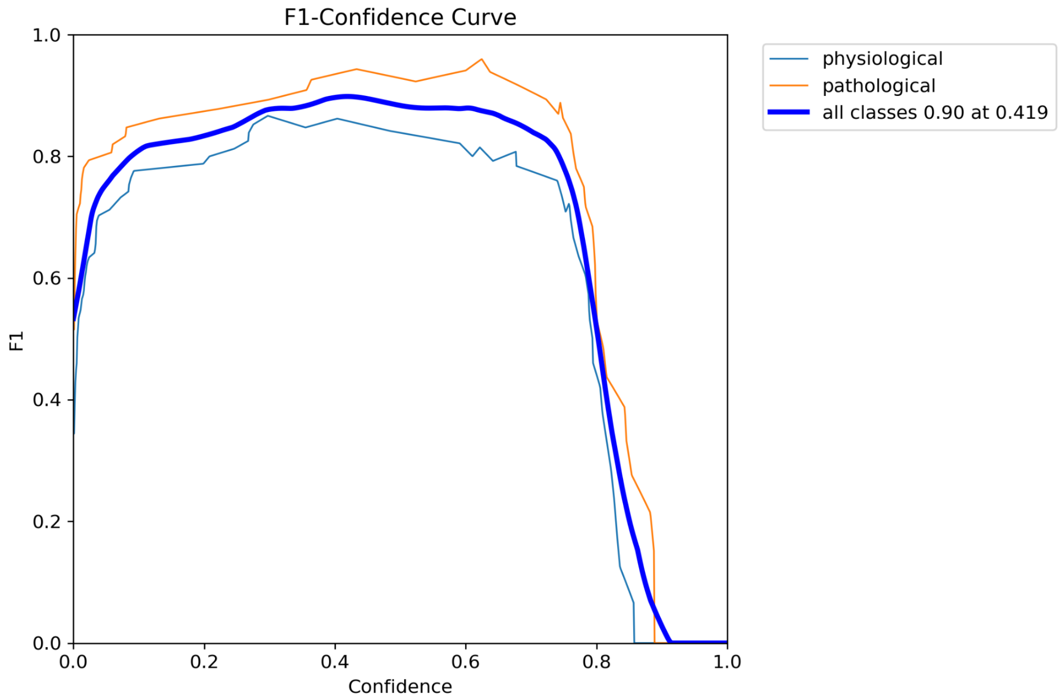

Figure 5: F1-curve for augmented dataset

The highest-performing configuration used the original dataset enriched with synthetic data and geometric augmentation. This combination produced consistent improvements across all metrics compared with the model trained exclusively on real images. We report below the normalized confusion matrix obtained on the test set.

Table 1: Normalized confusion matrix on the test set

Pred/Real	Physiological	Pathological	Background
Physiological	0.93	0.08	0.27
Pathological	0.00	0.92	0.73
Background	0.07	0.00	0.00

The normalized confusion matrix shows good classification capabilities. The physiological class was correctly identified in 93% of cases, while the pathological class was correctly identified in 92% of cases, with a residual overlap between the two. However, a residual overlap remains between classes, particularly regarding pathological areas occasionally misidentified as background. This suggests an underestimation of pathological lesions when they closely resemble surrounding tissue.

4. Insights and Future Work

Although the results are encouraging, several limitations remain. In particular, the current dataset still offers moderate size and variability, which may reduce generalizability to more heterogeneous clinical contexts. To address this, the acquisition of approximately 400 additional post-operative wound images from varied body regions is planned in collaboration with Ospedale Policlinico San Martino (Genoa, Italy). This expansion aims to strengthen dataset representativeness and improve model robustness. Nevertheless, further investigation will be necessary, such as integrating more anatomical sites and including currently under-represented skin phototypes.

Access to images of the same wound over time paired with clinical outcomes, would also be highly valuable. Such temporal information could support the development of predictive models for wound evolution and healing, broadening the clinical relevance of this approach. While still in an active development phase, this work shows promising potential to support clinicians and enhance post-operative patient care.

Author contributions: Conceptualization, Isabel Carozzo, Elisa Bruzzo, Luca Giulio Brayda, Ennio Ottaviani; methodology, Isabel Carozzo; analysis & validation, Isabel Carozzo, Michele Minuto; data curation, Isabel Carozzo, Elisa Bruzzo, Matteo Parodi; writing draft preparation, Isabel Carozzo, Elisa Bruzzo, Matteo Parodi; writing review, Luca Giulio Brayda, Michele Minuto, Ennio Ottaviani.

Funding statement: This work was partially supported by "Programma Operativo Por FSE Regione Liguria 2021-2027" through the project no. G39J23001280009.

Data availability statement: The real-world surgical wound images used in this study are not publicly available due to copyright restrictions (Redscar© database) and privacy and ethical restrictions (images collected by Casa di Cura Villa Montallegro, Genoa, Italy).

Conflicts of interest: The authors declare no conflicts of interest.

Ethics declaration: This study is conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee Genoa (Italy) (Protocol Code 141/2025 & 142/2025 —DB id 14199 & 14369, date 21 May 2025). Informed consent statement: Images provided by Casa di Cura Villa Montallegro (Genoa, Italy) were obtained with the patient's informed consent. The survey conducted among healthcare professionals in Liguria, cited in the Introduction, was anonymous.

AI declaration: ChatGPT (https://chatgpt.com/) was used in the preparation of this manuscript for assistance with English syntax and grammar correction. The content and scientific contributions of the paper were entirely developed by the authors. All outputs generated by the AI were reviewed and edited to ensure accuracy, appropriateness, and alignment with the intended meaning. The authors take full responsibility for the content of the publication.

References

Badia, J.M., Casey, A.L., Petrosillo, N., et al. (2017). Impact of surgical site infection on healthcare costs and patient outcomes: a systematic review in six European countries. Journal of Hospital Infection, 96, 1–15. https://doi.org/10.1016/j.jhin.2017.03.004

Fateen, I. Wound classification [Kaggle dataset]. Retrieved Feb 1st 2025 from https://www.kaggle.com/datasets/ibrahimfateen/wound-classification

González-Hidalgo, M., Munar, M., Bibiloni, P., et al. (2019a). Detection of infected wounds in abdominal surgery images using fuzzy logic and fuzzy sets. 2019 International Conference on Wireless and Mobile Computing, Networking and Communications, Barcelona, Spain, 99–106. https://doi.org/10.1109/WiMOB.2019.8923289

González-Hidalgo, M., Moyà-Alcover, G., Munar, M., et al. (2019b). Detection and automatic deletion of staples in images of wound of abdominal surgery for m-Health applications. In: VipIMAGE 2019. Lecture Notes in Computational Vision and Biomechanics, vol. 34. Springer, Cham. https://doi.org/10.1007/978-3-030-32040-9_23

Haque, M., Sartelli, M., McKimm, J., et al. (2018). Health care-associated infections: an overview. Infection and Drug Resistance, 11, 2321–2333. https://doi.org/10.2147/IDR.S177247

Kazerouni, A., Khodapanah Aghdam, E., Heidari, M., et al. (2023). Diffusion models in medical imaging: a comprehensive survey. Medical Image Analysis, 102846. https://doi.org/10.1016/j.media.2023.102846

Mou, L., Li, Y., Song, J., & Yang, J. (2023). Generating synthetic skin lesion images with diffusion models for data augmentation in dermatology. Medical Image Analysis, 87, 102823. https://arxiv.org/html/2507.23001v1

Oota, S.R., Rowtula, V., & Mohammed, S. (2023 January). WSNet: Towards an effective method for wound image segmentation [GitHub repository]. Retrieved August 8th 2025 from https://github.com/subbareddy248/WSNET