Cosmetic assessment in the UNICANCER HypoG-01 trial: a deep learning approach

Alexandre Cafaro [1], Amandine Ruffier [2], Gabriele Bielinyte [3], Y. Kirova [4]; S. Racadot [5]; M. Benchalal [6]; JB. Clavier [7]; C. Charra-Brunaud [8]; ME. Chand-Fouche [9]; D. Argo-Leignel [10]; K. Peignaux [11]; A. Benyoucef [12]; D. Pasquier [13]; P. Guilbert [14]; J. Blanchecotte [15]; A. Tallet [16]; A. Petit [17]; G. Bernadou [18]; X. Zasadny [19]; C. Lemanski [20]; J. Fourquet [21]; E. Malaurie [22]; H. Kouto [23]; C. Massabeau [24]; A. Henni [25]; P. Regnault [26]; A. Belliere [27]; Y. Belkacemi [28]; M. Le Blanc-Onfroy [29]; J. Geffrelot [30]; JB. Prevost [31]; E. Karamouza [32]; Stefan Michiels [32]; Marie Bergeaud [33], Assia Lamrani-Ghaouti [33], Sami Rhomdani [1], Alexis Bombezin--Domino [1], Nikos Paragios [1], Sofia Rivera [3]

[1] TheraPanacea, France, [2] Clinique Victor Hugo-Centre Jean Bernard, Le Mans, France, Bureau de Biostatistique et d'Epidémiologie, [3] Institut Gustave-Roussy, Département d'oncologie radiothérapie, France, Gustave Roussy, [4] Oncologie-radiothérapie, Hôpital Institut Curie - Université PSL, Paris, France, [5] Radiotherapie, Centre Léon Bérard, Lyon, France, Radiothérapie, [6] Centre Eugène Marquis, Rennes, France, [7] Radiothérapie, ICANS - Institut de cancérologie Strasbourg Europe, Strasbourg, France, [8] Radiothérapie, Institut de Cancérologie de Lorraine – Alexis Vautrin, Vandoeuvre-lès-Nancy, France, [9] Département de radiothérapie, Centre Antoine Lacassagne, Nice, France, [10] Radiothérapie, Hôpital du Scorff, Lorient, France, [11] Radiothérapie, Centre Régional De Lutte Contre Le Cancer Georges-François Leclerc C.G.F.L, Dijon, France, [12] Radiotherapie, Centre Henri Becquerel, Rouen, France, [13] Radiotherapie, Centre Oscar Lambret, Lille, France, [14] Radiothérapie, Institut Godinot, Reims, France , [15] Radiotherapie, Institut de Cancérologie de l'Ouest (ICO) - Site d'Angers, Angers, France , [16] Radiothérapie, Institut Paoli-Calmettes, Marseille, France, [17] Radiothérapie, Institut Bergonié, Bordeaux, France, [18] Radiothérapie, CORT37, Chambray-lès-Tours, France, [19] Radiothérapie, Site Chénieux - Polyclinique de Limoges, Limoges, France, [20] Radiothérapie, ICM - Montpellier, Montpellier, France, [21] Radiothérapie, C.H. de Lens, Lens, France, [22] Radiothérapie, C.H. Intercommunal Créteil, Créteil, France, [23] Radiothérapie, Centre Galilée, Lille, France , [24] Radiothérapie, Institut Universitaire du Cancer Toulouse, Oncopole, Toulouse, France, [25] Radiotherapie, Institut de cancerologie des haut de France, Beuvry, France, [26] Radiothérapie, Clinique Bordeaux Tivoli-Ducos, Bordeaux, France, [27] Radiothérapie, Centre Jean Perrin, Clermont-Ferrand, France, [28] Radiothérapie, C.H. Henri Mondor, Aurillac, France, [29] Département de Radiothérapie, Institut de Cancérologie de l'Ouest, ICO - René Gauducheau, Nantes, France, [30] Radiothérapie, Centre François Baclesse, Caen, France, [31] Radiothérapie, Centre Pierre Curie, Beuvry, France, [32] Bureau de Biostatistique et d'Epidémiologie, Gustave Roussy, Oncostat U1018, Inserm, Paris-Saclay University, Paris; [33] UNICANCER, Unitrad, Paris, France

San Antonio Breast Cancer Symposium
December 5-9, 2023

Abstract

Patients receiving radiotherapy in clinical trials have the cosmetic effects of treatment evaluated, on a 4-classes scale. This evaluation is subject to bias and lacks for an automated, standard and fast tool.

We propose a nearly automatic tool (needs only the position of 2 points in the image), based on Machine Learning, and compare it to BCCT.core an open tool for cosmetic evaluation.

Our results were comparable to BCCT, with better F1-score and accuracy over the 4 classes of evaluation.
This paves the way for standard, automatic cosmetic breast evaluation through AI.

Introduction

Cosmetic evaluation after breast cancer treatment is a clinical indicator of toxicity. User bias and inter-subject variability hamper this objective. To address this limitation, a deep learning approach was developed on the basis of the HYPOG-01 trial (NCT03127995), a phase III trial comparing hypo vs normo-fractionated radiotherapy (RT) in breast cancer patients requiring nodal irradiation (See PO5-19-10).

Methods

- Data from 581 female patients included in the intention-to-treat population of the HYPOG-01 study analysis (exclusion of mastectomy/pamectomy)
- Images: 2,346 front images with arms along the body at baseline before RT, 3-weeks after RT start, end of RT, 6 months and every year after randomization up to 5 years.
  These were divided into training (1,661), validation (308), and testing (377) datasets.
- On each image, cosmetic outcome was evaluated by an independent radiation oncologist using Harris score (Excellent / Good / Fair / Poor). This evaluation was used as reference. The distribution of Harris scores in the dataset was highly imbalanced: 7% excellent, 33% good, 45% fair, and 15% poor.
- Additionally, all images were assessed using the BCCT.core© software.

Nipples were used as landmarks to address picture acquisition variations by cropping and resizing images to 224x224 resolution. Feature extraction was performed using a pre-trained Swin-TransformerV2 model.
The model was finetuned for 300 epochs, and the highest F1-score model was selected.

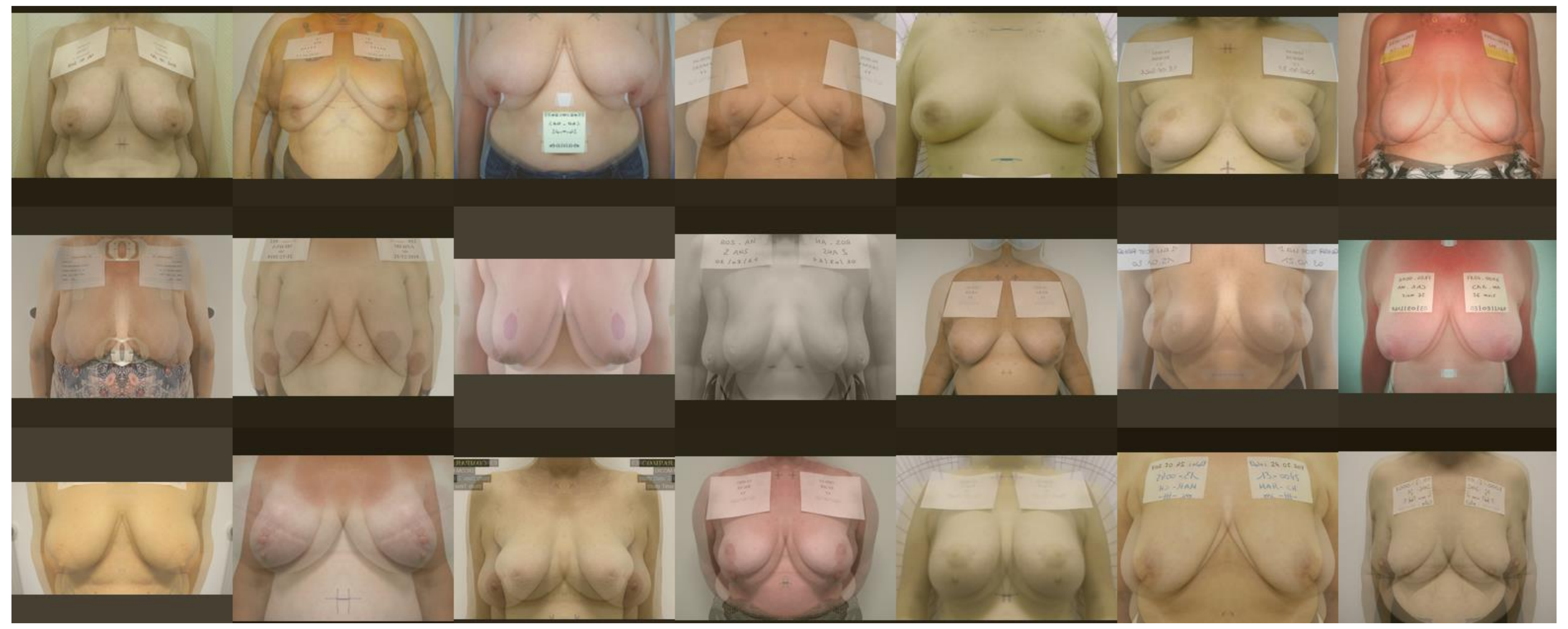

Figure 1. Examples of augmented images (contrast modulation, lighting adjustments, and geometric transformations), superimposed with their symmetric, used to improve model's generalization and accuracy.

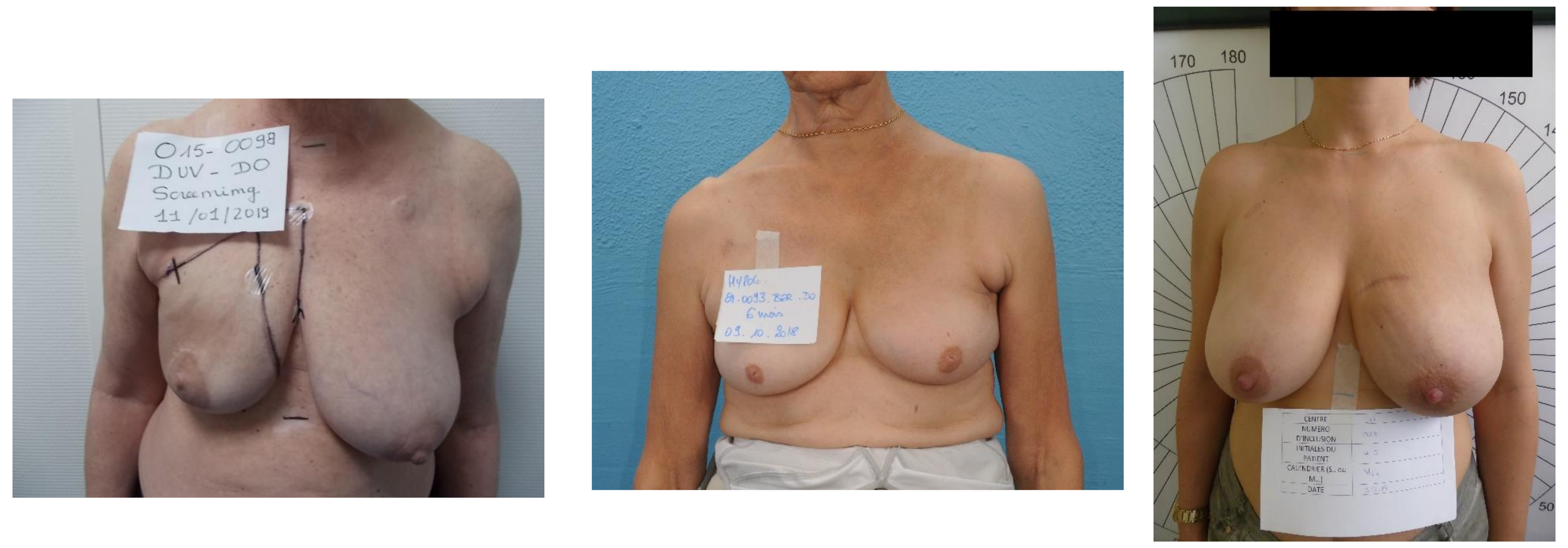

Figure 2. Timestamps from image captures were integrated as an additional influencing factor. This example shows the effect of these timestamps (FAIR baseline vs FAIR M24), but also variability between centers.

Results

The performance of our model was evaluated using balanced binary classification, multi-class accuracy, and F1-score. Comparatively, our model performed similarly to BCCT in terms of overall accuracy but demonstrated better performance in separating multiple classes, as indicated in Table 1. In Table 2, we present a confusion matrix that provides insights into the model's performance, showing effective discrimination of poor results and some ambiguity between good and fair classes, as seen in BCCT.

Figure 3. Representative examples of Correct and Incorrect predictions.
Our model's precision seems helped by better grasping of more subtle details compared to BCCT

	F1** score (4 classes)	Balanced Multi-class Accuracy (4 classes)	Balanced Binary Accuracy*
BCCT	0.41	0.49	0.69
Tested model	**0.42**	**0.54**	0.68

* (Excellent & Good vs Fair & Poor)
**F1 score balances precision (true positives out of all positive predictions) and recall (true positives out of actual positives)

Table 1. Evaluation of performance on the test set (only on cases with evaluation by BCCT (327 images))

Prediction \ True**	Excellent (12)	Good (109)	Fair (167)	Poor (39)
Excellent	**50%**	24%	12%	0%
Good	50%	**49%**	32%	5%
Fair	0%	26%	**43%**	18%
Poor	0%	1%	13%	**77%**

* The independent radiation oncologist evaluation was considered as "true result"

Table 2. Confusion matrix between our predictions and the labels on the test set (only on cases with evaluation by BCCT (327 images))

Conclusions

Performant AI based cosmetic evaluation is feasible. The proposed solution could simplify and accelerate the evaluation process by utilizing only two nipple landmarks, surpassing manual and semi-automated tools. This advancement opens doors for automated, large-scale cosmetic toxicity evaluation. Continuous improvement and validation contribute to its robustness and reinforce its significant impact in assessing cosmetic outcomes after breast cancer treatment.

Contact Alexandre Cafaro, TheraPanacea, a.cafaro@therapanacea.eu