CTU Journal of Innovation and Sustainable Development
ISSN 2588-1418 | e-ISSN 2815-6412

DOI:10.22144/ctujoisd.2025.051

A clinically-oriented 2D image-analysis system for appearance-based aesthetic outcome evaluation of breast reconstruction surgery

Takumi Sonoi[1], Nam Phong Duong[1], Yoshihiro Sowa[2], and Masayuki Fukuzawa[1*]

[1]*Graduate School of Science of Technology, Kyoto Institute of Technology, Japan*

[2]*Department of Plastic Surgery, Jichi Medical University, Japan*

**Corresponding author (fukuzawa@kit.ac.jp)*

Article info.

Received 10 Jul 2025
Revised 7 Aug 2025
Accepted 8 Oct 2025

Keywords

Aesthetic outcome, breast reconstruction surgery, image-based grading, image processing, machine learning

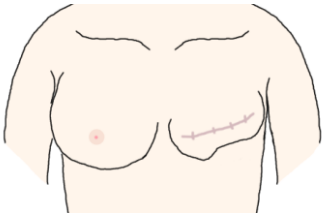
ABSTRACT

Aesthetic outcome of reconstructed breasts is currently rated subjectively by plastic surgeons, which introduces inter-rater bias and variability; thus, an image-based objective technique is desired. Developing such a technique, however, has been challenging due to the limited availability of reconstructed breast images under standardized conditions, and the complexity of assessing multiple aesthetic viewpoints. In this study, we propose a clinically-oriented two-dimensional (2D) image-analysis system where small fingerprint pairs representing the left and right breasts are extracted from conventional 2D chest images obtained in clinical settings and used as training data for a simple convolutional neural network (CNN), aiming for high effectiveness even with a limited number of cases. We extracted 16 type variations of fingerprints from 170 cases, and evaluated their influence on CNN performance. The optimal fingerprint types varied depending on the aesthetic viewpoint. The overall aesthetic score, calculated by aggregating the best-performing model scores across all viewpoints, showed a strong correlation ($r > 0.9$) with the average rater scores. Although 2D images capture only partial breast appearances and may not fully represent intrinsic three-dimensional (3D) features, the experimental results strongly support the potential of the proposed system for developing appearance-based models for aesthetic evaluation in clinical settings.

1. INTRODUCTION

Plastic surgery aims to restore body shape and function, or to enhance appearance. One of the common surgeries is breast reconstruction surgery, performed to restore the natural breast shape, which has been damaged after mastectomy. An important measure of reconstruction quality is the degree to which breast shape and appearance are restored, referred to as the "aesthetic outcome", which has a significant impact on the patient's quality of life (QOL). Typically, this aesthetic outcome is subjectively rated by experienced plastic surgeons by evaluating several aesthetic viewpoints including shape, appearance, and softness. Figure 1 shows (a) an example of a reconstructed breast and (b) its aesthetic viewpoints (shape and appearance) recognized among Japanese plastic surgeons (Yano, 2007). The viewpoints include the breast volume (BV), shape (BS), location (BL), conspicuity of scars and color (SC), inframammary fold symmetry (IMF), nipple location (NL), nipple-areola complex (NAC) size and shape (NS), and NAC color (NC). These viewpoints are evaluated based on the symmetry and equivalence to the contralateral

healthy breast rather than the pre-reconstruction condition, except for scars. However, subjective rating using a scoring system is prone to bias and variation between raters. Therefore, an objective method is necessary for aesthetic grading of reconstructed breasts.



(a)

Abbreviation	Evaluation viewpoint	Evaluation score (Scoring criteria)
BV	Breast volume	2 (equivalent), 1 (slightly different), 0 (considerably different)
BS	Breast shape	2 (equivalent), 1 (slightly different), 0 (considerably different)
BL	Breast location	2 (equivalent), 1 (slightly different), 0 (considerably different)
SC	Scars and color (Non-symmetric)	2 (inconspicuous), 1 (slightly conspicuous), 0 (conspicuous)
IMF	Inframammary fold symmetry (height)	1 (less than 2cm), 0 (2cm and over)
NL	Nipple location (nipple-to-sternal notch distance)	1 (less than 2cm), 0 (2cm and over)
NS	Nipple-Areola Complex (NAC) size and shape	1 (equivalent), 0 (different)
NC	NAC color	1 (equivalent), 0 (different)

(b)

Figure 1. (a) Example of reconstructed breast and (b) evaluation viewpoints of aesthetic outcome recognized among Japanese plastic surgeons (Yano, 2007)

There have been several attempts made to improve the aesthetic grading of reconstructed breasts using two-dimensional (2D) breast images obtained by conventional digital cameras due to their high clinical availability, despite the fact that they only capture a partial appearance of the breast. The software BCCT.core was developed to assist in aesthetic grading by interactively analyzing these 2D images and has demonstrated a significant correlation with subjective measure based on the Harris scale (Cardoso et al., 2007; Heil et al., 2012; Preuss et al., 2012). However, most of its procedures require manual operations, such as outlining the breast boundary and locating the NAC and scars, which limits efficiency for large-scale data collection and analysis. In addition, image-based aesthetic grading faces two major challenges: (1) acquiring reconstructed breast images under standardized clinical conditions, and (2) addressing multiple aesthetic viewpoints, which is an essential part of the evaluation process. These challenges have created significant obstacles in developing machine learning models for aesthetic grading, as seen in a recent study (Guo et al., 2022), because a limited amount of quality training data must support the development of multiple models to address all aesthetic viewpoints.

Recently, Harada et al. (2023) have developed an image processing system applicable to 2D and three-dimensional (3D) breast images to reduce the effort required for case collection. This system facilitated anonymization, annotation support, and feature extraction of 3D breast images, enabling effective collection of reconstructed breast cases in clinical settings.

In this study, we propose a clinically-oriented 2D image-analysis system for appearance-based aesthetic grading of reconstructed breasts, extending the Harada's system (2023) with several new functions. Effective image analysis can be achieved even with a limited number of cases by extracting pairs of small fingerprints representing the left and right breasts from 2D chest images, and using them to develop machine learning models to predict aesthetic score. To address the assessment of different viewpoints, multiple types of fingerprints were extracted based on different breast division criteria, with each type tailored to and expected to yield varying performance across each viewpoint-specific model. To assess the potential of these fingerprint types and their influence on model performance, we employed a simple convolutional neural network (CNN). Model-predicted aesthetic scores then underwent correlation analysis with actual scores to assess the overall potential of the proposed system.

The remainder of this paper is as follows. Section 2 describes the design and implementation of our proposed system, including the details of the fingerprints and the CNN models. Section 3 presents the experimental results of fingerprint extraction using several breast division criteria, as well as the performance of fingerprint-based CNN. Finally, Section 4 summarizes the key findings and concludes this paper.

2. MATERIALS AND METHOD

2.1. Expected 2D chest images

To extract and examine fingerprints across many cases of breast reconstruction, 2D chest images must be acquired under standardized conditions in clinical settings. We adopted the image-acquisition conditions proposed by Harada et al. (2023), which used a conventional camera.

As a pretreatment, plastic surgeons drew two types of markers on the patient's body surface before image acquisition. One type is a pair of cross marks at the center of the clavicle and 25 cm below it along the center line of the body. The other type is a pair of outlines as closed curves representing the region boundaries of the left and right breasts.

A 2D chest image is acquired from a camera position and orientation to include the upper body but to exclude the face of the patient. Although the proposed acquisition conditions included six different camera positions and orientations, this study only assumes a central camera position and 2D chest images in the horizontal frontal orientation.

2.2. Fingerprint extraction

A fingerprint is defined as a small $M \times N$ pixels image extracted from a one-sided breast image, cropped from the left or right breast region in a 2D chest image. Each pixel of the fingerprint represents a small region of the one-sided breast image, divided radially and angularly. In this study, fingerprints of size 16 × 16 pixels were used.

Figure 2 presents the functional block diagram of the fingerprint extraction from the 2D chest images. The pipeline combines both manual operations and automated processes applied to the dataset and consists of two main stages: preprocessing and region-representative determination.

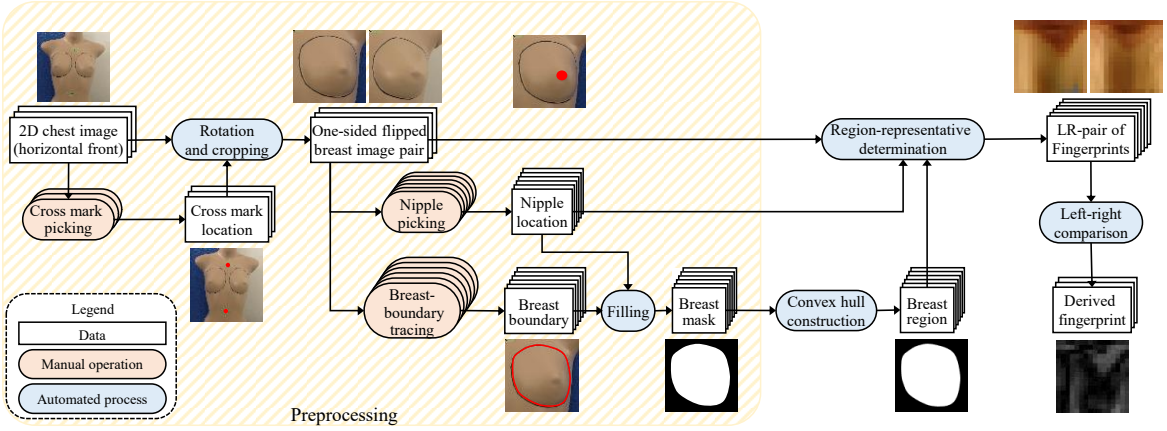


Figure 2. Functional block diagram of the fingerprint extraction from 2D chest images

2.2.1. Preprocessing

The preprocessing consists of manual operations and their complementary automated processes. The manual operations include cross mark and nipple picking, and breast-boundary tracing assisted by ImageJ, an in-house Python script, and GIMP (Rasband, W. S., n.d.; The GIMP Development Team, n.d.). The exception is for breasts without nipple reconstruction, in which the same nipple height as in the healthy breast was used. Furthermore, a plausible nipple location was subjectively identified if the healthy nipple height was outside the reconstructed breast, or if neither breast had nipple reconstruction. The breast boundary was extracted by hand tracing the pre-drawn breast outlines on the body surface. If a part of the pre-drawn outline, such as the inframammary fold, was hidden by the sagging breast, the lower edge of the sagging breast was traced instead.

The automated processes include cropping and filling steps. The cropping step was performed using an in-house Python script. A pair of one-sided breast images were cropped from the 2D chest image so that the origin was at the upper cross mark and the Y-axis passed through the two cross marks by rotating the 2D chest image if necessary. One of the cropped images corresponding to the right breast was flipped left to right in order to handle all the one-sided breast images in a common manner during the subsequent processes. The filling step was performed using an in-house GIMP plug-in. A breast mask was generated as a binary image from

each one-sided breast image by filling the inside of the extracted breast boundary with white and the outside with black, starting from the nipple location.

2.2.2. *Region-representative determination and left-right comparison*

The rest of fingerprint extraction after preprocessing mainly includes two steps of region-representative determination and left-right comparison. The step of region-representative determination begins with obtaining a breast region by calculating the convex hull of the boundary from a breast mask of one-sided breast. Then, the breast region was divided radially and angularly centered at the nipple location. There are some division criteria, which are detailed in Section 2.2.3. A fingerprint was extracted by determining a region-representative of every divided region in the one-sided breast image and arranging all the representatives as pixels in a Cartesian coordinate system of radius and angle. This extraction was performed on both left and right one-sided breast images to obtain a left and right (LR) pair of fingerprints.

A derived fingerprint was also extracted as an alternative category by comparing the left and right fingerprints. We adopted ΔV as a derived representative defined as the absolute difference of the V component after transforming each LR-pair of fingerprints into the HSV color space.

2.2.3. *Criteria of breast division*

Figure 3 illustrates the pseudo-fan and subring regions divided from a breast region and their corresponding pixels in the fingerprint. The breast region was first divided into M regions of pseudo-fans centered at the nipple location. The i^{th} pseudo-fan region corresponds to the i^{th} column of the fingerprint. The division radii were arranged counterclockwise so that the initial radius points vertically upward from the nipple. There are two pseudo-fan division criteria: common fan size ($\mathrm{C_{fs}}$) or common central angle ($\mathrm{C_a}$).

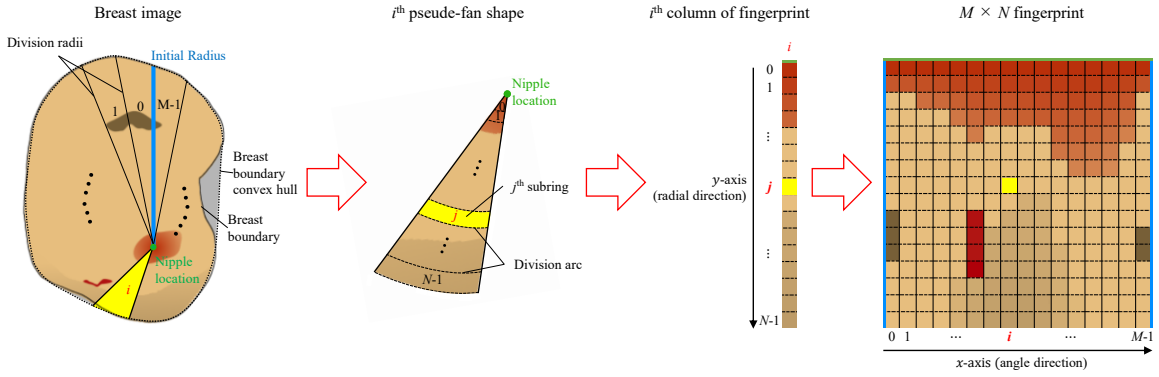


Figure 3. Pseudo-fan and subring regions divided from the breast region and corresponding pixels

A pseudo-fan region was further divided into N subregions of subring surrounded by two common division radii and two different division arcs. The j^{th} subring region in ascending order outward from the nipple corresponds to the j^{th} row of the fingerprint. There are also two types of subring division criteria: common subring size ($\mathrm{C_{ss}}$) or common subring interval based on a certain reference radius. Three reference radii were considered: the radius pair of each pseudo-fan ($\mathrm{R_f}$), the maximum radius in a single breast ($\mathrm{M_{sb}}$), or the maximum radius among the breast pair ($\mathrm{M_{bp}}$).

Fingerprints were extracted using four promising combinations of division criteria from a total of eight, including two pseudo-fan divisions and four subring divisions. These four promising combinations and their corresponding expected fingerprints are demonstrated in Figure 4.

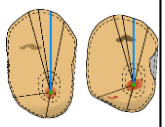
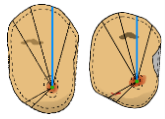
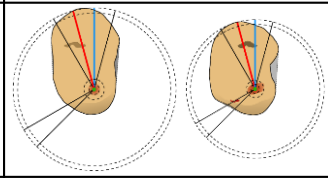
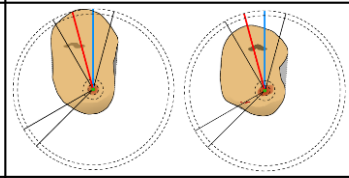
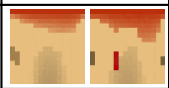
Division criteria of pseudo-fan	Common fan size	Common central angle		
Division criteria of subring	Common subring size	Common subring interval		
		Radius pair of each pseudo-fan	Maximum radius in each single breast	Maximum radius among the breast pair
Symbol	$C_{fs}C_{ss}$	C_aR_f	C_aM_{sb}	C_aM_{bp}
Division sketch				
Typical fingerprint				

Figure 4. Four promising combinations of division criteria and expected fingerprints

- Common fan size and common subring size ($C_{fs}C_{ss}$)
  This combination aimed to reflect the effect of nipple and scars based on their size rather than their location, by iso-area division of the breast region. The central angle of pseudo-fan varies depending on its radius. The subring thickness increases near the nipple and decreases toward the breast boundary.
- Common central angle and common subring interval based on the radius pair of each pseudo-fan (C_aR_f)
  It aimed to enlarge pseudo-fans with short radii by adjusting all the division radii of pseudo-fans uniformly. The relative location of the divided region is kept, but its occupied area varies significantly. Especially, the pseudo-fans in the lower breast are often enlarged radially, which results in thick NAC-contributed pixels in the fingerprint.
- Common central angle and common subring interval based on the maximum radius in each single breast (C_aM_{sb})
  It aimed to enhance the LR difference in shape rather than the inter-case difference by normalizing the distance of each divided region from the nipple by the maximum radius of each one-sided breast. Some divided regions, mainly in the lower breast, may be located outside the breast region. Their corresponding region-representatives become zero and are revealed as black regions in the fingerprint. Since the normalization was performed for each one-sided breast, at least one column consists of valid region-representatives in both fingerprints.
- Common central angle and common subring interval based on the maximum radius among the breast pair (C_aM_{bp})
  It aimed to enhance the LR difference in shape and size rather than the inter-case difference by normalizing the divided-region distance by the maximum radius of both breasts. In a same manner as the C_aM_{sb}, black regions will be found in the fingerprint where the divided regions are located outside the breast region. There may be only one column containing valid region-representatives across both fingerprints because the normalization was performed for both breasts.

2.2.4. Region-representatives

Two different region-representatives, which are denoted as μRGB and RGB@V_{min}, were proposed. The definitions and calculation procedures for each subring region are as follows.

- μRGB
  The average color in a subring region. The average value of R, G, and B components are separately calculated over all the pixels in the subring region, and concatenated to form an RGB value.
- RGB@V_{min}
  A pixel value of the darkest pixel in the subring region. All pixels in the subring region are converted to HSV color space, and the pixel with the minimum V component is selected to obtain its RGB value as representative.

Based on the detailed design of fingerprint extraction described above, there are 16 type

variations of fingerprints reflecting four promising combinations of division criteria, two region-representatives, and two categories.

2.3. Fingerprint-based simple CNN architecture for aesthetic score prediction

Figure 5 illustrates two types of simple CNN models that were designed to predict the aesthetic score for each aesthetic viewpoint from a certain category (LR-pair or derived) of fingerprint. The upper is one-input model for ΔV and the lower is two-input model for LR-pair of fingerprints. Both models contain four convolution layers with 3 × 3 kernels with channel-independent normalization and active function such as ReLU. The number of filters in each layer is 8, 16, 32, and 32, respectively. Two 2 × 2 max-pooling layers are applied after the first and second layers for downsampling. Extracted features are flattened into a vector via fully connected layer(s). The one-input model consists of only a final fully connected layer with a 50% dropout, while the two-input model incorporates two LR-dependent fully connected layers and a concatenation step. Finally, a regression layer outputs the aesthetic score as a single scalar value.

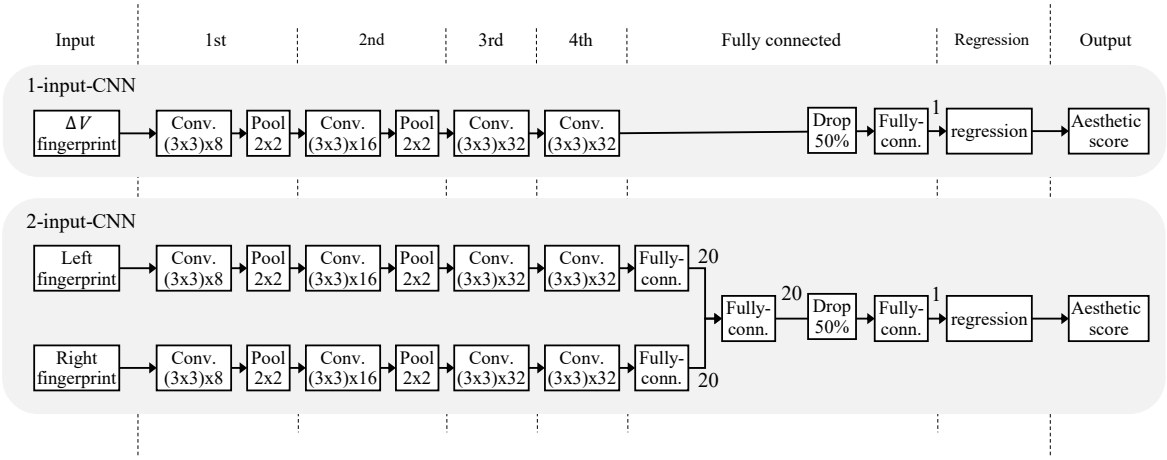


Figure 5. Architecture of two fingerprint-based CNN models for aesthetic score prediction

The proposed CNN models were implemented in MATLAB (The MathWorks, Inc., n.d.). Table 1 shows the hyperparameters of the simple CNN models.

Table 1. Simple CNN model hyperparameters

Hyperparameter	Value
Solver for training neural network	Adam
Decay rate of gradient moving average	0.9
Decay rate of squared gradient moving average	0.999
Denominator offset	10^{-8}
Learning rate	0.001
Maximum number of epochs	60
Size of mini-batch	20
Data shuffle	None
Factor for L2 regularization	0.0001
Neural network to return when training completes	Last-iteration

For each of the one-input and two-input model design, we trained eight CNN models to predict the aesthetic scores of eight different viewpoints (viewpoint-specific) shown in Figure 1 (b), and one CNN model to predict the total value of them (Total). The summary of eight prediction scores of the best viewpoint-specific models (8-models summary) was also calculated.

Two types of case groups were used for training: a full case group (Full) and a selected case group with reconstructed nipple (NAC) to examine the effect on the viewpoints with strong dependence on nipple reconstruction. In order to train the proposed CNN models, we prepared 32 fingerprint sets to cover 16 type variations of fingerprints and two case groups.

Model performance was evaluated by 5-fold cross-validation using root mean square error (RMSE), correlation coefficient (r), and their ratio r/RMSE. To ensure common test data across all models, cases were pre-selected from the NAC group for testing, while the remaining cases were divided into five equal parts, with one for validation and four for training.

2.4. Fingerprint augmentation

Rather than conventional augmentation to generate a modified copy of the original image, fingerprints were augmented by extracting multiple variants from a one-sided breast image through modulation of the division criteria, such as initial radius or nipple location. Figure 6 shows two augmentation techniques: (a) rotation of initial radius and (b) translation of nipple location.

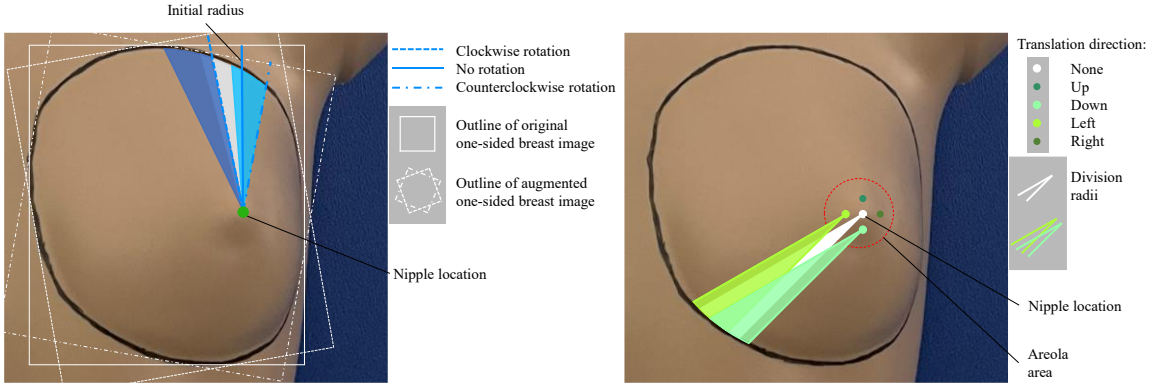


Figure 6. Fingerprint augmentation: (a) rotation of initial radius and (b) translation of nipple location

The initial radius, originally oriented vertically upward from the nipple, was rotated to obtain augmented fingerprints. Rotation beyond the typical center angle would merely shift the pseudo-fan order and shuffle the fingerprint columns, without increasing the variety of the fingerprint. In this study, the initial radius was rotated by ±7.5°, equivalent to one-third of a typical center angle of a pseudo-fan under C_a criteria in a 16 × 16 fingerprint. Since each i^{th} pseudo-fan in an augmented fingerprint partially overlaps with the original, simple column shuffling is effectively avoided.

The nipple location, acting as the center of the pseudo-fan, was translated up, down, left, and right within the areola region to obtain four augmented fingerprints. Ideally, the nipple should remain within the areola, but its size varies by case. Thus, a translation threshold was dynamically set based on breast area, with 5% of the positive square root suitable in most cases. If the nipple was outside the breast region under this displacement, the fingerprint augmentation was excluded.

By using the above two techniques, a maximum of six augmented fingerprints were extracted from a one-sided breast image, augmenting the number of fingerprints by approximately seven times.

3. RESULTS AND DISCUSSION

3.1. 2D image acquisition and aesthetic scoring

The 2D chest images used in this study were acquired between May 2019 and December 2022 at the former affiliated medical institutions (Kyoto Prefectural University of Medicine and Kyoto University) of co-author Sowa, Y.

Aesthetic scores were obtained from four raters of plastic surgeons through cross-annotation using the medical-image processing system of Harada et al. (2023). Annotation was performed for each viewpoint listed in Figure 1 (b) and a rater average score was calculated per case.

A total of 175 image–annotation pairs were collected, with successful fingerprint extraction in 170 cases.

3.2. Fingerprint extraction with several breast division criteria

Figure 7 presents 16 type variations of fingerprints for four representative cases with corresponding breast images, including cases with particularly poor BS (pBS), SC (pSC), and NC (pNC), and a case with excellent scores across all aesthetic viewpoints (exALL). Each fingerprint type is defined by a combination of division criteria, region-representative, and category.

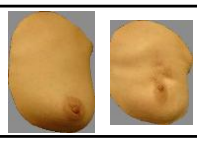
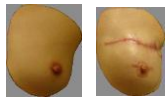
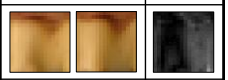
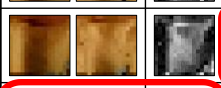
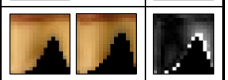
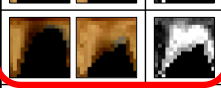
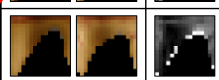
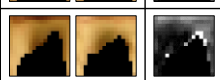
Typical case		pBS		pSC		pNC		exALL	
Aesthetic score	BS	0.00		1.00		1.75		2.00	
	SC	1.25		0.00		1.75		2.00	
	NC	0.75		0.75		0.00		1.00	
Breast images									
Division criteria	Region-representative	Category		Category		Category		Category	
		LR-pair	ΔV	LR-pair	ΔV	LR-pair	ΔV	LR-pair	ΔV
$C_{fs}C_{ss}$	μRGB								
	RGB@V_{min}								
C_aR_f	μRGB								
	RGB@V_{min}								
C_aM_{sb}	μRGB								
	RGB@V_{min}								
C_aM_{bp}	μRGB								
	RGB@V_{min}								

Figure 7. Examples of 16 type variations of fingerprints (combination of division criteria, region-representative, and category) for four typical cases

The$C_{fs}C_{ss}$ type fingerprints did not enhance the NAC at upper rows of the fingerprint compared the other division criteria, reflecting its iso-area division criteria as described in Section 2.2.3. On the other hand, the C_aR_f type fingerprints included more pixels reflecting the NAC region in the upper rows. They were enhanced especially in the right columns corresponding to the lower breast, reflecting its radial enlargement of the radially-short pseudo-fans. The C_aM_{sb} and C_aM_{bp} types of fingerprints displayed black pixels, indicating some divided regions located outside the breast due to the normalization by the maximum radius.

For exALL case, all LR-pair fingerprints revealed strong similarities regardless of fingerprint type. While for particularly poor viewpoints, including pBS, pSC, and pNC, their fingerprints displayed noticeable differences.

The pBS case revealed a clear difference in the overall fingerprint patterns between the left and right breasts. The difference was noticeable in the right columns corresponding to the lower breast, as well as in the shape of black pixels reflecting the breast boundary. The greatest difference was found in the C_aM_{sb}-ΔV type fingerprints mainly around the black pixels, but the overall difference was also exhibited well in the $C_{fs}C_{ss}$ fingerprint types. These fingerprint types are expected to be a potential in evaluating the BS viewpoint.

The pSC case exhibited local abnormalities in the left fingerprint reflecting the scars. Many abnormal pixels were found in the fingerprint types of $C_{fs}C_{ss}$ and C_aR_f with RGB@V_{min} representative. It implies the advantage of the division criteria to cover only the inside of the breast, and the region-representative sensitive to local color deviation in each divided region.

In the pNC case, focus is placed on the upper rows of fingerprint corresponding to the NAC region, as the evaluation of NC viewpoint relies on a difference of the NAC color. The C_aR_f type fingerprints revealed color difference over a relatively wide area in the upper rows, suggesting the advantage of C_aR_f criteria in enhancing the NAC region, especially in the lower breast.

3.3. Fingerprint-based CNN performance

The performance of eight viewpoint-specific CNN models and Total CNN models was examined. Each CNN model was trained by 32 different fingerprint sets to cover 16 type variations of fingerprints and two case groups. The best and worst models were identified based on the value of r/RMSE. The number of cases was 170 for the Full group, and 119 for the NAC group. The total number of successfully extracted fingerprints was 1185 in the Full group and 832 in the NAC group.

Figure 8 presents the best and worst models among eight viewpoint-specific and Total models with the normalized score of '8-models summary'. It includes performance metrics, fingerprint type variation identifiers, and scatter plots. The performance metrics, RMSE, r, and r/RMSE, were the average over 5-fold cross-validation.

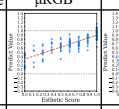
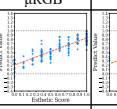
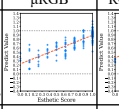
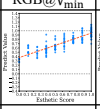
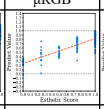
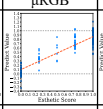
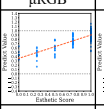
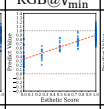
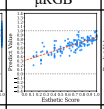
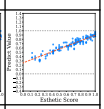
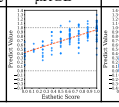
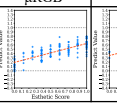
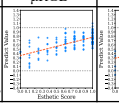
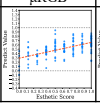
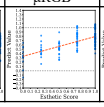
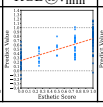
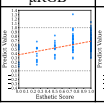
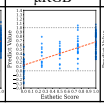
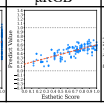
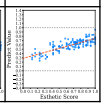
Viewpoint			BV	BS	BL	SC	IMF	NL	NS	NC	Total	8-models summary
Best model	RMSE		0.186	0.188	0.173	0.182	0.195	0.186	0.181	0.236	0.146	0.102
	r		0.822	0.847	0.831	0.842	0.832	0.888	0.805	0.851	0.806	0.943
	r/ RMSE		**4.432**	**4.597**	**4.934**	**4.778**	**4.264**	**4.805**	**4.477**	**3.657**	**5.565**	**9.216**
	Case group		NAC	Full	Full	NAC	NAC	NAC	NAC	NAC	NAC	—
	Fingerprint type	Category	ΔV	ΔV	ΔV	LR-pair	LR-pair	ΔV	ΔV	LR-pair	ΔV	—
		Division criteria	C_aR_f	$C_{fs}C_{ss}$	$C_{fs}C_{ss}$	C_aR_f	C_aR_f	C_aM_{bp}	$C_{fs}C_{ss}$	$C_{fs}C_{ss}$	$C_{fs}C_{ss}$	—
		representative	µRGB	µRGB	µRGB	RGB@V_{min}	µRGB	µRGB	µRGB	RGB@V_{min}	µRGB	—
	Scatter plot											
Worst model	RMSE		0.242	0.273	0.226	0.254	0.282	0.312	0.328	0.329	0.204	0.165
	r		0.694	0.741	0.658	0.594	0.678	0.695	0.415	0.647	0.635	0.809
	r/ RMSE		**2.935**	**2.770**	**2.921**	**2.351**	**2.475**	**2.243**	**1.291**	**2.000**	**3.134**	**4.902**
	Case group		Full	NAC	Full	Full	Full	Full	Full	Full	Full	—
	Fingerprint type	Category	LR-pair	LR-pair	LR-pair	ΔV	ΔV	ΔV	ΔV	ΔV	LR-pair	—
		Division criteria	C_aM_{sb}	C_aM_{sb}	C_aM_{sb}	C_aM_{sb}	C_aM_{sb}	$C_{fs}C_{ss}$	C_aM_{sb}	C_aM_{bp}	C_aM_{sb}	—
		representative	µRGB	µRGB	µRGB	µRGB	µRGB	RGB@V_{min}	µRGB	µRGB	µRGB	—
	Scatter plot											

Figure 8. RMSE and r of the best and worst CNN models

The fingerprint types that yield the best and worst models were different, and the optimal fingerprint types varied depending on the viewpoint. In BS, the best model was the $C_{fs}C_{ss}$-ΔV type instead of the C_aM_{sb} types discussed previously in Section 3.2. This suggests that the overall inner breast textures contribute more significantly than the features of breast boundary. The best fingerprint category was ΔV, indicating that the pre-calculated differences between fingerprint pairs is more suitable. In SC, the best model was the C_aR_f-RGB@V_{min} type, which aligns with the advantageous types discussed in Section 3.2. The best fingerprint category was LR-pair, highlighting the suitability of analyzing both breast side fingerprints simultaneously. In contrast, for the NC viewpoint, the best model was based on the $C_{fs}C_{ss}$ type, which contradicts the assumed advantageous types in Section 3.2. This result suggests that extensive representation of NAC shapes and sizes is not always advantageous. Instead, the iso-area division used in the $C_{fs}C_{ss}$ type may offer relative advantages.

Regarding different case groups, the NAC case group yielded better performance in NAC-related viewpoints such as NL, NS, and NC compared to the Full case group, which suggests that the NAC case group is a suitable subset of data for predicting NAC-related aesthetic viewpoints.

The r of eight viewpoint-specific models ranged from 0.805 to 0.888, indicating insufficient performance for clinical application. That of the 'Total' model was 0.806, revealing a limitation in predicting the total score with a single CNN model. Whereas the '8-models summary' score showed better prediction performances with r of 0.943 and a corresponding RMSE of 0.102. The scatter plot of the '8-models summary' demonstrated a strong

correlation between the model-predicted score and the rater average one.

Although these results are not yet sufficient for immediate clinical application, they strongly demonstrate the potential of the proposed system, particularly in its capacity to train multiple CNN models with a limited number of cases.

4. CONCLUSION

In this study, we proposed an image analysis system that is effective even with a limited number of cases by extracting pairs of small fingerprints representing the left and right breasts from conventional 2D chest images obtained in clinical settings, and using them to train a simple CNN model.

Fingerprints were extracted from clinical 170 cases with 16 type variations reflecting four promising combinations of division criteria ($C_{fs}C_{ss}$, C_aR_f, C_aM_{sb}, and C_aM_{bp}), two region-representatives (μRGB and RGB@V_{min}), and two categories (LR-pair and ΔV).

The LR-pair of fingerprints of the cases with good aesthetic scores showed strong similarity across all viewpoints regardless of fingerprint type, while those with poor scores in specific viewpoints exhibited clear differences. The fingerprint-based CNN model's prediction performance varied by fingerprint type, with the optimal type differing for each viewpoint. The overall aesthetic score, obtained by aggregating the best-model scores across all viewpoints, correlated strongly with the average rater score ($r > 0.9$).

Although 2D breast images capture only partial appearance and do not fully represent intrinsic 3D breast features, these results suggest the proposed system's potential for developing appearance-based aesthetic evaluation models from conventional 2D images suitable in clinical settings.

CONFLICT OF INTEREST

The authors have no competing interests to declare that are relevant to the content of this article.

ACKNOWLEDGMENT

The authors extend their gratitude to Mr. Kazuya Koyanagi, Ms. Kotori Harada, and Mr. Takahiro Yoshimoto for their invaluable contribution during the early development stage of this study. This work was supported by JST SPRING, Grant Number JPMJSP2107 and by JSPS Core-to-Core Program, Grant Number JPJSCCB20230005.

REFERENCES

Cardoso, M. J., Cardoso, J., Amaral, N., Azevedo, I., Barreau, L., Bernardo, M., Christie, D., Costa, S., Fitzal, F., Fougo, J. L., Johansen, J., Macmillan, D., Mano, M. P., Regolo, L., Rosa, J., Teixeira, L., & Vrieling, C. (2007). Turning subjective into objective: The BCCT.core software for evaluation of cosmetic results in breast cancer conservative treatment. *The Breast*, *16*(5), 456–461. https://doi.org/10.1016/j.breast.2007.05.002

Guo, C., Smith, T. L., Feng, Q., Benitez-Quiroz, F., Vicini, F., Arthur, D., White, J., & Martinez, A. (2022). A fully automatic framework for evaluating cosmetic results of breast conserving therapy. *Machine Learning with Applications*, *10*, 100430. https://doi.org/10.1016/j.mlwa.2022.100430

Harada, K., Yoshimoto, T., Duong, N. P., Nguyen, M. N., Sowa, Y., & Fukuzawa, M. (2023). A New Integrated Medical-Image Processing System with High Clinical Applicability for Effective Dataset Preparation in ML-Based Diagnosis. *Communications in Computer and Information Science*, *1950*, 41–50. Springer, Singapore. https://doi.org/10.1007/978-981-99-7666-9_4

Heil, J., Carolus, A., Dahlkamp, J., Golatta, M., Domschke, C., Schuetz, F., Blumenstein, M., Rauch, G., & Sohn, C. (2012). Objective assessment of aesthetic outcome after breast conserving therapy: Subjective third party panel rating and objective BCCT.core software evaluation. *The Breast*, *21*(1), 61–65. https://doi.org/10.1016/j.breast.2011.07.013

Preuss, J., Lester, L., Saunders, C. (2012). BCCT.core – Can a computer program be used for the assessment of aesthetic outcome after breast reconstructive surgery?. *The Breast*. *21*(4), 597–600. https://doi.org/10.1016/j.breast.2012.05.012

Rasband, W. S. (n.d.). *ImageJ*. https://imagej.net/ij/

The GIMP Development Team. (n.d.). *GIMP - GNU Image Manipulation Program.* https://www.gimp.org/

The MathWorks, Inc. (n.d.). *MATLAB*. https://www.mathworks.com/products/matlab.html

Yano, K. (2007). Cosmetic results after breast reconstruction (in Japanese). *Japanese Journal of Clinical Medicine*, *65*(6), 465–468