Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Deep learning-based prediction of mastectomy and reconstruction weights using three-dimensional CT and MRI: a retrospective proof-of-concept study

Sumin Jung
Core Research & Development Center, Korea University Ansan Hospital, Korea University College of Medicine

Han-Sang Song
Department of Plastic and Reconstructive Surgery, Korea University Ansan Hospital, Korea University College of Medicine

Dong Hwan Shin
School of Industrial and Management Engineering, Korea University

Hi-Jin You
Department of Plastic and Reconstructive Surgery, Korea University Ansan Hospital, Korea University College of Medicine

Deok-Woo Kim
Department of Plastic and Reconstructive Surgery, Korea University Ansan Hospital, Korea University College of Medicine

Taesu Cheong
School of Industrial and Management Engineering, Korea University

Tae-Yul Lee
tylee0919@korea.ac.kr

Department of Plastic and Reconstructive Surgery, Korea University Ansan Hospital, Korea University College of Medicine

Research Article

Keywords: Deep learning, Breast reconstruction, Mastectomy, Computed tomography, Magnetic resonance imaging, Three-dimensional imaging

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11239746/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Deep learning-based prediction of mastectomy and reconstruction weights using three-dimensional CT and MRI: a retrospective proof-of-concept study

Authors

Sumin Jung, PhD[1], Han-Sang Song, MD[2], Dong Hwan Shin, PhD[3], Hi-Jin You, MD, PhD[2, 4], Deok-Woo Kim, MD, PhD[2, 4] Taesu Cheong, PhD[3*] and Tae-Yul Lee, MD, PhD[2, 4#]

Affiliations

[1] Core Research & Development Center, Korea University Ansan Hospital, Korea University College of Medicine, Ansan, Republic of Korea

[2] Department of Plastic and Reconstructive Surgery, Korea University Ansan Hospital, Korea University College of Medicine, Ansan, Republic of Korea

[3] School of Industrial and Management Engineering, Korea University, Seoul, Republic of Korea

[4] Institute of Advanced Regeneration and Reconstruction, Seoul 02841, Republic of Korea

*** Co-corresponding authors**

[#] Corresponding author

Tae-Yul Lee, MD, PhD

Department of Plastic and Reconstructive Surgery
Korea University Ansan Hospital
Korea University College of Medicine
123 Jeokgeum-ro, Danwon-gu
Ansan 15355, Republic of Korea

E-mail: tylee0919@korea.ac.kr
ORCID: 0000-0001-6012-1694

Tae-Yul Lee and Taesu Cheong are corresponding authors.

Running title

Deep Learning Prediction of Breast Surgical Weights

Word count

Abstract: 190 words
Main text: 2336 words

Number of tables and figures

Main tables: 4
Main figures: 4
Supplementary tables: 1

Statements and Declarations

Funding

This work was supported by the Korea University Grant (grant no. K2426791).

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Author contributions

Sumin Jung contributed to data analysis, interpretation of the findings, and manuscript preparation. Han-Sang Song contributed to data collection and manuscript preparation. Dong Hwan Shin contributed to data collection and data curation. Hi-Jin You and Deok-Woo Kim contributed to the surgical treatment, interpretation of the findings, and critical revision of the manuscript. Taesu Cheong contributed to the study conception and design, data acquisition, interpretation of the findings, and supervision. Tae-Yul Lee contributed to the study conception and design, reconstructive treatment, data acquisition, interpretation of the findings, and supervision. All authors read and approved the final version of the manuscript.

Ethics approval and consent to participate

This study was approved by the Institutional Review Board (IRB) of Korea University Ansan Hospital (IRB no. 2025AS0246) and conducted in accordance with the ethical principles of the Declaration of Helsinki. The requirement for informed consent for study participation was waived by the IRB because of the retrospective nature of the study.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed in the current study are available from the corresponding authors upon reasonable request and are subject to institutional and ethical restrictions.

Acknowledgements

Not applicable.

Clinical trial number

Not applicable.

Abstract

Background: To evaluate the feasibility of deep learning-based prediction of mastectomy specimen and final reconstruction weights from three-dimensional CT and MRI, including multimodal and clinical-feature fusion.

Methods: This retrospective single-center study included 200 patients who underwent total mastectomy and TRAM (transverse rectus abdominis myocutaneous) flap breast reconstruction. Preoperative imaging was used for mastectomy-weight prediction and postoperative imaging for reconstruction-weight prediction. Three-dimensional deep learning models and CT-MRI feature fusion were evaluated with patient-level five-fold cross-validation, and clinical-feature fusion was additionally explored.

Results: The paired cohorts comprised 184 patients for mastectomy-weight prediction and 146 for reconstruction-weight prediction. MRI with ResNet3D-18 showed the lowest error for mastectomy weight (mean absolute error [MAE], 134.3 g; 95% confidence interval [CI], 119.2−151.4; weighted absolute percentage error, 28.3%), whereas CT-MRI fusion showed the lowest error for reconstruction weight (MAE, 123.0 g; 95% CI, 103.3−144.0; weighted absolute percentage error, 23.2%). Adding clinical features minimally affected mastectomy-weight prediction but numerically improved reconstruction-weight prediction (MAE, 123.0 to 108.2 g), although the 95% CI for the improvement crossed zero. Predictions tended toward the cohort mean at extreme weights.

Conclusions: Three-dimensional CT and MRI enabled direct deep learning-based prediction of measured surgical weights without explicit volumetric measurement, demonstrating the feasibility of direct image-to-weight prediction in breast reconstruction.

Keywords: Deep learning, Breast reconstruction, Mastectomy, Computed tomography, Magnetic resonance imaging, Three-dimensional imaging

Background

For patients undergoing total mastectomy with breast reconstruction, estimating the volume of tissue to be removed can inform operative planning. In autologous reconstruction, the anticipated mastectomy defect can guide flap design and tissue allocation. Breast size is commonly approximated from anthropometric measurements or visual assessment, but these methods do not directly estimate mastectomy specimen weight [1–3].

Several approaches have been proposed to quantify breast size, including direct anthropometry, casting or water displacement, and three-dimensional surface imaging [1, 3]. These techniques can characterize external shape and volume, but the mass of a mastectomy specimen is not determined by external volume alone, because the relative amounts and spatial distribution of fibroglandular and adipose tissue alter the relationship between volume and weight. Magnetic resonance imaging (MRI)-based volumetry has shown that tissue composition contributes to discrepancies between estimated breast volume and actual specimen weight [2].

Unlike surface measurements, computed tomography (CT) and MRI depict both three-dimensional anatomy and internal soft-tissue composition, and these examinations may already be available in selected patients during staging, treatment planning, or postoperative follow-up. Previous studies have estimated breast or reconstructive volume from MRI, mammography, CT-derived measurements, or three-dimensional surface imaging [2–6], but most of these methods require manual or automated boundary definition before volume can be calculated.

Deep learning can learn predictive features directly from three-dimensional image volumes. Three-dimensional convolutional neural networks integrate spatial information across slices, whereas transformer-based models can capture longer-range image relationships [7–9]. Feature-level fusion further permits the joint analysis of CT, MRI, and clinical variables.

The primary objective was to predict mastectomy specimen weight from preoperative CT or MRI in patients who underwent total mastectomy and TRAM (transverse rectus abdominis myocutaneous) flap breast reconstruction. The secondary objective was to estimate final reconstruction weight from

postoperative CT or MRI. Exploratory analyses assessed CT-MRI feature fusion, the addition of selected clinical variables, and associations between clinical characteristics and prediction error.

Methods

Study design and participants

This retrospective single-center study included a source cohort of 200 patients treated at Korea University Ansan Hospital between January 2015 and May 2025. The study was approved by the Institutional Review Board of Korea University Ansan Hospital (IRB No. 2025AS0246). Patients underwent total mastectomy followed by TRAM flap breast reconstruction. The patient was the unit of analysis. Patients were excluded if surgical weight data, the required preoperative or postoperative imaging, or essential clinical information were unavailable, or if reconstruction failure occurred. Modality-specific analyses included patients with the corresponding CT or MRI examination, whereas direct modality comparisons were restricted to patients with both CT and MRI. No patient was excluded solely because of an extreme observed weight or prediction error.

Imaging and outcome definitions

Clinical records provided patient identifiers, examination dates, mastectomy specimen weight, and final reconstruction weight. Preoperative imaging was used for the primary analysis and postoperative imaging for the secondary analysis. Image folders were matched by patient identifier and accepted when the DICOM StudyDate agreed with the assigned examination date. The primary outcome was mastectomy specimen weight in grams. The secondary outcome was final reconstruction weight in grams, defined as the final flap weight retained for inset and calculated as the initial harvested flap weight minus the weight of discarded tissue, based on the operative record.

Image preprocessing

DICOM slices were sorted by ImagePositionPatient when available and otherwise by InstanceNumber. MRI voxel intensities were clipped to the first and 99th percentiles and then min-max normalized, and CT pixel values were converted to Hounsfield units (HU) using the rescale slope and

intercept. For CT-only models, a soft-tissue window with a center of 50 HU and a width of 400 HU was used, and for feature-level fusion, three CT channels represented soft-tissue (50/400 HU), lung (−600/1500 HU), and wide (100/800 HU) windows. Each volume was resampled by linear interpolation to 128 × 128 × 64 voxels and scaled to the range 0–1. Labels were divided by 1,000 during optimization and converted back to grams for evaluation.

Model architectures

Three single-modality regressors were evaluated for both CT and MRI: a three-dimensional ResNet-18 (ResNet3D-18), a patch-based three-dimensional vision transformer (ViT3D), and a compact three-dimensional convolutional neural network (SimpleCNN3D). The ResNet stem was adapted for one-channel MRI or one-channel CT input, and each classifier head was replaced by a single linear regression output. The CT-MRI feature-fusion model used separate compact three-dimensional encoders for the three-channel CT and one-channel MRI volumes. The resulting 128-dimensional feature vectors from each modality were concatenated and passed through a 128-unit fully connected layer, rectified linear activation, 0.2 dropout, and a scalar output. The architectures followed published residual, three-dimensional convolutional, and transformer designs [7–9].

Training procedure

Models were trained from random initialization with Adam optimization [10], a learning rate of 1×10^{-4}, Huber loss [11], and 100 epochs. No early-stopping rule was applied.

Batch size was 8 for all MRI, CT, and CT-MRI fusion models. The training data loader was shuffled within each epoch. The configuration with the lowest mean absolute error (MAE) across pooled out-of-fold predictions was designated the best model for each target.

Validation strategy

Patient-level five-fold cross-validation was performed in the paired CT-MRI cohorts. Folds were generated with shuffling and a fixed random seed of 42. Each patient served as validation data exactly once, and all out-of-fold predictions were pooled for evaluation. The paired cohorts comprised 184 patients for the resection target and 146 patients for the reconstruction target. The same fold assignments were used across the MRI, CT, and fusion models to support direct comparison.

Exploratory clinical-feature fusion

Clinical-feature fusion was evaluated with five-fold cross-validation using the same fold assignments as the image-only models. Age, height, and body weight were standardized using the means and standard deviations computed within each training fold. Mastectomy type and cancer type were one-hot encoded. BMI (body mass index) was not entered concurrently because it is mathematically derived from height and body weight. A clinical encoder with 32 and 16 hidden units was concatenated with the image feature representation. For resection, the image branch was MRI ResNet3D-18; for reconstruction, it was the CT-MRI feature-fusion network. Clinical-only prediction was beyond the scope of the present study.

Performance metrics and statistical analysis

MAE in grams was the primary performance metric because it is directly interpretable and gives each patient equal weight. Root mean square error (RMSE) in grams was used to emphasize larger errors. Weighted absolute percentage error (WAPE) was calculated as 100 multiplied by the sum of absolute errors divided by the sum of observed weights and was included as a scale-normalized complementary metric. Mean signed error (prediction minus observation) represented bias, with negative values indicating average underestimation and positive values average overestimation. Pearson correlation and the coefficient of determination were also reported. Patient-level bootstrap resampling with 2,000 iterations generated percentile-based 95% confidence intervals (CIs) for MAE, RMSE, WAPE, and correlation.

For clinical-feature fusion, image-only and image-plus-clinical predictions were paired by patient. A positive MAE difference indicated lower error after adding clinical variables, and a 2,000-iteration paired bootstrap provided the 95% CI. Exploratory associations between absolute error from the best image-only five-fold model and continuous clinical variables were assessed with Spearman correlation, whereas categorical variables were assessed with the Kruskal-Wallis test. Benjamini-Hochberg false-discovery-rate adjustment was applied across 12 exploratory tests. Analyses were conducted in Python 3.12.8 with PyTorch 2.5.1, torchvision 0.20.1, scikit-learn 1.6.1, SciPy 1.15.1, pydicom 3.0.1, pandas 2.2.3, and NumPy 2.1.3.

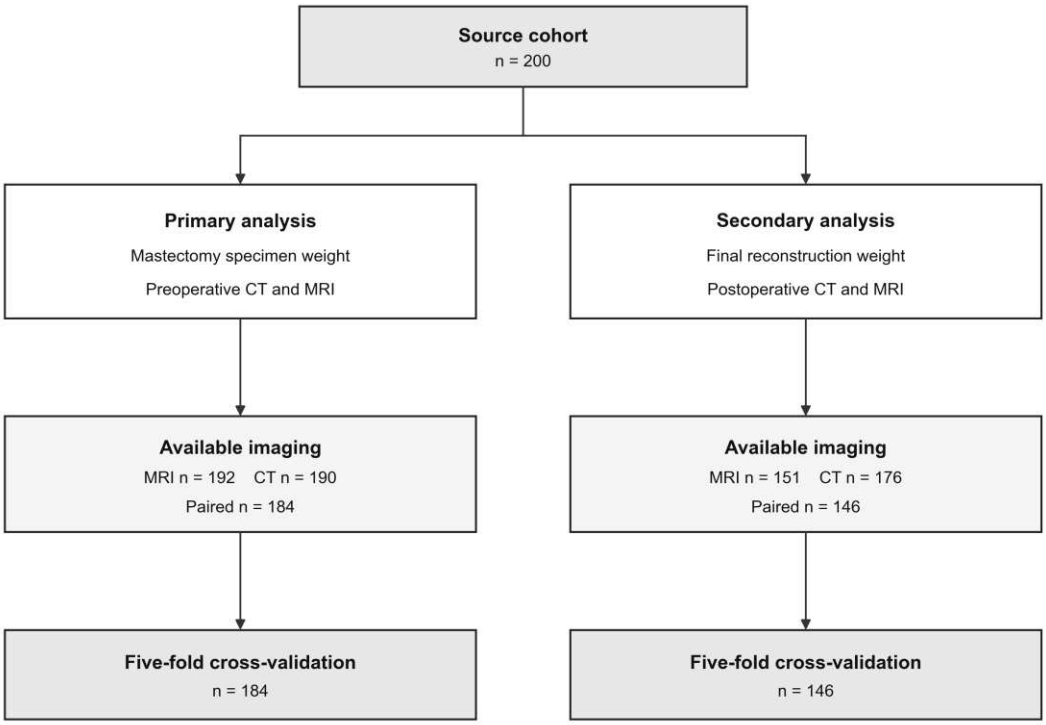
Results

Cohort and imaging availability

The source cohort comprised 200 patients. Preoperative imaging was available for 192 MRI and 190 CT examinations, with 184 patients having paired CT and MRI. Postoperative imaging was available for 151 MRI and 176 CT examinations, with 146 paired cases. The mean mastectomy specimen weight in the source cohort was 470.5 g (standard deviation, 232.5 g), and the mean final reconstruction weight was 540.2 g (standard deviation, 235.7 g).

Five-fold cross-validation included all 184 paired patients for the primary resection analysis and all 146 for the secondary reconstruction analysis (Figure 1 and Table 1).

Figure 1. Study flow and analytic cohorts for five-fold cross-validation.



From the source cohort of 200 patients, 184 were included in the primary analysis of mastectomy specimen weight and 146 in the secondary analysis of final reconstruction weight, with both cohorts

evaluated with five-fold cross-validation.

Table 1. Cohort characteristics

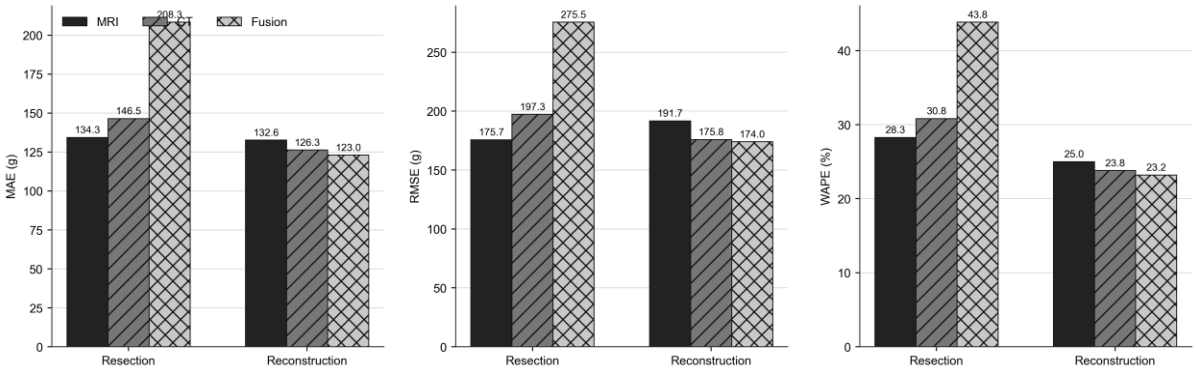
Characteristic		Source cohort	Resection evaluation	Reconstruction evaluation
N		200	184	146
Age, years		48.8 (7.4) 48.0 [44.0–53.0]	48.8 (7.4) 48.0 [44.0–53.0]	48.5 (6.6) 48.0 [44.0–52.0]
Height, cm		157.9 (6.1) 158.0 [153.6–162.2]	157.9 (6.2) 158.0 [153.2–162.5]	158.2 (6.4) 158.5 [153.7–163.2]
Body weight, kg		61.7 (11.0) 60.1 [54.2–67.1]	61.9 (11.2) 60.3 [54.2–67.3]	62.0 (11.4) 60.5 [54.9–67.1]
Body mass index, kg/m^2		24.7 (3.9) 24.2 [22.4–26.2]	24.8 (3.9) 24.2 [22.3–26.1]	24.7 (4.0) 24.1 [22.4–26.1]
Mastectomy specimen weight, g		470.5 (232.5) 428.0 [302.0–560.0]	475.3 (237.6) 425.5 [302.0–572.2]	459.2 (222.8) 423.5 [290.0–559.0]
Final reconstruction weight, g		540.2 (235.7) 494.0 [380.8–629.5]	543.6 (240.5) 495.5 [380.8–629.5]	530.4 (230.9) 479.0 [378.5–622.8]
Mastectomy type	Nipple-sparing mastectomy	91 (45.5%)	87 (47.3%)	72 (49.3%)
	Skin-sparing mastectomy	105 (52.5%)	94 (51.1%)	73 (50.0%)
	Modified radical mastectomy	4 (2.0%)	3 (1.6%)	1 (0.7%)
Cancer type	Ductal carcinoma in situ	23 (11.5%)	20 (10.9%)	13 (8.9%)
	Invasive carcinoma	167 (83.5%)	155 (84.2%)	127 (87.0%)
	Inflammatory carcinoma	4 (2.0%)	4 (2.2%)	3 (2.1%)
	Phyllodes tumor	2 (1.0%)	2 (1.1%)	0 (0.0%)
	Mucinous carcinoma	4 (2.0%)	3 (1.6%)	3 (2.1%)

Clinical and surgical characteristics of the source cohort and the subsets included in the resection and reconstruction evaluations.

Mastectomy specimen weight

MRI with ResNet3D-18 achieved the lowest pooled out-of-fold MAE among the seven image-model configurations, with an MAE of 134.3 g (95% CI, 119.2–151.4). The RMSE was 175.7 g, the WAPE was 28.3%, and the mean bias was 29.8 g. Pearson correlation was 0.70 (95% CI, 0.60–0.77), and R^2 was 0.45. The best CT model achieved an MAE of 146.5 g and a WAPE of 30.8%, whereas feature-level CT-MRI fusion yielded an MAE of 208.3 g and a WAPE of 43.8% (Figure 2 and Tables 2 and 4).

Figure 2. Performance of the best model in each imaging modality under five-fold cross-validation.



Comparison of the best-performing MRI, CT, and CT-MRI fusion models for mastectomy specimen weight and final reconstruction weight, evaluated with mean absolute error (MAE), root mean square error (RMSE), and weighted absolute percentage error (WAPE).

Table 2. Candidate model performance for resection

Modality	Model	*N*	MAE (g)	RMSE (g)	WAPE (%)	Bias (g)
MRI	**ResNet3D-18**	184	**134.3**	175.7	28.3	29.8
MRI	ViT3D	184	159.4	229.5	33.5	−33.2
MRI	SimpleCNN3D	184	159.7	225.2	33.6	28.8
CT	ResNet3D-18	184	146.5	197.3	30.8	−6.7
CT	ViT3D	184	156.3	205.2	32.9	20.1
CT	SimpleCNN3D	184	166.1	216.2	35.0	80.2
Fusion	CT-MRI feature fusion	184	208.3	275.5	43.8	14.4

Performance of candidate deep-learning models for predicting mastectomy specimen weight using MRI, CT, or CT-MRI feature fusion.

Table 4. Performance of the best model for each target

Metric	Resection	Reconstruction
Modality	MRI	Fusion
Model	ResNet3D-18	CT-MRI feature fusion
N	184	146
MAE, g (95% CI)	134.3 (119.2–151.4)	123.0 (103.3–144.0)
RMSE, g (95% CI)	175.7 (147.5–205.0)	174.0 (137.7–218.3)
WAPE, % (95% CI)	28.3 (25.1–31.8)	23.2 (20.1–26.2)
Median absolute error, g	117.7	89.4
Bias, g	29.8	−53.0
R^2	0.45	0.43
Pearson *r* (95% CI)	0.70 (0.60–0.77)	0.71 (0.64–0.79)

Performance of the best-performing imaging model for each prediction target, including error, bias, and correlation metrics.

Final reconstruction weight

Feature-level CT-MRI fusion achieved the lowest pooled out-of-fold MAE for final reconstruction

weight, at 123.0 g (95% CI, 103.3–144.0). The RMSE was 174.0 g, the WAPE was 23.2%, and the mean bias was −53.0 g. Pearson correlation was 0.71 (95% CI, 0.64–0.79), and R^2 was 0.43. The best MRI and CT configurations achieved MAEs of 132.6 g and 126.3 g, respectively (Figure 2 and Tables 3 and 4).

Table 3. Candidate model performance for reconstruction

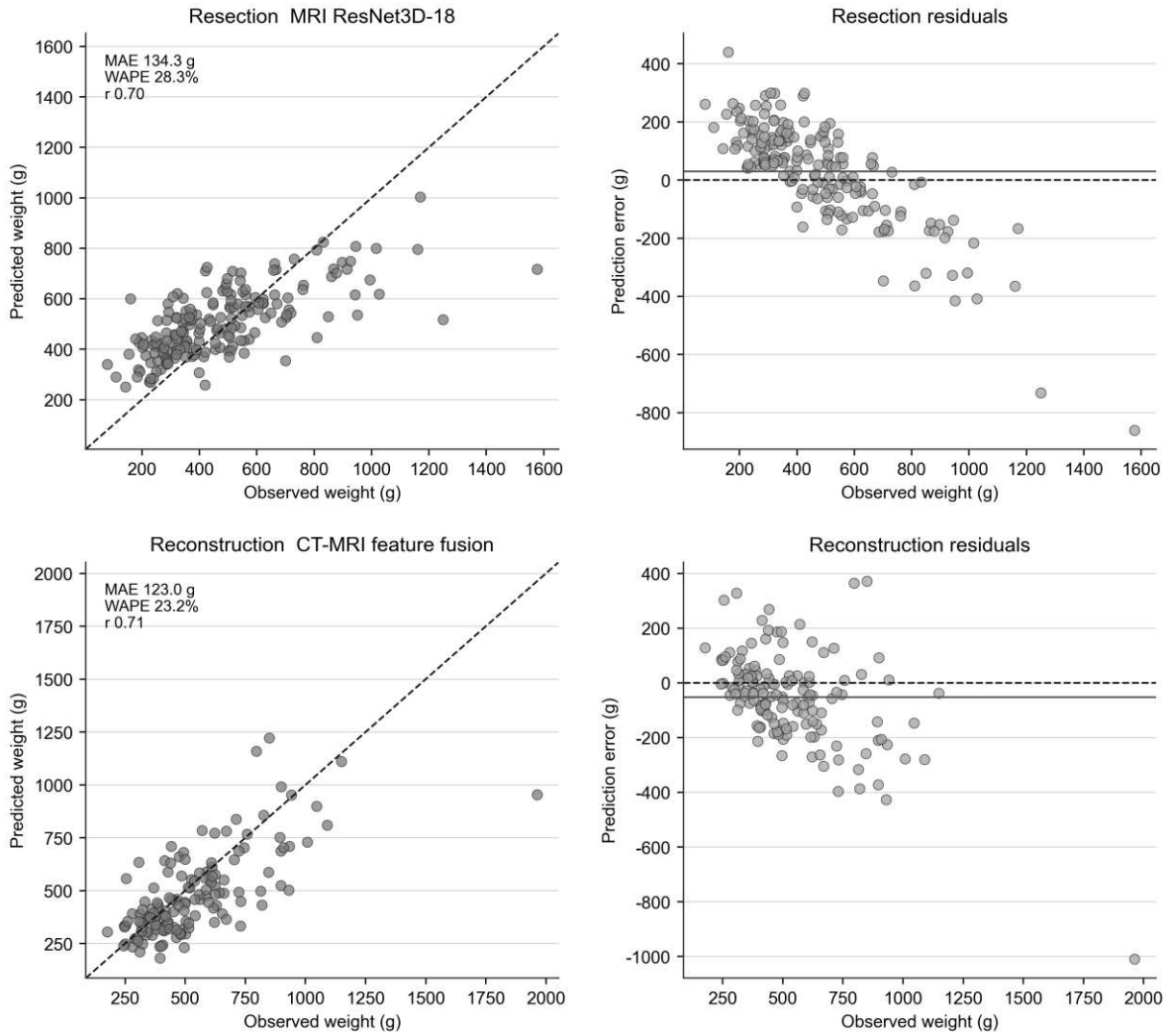
Modality	Model	*N*	MAE (g)	RMSE (g)	WAPE (%)	Bias (g)
MRI	ResNet3D-18	146	132.6	191.7	25.0	−25.4
MRI	ViT3D	146	133.2	174.4	25.1	4.4
MRI	SimpleCNN3D	146	183.2	251.4	34.5	101.4
CT	ResNet3D-18	146	126.3	175.8	23.8	4.5
CT	ViT3D	146	177.5	253.8	33.5	17.2
CT	SimpleCNN3D	146	234.6	301.3	44.2	71.2
Fusion	CT-MRI feature fusion	146	**123.0**	174.0	23.2	−53.0

Performance of candidate deep-learning models for predicting final reconstruction weight using MRI, CT, or CT-MRI feature fusion.

Observed-predicted agreement and error pattern

Observed and out-of-fold predicted weights were positively associated for both targets. Residual plots showed a pronounced tendency toward overestimation at lower observed weights and underestimation at higher observed weights, with several large errors among the highest-weight cases (Figure 3).

Figure 3. Observed and predicted weights and residuals for the best-performing models under five-fold cross-validation.

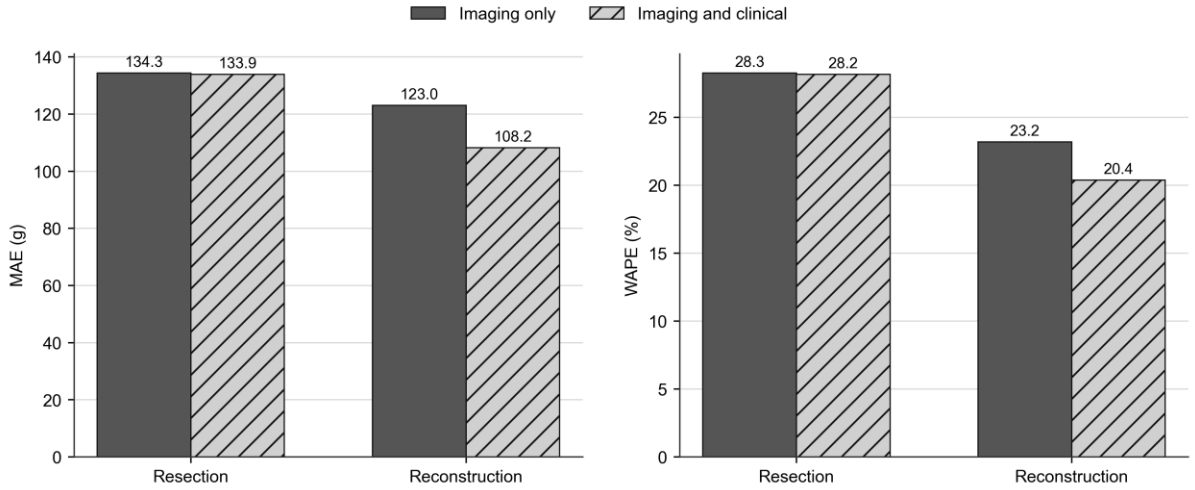


Comparison of actual and predicted weights and the corresponding prediction errors for mastectomy specimen weight and final reconstruction weight. Lower weights tended to be overestimated, whereas higher weights tended to be underestimated.

Exploratory clinical-feature fusion

For the resection target ($N = 184$), adding clinical features changed the MAE from 134.3 g to 133.9 g and the WAPE from 28.3% to 28.2%, and the paired MAE improvement was 0.4 g (95% CI, −8.8 to 9.5). For the reconstruction target ($N = 146$), the MAE decreased from 123.0 g to 108.2 g, the RMSE from 174.0 g to 145.1 g, and the WAPE from 23.2% to 20.4%, and the paired MAE improvement was 14.9 g (95% CI, −1.6 to 31.9). Adding clinical variables therefore did not change resection performance and lowered reconstruction error, although the 95% CI for the paired difference crossed zero (Figure 4 and Table 5).

Figure 4. Exploratory comparison of imaging-only and imaging-plus-clinical models in five-fold cross-validation.



Comparison of imaging-only and imaging-plus-clinical models for mastectomy specimen weight and final reconstruction weight. The clinical features comprised age, height, body weight, mastectomy type, and cancer type.

Table 5. Exploratory clinical-feature fusion

Target	*N*	Imaging only MAE (g)	Imaging and clinical MAE (g)	MAE difference* (95% CI), g	Imaging only WAPE (%)	Imaging and clinical WAPE (%)
Resection	184	134.3	133.9	0.4 (−8.8 to 9.5)	28.3	28.2
Reconstruction	146	123.0	108.2	14.9 (−1.6 to 31.9)	23.2	20.4

*MAE difference was calculated as imaging-only MAE minus imaging-plus-clinical MAE

Comparison of imaging-only models with models incorporating clinical features for resection and reconstruction weight prediction.

Exploratory clinical correlates of absolute error

No association between absolute prediction error and age, height, body weight, BMI, mastectomy type, or cancer type remained significant after false-discovery-rate adjustment (all $q \geq 0.07$; Additional file 1). The lowest adjusted q values were observed for body weight and BMI in the reconstruction task; however, these findings were exploratory and did not meet the adjusted $q < 0.05$ criterion.

Discussion

The present study evaluated the feasibility of direct image-based prediction of mastectomy specimen weight and final reconstruction weight. In five-fold cross-validation, preoperative MRI with ResNet3D-18 showed the lowest error for mastectomy specimen weight (MAE, 134.3 g; WAPE, 28.3%), whereas CT-MRI fusion showed the lowest error for final reconstruction weight (MAE, 123.0 g; WAPE, 23.2%). These findings provide proof-of-concept that three-dimensional CT and MRI carry information sufficient for direct prediction of surgically measured tissue weight. Ultimately, the approach should be extended to predict the final reconstruction requirement from preoperative imaging alone.

Previous studies have assessed breast size using anthropometric measurements, three-dimensional surface imaging, and CT- or MRI-based volumetry [1–6]. MRI-derived breast volume has been associated with actual mastectomy weight, with tissue composition contributing to discrepancies between volume and weight [2]. Three-dimensional surface imaging has also been used to estimate mastectomy weight [12]. However, these approaches generally rely on breast boundary identification and volume calculation, whereas our approach used three-dimensional CT or MRI as input to deep-learning regression models to predict surgical weight directly, without explicit volumetric calculation. Recent deep-learning studies have demonstrated automated MRI-based breast segmentation and volumetric assessment [13]; the present study extends this concept to direct image-to-weight prediction.

The optimal imaging approach differed according to the prediction target. For mastectomy wei ght, MRI with ResNet3D-18 performed best, and CT-MRI fusion did not improve prediction, wh ereas for reconstruction weight, CT-MRI fusion achieved the lowest error, although its improvem ent over CT alone was modest (MAE, 123.0 vs. 126.3 g). These findings suggest that the value of multimodal imaging may be task-dependent rather than universally advantageous. Larger datas ets are required to determine whether these modality-specific findings are reproducible.

Clinical variables also showed different effects between the two targets. Their addition produced essentially no improvement in mastectomy-weight prediction (MAE, 134.3 to 133.9 g), whereas reconstruction-weight MAE decreased from 123.0 to 108.2 g. However, the CI for this improvement crossed zero, and the result should be interpreted as a numerical rather than a statistically confirmed improvement. Previous machine-learning studies have demonstrated that anthropometric variables can

predict breast weight [14], suggesting that clinical variables may provide complementary predictive value in selected tasks.

Prediction errors showed regression toward the mean, with lower weights tending to be overestimated and higher weights underestimated. This pattern is clinically relevant, because larger breasts may require especially careful estimation of tissue requirements and flap allocation. Limited representation of extreme-weight cases and the relatively small dataset may have contributed to it. Future studies should therefore include more extreme-weight cases and investigate model optimization and calibration across the full weight range.

A potential advantage of this approach is that it can use CT or MRI obtained during routine clinical care, without requiring dedicated three-dimensional surface imaging or explicit volumetric measurement. Nevertheless, the present models represent a proof-of-concept rather than a clinically deployable planning tool. In particular, although the reconstruction target was the actual final TRAM flap weight calculated intraoperatively, reconstruction prediction was based on postoperative imaging. The clinically relevant next step is therefore to predict the required final reconstruction weight directly from preoperative imaging, which could ultimately assist flap design and tissue allocation before surgery.

This study has several limitations. First, it was a retrospective, single-center study without external validation, and the sample size was relatively small for three-dimensional deep learning. Second, a clinical-only baseline model was not evaluated, limiting direct comparison between imaging-based prediction and prediction from clinical variables alone. Finally, reconstruction-weight prediction was based on postoperative imaging and therefore does not yet represent true preoperative prediction of reconstruction requirements. Future studies with larger multicenter cohorts and external validation should evaluate direct prediction of final reconstruction weight from preoperative imaging.

Conclusions

This study demonstrated the feasibility of direct deep learning-based prediction of mastectomy

specimen weight and final reconstruction weight from three-dimensional CT and MRI without explicit volumetric measurement. MRI with ResNet3D-18 showed the lowest error for mastectomy-weight prediction, whereas CT-MRI fusion showed the lowest error for reconstruction-weight prediction. Together, these findings provide proof-of-concept for direct image-to-weight prediction in breast reconstruction.

List of abbreviations

BMI: body mass index; CI: confidence interval; CT: computed tomography; HU: Hounsfield units; MAE: mean absolute error; MRI: magnetic resonance imaging; RMSE: root mean square error; TRAM: transverse rectus abdominis myocutaneous; WAPE: weighted absolute percentage error

Additional files

Additional file 1: Clinical variables and absolute prediction error in five-fold cross-validation. Format: Microsoft Word document (.docx). Spearman correlation and Kruskal-Wallis results relating age, height, body weight, BMI, mastectomy type, and cancer type to absolute prediction error for the resection and reconstruction targets, with Benjamini-Hochberg false-discovery-rate-adjusted q values.

References

1. Choppin SB, Wheat JS, Gee M, Goyal A. The accuracy of breast volume measurement methods: a systematic review. Breast. 2016;28:121–9. https://doi.org/10.1016/j.breast.2016.05.010.
2. Yoo A, Minn KW, Jin US. Magnetic resonance imaging-based volumetric analysis and its relationship to actual breast weight. Arch Plast Surg. 2013;40(3):203–8. https://doi.org/10.5999/aps.2013.40.3.203.
3. Lee WY, Kim MJ, Lew DH, Song SY, Lee DW. Three-dimensional surface imaging is an effective tool for measuring breast volume: a validation study. Arch Plast Surg. 2016;43(5):430–7. https://doi.org/10.5999/aps.2016.43.5.430.
4. Zingaretti N, Miotti G, Maronese CA, Isola M, Guarneri GF, Albanese R, et al. A prospective investigation of predictive parameters for preoperative volume assessment in breast reconstruction. J Clin Med. 2021;10(22):5216. https://doi.org/10.3390/jcm10225216.
5. Kim YS, Cho HG, Kim J, Park SJ, Kim HJ, Lee SE, et al. Prediction of implant size based on breast volume using mammography with fully automated measurements and breast MRI. Ann Surg Oncol. 2022;29(12):7845–54. https://doi.org/10.1245/s10434-022-11972-9.
6. Yip JM, Mouratova N, Jeffery RM, Veitch DE, Woodman RJ, Dean NR. Accurate assessment of breast volume: a study comparing the volumetric gold standard (direct water displacement measurement of mastectomy specimen) with a 3D laser scanning technique. Ann Plast Surg. 2012;68(2):135–41. https://doi.org/10.1097/SAP.0b013e31820ebdd0.
7. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770–8.
8. Hara K, Kataoka H, Satoh Y. Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and ImageNet? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:6546–55.
9. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. International Conference on Learning Representations. 2021.
10. Kingma DP, Ba J. Adam: a method for stochastic optimization. International Conference on Learning Representations. 2015. arXiv:1412.6980.
11. Huber PJ. Robust estimation of a location parameter. Ann Math Stat. 1964;35(1):73–101. https://doi.org/10.1214/aoms/1177703732.
12. Yu M, Mahoney MH, Soon G, Pinchuk B, Somogyi R. Predictive value of 3D imaging to guide implant selection in immediate breast reconstruction. JPRAS Open. 2022;31:50–61. https://doi.org/10.1016/j.jpra.2021.10.007.
13. Chen M, Xing J, Guo L. MRI-based deep learning models for preoperative breast volume and density assessment assisting breast reconstruction. Aesthetic Plast Surg. 2024;48(23):4994–5006. https://doi.org/10.1007/s00266-024-04074-2.
14. Teo SP, See MH, Lai LL, Wong LK, Nagappan S, See J, et al. Machine learning models for volume and weight estimation in breast reconstruction planning. Health Inf Sci Syst. 2026;14:33. https://doi.org/10.1007/s13755-025-00408-9.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx