Developing a deep learning model to predict the breast implant texture types with ultrasonography image: feasibility study

Ho Heon Kim, Won Chan Jeong, Kyungran Pi, Angela Soeun Lee, Min Soo Kim, Hye Jin Kim, Jae Hong Kim

Submitted to: JMIR Formative Research
on: March 25, 2024

Table of Contents

Original Manuscript **5**
Supplementary Files **22**
Multimedia Appendixes **23**
Multimedia Appendix 1 ..... **23**
Multimedia Appendix 2 ..... **23**
Multimedia Appendix 3 ..... **23**

Developing a deep learning model to predict the breast implant texture types with ultrasonography image: feasibility study

Ho Heon Kim[1] MS; Won Chan Jeong[2] BS; Kyungran Pi[3]; Angela Soeun Lee[4] MD; Min Soo Kim[4] MD; Hye Jin Kim[4] MD; Jae Hong Kim[5] MD

[1]Department of Biomedical Informatics Medical School of Yonsei University Seoul KR
[2]3Billion Seoul KR
[3]Qauntic EMBA 80 M Street SE Suite 2-196 Washington US
[4]Korean Society of Breast Implant Research Seoul KR
[5]THE W Clinic Seoul KR

Corresponding Author:
Jae Hong Kim MD
THE W Clinic
596 Gangnam-daero, Gangnam-gu
Seoul
KR

Abstract

Background: Breast implants, including textured variants, have been widely used in aesthetic and reconstructive mammoplasty. However, the textured type, which is one of the shell types of breast implants, has been identified as a possible carcinogenic factor for lymphoma, specifically breast implant-associated anaplastic large cell lymphoma (BIA-ALCL). Identifying the texture type of the implant is critical to the diagnosis of BIA-ALCL. However, distinguishing the shell type can be difficult due to human memory or loss of medical history. An alternative approach is to use ultrasonography, but this method also has limitations in quantitative assessment.

Objective: The objective of this study is to determine the feasibility of using a deep learning model to classify the textured shell type of breast implants and make robust predictions from ultrasonography images from heterogeneous sources.

Methods: A total of 19,502 breast implant images were retrospectively collected from heterogeneous sources, including images from both Canon (D1) and GE (D2), images of ruptured implants (D3), and images without implants (D4), as well as publicly available images (D5). The Canon (D1) images were trained using Resnet-50. The performance of the model on D1 was evaluated using stratified 5-fold cross-validation. Additionally, external validation was conducted using D2 and D5. The AUROC and PRAUC were calculated based on the contribution of the pixels with Grad-CAM. To identify the significant pixels for classification, we masked the pixels that contributed less than 10%, up to a maximum of 100%. To assess model robustness to uncertainty, Shannon entropy was calculated for four image groups: Canon (D1), GE (D2), ruptured implant (D3), and without implants (D5).

Results: The deep learning model achieved an average AUROC of 0.98 and a PRAUC of 0.88 in the Canon dataset (D1). For images captured with GE (D2), the model achieved an AUROC of 0.985 and a PRAUC of 0.748. Additionally, the model predicted an AUROC of 0.909 and a PRAUC of 0.958 for a dataset available online. For quantitative validation, this model maintained PRAUC up to 90% masking of less contributing pixels, and the remnant pixels located in breast shell layers. Furthermore, the prediction uncertainty increased in the following order: Canon (D1), GE (D2), ruptured implant (D3), no implant (D5) (0.066; 0072; 0.371; 0.777, respectively).

Conclusions: We have demonstrated the feasibility of using deep learning to predict the shell types of breast implants. With this approach, the textured shell types of breast implants can be quantified, supporting the first step in the diagnosis of BIA-ALCL.

(JMIR Preprints 25/03/2024:58776)
DOI: https://doi.org/10.2196/preprints.58776

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Developing a deep learning model to predict the breast implant texture types with ultrasonography image: a feasibility study

Ho Heon Kim[1], MS; Won Chan Jeong[2]; Kyungran Pi[3], MS; Angela Soeun Lee[4], M.D; Min Soo Kim[4] M.D; Hye Jin Kim[4], M.D; *Jae Hong Kim[5], M.D[5]

[1]Department of Biomedical Informatics, Medical School of Yonsei University, Seoul, South Korea

[2]3Billion, Inc, Seoul, South Korea

[3]Qauntic EMBA

[4]Korean Society of Breast Implant Research, Seoul, 03186, Republic of Korea

[5]THE W Clinic, Seoul, Korea

*Address correspondence to: Jae Hong Kim, M.D.,

Current affiliation: The W Clinic

Address: 9F Kukdong B/D, 596 Gangnam-daero, Gangnam-gu, Seoul 06626 Korea,

Tel: +82-2-517-7617,

Fax: +82-2-518-7617,

E-mail: stenka@hanmail.net.

Keywords: Breast Implants; Breast Implants; Ultrasonography: AI-assisted diagnosis: Shell surface topography

Abstract

Introduction

Breast implants, including textured variants, have been widely used in aesthetic and reconstructive mammoplasty. However, the textured type, which is one of the shell types of breast implants, has been identified as a possible etiologic factor for lymphoma, specifically breast implant-associated anaplastic large cell lymphoma (BIA-ALCL). Identifying the texture type of the implant is critical to the diagnosis of BIA-ALCL. However, distinguishing the shell type can be difficult due to the loss of human memory or medical history. An alternative approach is to use ultrasonography, but this method also has limitations in quantitative assessment.

Objective

This study aims to determine the feasibility of using a deep learning model to classify the textured shell type of breast implants and make robust predictions from ultrasonography images from heterogeneous sources.

Methods

A total of 19,502 breast implant images were retrospectively collected from heterogeneous sources, including images from both Canon (D1) and GE (D2), images of ruptured implants (D3), and images without implants (D4), as well as publicly available images (D5). The Canon (D1) images were trained using Resnet-50. The performance of the model on D1 was evaluated using stratified 5-fold cross-validation. Additionally, external validation was conducted using D2 and D5. The AUROC and PRAUC were calculated based on the contribution of the pixels with Grad-CAM. To identify the significant pixels for classification, we masked the pixels that contributed less than 10%, up to a maximum of 100%. To assess model robustness to uncertainty, Shannon entropy was calculated for four image groups: Canon (D1), GE (D2), ruptured implant (D3), and without implants (D5).

Result

The deep learning model achieved an average AUROC of 0.98 and a PRAUC of 0.88 in the Canon dataset (D1). For images captured with GE (D2), the model achieved an AUROC of 0.985 and a PRAUC of 0.748. Additionally, the model predicted an AUROC of 0.909 and a PRAUC of 0.958 for a dataset available online. For quantitative validation, this model maintained PRAUC up to 90% masking of less contributing pixels and the remnant pixels located in breast shell layers. Furthermore, the prediction uncertainty increased in the following order: Canon (D1), GE (D2), ruptured implant (D3), and no implant (D5) (0.066; 0072; 0.371; 0.777, respectively).

Conclusion

We have demonstrated the feasibility of using deep learning to predict the shell types of breast implants. With this approach, the textured shell types of breast implants can be quantified, supporting the first step in the diagnosis of BIA-ALCL.

Introduction

Breast implants have been developed for aesthetic and reconstructive mammaplasty since 1962. The first textured breast implant was developed in 1968 to prevent capsular contracture after aesthetic or reconstructive implant-based mammaplasty [1,2]. Engraving and embossing types of textured implants have also been used in anatomical breast implants for natural shape. Since the first case of breast implant-associated anaplastic large cell lymphoma (BIA-ALCL) in 1997 by Dr. Keech, a total of 1264 cases of BIA-ALCL, including 59 deaths, have been reported by the Food and Drug Administration (FDA), according to a recent update as of June 30, 2023 [3,4]. Since the first case of BIA-ALCL, numerous investigations have been conducted to examine its etiology, prevalence rates, and clinical characteristics.

Several studies have demonstrated that the topography of a textured breast implant shell surface is associated with BIA-ALCL [5–11]. Classified as a rare T-cell lymphoma, BIA-ALCL is nevertheless a significant concern in the context of breast augmentation and reconstruction surgeries, with documented cases of mortality. BIA-ALCL is often treatable when detected early, underscoring the critical importance of timely and accurate diagnosis [6,9–11].Recent study found that prophylactic replacement can be indicated in asymptomatic, risk-stratified patients to reduce the risk of BIA-ALCL [12]. Identifying the inserted breast implant shell surface topography is the first step for diagnosing BIA-ALCL in follow-up breast examination and replacement cases. If a patient has a history of primary aesthetic or reconstructive mammaplasty utilizing a smooth-type breast implant, there is generally no cause for concern regarding BIA-ALCL.

Nevertheless, a substantial number of patients may not be aware of the specific type of breast implant shell inserted during surgery over a long period of time, and medical records, especially within private clinics, may not be well preserved. Additionally, there are flaws in government policy in the regulation of medical devices in Korea [13]. As the medical community deepens its understanding of the complexities of this condition, the significance of diagnosing the surface topography of breast implant shells becomes increasingly apparent. Traditional diagnostic methods for assessing breast implants have relied heavily on subjective human evaluations. Mammograms and MRIs, which are used to monitor breast implant-related complications, cannot identify the implant shell surface topography. Only ultrasonography can identify the inserted implant shell surface topography [14,15].

However, ultrasonography also has many limitations. The generalizability of the results may be limited due to the real-time and operator-dependent nature of ultrasonography, which may result in

inter- and intra-observer variability. Although useful in overcoming these limitations of ultrasound, we conducted this study to determine whether it is useful in distinguishing breast implant shell surface topography using our algorithm. Additionally, the development of artificial intelligence (AI) programs aimed at accurately diagnosing breast implant shell surface topography holds promise in revolutionizing early detection and management strategies for BIA-ALCL. The development of AI algorithms aims to eliminate subjectivity and provide a more accurate and standardized approach to identifying the breast implant shell surface topography as textured and smooth type.

Methods

Study design

In this study, we retrospectively collected anonymous and de-identified medical records containing information on shell types of implants, ultrasonographic images, and demographic characteristics. We built multiple datasets as follows: 1) Canon dataset (D1), 2) GE dataset (D2), 3) ruptured implant dataset (D3), 4) no implant image dataset (D4), and publicly available dataset (D5). We used the Canon dataset (D1) for training, interval validation, and testing, GE, and a publicly available dataset for external validation. The ruptured implant dataset and no implant images were also used as out-of-distribution datasets to identify model interpretation.

First, the Canon and GE datasets (D1, D2) include the ultrasonography images with medical data generated from patients who underwent aesthetic or reconstructive implant-based mammaplasty without implant rupture at a single institution in South Korea. All patients underwent both breast cancer examination and ultrasonography-assisted examination at the institution. Ultrasonography-assisted examinations were conducted with an Aplio i600 (Canon Medical System, Otawara, Tochigi, Japan) system with a 7-18-MHz linear transducer, General Electric LOGIQ™ E10. We used these retrospective data confirmed by a surgeon between 31 August 2017 and 31 November 2022. We finally obtained the ultrasonography with medical data from 1,043 patients (Supplementary data 1). Multiple ultrasonography images from each patient were captured to assess the shell types of implants and saved in PACS-rendered JPEG format. We retained unique images for model development by checking 128-bit MD5 hash algorithms to rule out data leakage between the train and test datasets. A breast surgeon with 14 years of breast implant ultrasonography experience labeled all the ultrasonographic images of shell surface topography.

In our problem, shell surface topography was divided into texture and smooth. The smooth type included micro-texture as conventional clinical classification [15]. Micro-texture types show almost the same shell surface topography with smooth type in high-resolution ultrasonography and light microscopy [15]. As some retrospective data were not stored as a DICOM (Digital Imaging and Communications in Medicine), we used only the centered PCAS-rendered image (Table 1), discarding the top 12%, bottom 10%, and left and right 7% of pixels.

Secondly, 131 ultrasonography images of the ruptured implant (D3) were collected using the Canon Aplio i600. Because of the damaged shell integration, the shell type of implant would be less easily identified from ultrasonography images. We used these images as an out-of-distribution (OOD) dataset to identify the model's ability to estimate the model's uncertainty for two types of shells. Thirdly, 338 ultrasonography images without implants were also captured and used as an OOD dataset to determine the transparency of our model by estimating uncertainty (D4). Finally, we constructed a publicly available dataset for external validation by searching for ultrasonography images using the following keywords: “breast implant ultrasound” and “breast implant ultrasonography” (D5).

Table 1. Eligible ultrasonography dataset

Dataset name	Device	Objective	Shell integrity	Shell surface topography (N=19,502)	
				Texture	Smooth
Canon (D1)	Canon	Train, validation, test	Intact	2,420	14,976
GE (D2)	GE	External validation	Intact	113	1,844
Ruptured implant (D3)	Canon	OOD	Ruptured	101	30
Without implant (D4)	Canon	OOD	NA	NA (338)	
Publicly available(D5)	Heterogeneous	External validation	Intact	11	7

Model development

Convolutional neural networks were used to scale the parameter sizes to achieve high performance. In this feasibility study, we used an off-the-shell convolution neural network (CNN) architecture originally designed for natural images, ResNet-50 composed of 50 layers as the backbone [16]. To speed up the proof of concept, we chose a lightweight model instead of the models with large number of parameters such as Vision transformer or SwinTransformer, which require expensive computational costs with a large amount of data due to the lack of inductive bias [17,18].

We also trained our model by conducting transfer learning on pre-trained ResNet-50, which learned ImageNet classification. Then, we replaced the classifier layer of ResNet-50, which has a 1,000-dimensional vector for multiclass, with a binary classifier layer to return two-dimensional vector for shell surface topology such as smooth and texture. Weighted binary cross-entropy was used as the objective function for parameter optimization, which effectively trains the model by penalizing incorrect predictions of the minor class due to class imbalance between shell types (texture types being a minor class). The weight in the weighted binary cross-entropy for minor class was calculated with the inversed ratio of minor class in training dataset.

Cropped PACS rendered images were preprocessed with resized with 224 x 224 pixels using bilinear interpolation. The image in the training dataset was feed into the model without any augmentation. In the train time, the deep learning model trained with Adam optimizer with 0.0001 learning rate, 32 of batch size. Also, total training epoch was 20, however, actually training epoch was smaller than total epochs due to early stopping by monitoring validation loss with 7 of patience.

Performance evaluation and model interpretation.

We evaluated our model with four experiments as following 1) Classification performance using multiple dataset; 2) Quantitative validation using masked image; 3) Uncertainty estimation for multiple datasets; 4) post-hoc explainable interpretation (Figure 1).

Firstly, to report the out model transparently, we performed two types of validation: 1) stratified cross-validation and 2) external validation with both the GE dataset and publicly available ultrasonography image (D2, D3). One is stratified 5-fold cross-validation was conducted to identify generalized performance in a Canon dataset due to class imbalance between smooth and texture shells. Data split was performed by stratified random split ultrasonography images into training (60%), validation (20%), and test (20%) sets with shell surface topology labels. We evaluated the area under the curve (AUC) with different cut-offs of the receiver operating characteristic (ROC) curve and precision-recall (PR) curve. Also, we conducted an external validation set to reduce latent bias with online acquired ultrasonography. We also identified the AUC of both the ROC and PR curves with these data.

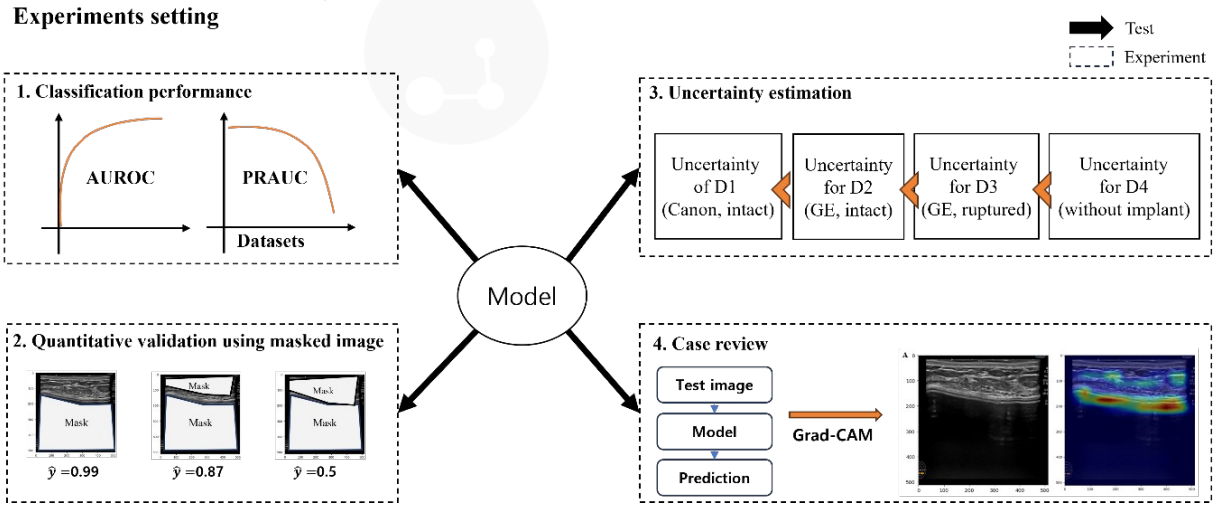
Secondly, we used an explainable AI approach (XAI) to determine whether our model accurately classified the shell types of implants from the features of echogenicity or layers from ultrasonography images and no other unexpected factors by assessing the classification performance according to the masking part of the image. For the quantitative validation, we hypothesized that the

important pixels of the image that distinguished the type of implant were on the layers of the implant. In addition, we also hypothesized that there would be no performance degradation if some non-layered pixels were erased. Therefore, we calculated the pixel importance using Grad-CAM, a method to quantify a pixel's contribution to the classification [19]. Both AUROC and PRAUC were calculated by removing 10% of the least contributing pixels from the total number of pixels in the image and replacing them with zeros.

Thirdly, for the uncertainty estimation, we calculated the Shannon entropy to estimate the predictive uncertainty for each image in OOD [20,21]. Entropy ranges from 0 to 1, with a large value indicating greater predictive uncertainty. We also hypothesized that entropies in the ruptured implant dataset (D3) would be larger than those in the test set because our model trained on only ultrasonography images from a patient without damaged shell integrity. Furthermore, we hypothesized that the entropies in the absence of breast implant images (D4) would be larger than those in the ruptured implant dataset (D3). Our model trained only ultrasonographic features from the implant image dataset to classify the shell types.

Finally, in order to gain further insight into the model's decision-making processes, we employed the Gradient-weighted Class Activation Mapping (Grad-CAM) technique. This interpretative method was employed to visually trace and affirm the alignment between the model's predictive patterns and established medical expertise concerning the diagnosis of different breast shell types. Through Grad-CAM, heatmaps were generated which highlighted the critical regions in the imaging data that influenced the model's diagnostic predictions, thus ensuring that these insights were consistent with conventional medical knowledge.

Figure 1. Illustration of 4 of experimental settings



Ethics

This retrospective study was approved by the Internal Institutional Review Board of the Korea National Institute of Bioethics Policy (IRB No. P01-202401-01-006), which waived the requirement for informed consent of medical records, including patients' images and characteristics. All procedures described herein were performed by the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards**.** None of the authors has a financial interest in any products, devices, or drugs mentioned in this manuscript.

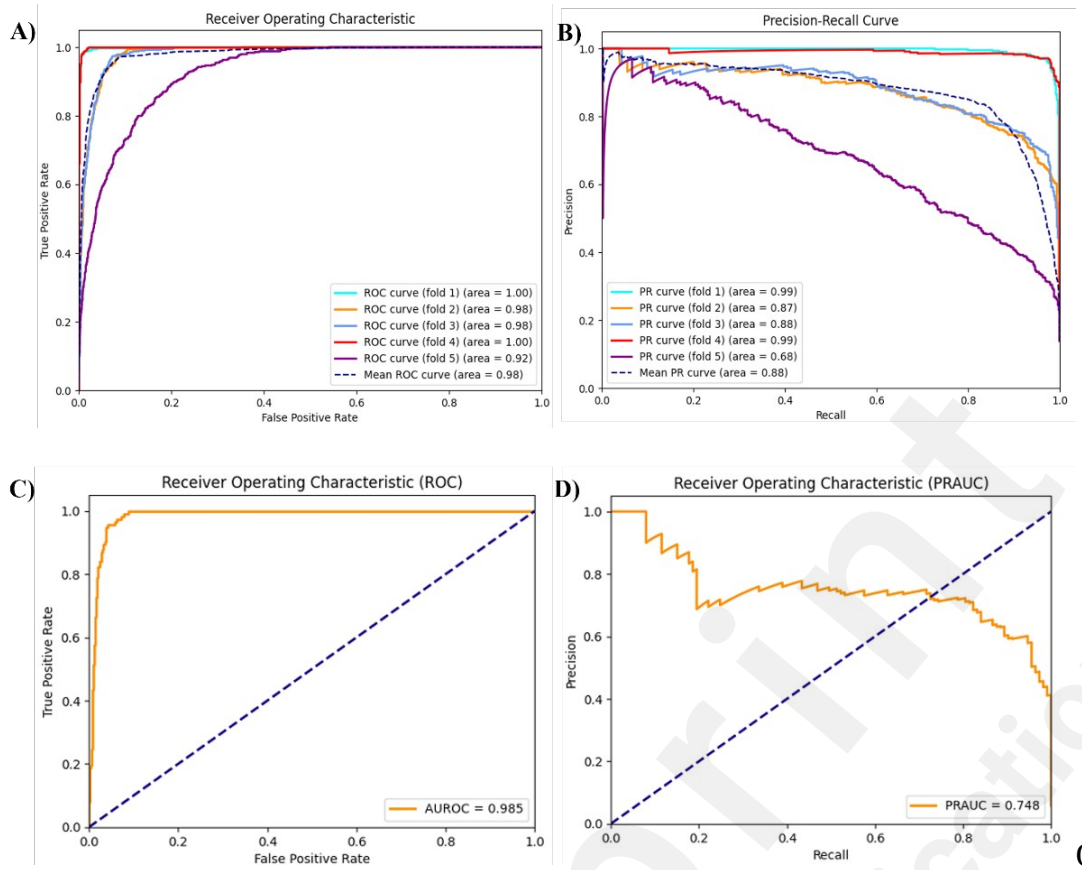
Results

Classification performance for shell surface type

Our model performed with an AUROC of 0.998 and PRAUC of 0.994 in the Canon dataset (D1) (Table 2). From stratified 5-fold cross-validation, our model showed an average AUROC of 0.98 and PRAUC of 0.88 in the Canon dataset captured with the Canon ultrasonography device (D1) (Supplementary data 2). Although the images were captured with a GE ultrasonography device (D2), we identified our model showed an AUROC of 0.985 and PRAUC of 0.748. (2-C, D). For the publicly available dataset (D5), the model showed 0.909 of AUROC and 0.958 of PRAUC (Supplementary data 2).

22**Table 2.** Performance metrics of classification for each dataset

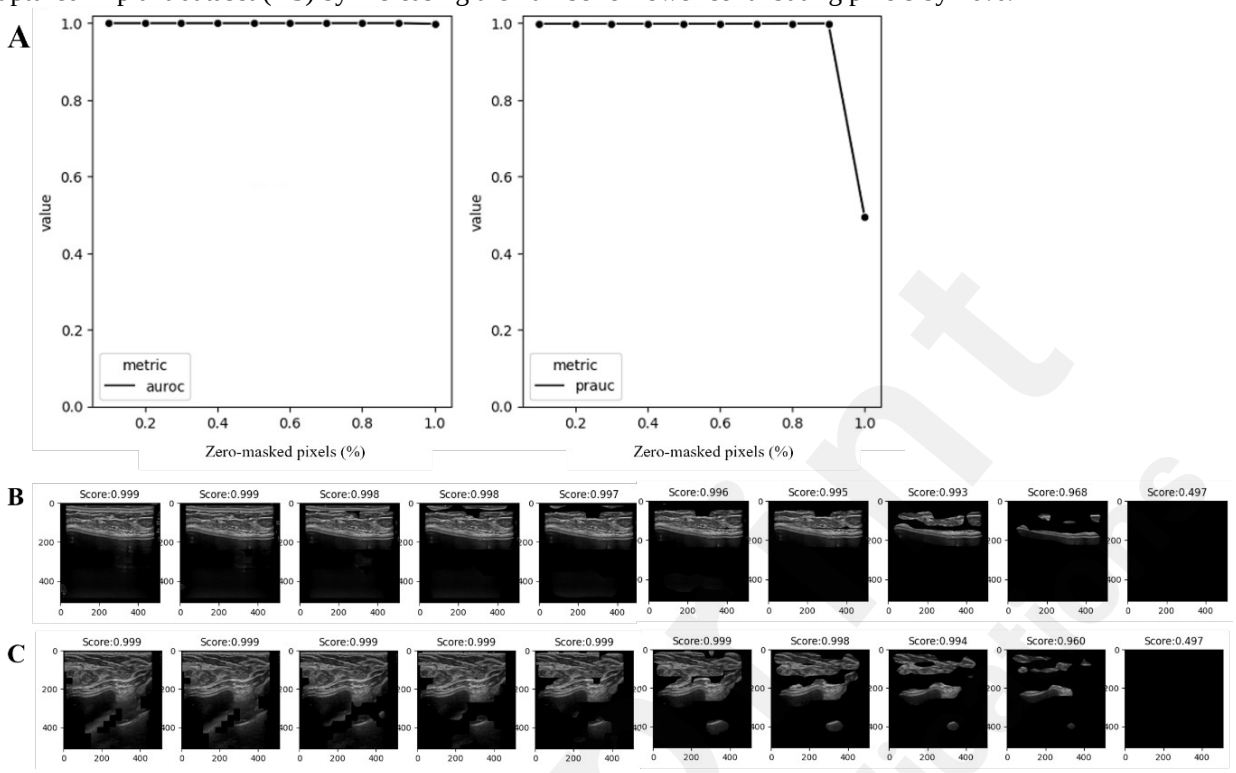
Metrics	Datasets			
	Canon (D1)	GE (D2)	Ruptured implant (D3)	Publicly available (D5)
AUROC	0.998	0.985	0.995	0.909
PRAUC	0.994	0.748	0.998	0.958



2 Quantitative validation

In the quantitative analysis to determine whether our model classifies ultrasonography images in accordance with medical knowledge, the model maintained an AUROC of 0.999 up to masking 90% of less contribution pixel to prediction and showed an AUROC of 0.997 when masking 100% of pixels. However, the PRAUC remained at 0.999 even after masking 90% of the pixels. After that, it decreased to 0.493 when all pixels were masked (Figure 3. Performance deterioration depends on masking non-contributing to prediction. A) Left. ROC curve for test dataset in Canon dataset; Right. PR curve for test dataset in Canon dataset; B) Texture shell implant prediction in Canon dataset (D1) by increasing the number of lower contributing pixels by 10%; C) Texture shell implant prediction in ruptured implant dataset (D3) by increasing the number of lower contributing pixels by 10%. For each individual case, the confidence for the texture shell type remained at 0.993, even when 80% or fewer contributing pixels were masked. At 90% of masking pixels, model confidence dropped to 0.968 and reached 0.497 when all of the pixels were masked. Similarly, the model confidence for another case with texture shell types was maintained at 0.994 until masking 80% of the pixels, decreased to 0.960 when masking 90% of pixels, and dropped to 0.947 when masking 100% of the pixels (Figure 3-B, C).

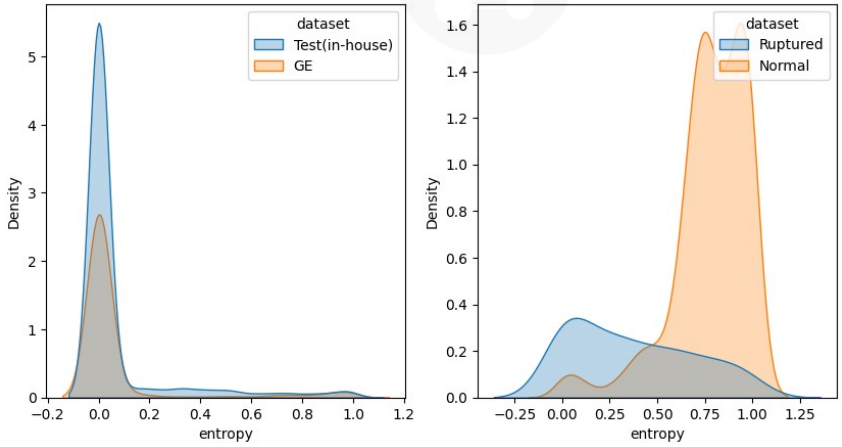
Figure 3. Performance deterioration depends on masking non-contributing to prediction. A) Left. ROC curve for test dataset in Canon dataset; Right. PR curve for test dataset in Canon dataset; B) Texture shell implant prediction in Canon dataset (D1) by increasing the number of lower contributing pixels by 10%; C) Texture shell implant prediction in ruptured implant dataset (D3) by increasing the number of lower contributing pixels by 10%.



Uncertainty estimation

The model did not significantly produce lower entropies for the test dataset in Canon data (D1) than for the external validation set from the GE ultrasonography device (mean [SD], 0.072[0.201] *vs* 0.066 [0.21]; p=0.350). However, the average entropy for ruptured implant images was significantly higher than for the test dataset in Canon data (mean [SD], 0.371 [0.318] *vs* 0.072[0.201]; p<.001). Moreover, the model also predicted a statistically significant high for the absence of implant images than for ruptured implant images (mean [SD], 0.777 [0.199] vs. 0.371 [0.318]; p<.001) (Figure 4).

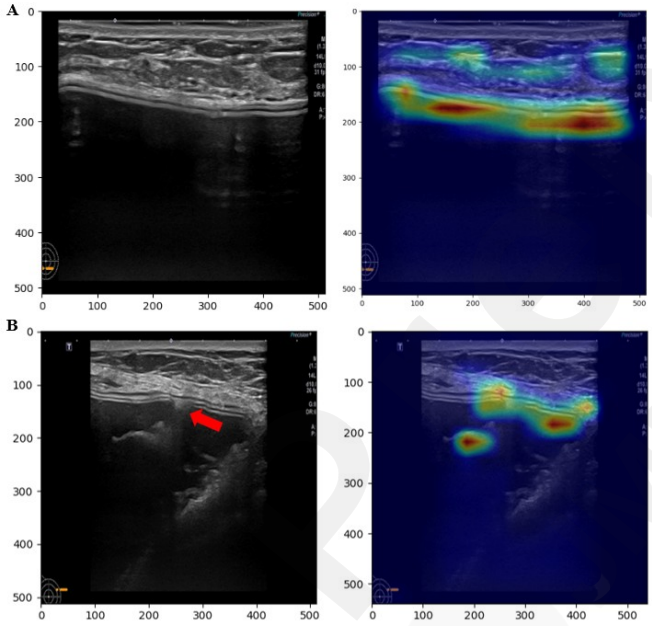
Figure 4. Entropy distribution of prediction from test in all datasets.



Individual case review

For a qualitative case review, we sampled two ultrasonography images, one from the test dataset (Canon, D1) and another from the ruptured implant datasets (D3), both captured by the same device. The model provided the model confidence of 0.998 as the texture type for the texture shell type image. The Grad-CAM score for texture type showed a high value at shell in a heatmap with Grad-CAM (white horizontal line in Figure 5-A). Also, for the image of the ruptured texture shell implant, the model provided the model confidence of 0.664 as the texture type. Although this score is higher than the classification threshold (0.5), it is lower than that of the intact texture shell-type implant. However, the Grad-CAM showed high in the adjacent intact layer from the ruptured shell area in the heatmap, despite the shell being ruptured due to a shell tear (Figure 5-B).

Figure 5. Preprocessed ultrasonography image with Grad-CAM for texture shell type prediction. Heatmap with Grad-CAM was bilinear interpolated to resize original image A) intact texture implant image captured with Canon ultrasonography device; B) damaged texture implant image captured with Canon ultrasonography device (the red arrow annotates shell tear).



Discussions

Principal Findings

Identifying breast implant shell types requires ultrasonographic examination, which can have inter- and intra-observer variability. Therefore, the generalizability of the results may be variable, leading to potentially missed diagnoses. However, no quantitative measurement or classification method distinguishes the two shell types. This feasibility study demonstrates that deep learning can be used to classify breast implant shell type quantitatively. Also, this study supports the use of echogenicity

from the shell layer of breast implants as an important region in classifying shell types. Furthermore, despite the use of different ultrasonography devices to capture images, our findings provide evidence that the deep learning model can classify the classes. Moreover, the model for ruptured breast implant ultrasonography images and ultrasonography images without the absence of an implant exhibit higher uncertainty than the mages from the intact shell type classification dataset, suggesting that the model could robustly quantify predictive uncertainty.

Clinical application

There are several classifications for breast implant shell surface topography; ISO 14607:2018 is a widely accepted classification [22]. Although there is a lack of standardized breast implant surface classification, ultrasonographic shell surface topographic images can divide the breast implant texture types into texture and smooth in high-resolution ultrasound [15]. Texture type shows roughness compared with smooth type in high-resolution ultrasonography (Supplementary data 3) [15]. Identifying the texture type of an inserted implant using ultrasonography without surgery is clinically important because the physician must consider the BIA-ALCL risk induced by the textured breast implant in patients, even in patients with no memory of the implant. As an extension of previous research on the feasibility of high-resolution ultrasonography for identifying the breast implant shell surface topography, this study was conducted to develop a deep-learning model to predict the breast implant texture types with ultrasonographic images [15]. In ultrasonography, smooth types include micro-texture types because micro-texture types show almost the same shell surface compared with smooth types.

This method offers a promising way to classify breast implants with respect to the risk of BIA-ALCL, a condition that remains underexplored in current research. Given its rarity and the association of certain texture types with BIA-ALCL, accurate identification of texture type emerges as a critical determinant in risk assessment. The use of ultrasonography to identify texture type allows for straightforward identification on ultrasound images, often eliminating the need for additional testing. In addition, the use of deep learning models has the potential to assist patients undergoing breast augmentation or reconstruction, particularly in cases of implant rupture. Given the limited familiarity of radiologists and breast physicians with the diverse landscape of breast implants, including both manufacturers and shell types, the integration of AI in clinical contexts is proving invaluable.

Reliable AI for clinical decision support

Reliable AI is essential for clinical decision support in the biomedical domain to avoid adverse patient outcomes[23,24]. This study includes multiple experiments on different datasets, such as

different devices and out-of-distribution datasets, to explore the model transparency. In AI research for Radiology, it was found that the deep learning models often showed deteriorated performance in external validation [25]. The study reveals that deep learning models may be vulnerable to medical images from heterogeneous sources due to unseen distribution. To eliminate biased evidence from these findings, we evaluated the model using an ultrasonography image from the heterogeneous device (2). In addition, this study showed uncertainty in the model's predictions, with the mean distribution being larger for images taken with the same device, images taken with a different device, images with ruptured implants, and images without implants (Figure 4). This can support the idea that the deep learning model classifies the shell type by learning the ultrasonographic features of breast implant shells. Further, ruptured implant images are consistent with those in the medical field, where determining the shell type of a ruptured implant is difficult due to the damaged surface of the implant (Figure 4). The entropies for ruptured implant images (D4) showed higher than intact implant images (D1, D2). This approach provides model confidence that can help clinicians make decisions that reflect the uncertainty in the diagnosis when uncertainty is high, for example, when the model consensus is close to 0.5, and more confident decisions when it is close to 1. Also, the important pixels can be provided to clinicians by conducting posthoc analyses such as Grad-CAM or Score-CAM [19,26].

Limitations

This study acknowledges several limitations that may introduce bias into interpretations. Primarily, the ultrasonography dataset did not represent all ultrasonography devices worldwide. Given the variability in the device resolution, configuration, and manufacturer, classification performance cannot be universally applied. To mitigate this, we performed internal and external validations on various ultrasound devices and incorporated out-of-distribution data to achieve less biased and more widely applicable results. In addition, the implant images collected did not include all types of shells used worldwide; rather, we focused only on implants from eight manufacturers licensed by the Ministry of Food and Drug Safety of the Republic of Korea. As a result, a multicenter study spanning multiple nations and including images of common implants in each region would allow for more generalized interpretations of the results. Lastly, as this research is at the feasibility stage, there are currently no existing studies that have classified breast shell types. Consequently, it is challenging to compare other state-of-the-art methodologies. This limits the ability to assess the objectivity of this study's findings or identify the best practice for the classification of the shell types. However, as this

is the first investigation into the classification of breast shell types, it can serve as the foundation for future studies in this area.

Conclusion

The feasibility study presented demonstrates the potential of deep learning to accurately classify breast implant shell types from ultrasound images, addressing the current lack of standardized methods. Our findings underscore the importance of differentiating implant texture types, particularly for assessing the risk of BIA-ALCL. In addition, the adaptability of the deep learning model to account for imaging device variations and navigate prediction uncertainties opens promising avenues for robust AI-driven clinical decision support in evaluating and managing breast implants.

References

1. Atlan M, Nuti G, Wang H, Decker S, Perry TA. Breast implant surface texture impacts host tissue response. J Mech Behav Biomed Mater. 2018;88:377–85.

2. Barnsley GP, Sigurdson LJ, Barnsley SE. Textured surface breast implants in the prevention of capsular contracture among breast augmentation patients: a meta-analysis of randomized controlled trials. Plast Reconstr Surg [Internet]. 2006 [cited 2024 Mar 6];117:2182–90. Available from: https://pubmed.ncbi.nlm.nih.gov/16772915/

3. Keech JA, Creech BJ. Anaplastic T-cell lymphoma in proximity to a saline-filled breast implant. Plast Reconstr Surg [Internet]. 1997 [cited 2024 Mar 6];100:554–5. Available from: https://pubmed.ncbi.nlm.nih.gov/9252643/

4. Medical Device Reports of Breast Implant-Associated Anaplastic Large Cell Lymphoma | FDA [Internet]. [cited 2024 Mar 6]. Available from: https://www.fda.gov/medical-devices/breast-implants/medical-device-reports-breast-implant-associated-anaplastic-large-cell-lymphoma

5. Blombery P, Thompson ER, Prince HM. Molecular Drivers of Breast Implant-Associated Anaplastic Large Cell Lymphoma. Plast Reconstr Surg [Internet]. 2019 [cited 2024 Mar 6];143:59S-64S. Available from: https://pubmed.ncbi.nlm.nih.gov/30817557/

6. Clemens MW, Jacobsen ED, Horwitz SM. 2019 NCCN Consensus Guidelines on the Diagnosis and Treatment of Breast Implant-Associated Anaplastic Large Cell Lymphoma (BIA-ALCL). Aesthet Surg J [Internet]. 2019 [cited 2024 Mar 6];39:S3–13. Available from: https://pubmed.ncbi.nlm.nih.gov/30715173/

7. Parthasarathy M, Orrell J, Mortimer C, Ball L. Case Report: Chemotherapy-resistant breast implant-associated anaplastic large cell lymphoma. BMJ Case Rep [Internet]. 2013 [cited 2024 Mar 6];2013. Available from: /pmc/articles/PMC3847595/

8. Brody GS, Deapen D, Taylor CR, Pinter-Brown L, House-Lightner SR, Andersen JS, et al. Anaplastic large cell lymphoma occurring in women with breast implants: analysis of 173 cases. Plast Reconstr Surg [Internet]. 2015 [cited 2024 Mar 6];135:695–705. Available from: https://pubmed.ncbi.nlm.nih.gov/25490535/

9. Doren EL, Miranda RN, Selber JC, Garvey PB, Liu J, Medeiros LJ, et al. U.S. epidemiology of breast implant-associated anaplastic large cell lymphoma. Plast Reconstr Surg [Internet]. 2017 [cited 2024 Mar 6];139:1042–50. Available from: https://journals.lww.com/plasreconsurg/fulltext/2017/05000/u_s__epidemiology_of_breast_implant_associated.3.aspx

10. Tripodi D, Amabile MI, Varanese M, D'Andrea V, Sorrenti S, Cannistrà C. Large cell anaplastic lymphoma associated with breast implant: a rare case report presentation and discussion of possible management. Gland Surg [Internet]. 2021 [cited 2024 Mar 6];10:2076–80. Available from: https://pubmed.ncbi.nlm.nih.gov/34268093/

11. Marra A, Viale G, Pileri SA, Pravettoni G, Viale G, De Lorenzi F, et al. Breast implant-associated anaplastic large cell lymphoma: A comprehensive review. Cancer Treat Rev [Internet]. 2020 [cited 2024 Mar 6];84. Available from: https://pubmed.ncbi.nlm.nih.gov/31958739/

12. Fabio Santanelli Di Pompeo, Demosthenes Panagiotakos, Guido Firmani, Michail Sorotos, *Aesthetic Surgery Journal*, Volume 43, Issue 5, May 2023, Page 621, https://doi.org/10.1093/asj/sjad036

13. Seong Ji K. Loss of Data from 13,000 Patients Receiving a Breast Implant. HiTNew. 2021.

14. Nam SE, Bang BS, Lee EK, Sung JY, Song KY, Yoo YB, et al. Use of High-Resolution Ultrasound in Characterizing a Breast Implant and Detecting a Rupture of the Device. Plast Reconstr Surg [Internet]. 2023 [cited 2024 Mar 6];152:39–43. Available from: https://pubmed.ncbi.nlm.nih.gov/36688630/

15. Kim YH, Park DW, Song KY, Lim HG, Jeong JP, Kim JH. Use of High-Resolution Ultrasound in Characterizing the Surface Topography of a Breast Implant. Medicina (Lithuania). 2023;59.

16. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition [Internet]. 2015 [cited 2024 Mar 6];2016-December:770–8. Available from: https://arxiv.org/abs/1512.03385v1

17. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin Transformer: Hierarchical Vision

Transformer using Shifted Windows. Proceedings of the IEEE International Conference on Computer Vision [Internet]. 2021 [cited 2024 Mar 6];9992–10002. Available from: https://arxiv.org/abs/2103.14030v2

18. 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. ICLR 2021 - 9th International Conference on Learning Representations [Internet]. 2020 [cited 2024 Mar 6]; Available from: https://arxiv.org/abs/2010.11929v2

19. Kim JH, Hong J, Choi H, Kang HG, Yoon S, Hwang JY, et al. Development of Deep Ensembles to Screen for Autism and Symptom Severity Using Retinal Photographs. JAMA Netw Open [Internet]. 2023 [cited 2024 Mar 6];6:e2347692–e2347692. Available from: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2812964

20. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. Int J Comput Vis [Internet]. 2016 [cited 2024 Mar 6];128:336–59. Available from: http://arxiv.org/abs/1610.02391

21. Shannon CE. A Mathematical Theory of Communication. Bell System Technical Journal. 1948;27:379–423.

22. Fabio Santanelli di Pompeo, Guido Paolini, Guido Firmani, Michail Sorotos History of breast implants: Back to the future JPRAS Open. 2022 Mar 11:32:166-177.
doi: 10.1016/j.jpra.2022.02.004.

23. Lötsch J, Kringel D, Ultsch A. Explainable Artificial Intelligence (XAI) in Biomedicine: Making AI Decisions Trustworthy for Physicians and Patients. BioMedInformatics 2022, Vol 2, Pages 1-17 [Internet]. 2021 [cited 2024 Mar 7];2:1–17. Available from: https://www.mdpi.com/2673-7426/2/1/1/htm

24. Juang WC, Hsu MH, Cai ZX, Chen CM. Developing an AI-assisted clinical decision support system to enhance in-patient holistic health care. PLoS One [Internet]. 2022 [cited 2024 Mar 7];17. Available from: https://pubmed.ncbi.nlm.nih.gov/36315554/

25. Yu AC, Mohajer B, Eng J. External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. Radiol Artif Intell. 2022;4.

26. Wang H, Wang Z, Mardziel P, Du M, Yang F, Hu X, et al. Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks [Internet]. Available from: https://github.com/haofanwang/Score-CAM

Supplementary Files

Multimedia Appendixes

Demographic characteristics of 1,043 patient.
URL: http://asset.jmir.pub/assets/c7d259f45758ce8f46ae2dd5be9ba270.docx

Model performance in Canon dataset (D1) using 5-fold stratified cross-validation.
URL: http://asset.jmir.pub/assets/2b19cc3ce662d2e9c64bd01bbbf29849.docx

Example images of texture-type breast implants. A) Allergan Texture 5, B) Bellagle macro Texture 1, C) Eurosilicone Texture 1, D) Polytech Texture 2, E) Sebbin Texture 1, and F) Silimed texture 3.
URL: http://asset.jmir.pub/assets/db61a342015e12442a42cb9425809adf.docx