Augmenting Tele-Postpartum Care with Vision-Based Detection of Breastfeeding-related Conditions: Algorithm Development and Validation

Jessica De Souza, Varun Kumar Viswanath, Jessica Maria Echterhoff, Kristina Chamberlain, Edward Jay Wang

Submitted to: JMIR AI
on: November 22, 2023

Table of Contents

Augmenting Tele-Postpartum Care with Vision-Based Detection of Breastfeeding-related Conditions: Algorithm Development and Validation

Jessica De Souza[1] MSc; Varun Kumar Viswanath[1] MSc; Jessica Maria Echterhoff[2] MSc; Kristina Chamberlain[3] CNM, ARNP, IBCLC; Edward Jay Wang[1] PhD

[1]Department of Electrical and Computer Engineering University of California, San Diego La Jolla US
[2]Department of Computer Science and Engineering University of California, San Diego La Jolla US
[3]Division of Extended Studies University of California, San Diego La Jolla US

Corresponding Author:
Jessica De Souza MSc
Department of Electrical and Computer Engineering
University of California, San Diego
9500 Gilman Drive
La Jolla
US

Abstract

Background: Breastfeeding benefits both mother and infant and is a topic of attention in public health. After childbirth, untreated medical conditions or lack of support lead many mothers to discontinue breastfeeding. For instance, nipple damage and mastitis affect 80% and 20% of US mothers, respectively. Lactation Consultants (LCs) help mothers with breastfeeding, providing in-person, remote, and hybrid lactation support. LCs guide, encourage and find ways for mothers to have a better experience breastfeeding. Current telehealth services help mothers seek LCs for breastfeeding support, where images help them identify and address many issues. Due to the disproportional ratio of LCs and mothers in need, these professionals are often overloaded and burned out.

Objective: We investigate the effectiveness of five distinct convolution neural networks (CNNs) in detecting healthy lactating breasts and six breastfeeding-related issues by only using RGB images. Our goal is to assess the applicability of this algorithm as an auxiliary resource for LCs to identify painful breast conditions quickly, better manage their patients through triage, respond promptly to patient needs, and enhance the overall experience and care for breastfeeding mothers.

Methods: We evaluate the potential for five classification models to detect breastfeeding-related conditions using 1,078 breast and nipple images gathered from online and physical educational resources. We used the CNNs Resnet50, VGG16, InceptionV3, EfficientNetV2, and DenseNet169 to classify the images across seven classes: healthy, abscess, mastitis, nipple blebs, dermatosis, engorgement, and nipple damage by improper feeding or misuse of breast pumps. We also evaluate the models' ability to identify between healthy and unhealthy images. We present an analysis of the classification challenges, identifying image traits that may confound the detection model.

Results: The best model achieves an average area under the ROC curve (AUC) of 0.93 for all conditions after data augmentation for multi-class classification. For binary classification, we achieved with the best model an average AUC of 0.96 for all conditions after data. Several factors contributed to the misclassification of images, including (1) similar visual features in the conditions that precede other conditions (such as the mastitis spectrum disorder), (2) partially covered breasts and/or nipples, and (3) images depicting multiple conditions in the same breast.

Conclusions: This vision-based automated detection technique offers an opportunity to enhance postpartum care for mothers and can potentially help alleviate the workload of LCs by expediting decision-making processes.

(JMIR Preprints 22/11/2023:54798)
DOI: https://doi.org/10.2196/preprints.54798

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

Original Paper

Augmenting Tele-Postpartum Care with Vision-Based Detection of Breastfeeding-related Conditions: Algorithm Development and Validation

Jessica de Souza, MSc; Varun Kumar Viswanath, MSc; Jessica Maria Echterhoff, MSc; Kristina Chamberlain, CNM, ARNP, IBCLC; Edward Jay Wang, PhD

Corresponding Author:
Jessica de Souza, MSc
Department of Electrical and Computer Engineering
University of California, San Diego
9500 Gilman Dr
La Jolla, CA, 92093
United States
Phone: (858) 534-7013
Email: jdesouza@ucsd.edu

Abstract

Background: Breastfeeding benefits both mother and infant and is a topic of attention in public health. After childbirth, untreated medical conditions or lack of support lead many mothers to discontinue breastfeeding. For instance, nipple damage and mastitis affect 80% and 20% of US mothers, respectively. Lactation Consultants (LCs) help mothers with breastfeeding, providing in-person, remote, and hybrid lactation support. LCs guide, encourage and find ways for mothers to have a better experience breastfeeding. Current telehealth services help mothers seek LCs for breastfeeding support, where images help them identify and address many issues. Due to the disproportional ratio of LCs and mothers in need, these professionals are often overloaded and burned out.
Objective: We investigate the effectiveness of five distinct convolution neural networks (CNNs) in detecting healthy lactating breasts and six breastfeeding-related issues by only using RGB images. Our goal is to assess the applicability of this algorithm as an auxiliary resource for LCs to identify painful breast conditions quickly, better manage their patients through triage, respond promptly to patient needs, and enhance the overall experience and care for breastfeeding mothers.
Methods: We evaluate the potential for five classification models to detect breastfeeding-related conditions using 1,078 breast and nipple images gathered from online and physical educational resources. We used the CNNs Resnet50, VGG16, InceptionV3, EfficientNetV2, and DenseNet169 to classify the images across seven classes: healthy, abscess, mastitis, nipple blebs, dermatosis, engorgement, and nipple damage by improper feeding or misuse of breast pumps. We also evaluate the models' ability to identify between healthy and unhealthy images. We present an analysis of the classification challenges, identifying image traits that may confound the detection model.
Results: The best model achieves an average area under the ROC curve (AUC) of 0.93 for all conditions after data augmentation for multi-class classification. For binary classification, we achieved with the best model an average AUC of 0.96 for all conditions after data. Several factors contributed to the misclassification of images, including (1) similar visual features in

the conditions that precede other conditions (such as the mastitis spectrum disorder), (2) partially covered breasts and/or nipples, and (3) images depicting multiple conditions in the same breast.
Conclusions: This vision-based automated detection technique offers an opportunity to enhance postpartum care for mothers and can potentially help alleviate the workload of LCs by expediting decision-making processes.

Keywords: Remote consultations; AI for healthcare; deep-learning; detection model; breastfeeding; telehealth; perinatal health; image analysis; women's health; artificial intelligence

Introduction

Background

The benefits of breastfeeding for both mother and baby, such as lower gastrointestinal infections in the child, more rapid maternal weight normalization after birth, and prolonged amenorrhea for the mother, are just a few examples of why physicians recommend breastfeeding for at least six months [1-5]. Breastfeeding rates are on the rise in the US, with 83.2% of newborn infants being breastfed in 2019, thanks to increased education and promotion of its benefits [6]. Despite the compelling evidence, many families struggle to continue breastfeeding. Although 95% of mothers initiate breastfeeding, the continuation rate drops to below 41% and 19% for exclusive breastfeeding at three and six months, respectively [7]. Parents who breastfeed may face issues such as low milk supply, fatigue, medical problems, difficulties with feeding techniques or pain, and lack of social support [8-10].

Lactation Consultant (LC) professionals specialize in breastfeeding, milk supply, breast and nipple issues, breast milk management, and prenatal education. LCs ensure a mother's smooth and painless transition into breastfeeding and increase the possibility of continued breastfeeding through six months or longer [11,12]. The availability of International Board-Certified Lactation Consultants (IBCLCs) globally is limited. In 2021, there were 3.6 million births in the U.S. and only 18.5 thousand LCs with IBCLC certification, a rate of 194 babies per LC a year. In low and middle-income countries (LMICs) like Brazil, for instance, there were 2.6 million births in the same year but only 154 certified LCs, resulting in a rate of 16,883 babies per LC per year. The high demand for LCs, coupled with geographic and financial barriers, underscores the need for better tools to improve access to specialized lactation services, especially in less urbanized areas where such resources are scarce, leading to decreased breastfeeding support [13-18, 20].

Another issue is professional availability itself, since LCs often combine their practice with midwife nursing, splitting their time between prenatal visits, attending births, lactation consultations, and managing their patients, which can lead to professional exhaustion, burnout, and emotional stress [21 - 23]. Moreover, the predominantly independent practice of LCs outside the U.S., without the support of clinics with sophisticated patient management and triage systems, further complicates their time management and patient organization [22,24].

Supporting Lactation Consultants through Tele-Lactation Services

Tele-lactation services facilitate text, audio, and video communication. This enables LCs to consult with patients from any location, reduces travel time, helps balance their workload, increases their availability to receive new patients, and provides quicker responses to their patients [20]. Complementing tele-lactation services, patient triaging using information systems allow LCs to prioritize in-person visits for severe cases requiring physical assessment, while less critical cases can be handled remotely [25,26]. Prior research suggests that LCs would benefit from time-saving tools for efficient patient information delivery while focusing on mitigating prolonged interactions, helping alleviate the burden on these professionals with a load of patients [27,22]. As LCs often follow up with their patients up to weeks after birth to ensure positive breastfeeding outcomes, an easy-to-access system to monitor patient progress is essential for effective patient triage, facilitating consultation scheduling, holding remote consultations, or providing reassurance. However, LC's current access to remote consultation systems lacks patient triaging tools and is not time-efficient, indicating an area in need of development.

Our work proposes a novel method for the identification of breastfeeding-related conditions using Convolutional Neural Networks (CNNs). We evaluated a self-curated dataset containing seven different breastfeeding conditions on five distinct CNN models. The assessment of breast conditions is vital as pain and discomfort experienced during breastfeeding is a major barrier faced by parents who want to continue to provide their child with breastfeeding. About 80% of mothers are estimated to suffer from nipple pain and fissures, while 20% are estimated to experience mastitis [28,29]. Our pipeline incorporates automatic detection of visually discernible painful breastfeeding-related conditions such as nipple cracks and fissures related to poor latching and positioning, skin conditions such as dermatitis, eczema, thrush, or herpes, and risk of mastitis spectrum issues such as engorgement, abscess, and nipple blebs. The CNN model is used for automatic detection of breast conditions, which can benefit the triaging of remote lactation patients for faster and more efficient patient response based on their conditions.

Our work evaluates five distinct CNN models' ability to differentiate between healthy and various unhealthy breast conditions (including breast abscesses, dermatoses, engorgement, mastitis, nipple blebs, and nipple damage) by performing both multi-class and binary evaluations on 1078 breast images. We evaluate the model's performance using the dataset with and without data augmentation techniques. The data was divided into training, validation, and testing sets, employing k-fold cross-validation for robustness. Performance evaluation on the best model includes an average AUC of 0.93 for all conditions after data augmentation and precise detection of healthy breasts (84.4%) and unhealthy breasts (average precision of 66%, SD = 12.8%) for six conditions. For binary classification, we achieved with the best model an average AUC of 0.96 for all conditions after data augmentation and precise detection of healthy breasts (93.8%) and unhealthy breasts (average precision of 83.5%). The breast images have been curated from perinatal education resources such as images and video recordings under various lighting, environments, and image-taking conditions, where we examine potential issues around how the images are taken and their impacts on performance. Finally, we provide insights into future designs of user interfaces and guidance needed for the proper application of the system.

Related Work

Lactating Care Pipeline: In-person, remote, and hybrid

Healthcare providers introduce breastfeeding options to expectant mothers, including educational materials in print or online, during prenatal care. The initiation of breastfeeding post-delivery is timed according to the type of birth. Many hospitals worldwide follow the UNICEF/WHO baby-friendly initiative, prioritizing maternal and infant health and supporting mothers facing challenges [30, 31]. After a child's birth, families often seek breastfeeding support with LCs, who typically offer hands-on consultations from birth until support is no longer required [18]. They conduct visual and physical evaluations of both mother and baby, assessing the baby's internal mouth structure, breast and nipple anatomy, milk supply and ensuring proper attachment or repositioning of the baby to prevent nipple fissures. LCs may also introduce laser therapy as a treatment option for damaged nipples from breast pump misuse or issues with baby attachment [8]. The immersive approach of LCs is crucial for providing personalized and effective lactation support to mothers and infants.

Remote lactation care

The widespread adoption of smartphone communication apps, particularly WhatsApp, has transformed public health facilities, including family clinics in developing countries, offering various patient services such as appointment scheduling, health guidance, and vaccine campaign notifications [32 - 34]. WhatsApp has become a popular communication tool between LCs and patients, facilitating breastfeeding education and family support during the neonatal period [35, 36]. During the COVID-19 pandemic, LCs transitioned to virtual consultations using established smartphone applications like WhatsApp, Instagram, and Facebook. LCs adapted their approach to maintain quality care despite resource limitations in remote consultations [37, 38]. Like other practices requiring physical evaluation, LCs reimagined their methods when shifting from in-person to remote consultations, using communication and social media apps to reach and educate parents while having broader visibility in their community [37, 39].

Remote lactation care presents challenges, including limited visibility during video calls, communication difficulties, and technical issues [18, 40, 41]. Despite challenges, remote care offers benefits, reducing the mother's sense of isolation, enabling faster feedback, and promoting effective communication and patient engagement for improved independent learning [17, 18, 22]. These benefits positively impact mothers' intentions in exclusive breastfeeding for up to six months and reduce the risk of breastfeeding cessation at three months by 25% [42].

Hybrid lactation care

Previous research showed that fully remote consultations work well for cases where geographic distance, transportation issues, or patient disease prevent in-person meetings between patients and providers. LCs often conduct remote consultations from their workplaces, including personal offices, clinics, or hospitals, especially when they are also midwives with on-call responsibilities [37]. They provide consultations for patients who seek them pre-birth, post-birth, and in emergency cases where the mother is facing breastfeeding challenges [22]. Depending on the nature of the consultation, in-person or remote visits are chosen to meet the patient's specific needs. In summary, remote care complements in-person

care, being a valuable resource for mothers seeking guidance, reassurance, and confidence, particularly in the absence of a supportive home environment [38].
LCs, especially those who are also midwives, have limited time availability due to demanding schedules and receiving numerous remote messages from patients daily, some requiring higher priority attention [22, 43]. Manually sorting through patient messages to determine priority can be time-consuming and inconvenient for mothers with urgent needs. Our work proposes a computer vision-based system to triage breast conditions, facilitating telehealth and assisting LCs in identifying patients who require immediate responses in remote settings.

Issues associated with breastfeeding

Breastfeeding pain is one of the reasons associated with breastfeeding cessation, which can be caused by issues such as poor attachment of the baby onto the breast, physical conditions of the mother or baby, misuse of breast pumps, oversupply of breast milk, and even environmental conditions [44]. These issues, if left untreated in the first few days after birth, can persist for weeks and pose a threat to breastfeeding continuity beyond six months. Some conditions can be fully mitigated when the mother receives orientation and education on the topic. In contrast, other conditions can be alleviated and managed for a better experience for the mother in the case of physical conditions, including nipple physiology, baby tongue-tie, jaw clenching, and excessive milk supply [28, 45].
This work concentrates on conditions leading to breastfeeding pain and potential interruption. The first condition is the mastitis spectrum disorder, where about 20% of mothers who breastfeed may face it during their time breastfeeding. This disorder starts with the overproduction of milk and/or breast engorgement, which can cause milk passage obstruction in the form of galactoceles and nipple blebs. When not properly treated, a case of milk bleb or galactocele can evolve into phlegmon, bacterial, or inflammatory mastitis, which may require patients to treat it with medications and sometimes medical procedures to drain the inflammation fluids from the breast in case it becomes an abscess [46,47]. Conditions associated with mastitis are painful and include symptoms such as redness in the breast, flu-like symptoms, hardened skin surface in the location of the milk blockage, formation of blisters in the nipple, and even blood in the milk [29,48].
The second condition is nipple damage caused by improper latching and positioning from the infant, excessive pressure from breast pumping devices, infant tongue-tie or palate abnormality, infant's arrhythmic milk expression, and even infant biting or jaw clenching [9, 44]. Considering the cause of nipple damage, 80% of mothers are expected to face some level of nipple issues during breastfeeding, which, if not treated, may cause an average of 35% of these mothers to cease breastfeeding before one month [28, 45]. Nipple damage is painful and may be visible or invisible. When visible, it can present features at the skin surface, such as fissures, cracks, pus, blood, scarring, or crusting. Some skin dermatoses such as thrush, herpes, eczema, and psoriasis are also responsible for discomfort and pain during breastfeeding. These conditions can be caused by friction, weather, and temperature changes, and using medications or ingredients that can make the skin prone to these disorders. Dermatoses conditions present on both breast and nipple and can have visible features such as scarring, crusting formations, redness, and thickened skin regions [44]. Our work incorporates breast and nipple images from the following disorders: breast abscess, dermatoses, breast engorgement, inflammatory and bacterial mastitis, nipple blebs, and nipple damage.

Current research supporting lactating mothers

Extensive literature has highlighted the efficacy of deep learning in assessing breast images, helping detect malignant and benign breast tumors for both lactating and non-lactating women [49 - 54]. This has helped improve the precision of breast ultrasound and mammogram examinations, involving the use of medical imaging previously taken in medical facilities to enhance the evaluation of breast-related illnesses and allow better accuracy in diagnosis for medical personnel [53]. Nevertheless, these studies relied on images gathered from specialized equipment found only in healthcare facilities, not extending their evaluation on external body images, making their main focus on helping healthcare practitioners in diagnosis. Our work diverges from previous contributions by primarily focusing on using external breast images gathered from personal devices, such as smartphones or cameras from lactating patients, to identify breastfeeding-related conditions in the early stages and evaluate the necessity of further examination and medical intervention.

In the context of breastfeeding disorders, there is a lack of research regarding using deep learning algorithms to evaluate real breast images and identify abnormalities such as mastitis, nipple fissures, dermatoses, and abscesses. To illustrate, literature addressing the early prediction of mastitis mainly originates from agricultural studies, in which the risk of mastitis is constantly assessed to prevent a reduction in animal milk production, which significantly impacts the dairy industry [55, 56]. This shows a need for research to adapt these technologies for detecting and preventing breastfeeding disorders in humans. Our study is crucial in settings where access to medical professionals and LCs is limited, as it can help prevent breastfeeding cessation, promote maternal-infant bonding, and improve the overall health and well-being of mothers and infants.

Materials and Methods

In this section, we detail the dataset collection process, including inclusion and exclusion criteria, data sources, and the characteristics of the images. The chapter will also discuss the AI algorithms used in the study, including the models and their training and validation process, and performance metrics used during evaluation.

Dataset Collection

This study uses a breast image dataset (see Table 1), a compilation of physical and digital images specifically curated to train and validate our deep learning model's ability to distinguish between healthy and unhealthy lactating breasts. The dataset includes images categorized according to their respective conditions: healthy lactating breast, nipple injuries due to various causes, nipple blebs due to plugged ducts, breast or nipple with signs of dermatoses, and breasts in the conditions of engorgement, mastitis, and abscess.

Data Inclusion/Exclusion Criteria: To be included in the dataset, images must meet the following criteria: (1) the image must be in RGB format, either as PNG or JPEG; (2) it must visually have at least one of the seven conditions; (3) the breast and/or nipple should be visible; (4) the image should be hosted in a trustworthy source (from medical professionals such as physicians, midwife nurses, and IBCLCs), in which the image must have a word or description identifying its condition among the seven classes to be included as its label; (5) the visual condition present in the image and the label provided describing the condition should match. Images were excluded from the dataset if (1) the breast or nipple were from non-lactating female patients, (2) if the condition described on the label and the visual features of the image did not match, (3) the breast or nipple was not visible in the image, and (4) if the image did not have any label describing it. A board-certified nurse practitioner (CNM, ARNP, IBCLC) with more than 15 years of experience performed a final review of the dataset to ensure that images and labels had no discrepancies.

Data Source: We collected images from diverse sources such as breastfeeding-related books, articles, online blogs for mothers and physicians, YouTube from educative organizations, and social media platforms (e.g., Instagram, Facebook, Twitter) from certified healthcare providers who would have educative resources for mothers. To ensure diversity in geographical and racial representation, we conducted image searches using multiple languages (e.g., English, Portuguese, Spanish, French, Chinese) and used search engines adjusted for other countries.

Table 1. Dataset description.

Dataset size	393.7 MB, each image: [min: 0.015, average: 0.360, max: 3.575] MB
Dimensions	Width [min: 68, average: 606, max: 2448] Height [min: 68, average: 607, max: 2448]
Number of images	1078
Number of classes	7
Number of unique subjects	586
Number of Images per class	**Abscess:** 115 **Dermatoses:** 123 **Engorgement:** 63 **Mastitis:** 180 **Nipple bleb:** 82 **Nipple damage:** 197 **Healthy:** 318
Visual features per class	**Abscess:** swelling and redness, area with palpable fluid collection, pus **Dermatoses:** rash, discoloration, flaky skin, uneven skin tone, crusting, redness **Engorgement:** swelling, redness, skin stretched and shiny, enlarged nipple **Mastitis:** red patches on breast and/or nipple, swelling, pus, or blood discharge **Nipple bleb:** small white or yellow bumps on nipple or areola, similar to a blister **Nipple damage:** nipple swelling, redness, peeling/flaking skin, bleeding, shape differences **Healthy:** regular breast and nipple color, may have visible veins
Number of images per source	**Physical:** 178 (books, magazines, articles) **Physician Websites:** 366 **YouTube:** 65 (educational channels on women's health) **Other:** 469 (received by LCs, IBCLC's Instagram, Google Images, and Flickr, support groups mediated by LCs on social media and other educational websites)
Number of images per skin tone per class (Fitzpatrick skin type (FST) [57])	(see below)

Class Name	FST I	FST II	FST III	FST IV	FST V	FST VI	N/A*
Abscess	28	35	20	8	14	8	2
Dermatoses	17	37	48	13	3	3	2
Engorgement	4	6	18	30	4	0	1
Mastitis	44	69	51	11	1	4	0
Nipple bleb	9	16	18	8	6	3	22
Nipple damage	40	59	22	15	11	5	45
Healthy	61	90	92	21	28	21	5

Total per FST	**203**	**312**	**269**	**106**	**67**	**44**	**77**

* N/A: Not classified due to the absence of breast tissue around the nipple in the image.

The images come from a diverse group of female patients with several skin colors, breast and nipple sizes, with unstandardized image sizes, orientations, backgrounds, and light sources. In total, the dataset consists of 1078 images, with 318 images of healthy breasts, 115 images of breast abscesses, 123 images of dermatoses, 63 images of breast engorgement, 180 images of mastitis, 82 images of nipple blebs, and 197 images of nipple damage. As shown in Figure 8 (G) and Table 1, a healthy lactating breast presented a uniform color, was free of redness, and had no signs of discharge. Nipples are expected to exhibit a variety of shapes, including flat, protruded, or inverted, and varying in sizes. In engorgement, images showed breast and nipple swelling, skin stretched and shiny, and some light redness due to high milk production. For nipple blebs or nipple damage, signs of laceration, blood, blisters, and redness are expected. Mastitis showed swelling, redness, and discharge of pus or blood in the nipple. Abscess shares similarities with mastitis but involves worsened redness and pus in the infected region and may display signs of rupture. Finally, dermatosis images contained signs of skin rash, breast or nipple uneven skin tone, and crusting.

AI Algorithms

We examined the performance of five CNNs commonly used in computer-vision problems: VGG16[58], Resnet50 [59], InceptionV3 [60], EfficientNetV2 [61], and DenseNet169 [62]. All models were built with the PyTorch library for image classification, in which the models had all layers frozen except for the last layer, which was replaced with a fully connected layer adapted to the number of classes—two for binary classification and seven for the multi-class task. All models were trained for 100 epochs using the AdamW optimizer with a learning rate of 3e-4, weight decay of 0.1, and batch size of 20. We chose 100 epochs since it was a converging point where the accuracy no longer increased or decreased. For the loss functions, we applied Binary Cross-Entropy with Logits Loss for binary classification tasks, and for multi-class tasks we used Cross-Entropy Loss, both fine-tuned with class weights to strategically adjust for class imbalances by proportionally penalizing misclassifications in less represented classes. These models were evaluated using stratified K-Fold cross-validation with 10 folds. To ensure the robustness of our cross-validation process, we reset any learned parameters by initializing the models from scratch at the beginning of each fold. Instead of using the entire image dataset to train the model, we did feature extraction to optimize the training process (detailed in the Feature Extraction section). We compare the performance of the five models across the same data and keep the hyperparameters the same: learning rate, weight decay, batch size, and number of epochs.

Dataset Preprocessing

Before using the images as inputs for the deep-learning models, the images were manually cropped to ensure images were deidentified and had no irrelevant content, such as unrelated body areas, clothes, jewelry, identifiable tattoos, or backgrounds, enhancing the model's accuracy and performance. The images were cropped in a 1:1 ratio to prevent image flattening or warping during resizing and loss of important features. Most images have breast and nipple tissue concentrated in the center of the image, thereby focusing the model's evaluation on the most relevant areas. Our image preprocessing guidelines follow similar works in dermatology for AI disease detection and telehealth applications [63-65], which aims to objectively show the area of interest for optimized detection and reduce risks of poorly triaged images.

After cropping the images in a 1:1 ratio and before entering the deep learning pipeline, we apply some standard transformations in the data, starting with image resizing. In this paper we train, validate, and test our dataset using five different models. Four of the chosen models (VGG16, Resnet50, EfficientNetV2, and DenseNet169) specify the input images to be resized to 224x224 pixels, and the model InceptionV3 model requires input images to be resized to 299x299 pixels. Therefore, we proceeded with the image resizing according to each model's requirements. The last transformation step incorporates normalization of the images, a procedure where the pixel intensity values are standardized across the dataset. To help the models generalize better for our dataset, we calculated the mean and standard deviation of all images in the dataset to use in the normalization process instead of using the ImageNet dataset pretrained parameters, inspired by the previous work involving skin disease classification [66].

Dataset Augmentation

In the process of curating the dataset, we recognized that the number of images per class was constrained, given the complexity of gathering images and variability in the clinical features of each class. We implemented data augmentation techniques to mitigate these limitations, reduce the risk of overfitting, and enrich the dataset. These techniques artificially expanded the dataset by generating realistic transformations of the existing images. We implemented the following six data augmentations that were previously used in datasets involving skin lesions [63, 67]: center zoom, random rotation, brightness, shear, vertical and horizontal flip. Samples of augmentation are shown in Figure 1. Before data augmentation our dataset consisted of 1078 images. After the augmentation, the dataset consists of 6478 images. The detailed number of samples before and after augmentation is shown in Table 2.

Figure 1. Samples of Augmented Data: (a) Original (b) Brightness (c) Center zoom (d) Horizontal flip (e) Rotation (f) Shear and (g) Vertical Flip.

We evaluate our dataset before and after data augmentation. In the original dataset, the 1000 images are allocated for training and validation, split using stratified k-fold cross-validation [68] with 10 folds. In this process, 90% of the data are used for training and 10% for validation within each fold, as described in Figure 2. The stratified k-fold maintains the proportion of images in each class in both train and validation splits, making sure each fold will be representative of the overall dataset. The remaining 78 images are completely excluded from these folds and reserved exclusively for final testing to assess the model's performance on unseen data. After augmenting the original dataset, we expanded it to 6000 images for training and validation. Similarly, we increased our test set to 468 images to maintain consistency with the expanded training data while ensuring the model's evaluation on unseen examples remains with good performance.

Figure 2. Graphical diagram of Stratified K-fold Cross Validation on a 7-class dataset.

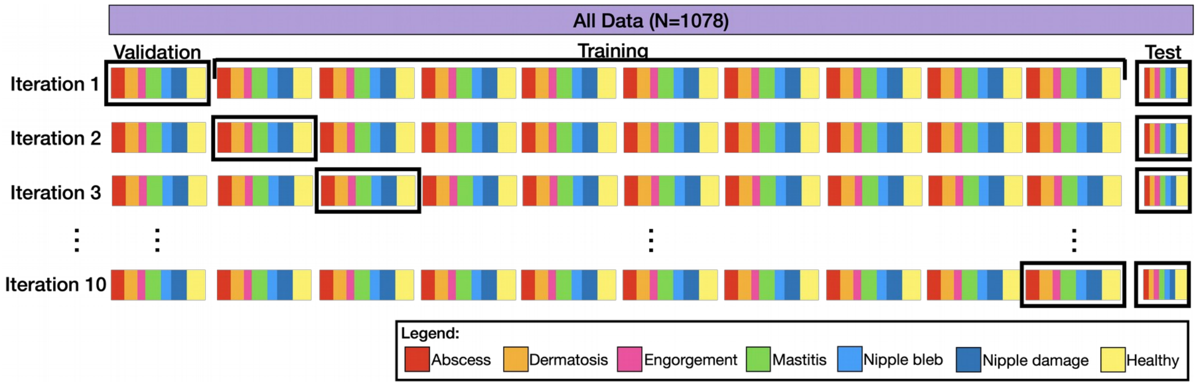


Table 2. Detailed number of samples on the dataset.

Dataset	Classes	Train Samples	Test Samples	Train Samples (Augmented)	Test Samples (Augmented)
7-class	Abscess	108	7	648	42
	Dermatoses	115	8	690	48
	Engorgement	55	8	330	48
	Mastitis	171	9	1026	54
	Nipple bleb	75	7	450	42
	Nipple damage	188	9	1128	54
	Healthy	288	30	1728	180
Binary	Unhealthy*	657	40	3942	240
	Healthy*	343	38	2058	228

* Unhealthy class combines the classes abscess, dermatoses, mastitis, nipple bleb and nipple damage, while the healthy class combines healthy and engorgement, all from the 7-class dataset.

Feature Extraction

We performed feature extraction using five models pre-trained on the ImageNet dataset. This process helped to reduce the number of computational resources necessary for processing the dataset by transforming images into numerical features, without losing relevant information. The models are set to evaluation mode, in which the feature maps are extracted from the final convolutional layers. These maps were then processed through adaptive pooling, and flattened into one-dimensional arrays. The extracted features were saved and used as input for the model classifiers.

Training and Evaluation

As previously mentioned in the AI Algorithms section, a total of five CNNs were trained on the dataset. We proposed four tasks in this study, which evaluates the CNNs in the following datasets: (1) multi-class not augmented, (2) multi-class augmented, (3) binary not augmented, and (4) binary augmented. As described in Table 2, we perform an additional two evaluations considering a binary model to assess the models' capacity to differentiate healthy from unhealthy images. The unhealthy class consolidates five of the previous conditions: abscess, dermatoses, mastitis, nipple bleb and nipple damage. The healthy class consolidates the original healthy and engorgement conditions. For this binary evaluation, we included engorgement images in the healthy condition because it is not inherently indicative of disease and often resolves without medical intervention. Also, engorgement shares visual characteristics with healthy breast conditions which might not be distinguishable at an early, non-problematic stage. All models underwent k-fold cross-validation, where we collected performance metrics from each fold and computed their average. We assessed the models' performance for the multi-class and binary datasets using the same metrics: accuracy, precision, recall, F1-score, and the Receiver Operating Characteristic Area Under the Curve (ROC-AUC).

Ethics

This study was approved by the University of California, San Diego Institutional Review Board (801904). We did not incorporate any personally identifiable data from the subjects into this research.

Results

We collected 1078 unique breast images from online and physical resources, 1000 images as part of the training and validation set, and 78 images as part of the testing set. The augmented dataset has 6000 images for training and validation and 468 images for testing. Below, we show evaluation results from the multi-class and binary datasets, which we evaluated before and after data augmentation. There was no hyperparameter tuning between each fold, and all models had the same optimizer, learning rate, weight decay, and batch size.

Multi-class Image Detection Evaluation

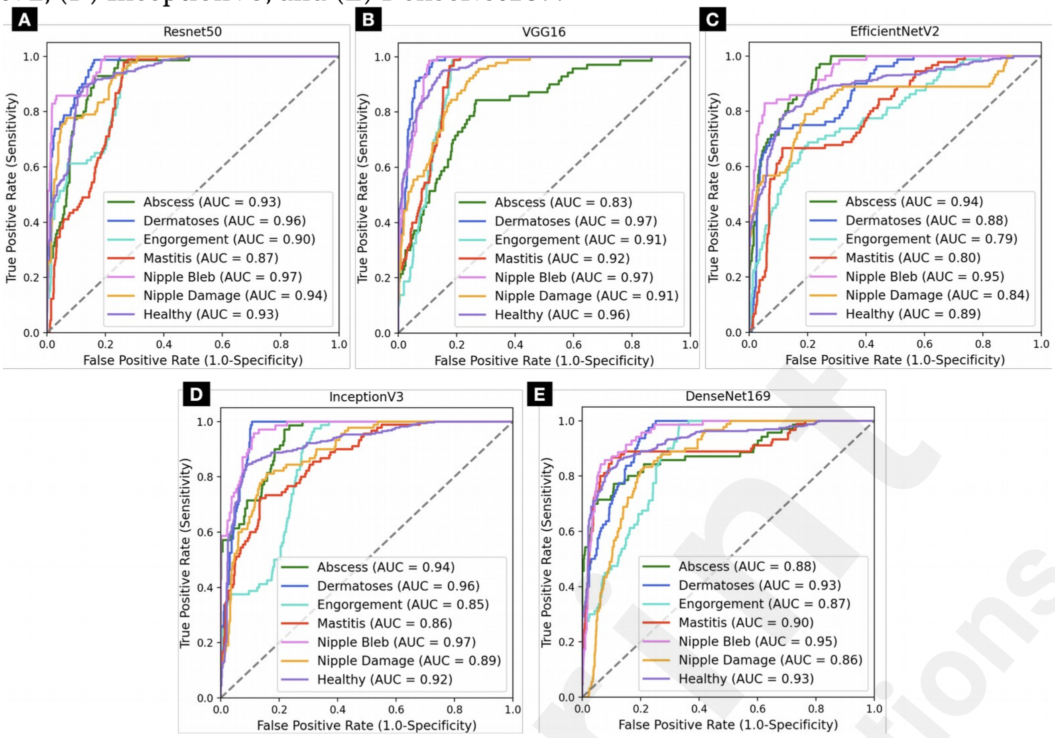
We evaluated five CNNs on their ability to distinguish between healthy and six other breastfeeding-related issues. Table 3 presents the aggregated evaluation metrics for each model sorted based on the test accuracy. The precision, recall, F1-score, and overall area under ROC-AUC are reported as weighted averages to account for the class imbalance within the datasets, ensuring that each class contributes to the final metric in proportion to its prevalence. For each fold in the cross-validation, a separate test set was used to evaluate the model, and the metrics presented are the mean of these evaluations. The best-performing model was the Resnet 50, as it managed to contain the best testing accuracy, followed by VGG16 and EfficientNetV2 on a small performance difference. With a similar weighted average setting, in a one-vs-rest fashion, the models achieved an overall receive operator curve area under the curve (ROC-AUC) of 0.934 for VGG16, 0.929 for Resnet50, 0.912 for InceptionV3, 0.908 for Densenet169, and 0.872 for EfficientNetV2. The detailed ROC-AUC per class for each model is shown in Figure 3.

Table 3. Average evaluation metrics for the trained models on the not augmented dataset (sorted based on performance).

Dataset	Model	Training Accuracy	Validation Accuracy	Test Set Metrics			
				Accuracy	Precision	Recall	F1-score
7-class	Resnet50	0.907	0.737	**0.608**	0.675	**0.623**	**0.637**
	VGG16	0.818	0.678	0.604	0.674	0.589	0.600
	EfficientNetV2	0.779	0.626	0.604	0.658	0.582	0.593
	InceptionV3	0.903	0.727	0.574	**0.680**	0.607	0.622
	DenseNet169	**0.932**	**0.771**	0.507	0.659	0.596	0.572

Figure 3. Performance of the five CNNs on the 7-class dataset: (A) Resnet50, (B) VGG16, (C)

EfficientNetV2, (D) InceptionV3, and (E) DenseNet169.

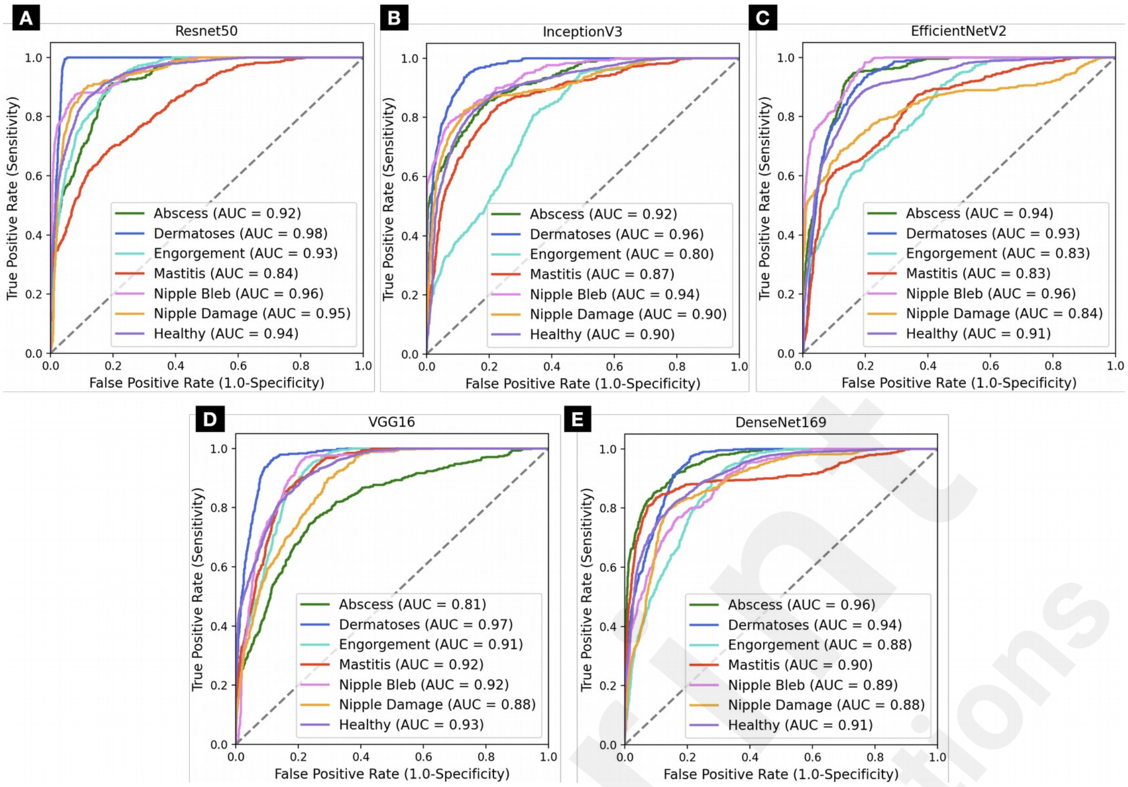


When applying data augmentation to the multi-class model, we provided a wider variety of images to help the model better generalize from the training data while not altering the original class distribution. In Table 4 and Figure 4, we show the results across the CNNs after data augmentation, where most of the models showed improved metrics, with Resnet50 being the leading model. The models achieved a ROC-AUC of 0.934 for Resnet50, 0.912 for VGG16, 0.909 for Densenet169, 0.898 for InceptionV3, and 0.893 for EfficientNetV2.

Table 4. Average evaluation metrics for the trained models on the augmented dataset (sorted based on performance).

Dataset	Model	Training Accuracy	Validation Accuracy	Test Set Metrics			
				Accuracy	Precision	Recall	F1-score
7-class augmented	Resnet50	0.953	**0.907**	**0.672**	**0.717**	**0.715**	**0.713**
	InceptionV3	0.920	0.844	0.617	0.692	0.637	0.649
	EfficientNetV2	0.803	0.808	0.602	0.650	0.586	0.599
	VGG16	0.755	0.801	0.585	0.644	0.561	0.563
	DenseNet169	**0.954**	0.889	0.506	0.639	0.611	0.563

Figure 4. Algorithm performance on the 7-class dataset after data augmentation.

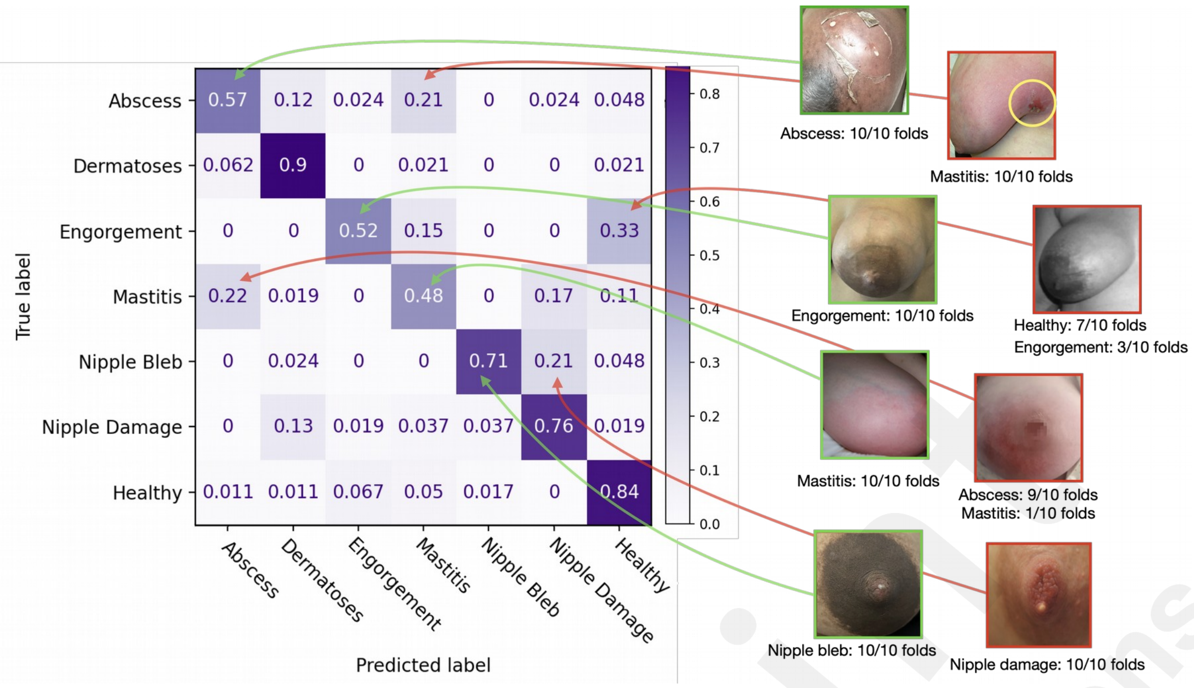


Looking into the performance of the best model, the Resnet50 with the augmented dataset, we can look closer at the metrics per class of this CNN. Table 5 shows the results for 10-fold cross-validation, in which the model had an overall consistent performance across the iterations. Figure 5 presents the aggregated confusion matrix for the Resnet50 model, in which we consolidated the predictions across all ten iterations applied to the augmented dataset. We achieved this aggregation by taking the median predicted class for each instance over the multiple folds, synthesizing a singular prediction representing the consensus of the model's behavior across the test set.

Table 5. Results of 10-fold Cross-Validation for the augmented dataset on Resnet50.

10-Fold Iterations	Accuracy	Precision	Recall	F1-score
Iteration 1	0.699	0.705	0.699	0.699
Iteration 2	0.714	0.715	0.714	0.712
Iteration 3	0.709	0.713	0.709	0.709
Iteration 4	0.729	0.730	0.729	0.727
Iteration 5	0.718	0.719	0.718	0.716
Iteration 6	0.733	0.734	0.733	0.730
Iteration 7	0.720	0.722	0.720	0.718
Iteration 8	0.707	0.711	0.707	0.706
Iteration 9	0.707	0.707	0.707	0.705
Iteration 10	0.720	0.715	0.720	0.713

Figure 5. Aggregated confusion matrix for the Resnet50 model for the augmented dataset with example images from the augmented dataset that were correctly and incorrectly classified across all folders.



Out of the 468 images used in the testing set, the model could correctly classify 341 images. The total images correctly classified by category are as follows: abscess (N = 24, Acc = 57%), dermatoses (N = 43, Acc = 90%), engorgement (N = 25, Acc = 52%), mastitis (N = 26, Acc = 48%), nipple bleb (N = 30, Acc = 71%), nipple damage (N = 41, Acc = 76%), and healthy (N = 152, Acc = 84%). The remaining images that were incorrectly classified happened throughout visually similar conditions and the conditions that can precede each other. Table 6 summarizes the selected model's performance per class on the augmented test set. The model had difficulty categorizing between abscesses, which had false positives on dermatoses and mastitis for 12% and 21% of the images, respectively. Breast engorgement had false positives on mastitis and healthy breasts for 15% and 33% of the images, respectively. Mastitis had false positives in abscess (22%), nipple damage (17%), and healthy breasts (11%). About 21% of the nipple bleb images were confused as nipple damage.

Table 6. Summary of the detection results per class: accuracy, precision, recall, F1-score and support (number of samples per class) using the Resnet50 architecture.

Class	Accuracy	Precision	Recall	F1-score	Support
Abscess	0.571	0.585	0.571	0.578	42
Dermatoses	0.895	0.729	0.895	0.804	48
Engorgement	0.520	0.641	0.520	0.575	48
Mastitis	0.481	0.481	0.481	0.481	54
Nipple bleb	0.714	0.857	0.714	0.779	42
Nipple damage	0.759	0.683	0.759	0.719	54
Healthy	0.844	0.844	0.844	0.844	180

Binary Image Detection Evaluation

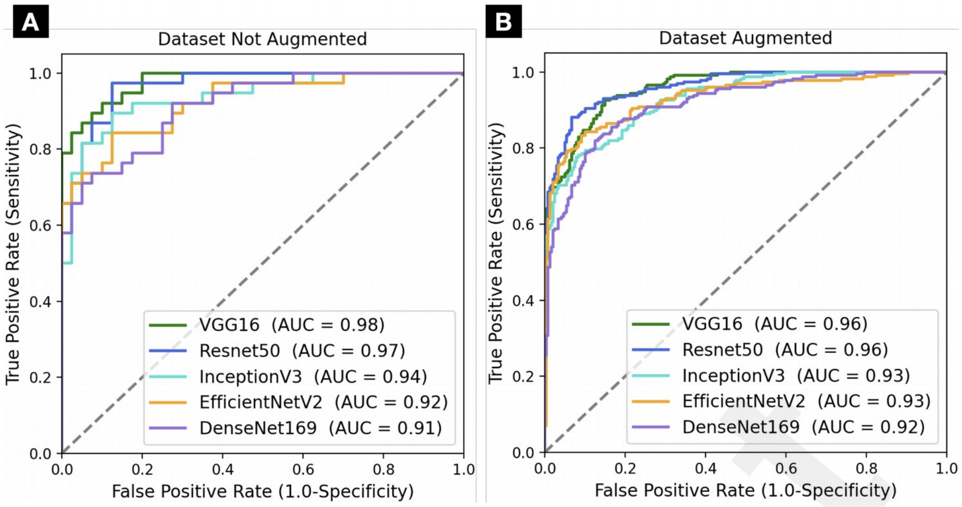
In order to improve the accuracy of our clinical predictions and reduce the chances of incorrect results, we simplified our dataset of seven categories to just two: healthy and unhealthy. The unhealthy category now includes five conditions: abscess, dermatoses, mastitis, nipple bleb, and nipple damage. The healthy category now includes the original healthy conditions and engorgement. Engorgement shares many visual similarities with healthy breast conditions, which made it difficult for the multi-class models to identify engorgement accurately. As presented previously, 33% of the images of engorgement were classified as healthy. Table 7 presents the aggregated evaluation metrics for five models sorted based on the test accuracy.

The accuracy is reported as a balanced score to address class imbalance, ensuring that each class contributes equally to the final metric. Precision, recall, and F1-score are reported for the positive class, with the positive class label specified. For each fold in the cross-validation, we used a separate test set to evaluate the model, and the reported metrics are the average of these evaluations. The best-performing model was the VGG16, which contained the best testing accuracy, followed by Resnet50 and InceptionV3. The models achieved an overall receive operator curve area under the curve (ROC-AUC) of 0.977 for VGG16, 0.966 for Resnet50, 0.935 for InceptionV3, 0.921 for EfficientNetV2, and 0.910 for Densenet169. The detailed ROC-AUC for the not augmented and augmented dataset is shown in Figure 6 (A) and (B), respectively.

Table 7. Average evaluation metrics for the trained models on the not augmented binary dataset (sorted based on test accuracy).

Dataset	Model	Training Accuracy	Validation Accuracy	Test Set Metrics			
				Accuracy	Precision	Recall	F1-score
Binary	VGG16	0.901	0.877	**0.876**	0.990	**0.760**	**0.859**
	Resnet50	0.923	0.872	0.841	0.954	0.715	0.817
	InceptionV3	0.906	0.845	0.838	0.963	0.702	0.812
	EfficientNetV2	0.866	0.831	0.811	**0.991**	0.629	0.769
	DenseNet169	**0.935**	**0.880**	0.761	0.990	0.529	0.688

Figure 6. Model performance on the binary dataset: (A) without and (B) with augmentation.



When applying data augmentation to the binary model, we provided a wider variety of images to help the model better generalize from the training data while not altering the original class distribution. In Table 8, we show the results across the CNNs after data augmentation, where most of the models showed improved metrics, with Resnet50 being the leading model. The models achieved a ROC-AUC of 0.962 for Resnet50, 0.956 for VGG16, 0.931 for EfficientNetV2, 0.929 for InceptionV3, and 0.915 for Densenet169.

Table 8. Average evaluation metrics for the trained models on the augmented binary dataset (sorted based on performance).

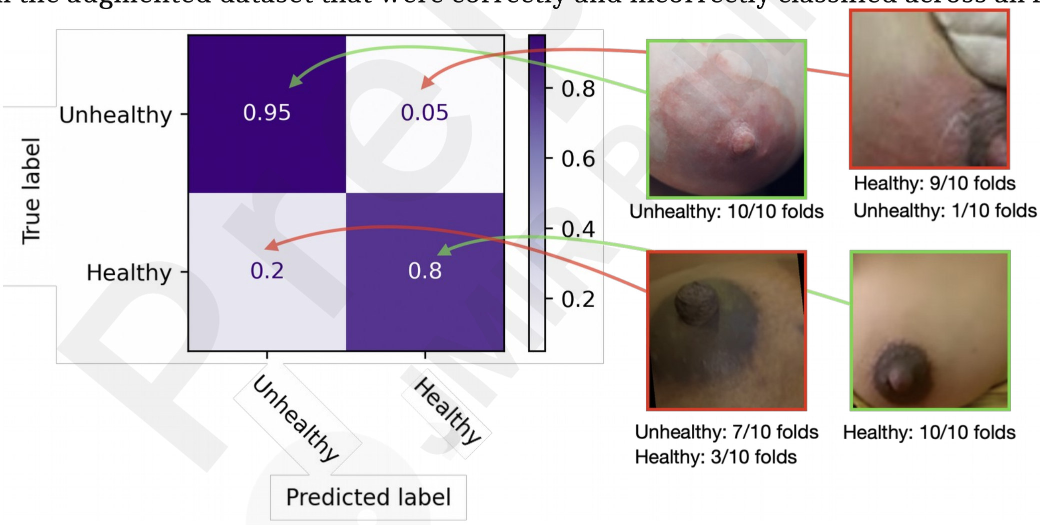
Dataset	Model	Training Accuracy	Validation Accuracy	Test Set Metrics			
				Accuracy	Precision	Recall	F1-score
Binary augmented	Resnet50	**0.952**	**0.933**	**0.877**	0.941	**0.801**	**0.865**
	VGG16	0.877	0.897	0.832	0.965	0.688	0.802
	InceptionV3	0.920	0.893	0.831	0.927	0.715	0.807
	EfficientNetV2	0.885	0.891	0.825	**0.975**	0.666	0.791
	DenseNet169	0.946	0.927	0.771	0.952	0.570	0.713

Looking into the performance of the best model, the Resnet50 with the augmented dataset, we can look closer at the metrics per class of this CNN. Table 9 shows the results for 10-fold cross-validation, in which the model had an overall consistent performance across the iterations. Figure 7 presents the aggregated confusion matrix for the Resnet50 model, in which we consolidated the predictions across all ten folds applied to the augmented dataset. This aggregation was achieved by taking the median predicted class for each instance over the multiple folds, synthesizing a singular prediction representing the consensus of the model's behavior across the test set.

Table 9. Results of 10-fold Cross-Validation for the augmented binary dataset on Resnet50.

Iteration of 10-Fold	Accuracy	Precision	Recall	F1-score
Iteration 1	0.769	0.948	0.557	0.702
Iteration 2	0.761	0.960	0.531	0.684
Iteration 3	0.791	0.951	0.601	0.737
Iteration 4	0.782	0.970	0.570	0.718
Iteration 5	0.782	0.932	0.596	0.727
Iteration 6	0.778	0.943	0.579	0.717
Iteration 7	0.793	0.928	0.623	0.745
Iteration 8	0.767	0.961	0.544	0.695
Iteration 9	0.778	0.963	0.566	0.713
Iteration 10	0.771	0.969	0.548	0.700

Figure 7. Aggregated confusion matrix for the Resnet50 model for the augmented dataset with example images from the augmented dataset that were correctly and incorrectly classified across all folders.



Out of the 468 images used in the testing set, the model could correctly classify 411 images. The total images correctly classified by category are as follows: unhealthy (N = 228, Acc = 95%) and healthy (N = 183, Acc = 80%). The remaining images that were incorrectly classified presented redness (for engorgement cases misclassified as unhealthy, N=26), and incomplete images (too close, nipple and breast not fully visible, N=12). Table 10 summarizes the trained model's performance per class on the augmented test set.

Table 10. Summary of the detection results per class: accuracy, precision, recall, F1-score and support (number of samples per class) using the Resnet50 architecture.

Class	Accuracy	Precision	Recall	F1-score	Support
Unhealthy	0.950	0.835	0.950	0.889	240
Healthy	0.802	0.938	0.802	0.865	228

Discussion

Among the issues that caused model misclassification, we found (1) wrong positioning of the breast in the image, (2) common visual features in the images between the classes, (3) a lack of variety of images belonging to specific cases in the dataset due to variety limitations, and (4) presence of an extraneous object in the frame. Figure 8 presents the correct prediction from the seven classes.

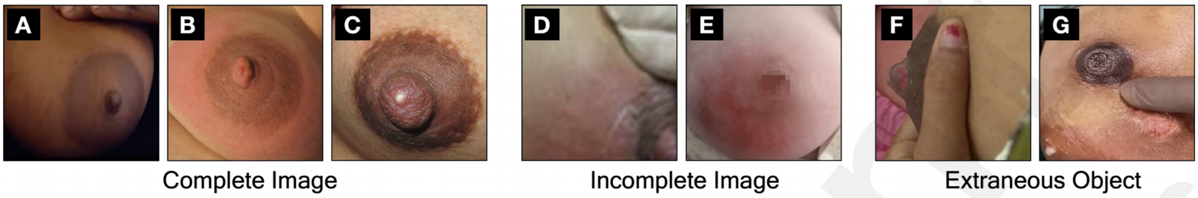
Figure 8. Example images from the testing set that were correctly classified and show features of each breastfeeding-related condition. (a) Abscess, (b) Dermatoses, (c) Engorgement, (d) Mastitis, (e) Nipple bleb, (f) Nipple damage, and (g) Healthy.

Image quality

When examining misclassification results in our image dataset study, we found many image quality issues that likely contributed to the model's diminished performance. In the example images from the testing set shown in Figure 9, examples (A), (B), and (C) demonstrate good image samples that allow a complete evaluation of the breast's condition and, therefore, can be used for the model's evaluation. These images fully or almost entirely show the nipple at a distance that allows diagnosis and does not show information about the person's surroundings or extraneous objects that the model might misinterpret. In Figure 9 (D) and (E), the main issue in both examples is the lack of nipple or breast presence or only partial presence, making it difficult for the model to assimilate them with breast figures, even if there are signs of mastitis or engorgement in both images, the image is incomplete. For Figures 9 (F) and (G), the presence of hands or fingers, nail polish, and partially occluded areas with extraneous objects also affects the model interpretation, especially because we did not train the model was with such extra components.

Other issues noted in the preprocessing phase were causing issues in training and validation loss, as well as false positive and negative detections. For example, having the image of both breasts instead of one affect prediction accuracy, especially in cases where one breast has a different condition compared to the other. The model did not have a large variety of images showing both breasts. Therefore, we improved the training and test results metrics once we separated the breasts into different figures. Additionally, we encountered classification problems with extracted images that show some background components, such as clothes surrounding the breast, breast pumps, or segments of the baby's face or hands. The issues were corrected for these cases by cropping the image to the area of interest. If an object is too similar, such as a hand or a baby, we manually apply blurriness filters in the area and remove saturation so that only the breast is recognizable. Images with low resolution also affect the model's performance, especially if they are originally smaller than the size determined by the data augmentation algorithm and were stretched later. Some images that belong to this case and were misclassified had their size manually corrected afterward, and the model properly classified them afterward.

Figure 9. Example images from the testing set. Images (A), (B), and (C) exemplify high-quality images, with a full view of the breast and nipple. (D) shows an image in which the full breast does not appear, making it hard to classify which condition it belongs to. In (E), even though the condition is clear, and the full breast is visible, the nipple is pixelated in the photo, altering the original features that the model is not used to. (F) and (G) show breasts partially occluded, and the presence of nail polish in the color of the wound also impacts the model's performance in those cases. The examples of low-quality data provide details about how to improve data acquisition for future development.



Visual similarities between conditions

Conditions that present common features and can cause confusion in the diagnosis are mastitis, engorgement, and healthy. Mastitis shows redness throughout the entire breast, showing little skin tone differences and having breasts appear fuller. Some of these features are commonly found in breast engorgement. However, there are fewer signs of intensified redness, sometimes no redness at all, but with the presence of veins and stretched nipples, and being visually similar to healthy. Due to the limited availability of images of breast engorgement for a separate class and the fact that engorgement is not necessarily an issue but can become mastitis when not alleviated, the model classified some engorged breasts as mastitis. When we included engorgement in the healthy class

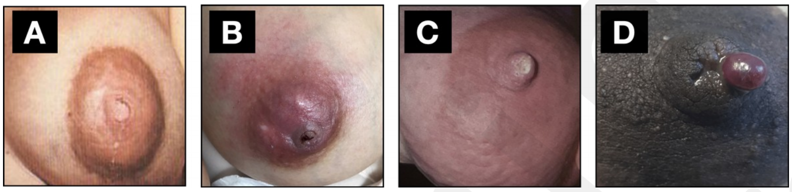
This highlights the need for (1) increasing the engorgement dataset, (2) working closely with LCs to investigate the need to categorize conditions that can be a problem but indicate false positive cases of more serious issues, and (3) exploring the possibility of using these conditions that have higher errors as a base for following patient condition progression, where there is a transition between conditions for improving or worsening a patient's situation.

Lack of variety of images belonging to specific cases in the dataset

Another example of misclassification includes conditions that occur together, which is the case of Figure 10 (b), showcasing a breast abscess concentrated behind the nipple and with signs of nipple damage. Such an example was one of the very few occurrences of simultaneous conditions in the dataset and emphasized the reality that LCs have patients with similar cases, bringing the need to think about systems that (1) recognize multiple conditions or (2) decide between the most severe one for patient priority. Figure 10 (c) is a case of granulomatous mastitis that was classified as nipple damage due to the presence of nipple scarring, highlighting the fewer occurrences of such a specific case in the dataset.

Additionally, in Figure 10, images (a) and (d) show breasts in the conditions of engorgement and nipple damage, respectively. For the case of (a), the engorged breast occurs in an inverted nipple, showing its center lighter and misclassifying it as a nipple bleb. For (d), due to the proximity and nature of the nipple damage with a blood blister, the reflection on the dot suggests that it could be a nipple bleb, also misclassifying the image. These misclassified images with distinct features can also be complex to classify for humans, mainly because some of these conditions rarely occur. Given the nature of the images and the lack of images publicly available with the variety of cases across different skin tones, breasts, and nipple sizes, we believe that working with more images involving rare disorders and providing more data augmentation alternatives can improve the model's classification significantly. Additionally, (d) highlights the issue with image angle and proximity. The picture was taken too close to the breast, having a higher chance of misclassification.

Figure 10. Images incorrectly classified due to dataset variety limitations. (a) is a breast with an abscess but also has nipple damage, (b) is an engorged breast with an inverted nipple classified as nipple bleb, (c) is a breast with granulomatous mastitis classified as nipple damage, and (d) is nipple damage classified as nipple bleb.



Limitations

Our findings emphasize the need for improvement in several areas. As demonstrated in our evaluation, naturalistic images captured by users have several image quality issues that can impede the classification system from proper functioning. Thus, future systems must implement a user interface to properly guide parents in taking pictures to input the AI triaging system. This system should provide basic guidelines around how to frame the breast such that no occlusion is present, not use the finger to point out parts of interest, and make sure the camera framing can see the entire breast so that the nipple, areola, and breast tissue are all visible. Previous works explore the importance of implementing guidelines for image assessment of external diseases, such as in dermatology disease assessments, and its benefits for better professional evaluation and higher accuracy in diagnosing conditions [64, 65, 69]. Guidelines may be implemented as a set of easy instructions, and more advanced systems could provide immediate image quality feedback.

Moreover, our system only uses RGB images to triage breastfeeding-related conditions, not incorporating patient input regarding pain onset, location, symptoms, and pain levels. These are critical data for diagnosing with higher accuracy and providing more effective feedback to patients experiencing breastfeeding-related pain [70]. Furthermore, automating patient responses [71 - 73] and using large language models [74] can help categorize issues based on their problem description and image inputs, streamlining the care process, and ensuring prompt patient attention.

Finally, the most significant limitation of this work is how this evaluation was limited in having a properly balanced dataset to help achieve close-to-perfect performance scores from the model. Despite these limitations, we addressed imbalance issues and proved it possible to obtain satisfactory results in detecting and differentiating the conditions we tested.

Applications and future work

This work showcases the potential for high-accuracy breastfeeding-related condition detection to manage postpartum challenges better. In addition, we demonstrate the feasibility of implementing patient support and condition triaging for smartphone-based applications by using deep-learning RGB image recognition. The model can be integrated into a telehealth pipeline for postpartum lactation care, helping LCs classify and organize patients based on the severity of their condition or the level of certainty that they are experiencing health issues. Additionally, the system can help track patient disease progression and aid newly qualified LCs by providing faster decision-making support.

The evaluation will serve as a baseline for performing a co-design study with mothers and LCs to evaluate the system requirements regarding data gathering and privacy concerns regarding sensitive data sharing. Understanding the benefits of such a system and recognizing its challenges is essential to building effective tools that will meet patients' and healthcare providers' needs. Also, a comprehensive approach is needed to determine the threshold for flagging a patient as unhealthy in the AI-mediated lactation care system, combining quantitative measures (image detection and pain assessment) with clinical expertise. These improvements will allow this work to compose applications for (1) patient self-assessment tools for actionable feedback for breastfeeding pain, (2) reliably identifying cases that require immediate attention and flagging them for LCs, and (3) enabling timely interventions and improved patient outcomes in lactation care. Future work envisions a fully developed hybrid-remote consultation system where patients answer questions for the assessment stage, and images are shared between the patient and provider to visualize the severity of the issue before care is provided. Integrating visual information and pain assessment in remote consultations enhances the diagnostic process and enables LCs to deliver tailored care promptly [74] and help overcome burnout from these professionals.

Conclusions

This paper demonstrates the feasibility of AI-mediated detection of breast conditions for lactating women. We take the first step in this domain by using RGB breast images to triage healthy from unhealthy breasts in mastitis spectrum disease conditions like nipple blebs, engorgement, abscess, and mastitis, in nipple damage caused by poor breastfeeding techniques, breast pumps, and other conditions, and dermatoses caused by a variety of conditions. We implemented five distinct CNN models to classify images in two different datasets, identifying seven breast conditions and between healthy and unhealthy conditions. The evaluation of the models based on our dataset demonstrated the feasibility of using CNNs to classify and intervein with patients who seek remote guidance and management of their symptoms. Although this model's performance was good, it can be improved by increasing the variety of images and conditions in the dataset and implementing the best practices for image posing for proper image classification, leaving significant room for improvement. The feasibility of this work is the initial step towards building tele-lactation services with better data for lactation consultants. We hope our work will inspire future exploration into applying technologies to help lactation support research that can reach more people globally and investigate ideas beyond laboratory settings. This will allow a more comprehensive understanding of breast health for postpartum mothers and empower them to take proactive steps in maintaining their well-being.

Acknowledgments

We thank the Google Health Equity Research Initiative that supported this research through

their program to advance health equity research and improve health outcomes for groups disproportionately impacted by health disparities.

Authors' Contributions

JDS conceptualized the research question, acquired the data, analyzed the data, wrote the manuscript, and take responsibility for the integrity of the data and the accuracy of the data analyses. JME provided guidance and assisted with the cross-validation and data augmentation strategies. KC provided guidance during the study design and material support and data consistency. EW and VV provided guidance, data analysis and technical support during the study. All authors contributed to drafting the paper and its critical revision for important intellectual content.

Data Availability

The data sets analyzed in this study are not publicly available for confidentiality reasons but are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Abbreviations

AUC: Area Under the Curve
CNN: Convolutional Neural Network
FNR: False Negative Rate
FPR: False Positive Rate
IBCLC: International Board-Certified Lactation Consultant
IRB: Institutional Review Board
JPEG: Joint Photographic Experts Group
LC: Lactation Consultant
LMIC: Low and Middle-Income Country
PPV: Positive Prediction Value
PNG: Portable Network Graphics
ROC: Receiver Operating Characteristic
TNR: True Negative Rate
TPR: True Positive Rate
UNICEF: United Nations International Children's Emergency Fund
VGG16: Visual Geometry Group model with 16 layers
WHO: World Health Organization

References

1. Kramer MS, Kakuma R. Optimal duration of exclusive breastfeeding. Cochrane database of systematic reviews. 2012(8). [doi: 10.1002/14651858.CD003517.pub2] [PMID:

22895934]
2. Duijts L, Ramadhani MK, Moll HA. Breastfeeding protects against infectious diseases during infancy in industrialized countries. A systematic review. Maternal & child nutrition. 2009 Jul;5(3):199-210. [doi: 10.1111/j.1740-8709.2008.00176.x] [PMID: 19531047]
3. Kramer MS, Guo T, Platt RW, Sevkovskaya Z, Dzikovich I, Collet JP, Shapiro S, Chalmers B, Hodnett E, Vanilovich I, Mezen I. Infant growth and health outcomes associated with 3 compared with 6 mo of exclusive breastfeeding. The American journal of clinical nutrition. 2003 Aug 1;78(2):291-5. [PMID: 12885711] [DOI: 10.1093/ajcn/78.2.291]
4. Kent G. Child feeding and human rights. International breastfeeding journal. 2006 Dec;1(1):1-2. [PMID: 17176464] [DOI: 10.1186/1746-4358-1-27]
5. Dinour LM. Speaking out on "breastfeeding" terminology: Recommendations for gender-inclusive language in research and reporting. Breastfeeding Medicine. 2019 Oct 1;14(8):523-32. [PMID: 31364867] [DOI: 10.1089/bfm.2019.0110]
6. Centers for Disease Control and Prevention. Breastfeeding Report Card. 2022. URL: https://www.cdc.gov/breastfeeding/data/reportcard.htm [accessed 2023-11-13]
7. UNICEF. Breastfeeding, a mother's gift for every child. 2023. URL: https://data.unicef.org/topic/nutrition/breastfeeding/ [accessed 2023-11-13]
8. Coca KP, Marcacine KO, Gamba MA, Corrêa L, Aranha AC, de Vilhena Abrão AC. Efficacy of low-level laser therapy in relieving nipple pain in breastfeeding women: a triple-blind, randomized, controlled trial. Pain Management Nursing. 2016 Aug 1;17(4):281-9. [PMID: 27363734] [DOI: 10.1016/j.pmn.2016.05.003]
9. Brown CR, Dodds L, Legge A, Bryanton J, Semenic S. Factors influencing the reasons why mothers stop breastfeeding. Canadian Journal of Public Health. 2014 May;105:e179-85. [PMID: 25165836] [DOI: 10.17269/cjph.105.4244]
10. Friesen CA, Hormuth LJ, Petersen D, Babbitt T. Using videoconferencing technology to provide breastfeeding support to low-income women: Connecting hospital-based lactation consultants with clients receiving care at a community health center. Journal of Human Lactation. 2015 Nov;31(4):595-9. [PMID: 26297347] [DOI: 10.1177/0890334415601088]
11. Chaves AF, Vitoriano LN, Borges FL, Melo RD, de Oliveira MG, Costa AC. Percepção das mulheres que receberam consultoria em amamentação. Enfermagem em Foco. 2019;10(5). [DOI: 10.21675/2357-707X]
12. Patel S, Patel S. The effectiveness of lactation consultants and lactation counselors on breastfeeding outcomes. Journal of Human Lactation. 2016 Aug;32(3):530-41. [PMID: 26644419] [DOI: 10.1177/0890334415618668]
13. IBLCE. Current statistics on worldwide IBCLCs. Technical Report. 2022. URL: https://iblce.org/about-iblce/current-statistics-on-worldwide-ibclcs/ [accessed 2023-11-13]
14. Hamilton BE, Martin JA, Osterman MJK. Births: Provisional data for 2021. Vital Statistics Rapid Release; no 20. Hyattsville, MD: National Center for Health Statistics. May 2022. [DOI: https://dx.doi.org/10.15620/cdc:116027]
15. ARPENBR. 2022. Nascimentos - Brasil - 2021. URL: https://transparencia.registrocivil.org.br/registros [accessed 2023-11-13]
16. DeLeo A, Geraghty S. iMidwife: midwifery students' use of smartphone technology as a mediated educational tool in clinical environments. Contemporary nurse. 2018 Jul 4;54(4-5):522-31. [PMID: 29228874] [DOI: 10.1080/10376178.2017.1416305]
17. Tripp N, Hainey K, Liu A, Poulton A, Peek M, Kim J, Nanan R. An emerging model of

maternity care: smartphone, midwife, doctor?. Women and Birth. 2014 Mar 1;27(1):64-7. [PMID: 24295598] [DOI: 10.1016/j.wombi.2013.11.001]
18. Feinstein J, Slora EJ, Bernstein HH. Telehealth can promote breastfeeding during the COVID-19 pandemic. NEJM Catalyst Innovations in Care Delivery. 2021 Apr 29;2(2). [DOI: doi:10.1056/CAT.21.0076]
19. Haase B, Brennan E, Wagner CL. Effectiveness of the IBCLC: Have we made an impact on the care of breastfeeding families over the past decade? Journal of Human Lactation. 2019 Aug;35(3):441-52. [PMID: 31206324] [DOI: 10.1177/0890334419851805]
20. Ray KN, Demirci JR, Uscher-Pines L, Bogen DL. Geographic access to international board-certified lactation consultants in Pennsylvania. Journal of Human Lactation. 2019 Feb;35(1):90-9. [PMID: 29969344] [DOI: 10.1177/0890334418768458]
21. Hoddinott P, Britten J, Pill R. Why do interventions work in some places and not others: a breastfeeding support group trial. Social science & medicine. 2010 Mar 1;70(5):769-78. [PMID: 20005617] [DOI: 10.1016/j.socscimed.2009.10.067]
22. de Souza J, Calsinski C, Chamberlain K, Cibrian F, Wang EJ. Investigating interactive methods in remote chestfeeding support for lactation consulting professionals in Brazil. Frontiers in Digital Health. 2023 Apr 3;5:1143528. [PMID: 37077406] [DOI: 10.3389/fdgth.2023.1143528]
23. Donovan H, Welch A, Williamson M. Reported levels of exhaustion by the graduate nurse midwife and their perceived potential for unsafe practice: a phenomenological study of Australian double degree nurse midwives. Workplace Health & Safety. 2021 Feb;69(2):73-80. [PMID: 32812841] [DOI: 10.1177/2165079920938000]
24. Fraser HS, Blaya J. Implementing medical information systems in developing countries, what works and what doesn't. AMIA Annual Symposium Proceedings 2010 (Vol. 2010, p. 232). American Medical Informatics Association. [PMID: 21346975]
25. Busch DW, Logan K, Wilkinson A. Clinical practice breastfeeding recommendations for primary care: applying a tri-core breastfeeding conceptual model. Journal of Pediatric Health Care. 2014 Nov 1;28(6):486-96. [PMID: 24786581] [DOI: 10.1016/j.pedhc.2014.02.007]
26. Kern-Goldberger AR, Srinivas SK. Obstetrical telehealth and virtual care practices during the COVID-19 pandemic. Clinical obstetrics and gynecology. 2022 Mar;65(1):148. [PMID: 35045037] [DOI: 10.1097/GRF.0000000000000671]
27. Burns E, Fenwick J, Sheehan A, Schmied V. Mining for liquid gold: midwifery language and practices associated with early breastfeeding support. Maternal & child nutrition. 2013 Jan;9(1):57-73. [PMID: 22405753] [DOI: 10.1111/j.1740-8709.2011.00397.x]
28. Niazi A, Rahimi VB, Soheili-Far S, Askari N, Rahmanian-Devin P, Sanei-Far Z, Sahebkar A, Rakhshandeh H, Askari VR. A systematic review on prevention and treatment of nipple pain and fissure: are they curable? Journal of pharmacopuncture. 2018 Sep;21(3):139. [PMID: 30283701] [DOI: 10.3831/KPI.2018.21.017]
29. Mitoulas LR, Davanzo R. Breast pumps and mastitis in breastfeeding women: clarifying the relationship. Frontiers in Pediatrics. 2022 Jun 10;10:856353. [PMID: 35757121] [DOI: 10.3389/fped.2022.856353]
30. Gomez-Pomar E, Blubaugh R. The Baby Friendly Hospital Initiative and the ten steps for successful breastfeeding. a critical review of the literature. Journal of Perinatology. 2018 Jun;38(6):623-32. [PMID: 29416115] [DOI: 10.1038/s41372-018-0068-0]
31. VanDevanter N, Gennaro S, Budin W, Calalang-Javiera H, Nguyen M. Evaluating implementation of a baby-friendly hospital initiative. MCN: The American Journal of Maternal/Child Nursing. 2014 Jul 1;39(4):231-7. [PMID: 24978002] [DOI:

10.1097/NMC.0000000000000046]
32. Similarweb. Most Popular Messaging Apps Worldwide 2023. 2023. https://www.similarweb.com/blog/research/market-research/worldwide-messaging-apps/ [accessed 2023-06-22]
33. Coelho LS. Telefarmácia na Atenção Primária à Saúde: Relato de Experiência sobre a Implementação e Prática em um Centro de Saúde de Florianópolis. Universidade Federal de Santa Catarina. 2021 Sep 23;35. URL: https://repositorio.ufsc.br/handle/123456789/228419
34. Weaver NS, Roy A, Martinez S, Gomanie NN, Mehta K. How WhatsApp is Transforming Healthcare Services and Empowering Health Workers in Low-and Middle-Income Countries. In2022 IEEE Global Humanitarian Technology Conference (GHTC). 2022 Sep 8 (pp. 234-241). IEEE. [DOI: 10.1109/GHTC55712]
35. Trude AC, Martins RC, Martins-Silva T, Blumenberg C, Carpena MX, Del-Ponte B, Loret de Mola C. A WhatsApp-based intervention to improve maternal social support and maternal-child health in Southern Brazil: the Text-Message Intervention to Enhance Social Support (TIES) Feasibility Study. INQUIRY: The Journal of Health Care Organization, Provision, and Financing. 2021 Sep 1;58:00469580211048701. [PMID: 34619999] [DOI: 10.1177/00469580211048701]
36. Araújo JC, de Sousa Lima T, dos Santos JA, dos Santos Costa E. Use of whatsapp app as a tool to education and health promotion of pregnant women during prenatal care/Uso do aplicativo whatsapp como ferramenta de educação e promoção à saúde de gestantes durante o pré–natal/Uso del aplicación whatsapp como herramienta de educación y promoción a la salud.. InAnais do I Congresso Norte Nordeste de Tecnologias em Saúde 2018 Dec 13 (Vol. 1, No. 1). URL: https://revistas.ufpi.br/index.php/connts/article/view/7954/4682
37. Lima AC, Chaves AF, Oliveira MG, Lima SA, Machado MM, Oriá MO. Consultoria em amamentação durante a pandemia COVID-19: relato de experiência. Escola Anna Nery. 2020 Dec 16;24. [DOI: 10.1590/2177-9465-EAN]
38. Gavine A, Marshall J, Buchanan P, Cameron J, Leger A, Ross S, Murad A, McFadden A. Remote provision of breastfeeding support and education: Systematic review and meta-analysis. Maternal & Child Nutrition. 2022 Apr;18(2):e13296. [PMID: 34964542] [DOI: 10.1111/mcn.13296]
39. Nóbrega VC, Melo RH, Diniz AL, Vilar RL. As redes sociais de apoio para o Aleitamento Materno: uma pesquisa-ação. Saúde em Debate. 2019 Aug 5;43:429-40. [DOI: https://doi.org/10.1590/0103-1104201912111]
40. Hinman RS, Lawford BJ, Bennell KL. Harnessing technology to deliver care by physical therapists for people with persistent joint pain: telephone and video-conferencing service models. Journal of Applied Biobehavioral Research. 2019 Jun;24(2):e12150. [DOI: https://doi.org/10.1111/jabr.12150]
41. Candido NL, Marcolino AM, Santana JM, Silva JR, Silva ML. Remote physical therapy during COVID-19 pandemic: guidelines in the Brazilian context. Fisioterapia em Movimento. 2022 Mar 25;35:e35202. [DOI: https://doi.org/10.1590/fm.2022.35202]
42. Giglia R, Cox K, Zhao Y, Binns CW. Exclusive breastfeeding increased by an internet intervention. Breastfeeding Medicine. 2015 Feb 1;10(1):20-5. [PMID: 25358119] [DOI: 10.1089/bfm.2014.0093]
43. Krishnamurti T, Simhan HN, Borrero S. Competing demands in postpartum care: a national survey of US providers' priorities and practice. BMC Health Services Research. 2020 Dec;20:1-0. [PMID: 32252757][DOI: 10.1186/s12913-020-05144-2]
44. Berens P, Eglash A, Malloy M, Steube AM, Academy of Breastfeeding Medicine. ABM

clinical protocol# 26: persistent pain with breastfeeding. Breastfeeding Medicine. 2016 Mar 1;11(2):46-53. [PMID: 26881962] [DOI: 10.1089/bfm.2016.29002.pjb]
45. Douglas P. Re-thinking lactation-related nipple pain and damage. Women's Health. 2022 Mar;18:17455057221087865. [PMID: 35343816] [DOI: 10.1177/17455057221087865]
46. Mitchell KB, Johnson HM, Rodríguez JM, Eglash A, Scherzinger C, Widmer K, Berens P, Miller B, Academy of Breastfeeding Medicine. Academy of Breastfeeding Medicine Clinical Protocol# 36: the mastitis spectrum, revised 2022. Breastfeeding Medicine. 2022 May 1;17(5):360-76. [PMID: 35576513] [DOI: 10.1089/bfm.2022.29207.kbm]
47. Pevzner M, Dahan A. Mastitis while breastfeeding: Prevention, the importance of proper treatment, and potential complications. Journal of Clinical Medicine. 2020 Jul 22;9(8):2328. [PMID: 32707832] [DOI: [10.3390/jcm9082328]
48. Nakamura M, Asaka Y, Ogawara T, Yorozu Y. Nipple skin trauma in breastfeeding women during postpartum week one. Breastfeeding medicine. 2018 Sep 1;13(7):479-84. [PMID: 30074830] [DOI: 10.1089/bfm.2017.0217]
49. Aldhyani TH, Nair R, Alzain E, Alkahtani H, Koundal D. Deep Learning Model for the Detection of Real-Time Breast Cancer Images Using Improved Dilation-Based Method. Diagnostics. 2022 Oct 16;12(10):2505. [PMID: 36292194] [DOI: 10.3390/diagnostics12102505]
50. Yoon JH, Kim EK. Deep learning-based artificial intelligence for mammography. Korean journal of radiology. 2021 Aug;22(8):1225. [PMID: 33987993] [DOI: 10.3348/kjr.2020.1210]
51. Kim SY, Choi Y, Kim EK, Han BK, Yoon JH, Choi JS, Chang JM. Deep learning-based computer-aided diagnosis in screening breast ultrasound to reduce false-positive diagnoses. Scientific reports. 2021 Jan 11;11(1):395. [PMID: 33432076] [DOI: 10.1038/s41598-020-79880-0]
52. Calisto FM, Nunes N, Nascimento JC. BreastScreening: on the use of multi-modality in medical imaging diagnosis. InProceedings of the international conference on advanced visual interfaces 2020 Sep 28 (pp. 1-5). [DOI: 10.1145/3399715.3399744]
53. Zhou Y, Feng BJ, Yue WW, Liu Y, Xu ZF, Xing W, Xu Z, Yao JC, Wang SR, Xu D. Differentiating non-lactating mastitis and malignant breast tumors by deep-learning based AI automatic classification system: A preliminary study. Frontiers in Oncology. 2022 Sep 15;12:997306. [PMID: 36185190] [DOI: 10.3389/fonc.2022.997306]
54. Shen Y, Shamout FE, Oliver JR, Witowski J, Kannan K, Park J, Wu N, Huddleston C, Wolfson S, Millet A, Ehrenpreis R. Artificial intelligence system reduces false-positive findings in the interpretation of breast ultrasound exams. Nature communications. 2021 Sep 24;12(1):5645. [PMID: 34561440] [DOI: 10.1038/s41467-021-26023-2]
55. Abdul Ghafoor N, Sitkowska B. MasPA: A machine learning application to predict risk of mastitis in cattle from AMS sensor data. AgriEngineering. 2021 Aug 4;3(3):575-83. [DOI: 10.3390/agriengineering3030037]
56. Fadul-Pacheco L, Delgado H, Cabrera VE. Exploring machine learning algorithms for early prediction of clinical mastitis. International Dairy Journal. 2021 Aug 1;119:105051. [DOI: 10.1016/j.idairyj.2021.105051]
57. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. Archives of dermatology. 1988 Jun 1;124(6):869-71. [PMID: 3377516] [DOI: 10.1001/archderm.124.6.869]
58. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014 Sep 4. [DOI:

10.48550/arXiv.1409.1556]
59. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).
60. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 2818-2826).
61. Tan M, Le Q. Efficientnetv2: Smaller models and faster training. InInternational conference on machine learning 2021 Jul 1 (pp. 10096-10106). PMLR.
62. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 4700-4708).
63. Rafay A, Hussain W. EfficientSkinDis: An EfficientNet-based classification model for a large manually curated dataset of 31 skin diseases. Biomedical Signal Processing and Control. 2023 Aug 1;85:104869. [DOI: https://doi.org/10.1016/j.bspc.2023.104869]
64. Vodrahalli K, Daneshjou R, Novoa RA, Chiou A, Ko JM, Zou J. TrueImage: a machine learning algorithm to improve the quality of telehealth photos. InBIOCOMPUTING 2021: Proceedings of the Pacific Symposium 2020 (pp. 220-231). [DOI: https://doi.org/10.1142/9789811232701_0021]
65. Finnane A, Curiel-Lewandrowski C, Wimberley G, Caffery L, Katragadda C, Halpern A, Marghoob AA, Malvehy J, Kittler H, Hofmann-Wellenhof R, Abraham I. Proposed technical guidelines for the acquisition of clinical images of skin-related conditions. JAMA dermatology. 2017 May 1;153(5):453-7. [DOI: 10.1001/jamadermatol.2016.6214]
66. Jain S, Singhania U, Tripathy B, Nasr EA, Aboudaif MK, Kamrani AK. Deep learning-based transfer learning for classification of skin cancer. Sensors. 2021 Dec 6;21(23):8142. [DOI: 10.3390/s21238142]
67. Perez, F., Vasconcelos, C., Avila, S., Valle, E. (2018). Data Augmentation for Skin Lesion Analysis. In: Stoyanov, D., et al. OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis. CARE CLIP OR 2.0 ISIC 2018 2018 2018 2018. Lecture Notes in Computer Science, vol 11041. Springer, Cham. [DOI: 10.1007/978-3-030-01201-4_33]
68. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. InIjcai 1995 Aug 20 (Vol. 14, No. 2, pp. 1137-1145). [DOI: 10.5555/1643031.1643047]
69. Ukharov AO, Shlivko IL, Klemenova IA, Garanina OE, Uskova KA, Mironycheva AM, Stepanova YL. Skin cancer risk self-assessment using AI as a mass screening tool. Informatics in Medicine Unlocked. 2023 Jan 1;38:101223. [DOI: https://doi.org/10.1016/j.imu.2023.101223]
70. Lucas R, McGrath JM. Clinical assessment and management of breastfeeding pain. Topics in Pain Management. 2016 Oct 1;32(3):1-1. [DOI: 10.1097/01.TPM.0000502820.55789.3a]
71. Yadav D, Malik P, Dabas K, Singh P. Feedpal: Understanding opportunities for chatbots in breastfeeding education of women in India. Proceedings of the ACM on Human-Computer Interaction. 2019 Nov 7;3(CSCW):1-30. [DOI: 10.1145/3359272]
72. Gupta V, Arora N, Jain Y, Mokashi S, Panda C. Assessment on Adoption Behavior of First-time Mothers on the Usage of Chatbots for Breastfeeding Consultation. Journal of Mahatma Gandhi University of Medical Sciences & Technology. 2021 May;6(2):65. [DOI: 10.5005/jp-journals-10057-0161]
73. Bennett V. Could artificial intelligence assist mothers with breastfeeding? British

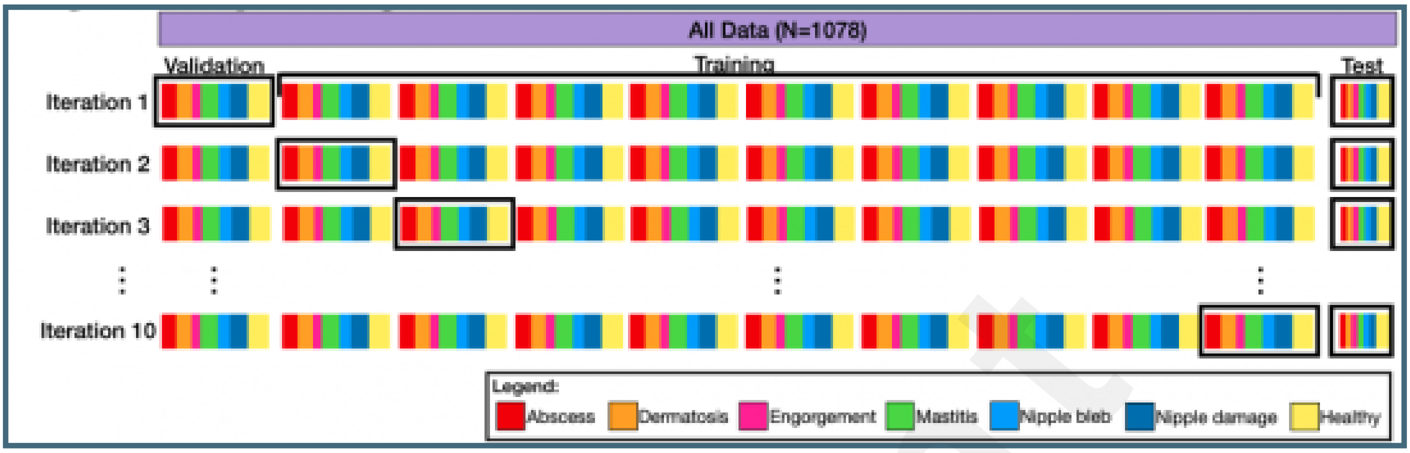
Journal of Midwifery. 2018 Apr 2;26(4):212-3. [DOI: 10.12968/bjom.2018.26.4.212]
74. Thirunavukarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. Nature medicine. 2023 Aug;29(8):1930-40. [PMID: 37460753] [DOI: 10.1038/s41591-023-02448-8]
75. De Souza J, Chamberlain K, Gupta S, Gao Y, Alshurafa N, Wang EJ. Opportunities in designing HCI tools for lactation consulting professionals. CHI Conference on Human Factors in Computing Systems Extended Abstracts 2022 Apr 27 (pp. 1-7). [DOI: 10.1145/3491101.3519762]

Supplementary Files

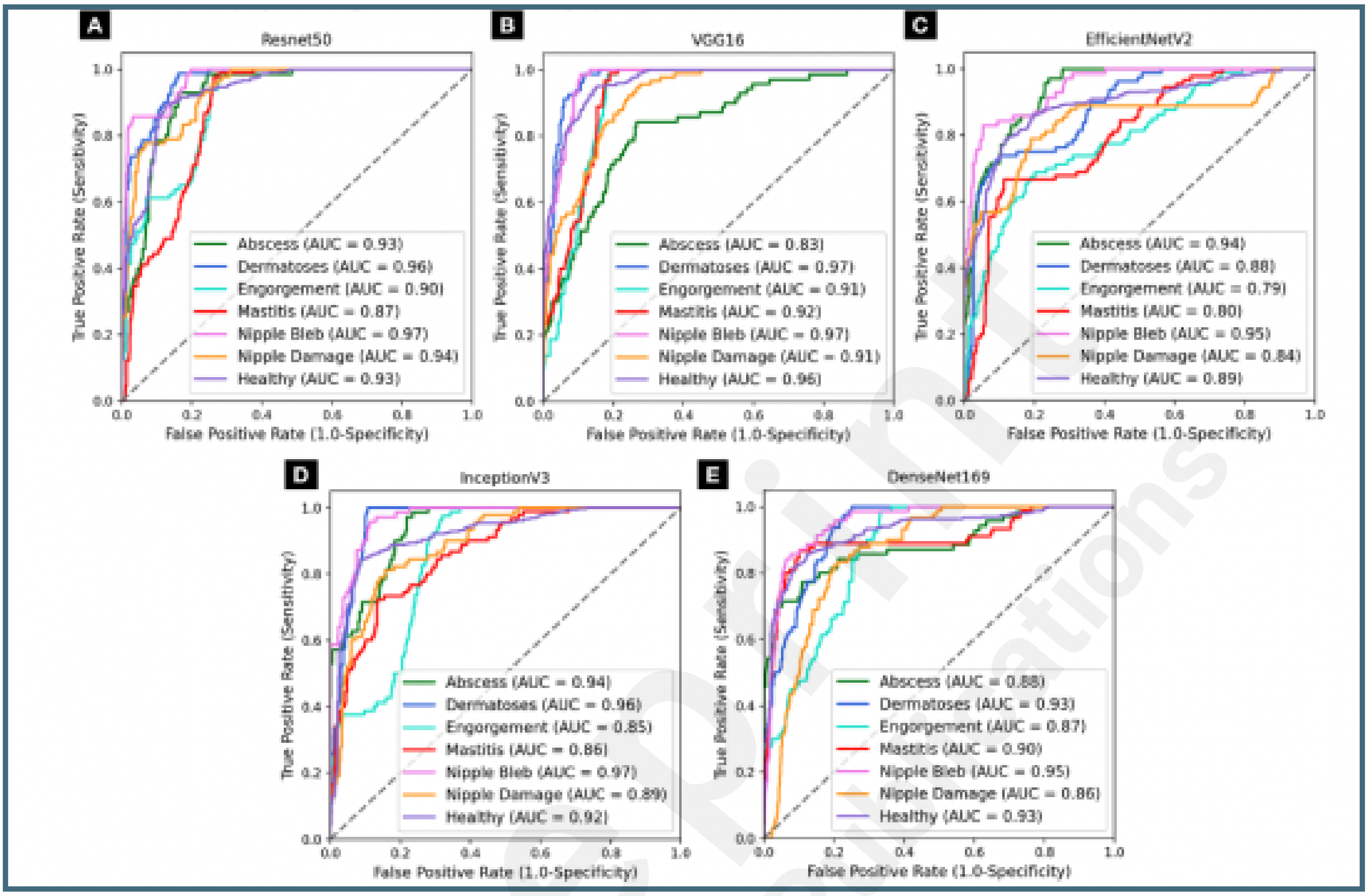
Figures

Samples of Augmented Data: (a) Original (b) Brightness (c) Center zoom (d) Horizontal flip (e) Rotation (f) Shear and (g) Vertical Flip.

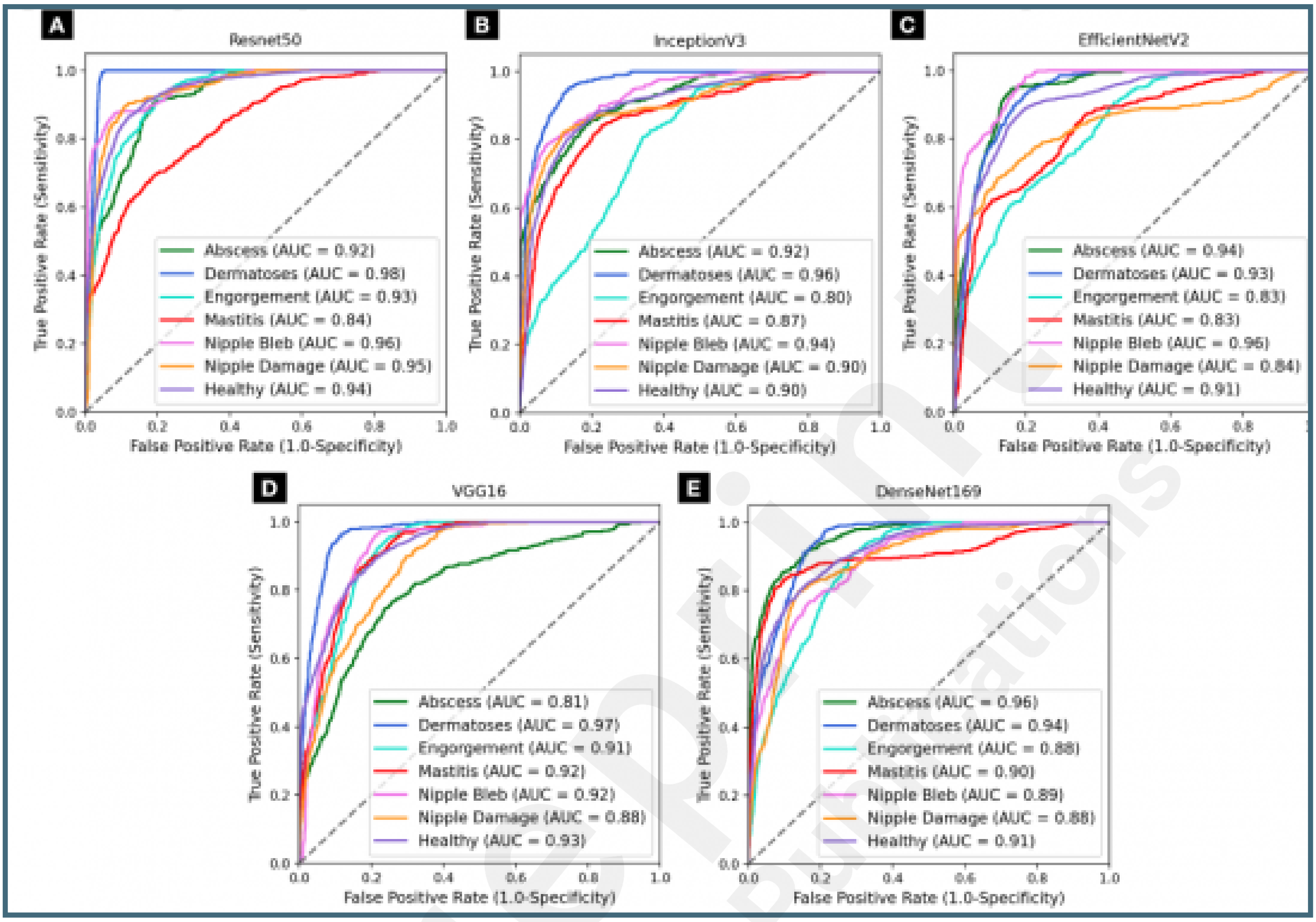
Graphical diagram of Stratified K-fold Cross Validation on a 7-class dataset.



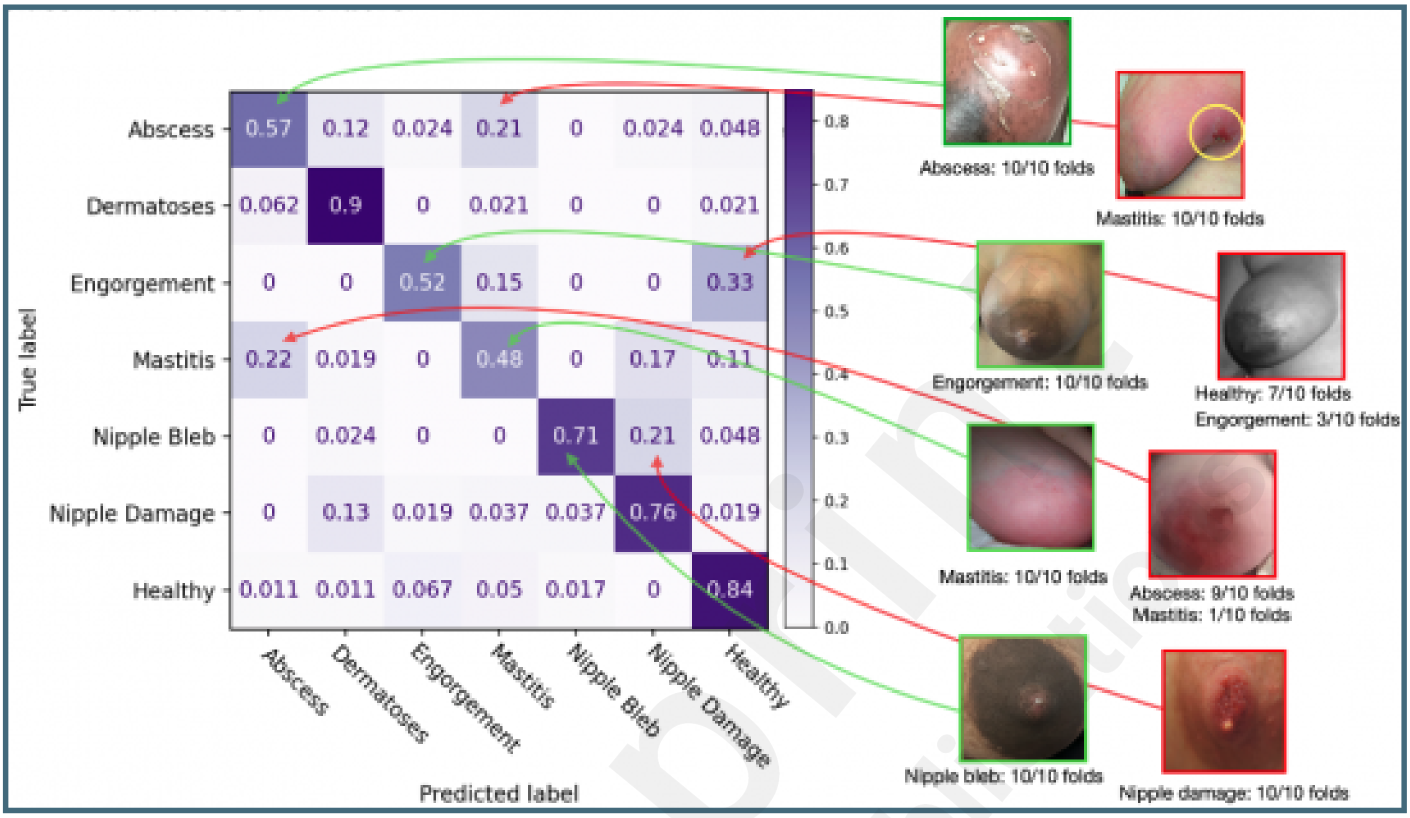
Performance of the five CNNs on the 7-class dataset: (A) Resnet50, (B) VGG16, (C) EfficientNetV2, (D) InceptionV3, and (E) DenseNet169.



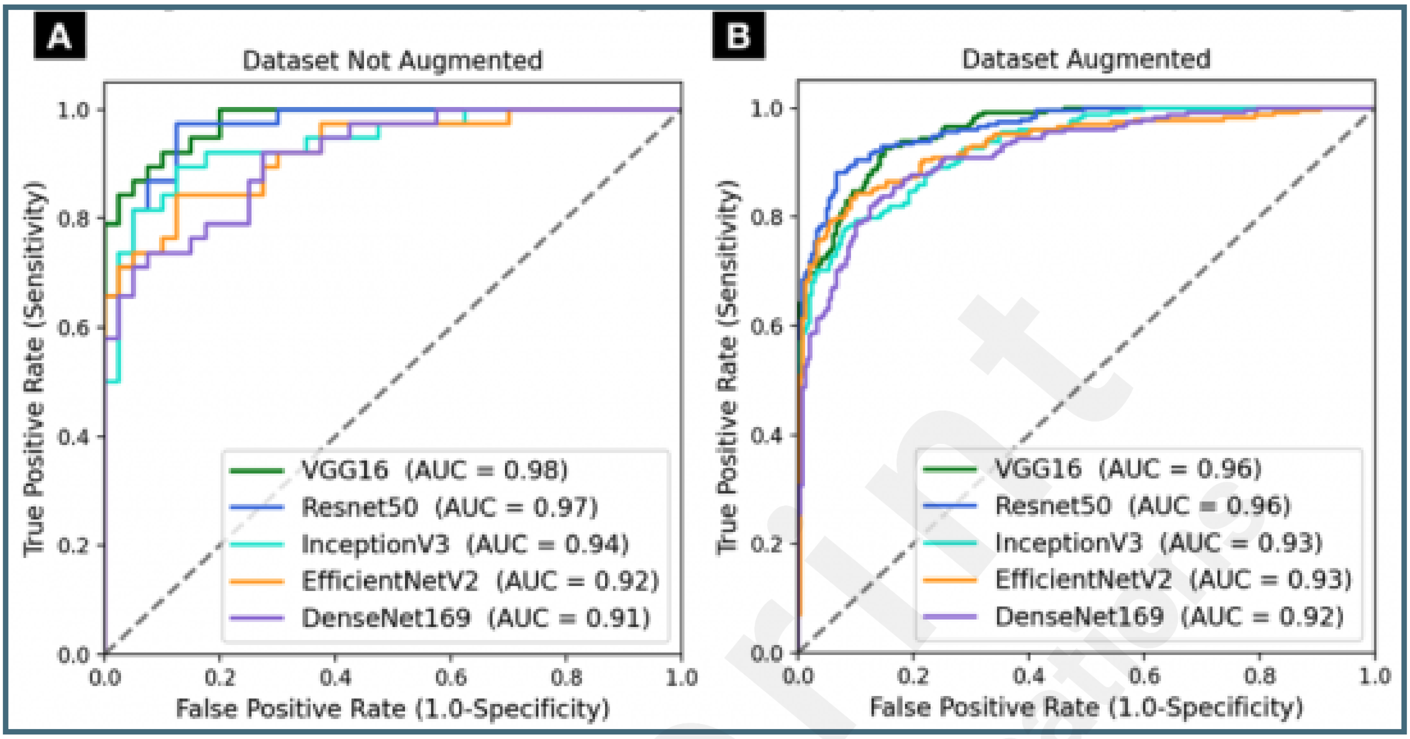
Algorithm performance on the 7-class dataset after data augmentation.



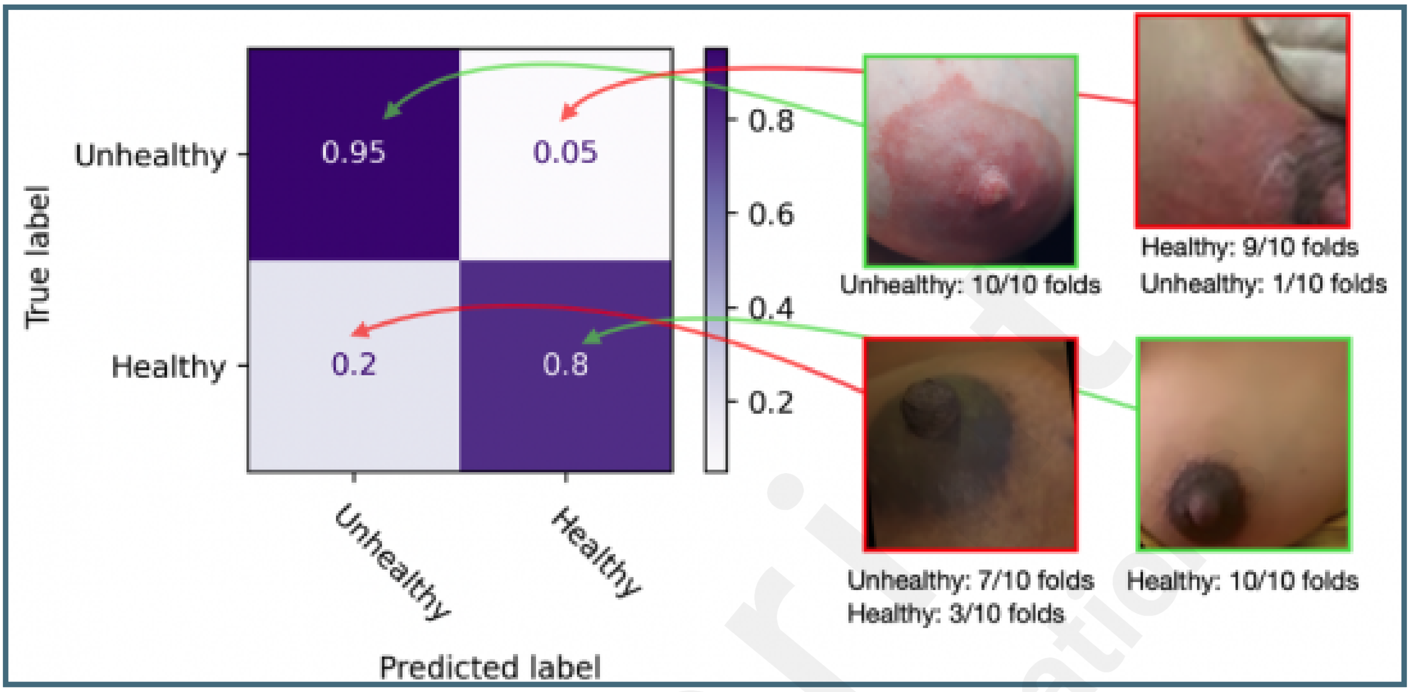
Aggregated confusion matrix for the Resnet50 model for the augmented dataset with example images from the augmented dataset that were correctly and incorrectly classified across all folders.



Model performance on the binary dataset: (A) without and (B) with augmentation.



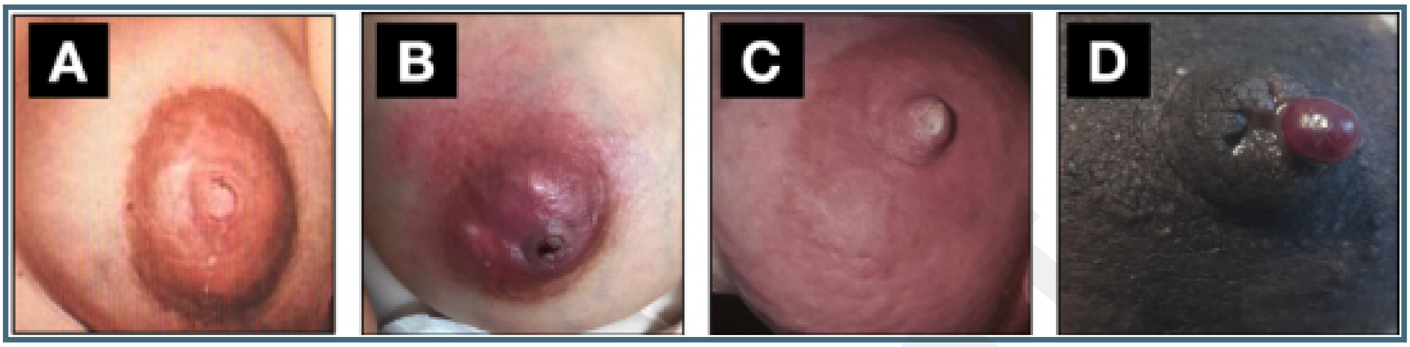
Aggregated confusion matrix for the Resnet50 model for the augmented dataset with example images from the augmented dataset that were correctly and incorrectly classified across all folders.



Example images from the testing set that were correctly classified and show features of each breastfeeding-related condition. (a) Abscess, (b) Dermatoses, (c) Engorgement, (d) Mastitis, (e) Nipple bleb, (f) Nipple damage, and (g) Healthy.

Example images from the testing set. Images (A), (B), and (C) exemplify high-quality images, with a full view of the breast and nipple. (D) shows an image in which the full breast does not appear, making it hard to classify which condition it belongs to. In (E), even though the condition is clear, and the full breast is visible, the nipple is pixelated in the photo, altering the original features that the model is not used to. (F) and (G) show breasts partially occluded, and the presence of nail polish in the color of the wound also impacts the model's performance in those cases. The examples of low-quality data provide details about how to improve data acquisition for future development.

Images incorrectly classified due to dataset variety limitations. (a) is a breast with an abscess but also has nipple damage, (b) is an engorged breast with an inverted nipple classified as nipple bleb, (c) is a breast with granulomatous mastitis classified as nipple damage, and (d) is nipple damage classified as nipple bleb.



TOC/Feature image for homepages

mother with breast concern using smartphone.