Using convolutional neural networks to discriminate between cysts and masses in Monte Carlo-simulated dual-energy mammography

Andrey Makeev[a)]
Division of Imaging, Diagnostics, and Software Reliability, Office of Scientific and Engineering Laboratories, Center for Devices and Radiological Health, U.S. Food & Drug Administration, Silver Spring, MD 20903, USA

Brian Toner
Program in Applied Mathematics, University of Arizona, Tucson, AZ 85721, USA

Marian Qian
Thomas Jefferson High School for Science and Technology, Alexandria, VA 22312, USA

Andreu Badal and Stephen J. Glick
Division of Imaging, Diagnostics, and Software Reliability, Office of Scientific and Engineering Laboratories, Center for Devices and Radiological Health, U.S. Food & Drug Administration, Silver Spring, MD 20903, USA

(Received 4 February 2021; revised 12 May 2021; accepted for publication 17 May 2021; published xx xxxx xxxx)

Purpose: A substantial percentage of recalls (up to 20%) in screening mammography is attributed to extended round lesions. Benign fluid-filled breast cysts often appear similar to solid tumors in conventional mammograms. Spectral imaging (dual-energy or photon-counting mammography) has been shown to discriminate between cysts and solid masses with clinically acceptable accuracy. This work explores the feasibility of using convolutional neural networks (CNNs) for this task.
Methods: A series of Monte Carlo experiments was conducted with digital breast phantoms and embedded synthetic lesions to produce realistic dual-energy images of both lesion types. We considered such factors as nonuniform anthropomorphic background, size of the mass, breast compression thickness, and variability in lesion x-ray attenuation. These data then were used to train a deep neural network (ResNet-18) to learn the differences in x-ray attenuation of cysts and masses.
Results: Our simulation results showed that the CNN-based classifier could reliably discriminate between cystic and solid mass round lesions in dual-energy images with an area under the receiver operating characteristic curve (ROC AUC) of 0.98 or greater.
Conclusions: The proposed approach showed promising performance and ease of implementation, and could be applied to novel photon-counting detector-based spectral mammography systems. [https://doi.org/10.1002/mp.15005]

Key words: breast cysts, Monte Carlo simulation, neural network, solid masses, spectral mammography

1. INTRODUCTION

The substantial reduction in breast cancer mortality observed over the past 30 yr has been attributed largely to screening mammography. Nonetheless, the recall rate (i.e., the percentage of screening mammograms interpreted as suspicious, requiring diagnostic follow-up imaging and/or tissue biopsy) in the United States is approximately 12–14%.[1] This relatively high recall rate was one factor contributing to the US Preventive Task Force's recommendation for biennial screening mammography instead of annual screening, as recommended by the American College of Obstetricians and Gynecologists (ACOG) and the American College of Radiology (ACR).

It has been estimated that of all probably benign lesions observed on screening mammograms, approximately 37% can be characterized as solitary, well-defined, noncalcified, circumscribed masses. Of these masses, only a small percentage are cancers (<2%[2]). Approximately 35%[3] turn out to be simple cysts after assessment with ultrasound, requiring no further clinical evaluation. If analysis of these circumscribed masses could be improved on screening mammograms, recall rates could be lowered, and patient anxiety and mammography screening program costs might be reduced.

Fredenberg et al.[4] used tissue specimen experiments to demonstrate that there are small differences in energy-dependent x-ray attenuation through solid and cystic masses, however, it is unlikely that conventional mammography could be used to differentiate between these masses. Spectral mammography, implemented using separate dual-energy acquisitions or with photon-counting detector (PCD)-based systems, has shown potential to differentiate among different breast tissue types.[5] Erhard et al.[3] recently reported that suspicious breast lesions visualized in dual-energy, photon-counting-based spectral mammography can be identified as cystic or solid masses with moderate accuracy, raising the possibility of reducing recalls for cystic lesions. The Erhard study used a

two-feature algorithm to discriminate solid masses from cysts on a patient dataset of 119 lesions (62 solid masses and 57 cystic masses) greater than 10 mm in diameter. The method required that expert radiologists delineate elliptical regions of interest defining the lesion as well as a surrounding reference background region. Results suggested that the area under the receiver operating characteristic curve (ROC AUC) was 0.88 with a specificity of 61% at the 99% sensitivity operating point. The authors reported that for each missed cancer predicted with the two-feature algorithm, at least 625 cystic lesions would be identified correctly as such and would not have to be recalled for further diagnostic workup.

Deep learning algorithms such as convolutional neural networks (CNNs) have been shown to help with classification problems in medical imaging. Our study investigated the potential use of a CNN to discriminate between solid and cystic masses visualized using dual-energy mammography. Since clinical dual-energy mammography data currently are scarce, we examined the method using a sophisticated Monte Carlo simulation study that realistically models many aspects of the imaging process. Solid and cystic masses of various sizes were modeled and inserted into compressed anthropomorphic breast phantoms of varying thickness.

Section 2 describes the methodologies used including the lesion models and the simulation process for acquiring training data, as well as the CNN architecture used to classify simulated mammograms of the two lesion types. Section 3 details the CNN classification performance, as reported with ROC curves and the AUC figure of merit. Implications of these results are discussed in Section 4, and conclusions are presented in Section 5.

2. MATERIALS AND METHODS

2.A. Monte Carlo simulations of dual-energy mammography acquisitions

MC-GPU[6] x-ray transport/simulation software and digital compressed breast phantoms[7] developed by the US Food and Drug Administration (FDA) were used to produce realistic dual-energy mammograms. The Monte Carlo experiment emulated a setup with an 85 μm pixel a-Se energy-integrating x-ray detector, mimicking a Siemens Mammomat Inspiration clinical FFDM system. A family of 350 BI-RADS *C* density 30%—fibroglandular anthropomorphic breast phantoms with compressed thicknesses of 30, 40, 50, 60, and 70 mm was generated for the study. The phantoms were divided into two sets populated with many spherical-shaped cysts or masses, with diameter of $6 \leq d_{\text{signal}} \leq 11$ mm. Figure 1 shows an MC-GPU simulated mammogram of a digital phantom with inserted solid masses. Each lesion in the phantom was assigned its own random x-ray attenuation profile $\mu_{i-\text{th lesion}} = f_i(E)$, according to the experimentally observed distribution described in the next section. Normal breast tissue and skin were modeled based on Woodard and White body tissue composition measurements.[8] Atomic compositions combined with density were used to generate energy-dependent

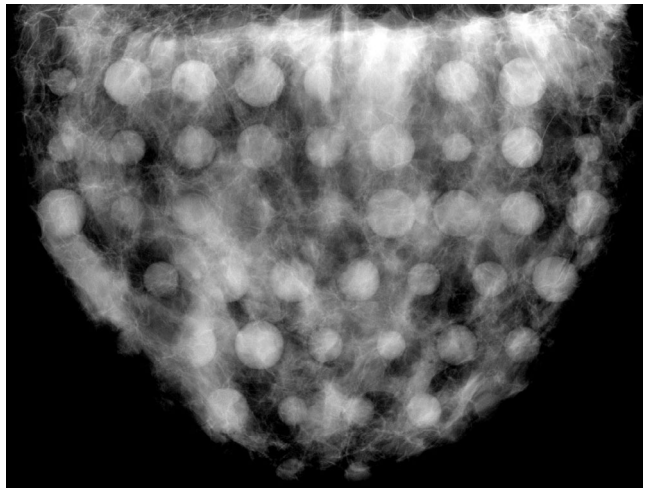

FIG. 1. Simulated breast mammogram with signals.

interaction cross-sections with the PENELOPE[9] (Monte Carlo algorithm) database, which includes all chemical elements in the periodic table. The same normal tissue/skin cross-sections were used both for solid and cystic mass simulations to minimize possible bias in classification performance.

Low- and high-energy exposures were modeled using 26 kVp W-Al$^{0.2\ \text{mm}}$ and 50 kVp W-Al$^{0.6\ \text{mm}}$ spectra, similar to that used by Ghammraoui et al.,[10] in a type-I/type-II breast calcification classification study. The total AGD was maintained via the number of photon histories to be ∽3 mGy across phantom thicknesses, and was split evenly between low- and high-energy shots.

Conventional FFDM simulations with a commonly used 30 kVp W-Rh$^{50\ \mu\text{m}}$ spectrum were carried out to compare single- vs dual-energy modality classification performance. No x-ray scattering processes were included in the simulations, assuming use of an imaging system with perfect scatter rejection. The AGD in conventional FFDM acquisitions was retained to be 3 mGy, equal to the total AGD in dual-energy mode. Table I summarizes the major Monte Carlo simulation parameters used in the study.

2.B. Lesion linear attenuation coefficient variability

Recently reported spectral attenuation measurements (Fredenberg et al.[4]) of breast mastectomy specimens with a poly (methyl methacrylate)(PMMA)/aluminum step

TABLE I. Exposure settings used in Monte Carlo experiments.

Phantom thickness, mm	30	40	50	60	70
Mode	Conventional FFDM	Dual-energy mammography			
Spectrum	W-Rh$^{50\ \mu\text{m}}$		Wh-Al$^{0.2\ \text{mm}}$		W-Al$^{0.6\ \text{mm}}$
kVp	30		26		50
AGD*, mGy	3.0		1.5		1.5

*Number of photon histories was adjusted to produce a target AGD for a given phantom thickness.

wedge were used to model cystic fluid and solid mass x-ray attenuation. In these experiments accurate data were obtained to represent x-ray attenuation of lesions in terms of a linear combination of equivalent PMMA and aluminum thicknesses normalized to a 10-mm-thick sample. Detector signal, obtained from such measurements, is identical for the tissue sample and for the equivalent combination of reference materials, independent of incident spectrum and detector response. The above methodology assumes full x-ray scatter rejection, which mostly is fulfilled in mammography systems with antiscatter grids. Equivalent thickness values included random fluctuations between measurements, sub-ROI and interimage variability, as well as, quantum noise in low- and high-energy acquisitions. In the simulations for each lesion in the phantom the attenuation curve was modeled using Eq. (1) with the base coefficients (A, B, C, D) sampled as normal random deviates $\mathcal{N}\{\mu, \sigma^2\}$ with the means and total variances as reported in Fredenberg.[4] Attenuation coefficients of PMMA and aluminum were modeled using the NIST Standard Reference Database[11]

$$\begin{aligned}\mu_{\text{cyst}}(E) &= A\cdot\mu_{\text{PMMA}}(E) + B\cdot\mu_{\text{Al}}(E)\\ \mu_{\text{mass}}(E) &= C\cdot\mu_{\text{PMMA}}(E) + D\cdot\mu_{\text{Al}}(E),\ \text{where}\\ A &= \mathcal{N}\{0.812,\ 0.61\%\},\ B = \mathcal{N}\{0.0323,\ 3.9\%\}\\ C &= \mathcal{N}\{0.859,\ 2.6\%\},\ D = \mathcal{N}\{0.0296,\ 9.5\%\}\end{aligned} \tag{1}$$

Figure 2(a) illustrates the spread of the attenuation curves for the two types of lesions. There is a significant overlap in the linear attenuation coefficient values for the lower photon energies, with slightly better separation observed at the higher energies. Distribution of the attenuation coefficient values for solid lesions is much broader for solid breast lesions than it is for cystic fluid and is due to the large spread of the equivalent material thicknesses mainly caused by natural variation among tumor samples in Fredenberg's measurements. Figures 2(b)–2(d) show distributions of μ for cysts and masses at 16, 26, and 40 keV correspondingly. It should be noted that although the two distributions at higher photon energies [Fig. 2(d)] are fully overlapping, the mean values of μ are distinctly different.

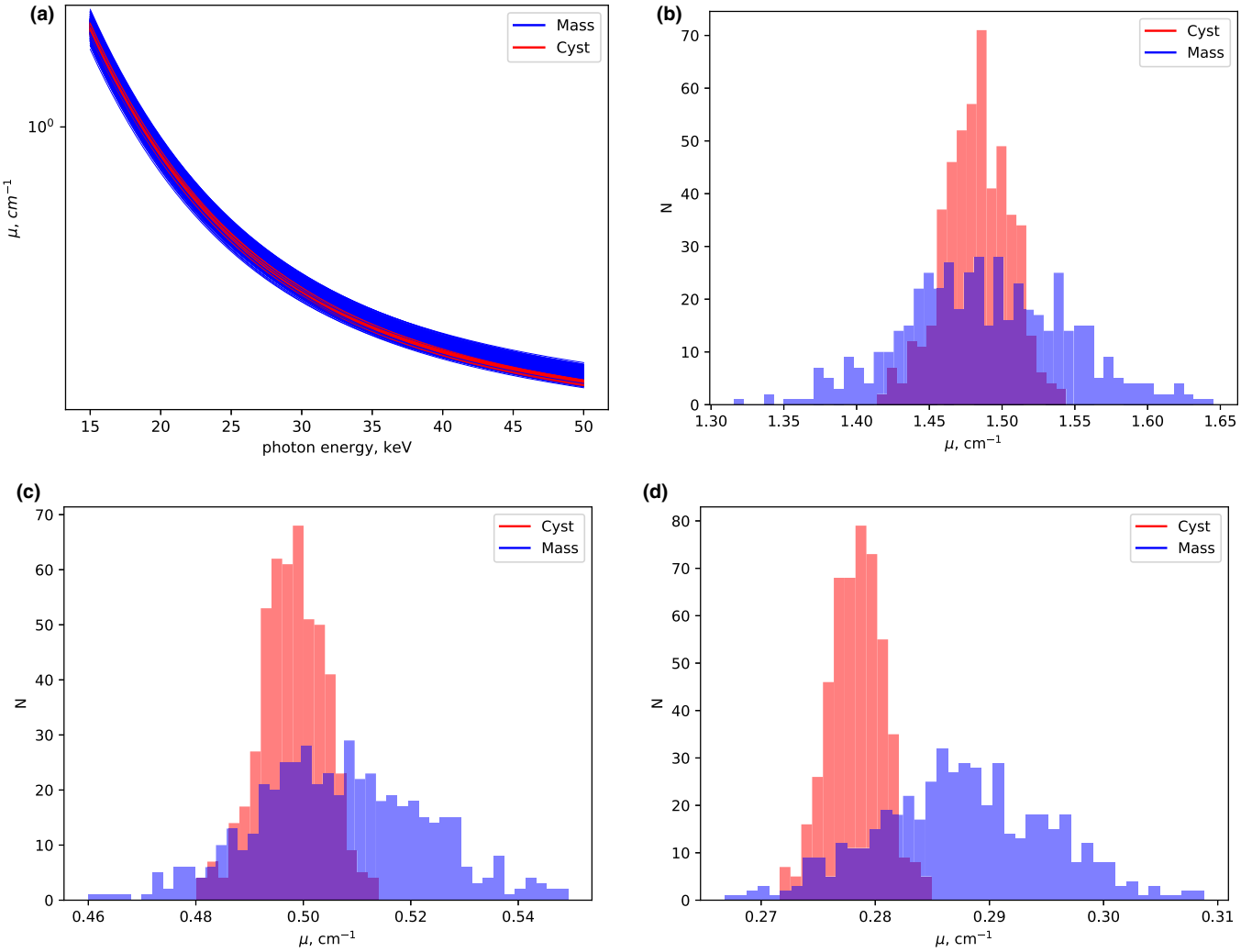

FIG. 2. Cyst/solid mass tissue linear attenuation coefficient variability based on measurements in Fredenberg et al. (2016).

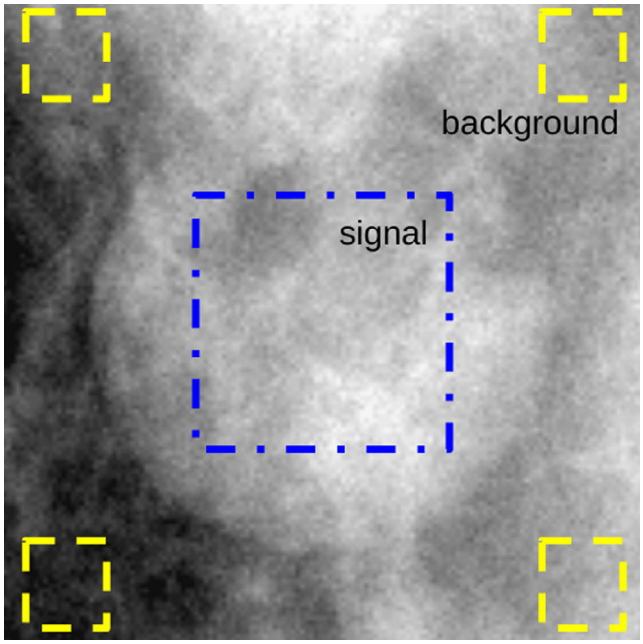

FIG. 3. Regions in the image used to compute λ.

2.C. X-ray density-based decision variable

X-ray density is one straightforward approach to defining an analytical classifier for distinguishing two tissue types with different x-ray attenuation properties using dual-energy spectral imaging[10,12]

$$D_{\mathrm{l,h}} = \ln\left(\frac{I_{\mathrm{l,h}}^{\mathrm{bkg}}}{I_{\mathrm{l,h}}^{\mathrm{sig}}}\right), \tag{2}$$

where the subscripts indicate either low- or high-energy acquisitions, and I is the average detector signal in the lesion or lesion-free regions of the mammogram, as indicated in Fig. 3. The decision variable is expressed by the ratio

$$\lambda = \frac{D_{\mathrm{h}}}{D_{\mathrm{l}}}, \tag{3}$$

and distributions of λ for cysts and masses can be used to construct an ROC curve.

2.D. Neural network binary image classifier

The ResNet-18 network, from the PyTorch deep learning library, was used for the binary image classification. The ResNet family of CNNs relies on a micro-architecture known as the residual module, in which the original input to the module is added to the output of a series of operations. This type of learning framework allows to train deeper networks with higher learning rates. Recent ResNet implementations include the pre-activation residual module with more optimal layer ordering for obtaining higher accuracy, as well as a more robust version of the original residual module. In our study, ResNet-18, used for image classification, consisted of an input layer, 16 hidden convolutional layers with different numbers of nodes, and one fully connected layer with a single binary output and a total of 11,689,512 trainable parameters.

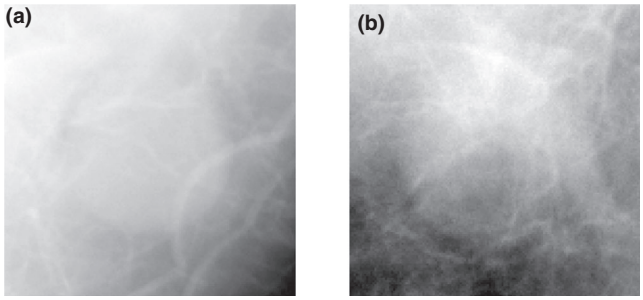

FIG. 4. Examples of ROIs with signals of approximately equal sizes (26 kVp acquisition shown).

For the input 140×140 px^2 (12 mm) regions of interest (ROIs), containing either cyst or mass, first were extracted from paired dual-energy acquisitions. The low-energy and high-energy ROIs of the same lesion then were arranged into a two-channel TIFF file of dimensions $140 \times 140 \times 2$ px^3 and passed to the neural network for training and testing. The network was trained from scratch with no normalization, standardization of pixel values, or other processing applied to the images. Sample cyst and mass regions of interest are shown in Fig. 4. A total of $6{,}860 + 6{,}860 = 13{,}720$ images were used to train the neural network, and $965 + 965 = 1{,}930$ were used for testing. Data augmentation in the form of horizontal and vertical flipping was used to reduce overfitting during training. The SGD optimizer with the initial learning rate of $\mathrm{LR}_0 = 0.005$, regularization parameter $R = 0.005$, and a learning rate scheduler with a 1/2 step-decay every 5 epochs was found to provide optimal convergence and stable training. The cross-entropy loss function was used for the two-class problem. The model was trained for 100 epochs, where the loss function attained a steady minimum. Training and testing sets included images from all phantom thicknesses, and their composition is shown in Fig. 5. Our training data included more ROIs from 40-mm-thick phantoms, because these phantoms already were available (Graff's[7] compressed breast phantom generation procedure uses 3D finite-element methods and is time-consuming). Such a distribution also partly resembled the compressed breast thicknesses observed clinically, that is, distributed approximately around 40 mm. The trained network was also tested separately on the images from only the 40-mm-thick phantoms.

3. RESULTS

3.A. Cyst vs mass classification using x-ray attenuation density method

Decision variable λ, defined in Eq. (3), was calculated for each cyst or mass ROI. The appeal of this figure of merit is that it is not dependent on breast thickness or signal size. ROIs with signals of both kinds (same as those used later for

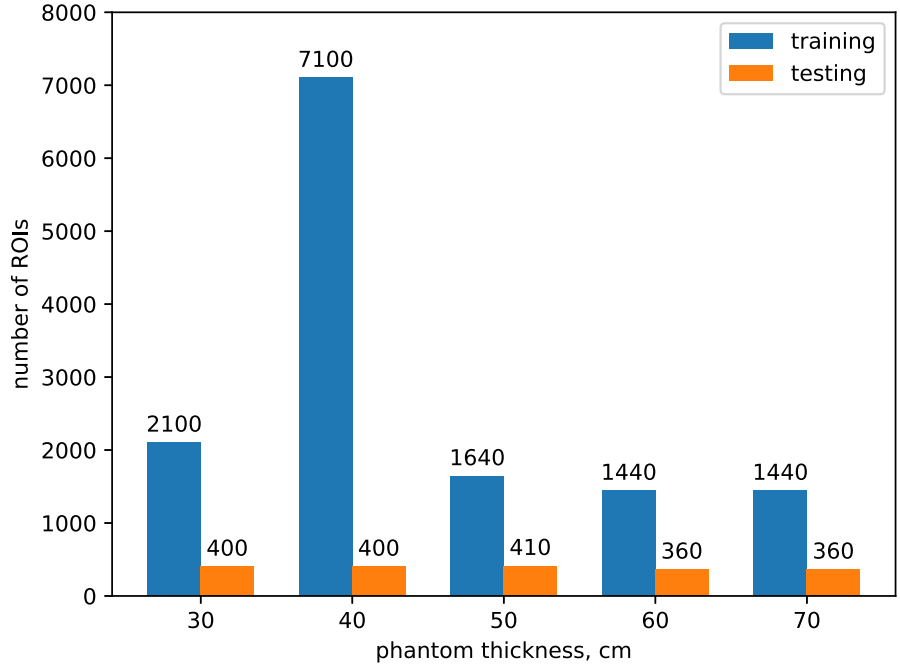

FIG. 5. Training/testing sets breast thickness composition.

the neural network testing) were used to calculate the distributions of λ for the ROC curve shown in Figs. 6(a) and 6(b) We then implemented the bootstrap method to obtain the 95% confidence interval bands on the ROC curve[13] and the error bars on the AUC values. The resulting AUC = 0.746 was unsatisfactory demonstrating the difficulty of the classification task. As a comparison, Ghammraoui et al. were able to achieve a perfect AUC = 1 to separate type-I/type-II microcalcifications with simulations and experimentally, using this method. It should be noted, however, that the difference in the linear attenuation coefficient between type-I/type-II microcalcifications was much greater than that between cystic fluid and solid masses.

3.B. Cyst vs mass classification using the neural network

3.B.1. Dual-energy mammography

Figure 7 summarizes our main results in the form of the ROC binary classification performance of the trained ResNet-18 network. The classifier was tested on the images from all phantom thicknesses equally represented (a conservative case), as well as on the images from 40-mm-thick phantoms only. Figs. 7(a)–7(b) show the respective ROC plots. The corresponding AUC values for the dual-energy mammography data are 0.98 and 0.987. Higher performance achieved on a

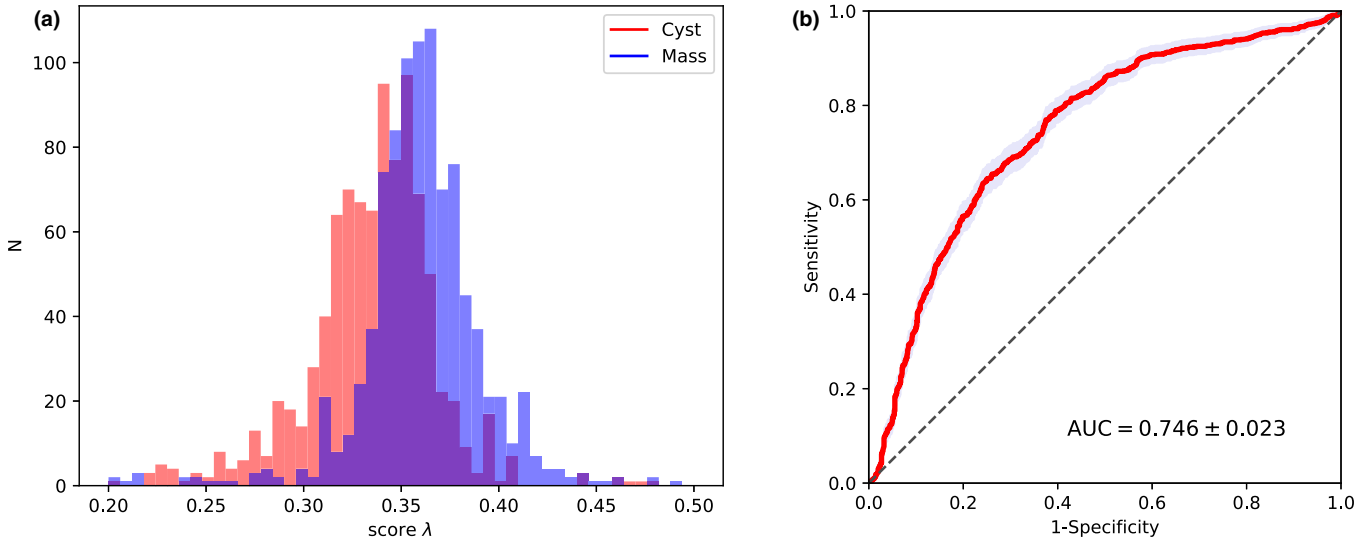

FIG. 6. Classification performance using x-ray density method.

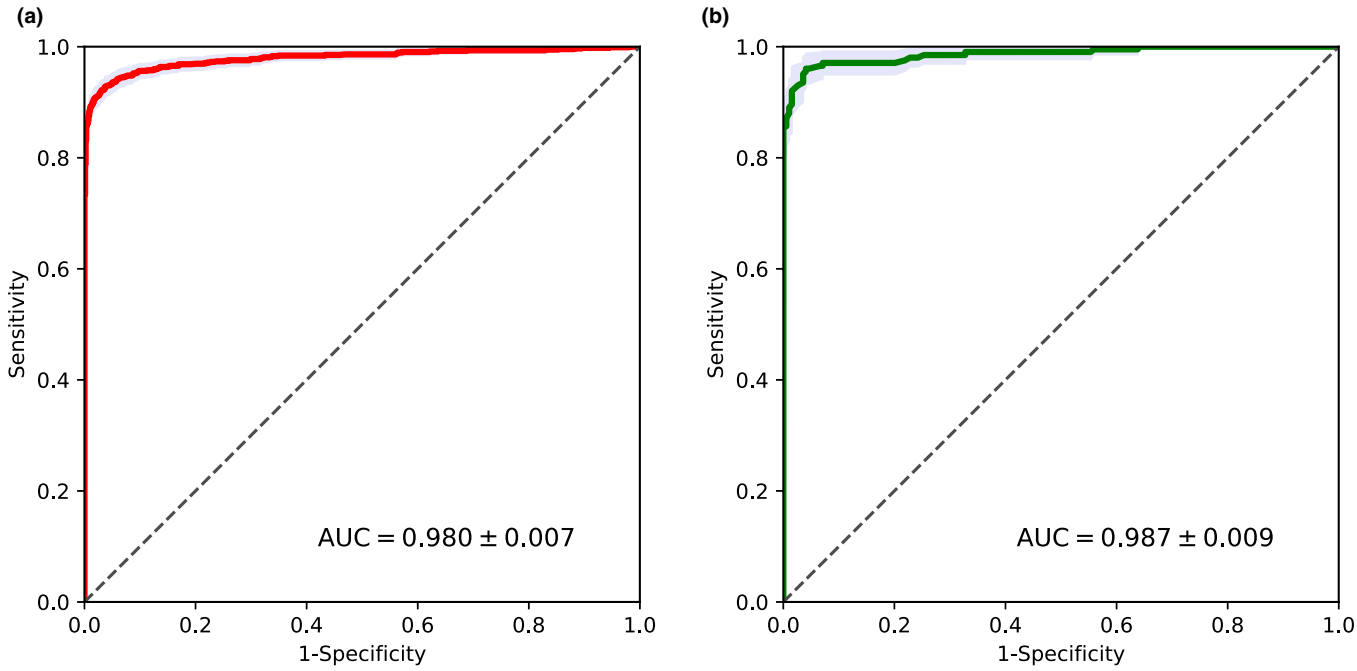

FIG. 7. Classification performance using the neural network.

40-mm-thick test set is explained by the amount of training images for this group and showed that more specific training would allow for better discrimination. Testing the trained network on the signal ROIs from 30, 50, 60, and 70 mm phantoms showed similar AUC performance, around 0.98, independent of phantom thickness.

Type classification ROC curves obtained from these simulations exhibited good behavior near the sensitivity = 1 line in the ROC plot, for example, enable to choose an operating point with very high sensitivity (to minimize the number of high-risk false negatives, when a mass is misclassified as a cyst), and acceptable specificity. For example, the median specificity at the 0.99 sensitivity level for all thicknesses and 40-mm-thick phantom test sets were 0.43 and 0.67, respectively. With sensitivity relaxed to 0.985 the model predicts 0.50 and 0.82 specificity for all data and 40-mm-thick phantoms, respectively. Since lesions were modeled as simple spheres in the simulations, we found it reasonable to assume that higher classification performance could be achievable if morphological differences (e.g., irregular shape, spiculations, leaky blood vessels in masses) between cysts and solid tumors were accounted for, when training the neural network. Likewise, using mammography systems with photon-counting detectors could provide improved spectral separation between low- and high-energy acquisitions and therefore would contribute to better signal discrimination.

3.B.2. Conventional (single-energy) mammography

At the same time we observed that the neural network trained on the single-energy (e.g., conventional mammography with a typical W-Rh 30 kVp spectrum) data was unable to learn the differences between the two types of lesions and showed 50% (random guessing) classification accuracy with the ROC curve close to a diagonal line. This was a somewhat anticipated result, given the large overlap in the linear attenuation coefficients and similar slopes of the attenuation curves for cysts and solid masses at lower photon energies.

3.C. Comparison to analytical classifier

Remarkable performance improvement with the CNN-based classifier in comparison to the x-ray density method ($\gtrsim$30% AUC increase) could, in our opinion, be explained by the latter model being relatively basic, relying on reduced information from the (high-energy/low-energy) ratio of x-ray density values. This analytical approach has no training aspect to it.

ResNet-18 network, on the other hand, uses a few thousand paired acquisitions of both lesion types to iteratively train several million parameters to recognize the differences in their x-ray attenuation. We speculate that the CNN does a significantly better job of estimating breast background and compression thicknesses, because it uses all available information in dual-energy data.

4. DISCUSSION

This study examined the possible use of the deep CNN approach to classifying cysts vs solid masses with dual-energy mammography. Since clinical dual-energy mammography data were not readily available, realistic Monte Carlo simulations were performed with digital anthropomorphic breast phantoms and embedded spherical lesions. Although solid mass lesions and cystic lesions typically have different

morphological characteristics, we assumed here that both lesion types were spherical. In doing this, the analysis focused on whether small differences in energy-dependent x-ray attenuation properties could be learned by the CNN to accurately classify these two types of lesions. To increase the difficulty of the classification task, variations in breast anatomical background, compression thickness (3–7 cm), lesion size (6–11 mm), and variability in lesion x-ray attenuation properties were modeled in the simulations. The CNN classifier was compared to a previously studied analytical classifier that used a classification decision variable related to x-ray density. It was reported to have an AUC of 0.980 (using all breast thicknesses), whereas the AUC for the analytical classifier was 0.746. It is important to note that the discrimination task studied here was between solid and cystic masses, not between malignant and benign lesions. Although most cystic masses are benign, in rare cases this is not true; for example, fluid-filled malignant cystic papillomas.

Kooi et al.[14] recently studied the use of a CNN-based CADx (computer-aided diagnosis) algorithm for discrimination of cysts from solid lesions in conventional mammography and reported a moderately accurate AUC of 0.80. Data used in their study included diagnostic exams of women who were recalled in screening based on suspicious mass lesions. Thus, the population of mass lesions was not restricted to round lesions. Worse performance with single-energy mammography might be expected if tested on round solid and cystic masses. Our study focused on use of dual-energy mammography using a training and testing set with no morphological differences between solid and cystic masses. The high accuracy achieved here (AUC = 0.98) would seem to suggest that additional information might be available to the CNN in dual-energy mammography as compared to conventional, single-energy mammography.

Another previously described clinical study by Erhard et al.[3] used a PCD-based dual-energy spectral mammography system and a two-feature discrimination algorithm to evaluate a set of 119 eligible ($>$10 mm diameter) cystic and solid mass breast lesions. Their approach was shown to achieve an AUC of 0.88 for this lesion classification problem. Although AUC is an important metric for evaluating classification performance, an operating point with high sensitivity for correctly diagnosing a lesion to be a cyst might be a more pertinent figure of merit, since the patient's clinical cost from misclassifying a malignant mass as a cystic lesion likely would be very high. For this reason, in addition to AUC, Erhard et al. reported a specificity of 61% at the 99% sensitivity level (on average) for all expert radiologists. The CNN classifier applied to the simulated data studied here performed comparably, resulting in specificity of 43% at the 99% sensitivity level for all breast thicknesses tested, and specificity of 67% at the 99% sensitivity level for the ensemble of 4-cm-thick compressed breasts tested.

Based on screening data from Erhard et al. (table 4 from[3]), this would correspond to a reduction of 24% and 38% of recalls, respectively, for well-defined solitary lesions. Given the large number of screening recalls in the United States, this reduction would be welcome. Thus, the CNN approach for classification of solid masses and cysts using dual-energy mammography appears feasible and should be investigated further with clinical data. Results from this and other studies[10,15–18] should motivate further development of spectral mammography systems.

This approach has several limitations. First and foremost, this is a simulation study, and whether the good classification performance observed here would also hold for dual-energy mammography data remains to be seen. Unfortunately, dual-energy mammography is not commonly used; thus testing data were not available. However, interest in dual-energy iodinated contrast mammography has increased recently, and three systems for this application have been approved by the FDA; thus, in the future validating the results of this study on clinical dual-energy mammography data might be possible. As Erhard et al.[3] have demonstrated, this classification technique also can be applied to PCD-based spectral mammography systems. In theory, photon-counting systems should show improved classification performance since there is less spectral overlap between the two energy windows, as well as less patient motion (dual-energy mammography requires two shots, whereas dual-energy imaging with a PCD requires only a single exposure). From the results of this study, one could hypothesize that CNN classification using data acquired from PCD-based mammography also should perform well.

Another limitation of this study is that the dual-energy acquisition parameters selected were not rigorously optimized. For example, dual-energy mammography was performed with equal dose of 1.5 mGy applied in the low- and high-energy windows, regardless of breast thickness. Commercial mammography systems typically set their automatic exposure algorithms (AECs) to increase the dose as the thickness of the compressed breast increases. Thus, more optimal dose settings might result in improved performance with various breast thicknesses (especially for thicker breasts). In addition, the kVp value and filter selections were based on previous findings for a different spectral mammography application.[10] It is expected that fine tuning of these dual-energy acquisition parameters would further improve performance over that reported here.

5. CONCLUSIONS

We explored the use of a common deep CNN (ResNet-18) for discriminating between solid masses and cysts in simulated dual-energy mammograms.

Results from realistic Monte Carlo experiments have shown that such a classifier can be trained to accurately discriminate between cystic fluid and solid mass tissue in dual-energy mammography by learning the differences in their energy-dependent x-ray attenuation properties. It can do this in the presence of several confounding factors, including variations in anatomical background, compressed breast thickness, lesion size, and variability in lesion x-ray attenuation. Our conservative modeling approach suggests that clinically usable classification performance (simulated AUC $>$ 0.95

with reasonable specificity at the 99% sensitivity level) can be achieved, provided that a sufficient amount of training data is available.

Although this work is a simulation study, the approach suggests promise for use in a clinical environment. This study focused on discriminating between round solid and cystic masses. If the study population included solid masses that were irregularly shaped and spiculated, we expect that additional benefit could be achieved by training the CNN to use morphological differences between the two lesion types.

ACKNOWLEDGMENTS

The authors acknowledge Dr. Frank Samuelson for helping with statistical analysis and Ms. Joanne Berger, FDA Library, for editing the manuscript.

CONFLICT OF INTEREST

The authors have no conflict to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

[a)]Author to whom correspondence should be address. Electronic mail: andrey.makeev@fda.hhs.gov

REFERENCES

1. Grabler P, Sighoko D, Wang L, Allgood K, Ansell D. Recall and cancer detection rates for screening mammography: finding the sweet spot. *Amer J Roentgen*. 2017;208:208–213.
2. Sickles EA. Probably benign breast lesions: when should follow-up be recommended and what is the optimal follow-up protocol? *Rad*. 1999;213:11–15.
3. Erhard K, Kilburn-Toppin F, Willsher P, et al. Characterization of cystic lesions by spectral mammography. *Investig Rad*. 2016;51:340–347.
4. Fredenberg E, Kilburn-Toppin F, Willsher P, et al. Measurement of breast-tissue x-ray attenuation by spectral mam- mography: solid lesions. *Phys Med Biol*. 2016;61:2595–2612.
5. Fredenberg E, Hemmendorff M, Cederstrom B, Aslund M, Danielsson M. Contrast- enhanced spectral mammography with a photon-counting detector. *Med Phys*. 2010;37:2017–2029.
6. Badal A, Badano A. Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit. *Med Phys*. 2009;36:4878–4880.
7. Graff CG. A new open-source multi-modality digital breast phantom. *SPIE Phys Med Imag*. 2016;9783:978309–978310.
8. Woodard HQ, White DR. The composition of body tissues. *Br J Radoil*. 1986;59:1209–1219.
9. Salvat F. The PENELOPE code system. specific features and recent improvements, *Ann Nucl Energy*. 2015;82:98–109.
10. Ghammraoui B, Makeev A, Zidan A, Alayoubi A, Glick SJ. Classification of breast microcalcifications using dual-energy mammography. *J Med Imag*. 2019;6:1–9.
11. N. I. of Standards and Technology, X-Ray Mass Attenuation Coefficients, https://www.nist.gov/pml/x-ray-mass-attenuation-coefficients, Online database version 1.4, July 12, 2004.
12. Breitenstein DS, Shaw CC. Comparison of three tissue composition measurement techniques using digital mammograms - a signal-to-noise study. *J Digit Imag*. 1998;11:137–150.
13. Macskassy SA, Provost FJ. Confidence bands for ROC curves: methods and an empirical study, Proc. of 1st intern. workshop ROC anal. In AI, ROCAI-2004, Spain; 2004.
14. Kooi T, van Ginneken B, Karssemeijer N, den Heeten A. Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network. *Med Phys*. 2017;44:1017–1027.
15. Ghammraoui B, Glick SJ. Investigating the feasability of classifying breast microcalcifications using photon-counting spectral mammography: a simulation study. *Med Phys*. 2017;44:2304–2311.
16. Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS. Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using Raman Spectroscopy. *Cancer Res*. 2002;62:5375–5380.
17. Veldkamp WJ, Karssemeijer N, Otten JD, Hendriks JH. Automated classification of clustered microcalcifications into malignant and benign types. *Med Phys*. 2000;27:2600–2608.
18. Kim H, Lee M, Kim D, Lee D, Kim H-J. Evaluation of photon-counting spectral mammography for classification of breast microcalcifications. *Rad. Phys Chem*. 2019;162:39–47.