International Journal of Computer Assisted Radiology and Surgery (2025) 20:1541–1549
https://doi.org/10.1007/s11548-025-03402-4

ORIGINAL ARTICLE

Uncertainty estimation for trust attribution to speed-of-sound reconstruction with variational networks

Sonia Laguna[1] · Lin Zhang[1] · Can Deniz Bezek[2] · Monika Farkas[3] · Dieter Schweizer[1] · Rahel A. Kubik-Huch[3] · Orcun Goksel[1,2]

Received: 25 February 2025 / Accepted: 14 April 2025 / Published online: 10 June 2025
© The Author(s) 2025

Abstract

Purpose Speed-of-sound (SoS) is a biomechanical characteristic of tissue, and its imaging can provide a promising biomarker for diagnosis. Reconstructing SoS images from ultrasound acquisitions can be cast as a limited-angle computed-tomography problem, with variational networks being a promising model-based deep learning solution. Some acquired data frames may, however, get corrupted by noise due to, e.g., motion, lack of contact, and acoustic shadows, which in turn negatively affects the resulting SoS reconstructions.

Methods We propose to use the uncertainty in SoS reconstructions to attribute trust to each individual acquired frame. Given multiple acquisitions, we then use an uncertainty-based automatic selection among these retrospectively, to improve diagnostic decisions. We investigate uncertainty estimation based on Monte Carlo Dropout and Bayesian Variational Inference.

Results We assess our automatic frame selection method for differential diagnosis of breast cancer, distinguishing between benign fibroadenoma and malignant carcinoma. We evaluate 21 lesions classified as BI-RADS 4, which represents suspicious cases for probable malignancy. The most trustworthy frame among four acquisitions of each lesion was identified using uncertainty-based criteria. Selecting a frame informed by uncertainty achieved an area under curve of 76% and 80% for Monte Carlo Dropout and Bayesian Variational Inference, respectively, superior to any uncertainty-uninformed baselines with the best one achieving 64%.

Conclusion A novel use of uncertainty estimation is proposed for selecting one of multiple data acquisitions for further processing and decision making.

Keywords Ultrasonography · Image reconstruction · Breast cancer differential diagnosis

Introduction

Breast cancer is a leading cause of cancer-related mortality in women [1]. Ductal carcinoma (CA) is the most frequent type of malignant breast lesion, while fibroadenoma (FA) is the most common type of benign breast lesion [2]. In breast cancer, early detection is the key in reducing mortality rate [1], emphasizing the need for fast, reliable, and efficient techniques to detect and classify lesions. Current gold-standard diagnostic and screening methods, such as mammography and MRI, have several limitations, e.g., ionizing-radiation, not being in real time, and reduced sensitivity with increased breast density or increased parenchymal background enhancement [3]. While biopsy can provide exact tissue information, it is localized only in several tissue points, and carries a risk of complications and places a burden on both the patient and the clinical pathology pipelines.

Ultrasound (US) offers a non-ionizing, real-time, and cost-effective imaging alternative, but conventional B-mode images do not provide sufficient information for differentiating malignant lesions. Speed-of-sound (SoS) is a promising quantitative biomechanical marker that can provide information about the pathological state of the tissue. Reconstructing the SoS distribution was proposed using through-

✉ Orcun Goksel
orcun.goksel@it.uu.se

1 Computer-assisted Applications in Medicine, ETH Zurich, Zurich, Switzerland

2 Department of Information Technology, Uppsala University, Uppsala, Sweden

3 Department of Radiology, Kantonsspital Baden (Affiliated Hospital for Research and Teaching of the Faculty of Medicine of the University of Zurich), Baden, Switzerland

transmission US computed tomography. However, such systems necessitate a complex hardware setup, submersion of the breast in water, and a trained operator—complicating the clinical routine. Pulse-echo methods eliminate the aforementioned limitations and enable SoS estimation with conventional ultrasound transducers. Some of these use echo profiles [4], optimize an image quality metric [5], relate the speckle shifts between different transmissions to SoS [6–11], or reconstruct SoS with black-box networks [12]. A conventional hand-held US transducer operated in pulse-echo mode has shown promising results in differentiating CA and FA, using the distinct SoS contrasts of these inclusions compared to background [13].

US imaging with pulse-echo SoS is an ill-posed limited-angle inverse problem, which is typically solved analytically with hand-crafted regularizers [8, 9]. Black-box deep learning [12, 14] or pre-/post-processing of analytical solutions with neural networks [15] have been proposed for image reconstruction. Yet, in cases where annotated in vivo data are not available, model-based deep learning through loop unrolling was found to be more effective [16]. Variational network (VN) is a model-based image reconstruction framework and enables prior learning in restricted analytical forms within algorithm loop unrolling [17]. This allows generalizability to in vivo from training from only simulated data, with its success shown for various imaging inverse problems, e.g., compressed sensing for MRI [18], limited-angle X-ray computed tomography [19], and imaging SoS [20]. As the output of a deep learning method may vary largely depending on various aspects, estimating uncertainty in such outputs is of utmost importance, especially for high-risk applications such as medical imaging and decision making, as was studied for MRI [21–23].

In this work, we study uncertainty estimation in SoS imaging, with the following key contributions: (1) we propose a novel use of uncertainty estimates as trust attribution to select the best acquisition frame in US imaging, leveraging its rapid and interactive acquisition capability; (2) we apply uncertainty estimation to SoS reconstruction for the first time; (3) we introduce a "relative" (normalized) uncertainty metric superior in trust attribution in regression-like tasks; (4) we demonstrate the clinical feasibility of this approach for breast cancer diagnosis with SoS imaging, including the first in vivo application of VN for SoS reconstruction; and (5) we present an algebraic reformulation for efficient uncertainty computation for large models on standard hardware. We implement this based on VN models and using well-studied uncertainty estimation frameworks, Monte Carlo Dropout and Bayesian Variational Inference [22, 24]. These innovations advance the reliability and clinical applicability of SoS. The framework is illustrated in Fig. 1.

Methods

SoS imaging and variational network

The approach in this work relies on measuring speckle shifts between beamformed frames obtained from different transmit/receive (Tx/Rx) sequences. A speckle observed under varying Tx/Rx parameters appears at shifted locations if the SoS changes along the propagation paths associated with these parameters. Our method relates local SoS variations, inferred from the beamforming SoS, to the observed echo shifts along different propagation paths, using an imaging model that accounts for the wave propagation. To reconstruct SoS distributions, beamformed images are generated using distinct Tx/Rx parameters, which are selected to ensure that the resulting echo shifts are sensitive to spatial SoS variations. Images are then beamformed from the acquired raw data using an assumed tissue SoS value (c_0). Apparent speckle shifts (displacements) are found between pairs of M beamformed frames (see Fig. 1a). Finally, a SoS map is reconstructed by solving the following inverse problem: $\hat{x} = \arg\min_x \|L(x - x_0) - d\|_p + \lambda\mathcal{R}(x)$, where $x \in \mathbb{R}^{N_x N_z}$ is the sought (vectorized) slowness (inverse SoS) map on a discretized $N_x \times N_z$ reconstruction grid, and x_0 is the initial beamforming slowness ($1/c_0$). Vectorized time delay (displacement) measurements on a $N_{x'} \times N_{z'}$ grid concatenated from M image pairs are denoted by $d \in \mathbb{R}^{MN_{x'}N_{z'}}$. Sensitivity of measurements to SoS map is then encoded by a sparse forward-model matrix $L \in \mathbb{R}^{MN_{x'}N_{z'} \times N_x N_z}$. The first term in the inverse problem corresponds to data fidelity, for which a norm p is chosen based on expected measurement noise, e.g., l_2-norm for Gaussian noise. The second term corresponds to regularization, with λ controlling its impact. The above optimization is typically solved with iterative analytical methods, with LBFGS [25] being an efficient option. The SoS distribution is computed as $\hat{c} = 1/(\hat{x} + x_0)$.

Conventional regularization priors penalize total variation of image gradients, e.g., via l_1-norm, i.e., $\lambda\|R_{\text{grad}}x\|_1$. A more general form can use multiple priors as field of experts [26], i.e., $\sum_{i=1}^{N} \lambda_i \|R_i x\|_{t_i}$. For instance, total variation can be expressed as $\lambda\|r_x * x\|_1 + \lambda\|r_y * x\|_1$ based on lateral and axial gradient filters r_x and r_y. A useful subset of such linear priors can be effectively represented using convolutions, i.e., as $\lambda_i\|r_i * x\|_{t_i}$ with $*$ denoting convolution and r_i the kernels. This form then can be extended with spatially-varying weighting (instead of a fixed λ for the entire image) and arbitrary penalty transformations (instead of hand-crafted norms). This can then model a wide range of useful regularizers in the form of $w_i \cdot \phi_i(r_i * x)$, where $w_i \in \mathbb{R}^{a_1 a_2}$ is a vector with weights for each image location and ϕ_i is a transformation function. To enable learning, ϕ_i for mapping (scaling) values is parameterized in a normalized space via

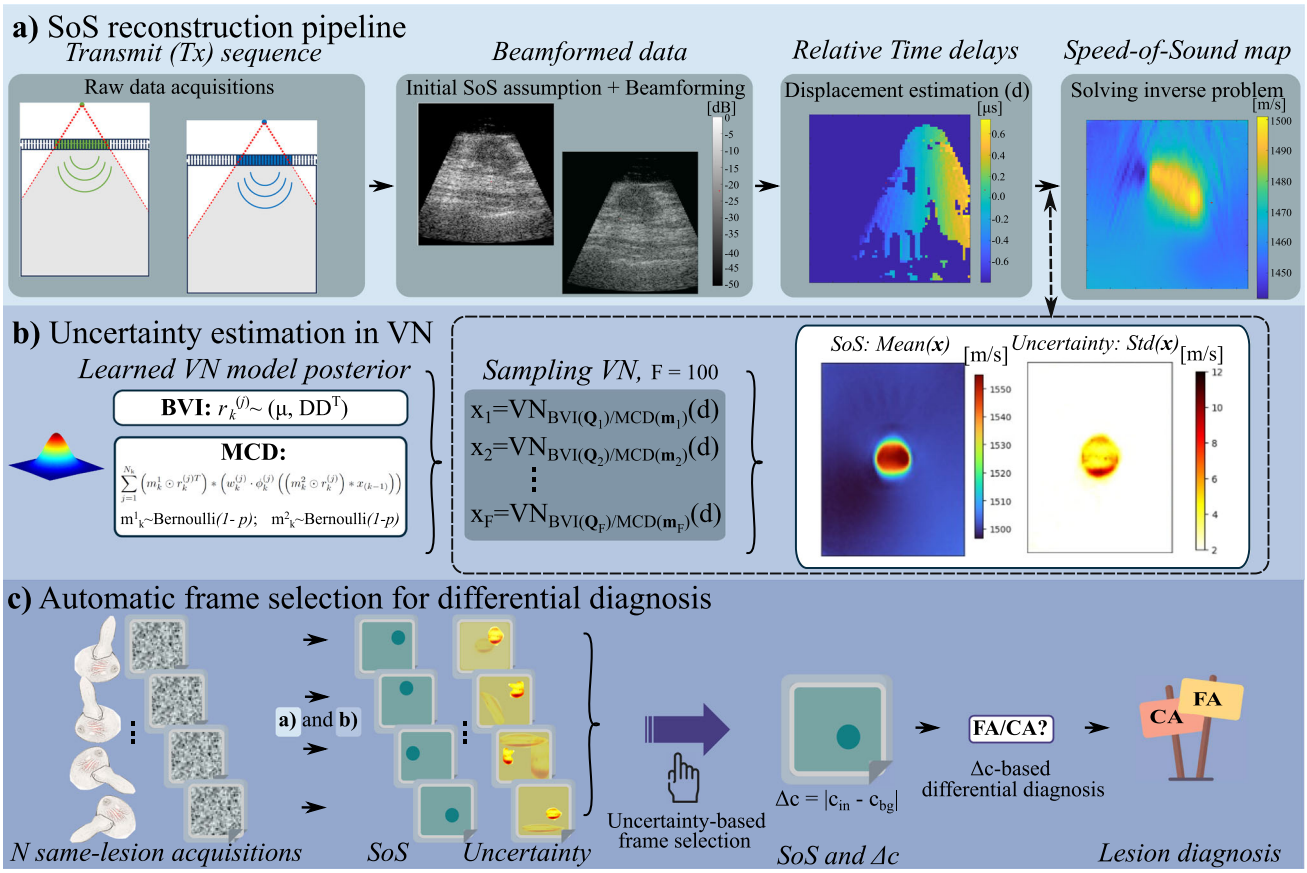

Fig. 1 **a** SoS reconstruction pipeline on a clinical example: Echo data from different transmit events are beamformed, between pairs of which the displacements are estimated. Using multiple displacement maps, a SoS map is reconstructed by solving an inverse problem with VN. **b** The VN posterior is learned using Bayesian variational inference or Monte Carlo Dropout. At inference, samples are drawn from each posterior, with their mean being the reconstructed SoS image and the standard deviation the uncertainty estimate. **c** Data from the same lesion is collected multiple times. Reconstruction uncertainty is used to select an optimal acquisition, later used for breast cancer differential diagnosis

some control points. This can then represent typical norms as well as non-norm functionals.

With its accuracy shown in various imaging inverse problems [10, 18–20, 27], VN is a model-based deep learning framework that learns data fitness and regularization in a restricted analytical form, via unrolling K iterations of an optimization algorithm, e.g., the gradient descent for SoS as follows:

$$x_k = x_{(k-1)} - \left[(s_k L)^{\mathrm{T}} \psi_k (s_k (L x_{(k-1)} - d)) + \lambda \nabla R(x_{(k-1)}) \right]. \tag{1}$$

The first term corresponding to data gradient involves the penalty transformation $\psi(.)$ and the spatial weight s_k (desired fidelity to each time delay measurement location) as learnable parameters at the k-th unrolled layer. The second term is the regularization gradient, in a restricted analytical form as:

$$\nabla R(x_k) = \sum_{j=1}^{N_k} r_k^{(j)T} * \left(w_k^{(j)} \cdot \phi_k^{(j)} \left(r_k^{(j)} * x_{(k-1)} \right) \right), \tag{2}$$

where N_k is the total number of priors in the VN field of experts model. A weighted exponential loss was applied across the unrolling layers [19], see Supplementary Material.

Uncertainty estimation

Monte Carlo Dropout (MCD)

Dropout is a technique that randomly deactivates a fraction of weights in a neural network with a specified probability, namely dropout rate. It is commonly used during training to reduce overfitting. Monte Carlo Dropout (MCD) applies dropout also at inference time, to sample multiple out-

come hypotheses and construct the posterior distribution of a model [24]. Generally, black-box learning methods apply dropout at the individual weight level. For the restricted analytical form of VN, however, dropping individual (e.g., kernel, spatial, transformation function control points) weights are not meaningful, since these are only relevant with respect to each other, e.g., within a kernel. Accordingly, although our MCD is based on kernel (filter) removal, it corresponds to removing the respective prior component. Analytically, the probability of a component remaining active in the VN follows a Bernoulli distribution with parameter $1-p$. Concretely, we choose to dropout the filters, $r_k^{(j)}$, which are used to model the regularization term. The filters are multiplied element-wise with the independent Bernoulli variables that act as a mask to the input variable:

$$\sum_{j=1}^{N_k}\left(m_k^1 \odot r_k^{(j)T}\right) * \left(w_k^{(j)} \cdot \phi_k^{(j)}\left(\left(m_k^2 \odot r_k^{(j)}\right) * x_{(k-1)}\right)\right), \tag{3}$$

where $\mathrm{m}_k^1 \sim \text{Bernoulli}(1-p)$ and $\mathrm{m}_k^2 \sim \text{Bernoulli}(1-p)$ are the layer dropout masks, with each element being dropped with probability p. To compute the final uncertainty and SoS reconstruction results, the models were sampled at test time.

Bayesian Variational Inference (BVI)

Bayesian variational inference (BVI) aims to model the posterior distribution of intrinsic model parameters. In particular, we learn the distribution of filters $r_k^{(j)}$. Following [22], we model the filters as a multivariate Gaussian distribution $N(\mu, \Sigma)$, with learned mean $\mu \in \mathbb{R}^t$ and learned covariance matrix $\Sigma \in \mathbb{R}^{t\times t}$, where t is the total number of all filters. Since Σ is a positive-definite and symmetric matrix, for computation and memory efficiency, we decompose it as $\Sigma = DD^\top \in \mathbb{R}^{t\times t}$, where $D \in \mathbb{R}^{t\times t}$ is the lower triangular Cholesky factor. With the assumption of filters being uncorrelated with each other, Σ admits a block diagonal structure, with $o^2 \times o^2$ blocks, where o is the filter size. Samples from $r_k^{(j)}$ can be directly obtained as $\mu + Dy$, with $y \sim \mathcal{N}(0, \mathrm{I_d})$.

An additional Kullback–Leibler (KL) loss is used to prevent degenerate models. Particularly, it measures the difference between two Gaussian distributions q_1 and q_2:

$$\mathrm{KL}(q_1\|q_2) = \frac{1}{2}\log\frac{|\Sigma_2|}{|\Sigma_1|} + \mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)^\top \Sigma_2^{-1}((\mu_2-\mu_1)-t), \tag{4}$$

where our sought distribution $q_1 = \mathcal{N}(\mu, \Sigma)$ is regularized by normal distribution $q_2 = \mathcal{N}\left(\mu, \alpha^{-1}\mathrm{I_d}\right)$, with $\alpha > 0$ being a hyperparameter as a prior. Neglecting constant terms for optimization, $\mathrm{KL}(q_1\|q_2) \approx \alpha\,\mathrm{tr}(DD^\top) - \log(\det(DD^\top))$. As the determinant of DD^T is computationally expensive and intractable, in this paper we propose to use the following reformulation:

$$\begin{aligned}\log\left(\det\left(DD^T\right)\right) &= \log\left(\det(D)^2\right) = 2\log\prod_i D_{ii} \\ &= 2\sum_i \log\left(D_{ii}\right) = 2\,\mathrm{tr}(\log(D)).\end{aligned} \tag{5}$$

Therefore, $\mathrm{KL}(q_1\|q_2) \approx \alpha\,\mathrm{tr}(DD^\top) - 2\,\mathrm{tr}(\log(D))$, and the final BVI loss becomes

$$\begin{aligned}\mathcal{L}_{\mathrm{BVI}} = &\sum_{k=1}^{K}\exp^{-\tau(K-k)}\left\|x_k - x^*\right\|_1 \\ &+ \lambda_r\sum_{k=1}^{K}\sum_{j=1}^{N_j^{\phi_k}}\sqrt{\left(y_{j-1}^{\phi_k} - 2y_j^{\phi_k} + y_{j+1}^{\phi_k}\right)^2 + \varepsilon} \\ &+ \beta\left(\alpha\,\mathrm{tr}(DD^\top) - 2\,\mathrm{tr}(\log(D))\right),\end{aligned} \tag{6}$$

where $\beta > 0$ weights the KL term to balance its effect in the final loss.

MCD and BVI were sampled at test time to estimate (1) the SoS reconstruction with their mean and (2) uncertainty with their standard deviation (see Fig. 1b). Computation steps are detailed in Algorithm S1 in Supplementary Material.

Automatic selection of acquisition data frame

In clinical practice, multiple data acquisitions are often performed to image one tissue location. Since these acquisitions may be subject to noise, such as that caused by hand/body motion affecting SoS reconstruction [9], selecting the best acquisition is crucial for diagnostic decisions and downstream tasks. We propose to use the prediction uncertainty as a surrogate for such trust attribution to image reconstructions and hence acquisitions, cf. Fig. 1c. Our goal is to automatically select one acquisition, and equivalently the corresponding reconstruction, out of N possibilities.

Given uncertainty estimations, one approach would be to select a reconstruction with minimum uncertainty within an inclusion (given its operator annotated location from B-mode). Frame *selection informed* by such *inclusion* uncertainty is hereafter referred as $\mathrm{SI}^{\mathrm{inc}}_{\mathrm{MCD}}$ or $\mathrm{SI}^{\mathrm{inc}}_{\mathrm{BVI}}$, depending, respectively, on the uncertainty assessment approach. Since absolute uncertainty value may vary significantly depending on the imaged tissue region, relying on the inclusion or total image uncertainty as a surrogate for assessing reconstruction confidence was found unreliable in our preliminary tests. To address this, we introduce a normalized (*relative*) uncertainty metric, defined as the absolute difference between the

inclusion uncertainty mean and the *background* uncertainty mean, with background defined as the 5 mm ring surrounding the inclusion. Frame selection using minimum *relative* uncertainty is hereafter referred as SI^{rel}_{MCD} or SI^{rel}_{BVI}, correspondingly. Computation of the above per-frame metrics and frame selection based on them are formalized in Algorithms S2 & S3 in Supplementary Material.

As baseline comparisons, we evaluate the diagnostic prediction performance *without* informed frame selection. For this, we use either simply the first acquisition (S1) or the third acquisition (S3) out of N. Additionally, we assess the results for choosing an acquisition at random (SR) for each patient. We study these baselines for the reconstructions from the analytical method ($S1_{LBFGS}$, $S3_{LBFGS}$, SR_{LBFGS}), from the original VN ($S1_{VN}$, $S3_{VN}$, AR_{VN}) and from the two uncertainty estimation approaches ($S1_{MCD}$ & $S1_{BVI}$, $S3_{MCD}$ & $S3_{BVI}$, and SR_{MCD} & SR_{BVI}).

Experimental setup

Simulated data

The VN to solve the inverse SoS problem requires supervision at training time. However, ground-truth local SoS distributions are unavailable and infeasible for in vivo data, as this would require voxel-wise biomechanical testing of excised samples. Therefore, while testing is conducted on in vivo patient data, during training we use only simulated data, as our utilized model-based reconstruction network enables us to bridge such large domain gap. For pulse-echo SoS reconstruction, we use virtual source transmit (VS) sequences. For each acquisition, 17 VS transmits are emitted, within 15 pairs of which the displacements are tracked [9]. Each simulated training sample, hence, contains 15 displacements maps and one ground-truth SoS map of the imaged area. For simulations and in vivo, there is a same, single imaging operator, built by vertically stacking discretized acoustic path integral matrices [8] for these said 15 pairs. The data are simulated in two ways.

First, we generate 10 000 training samples using a ray-based (RB) approach, applying the known forward imaging model. To avoid inverse crime by overfitting to the imaging operator, we incorporate discretization and noise effects by following the data preprocessing in [19], by actually simulating high-resolution noisy measurements $d_{hr} = L_{hr} x_{hr} + \epsilon$, using a high-resolution imaging model L_{hr} and slowness map x_{hr}, with ϵ Gaussian noise. SoS maps are then reconstructed on the original coarser grid, for the VN filters to learn to resolve noise and discretization artifacts.

Second, we generate 840 training samples using a sophisticated physics-based simulation approach, based on the k-Wave Matlab toolbox. We simulated multi-static full-matrix single-element transmission, recording the RF echo signals at each receive (Rx) element for each Tx. We then retrospectively employ simulated Tx beamforming on such RF data to obtain the data for the desired 17 VS, for each and every training sample. The Rx beamforming is subsequently performed on these signals using delay-and-sum beamforming, followed by displacement tracking for each of the 15 pairs with a normalized cross-correlation algorithm. This process more accurately represents real data acquisition settings but is also more computationally costly. For both cases, the ground-truth SoS maps were designed to contain randomly shaped inclusions, created from deformations of random ellipses. Half of these inclusions were filtered to generate smooth edges. The background SoS value was varied slightly, to make the network more robust to such changes. Additionally, 5% of the samples did not include any inclusion to prevent the network from always predicting an inclusion. Finally, using both approaches a test set of 32 samples was generated to verify the *reconstruction* accuracy of the *uncertainty-aware models* against the baseline VN. The data generation pipelines are illustrated in Figure S1 in Supplementary Material. For more details on data generation, we refer the reader to [20].

Clinical data

Clinical data was collected in vivo from breast lesions in a study conducted at Kantonsspital Baden, Switzerland, with ethics approval and informed consent (EKNZ, Switzerland, BASEC 2020-01962). Each inclusion was imaged $N = 3$ to $N = 5$ times from different directions, with 17 VS transmits yielding 15 displacement maps as input to the VN for SoS reconstruction. We focus on differentiating FA and CA in patients in BI-RADS 4 class, which involves suspicious cases challenging for classification. Out of 21 BI-RADS 4 patients, 8 had biopsy-confirmed CA and 13 had biopsy-confirmed FA.

Implementation details

The measurements and reconstructions are obtained on a 84×64 grid. As the analytical baseline, we employed the LBFGS algorithm. The models were implemented in Python and Tensorflow 1.11 and were ran on a 12 GB NVIDIA TITAN Xp GPU. The training used mini-batch optimization, ADAM optimizer, 120 000 iterations, and a batch size of 16 including a mix of 12 ray-based and 4 k-Wave samples. Dropout probability p for MCD was set to 0.25. For BVI, kernel size $o = 8$, $\alpha = 0.1$, and $\beta = 10$ were used. For the initialization of the Gaussian parameters, we sampled the means normally with standard deviation of 10^{-2} and the covariance matrix diagonal uniformly between 0.9 and 1.1. Uncertainty-based frame selection comes at a computation cost, mainly because inference needs to be run for

many sample repetitions, but also because Bayesian weights bring an overhead on the BVI model computation. We chose 100 samples, empirically based on a test image as a safe upper-bound to ensure convergence of uncertainty statistics. For constraints on computation time, a lower number of samples can be tested and may work as well. Table S1 in Supplementary Material lists the parameter dimensions and the hyperparameter values used.

Evaluation

Our selected frame is used for breast cancer differential diagnosis. For this diagnosis, we use the SoS contrast metric, which was shown to be effective in differentiating breast cancer. This metric is defined here as $\Delta c = |c_{\text{inc}} - c_{\text{bkg}}|$, i.e. the difference between the median inclusion SoS and the median background SoS, with the latter defined as the same 5 mm ring as for relative uncertainty estimation. For the in vivo data, inclusion masks were annotated by an expert on B-mode images.

In simulations where the ground-truth SoS map c is available, root-mean-square-error RMSE$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(c_i - \hat{c}_i)^2}$ over n reconstructed pixels was calculated to assess reconstruction accuracy. For the clinical data, we assess performance in differentiating FA and CA cases, assuming a higher inclusion SoS contrast Δc for the latter. Specifically, we analyze Δc distributions between the two lesion groups. To quantify the classification performance, we compute receiver-operating characteristic (ROC) curves, area under the curve (AUC) and F1 score, i.e. $\frac{\text{TP}}{\text{TP}+\frac{1}{2}(\text{FP}+\text{FN})}$, where TP are true positives, FP false positives, TN true negatives, and FN false negatives. We also report the sensitivity ($\frac{\text{TP}}{\text{TP}+\text{FN}}$) and specificity ($\frac{\text{TN}}{\text{TN}+\text{FP}}$), particularly at an operating point maximizing their sum. For statistical significance, we used Wilcoxon rank-sum as an unpaired nonparametric test, since the data was observed not to be parametric given Levene's test for homogeneity of variance and the Shapiro test for normality.

Results

Simulated data

We first tested whether adding uncertainty estimation on VN compromises its reconstruction quality. For this, we used the simulation test set with ground-truth maps, with the resulting RMSEs shown in Fig. 2. As seen, the reconstruction accuracies of VN augmented with the proposed uncertainty methods are not inferior to VN alone. Reconstruction from k-Wave data is expectedly more challenging (as with higher RMSE) than RB data. Sample SoS reconstructions of RB data are shown in Fig. 3.

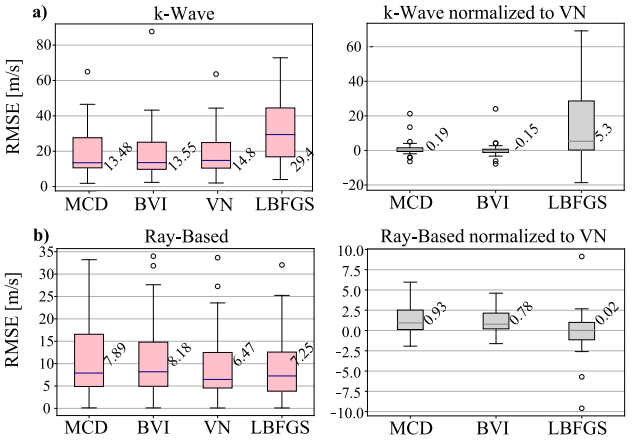

Fig. 2 Reconstruction errors in **a** the k-Wave test data, and **b** the RB test data. (Left) RMSEs of the different reconstruction methods, and (right) image-wise differential RMSEs with respect to VN

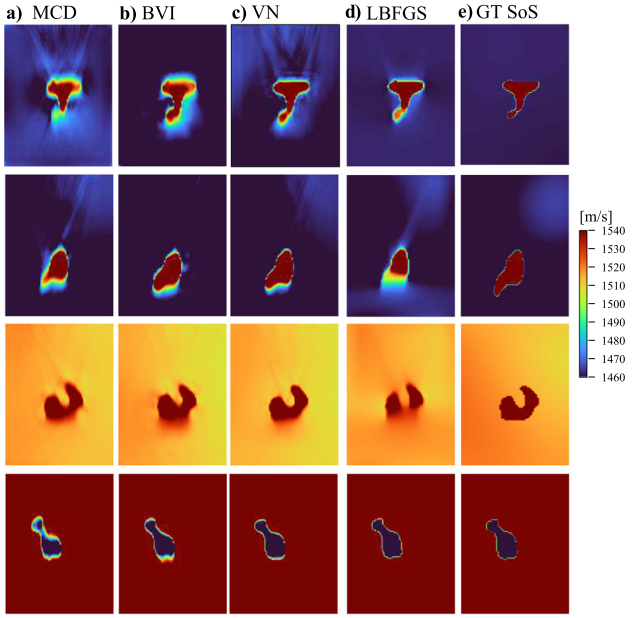

Fig. 3 Example reconstructions of four simulated ray-based numerical phantoms using each method (**a**–**d**), compared to the corresponding ground-truth (GT) SoS maps (**e**)

In vivo data

To use mean and standard deviation of uncertainty estimates, we first qualitatively checked the hypothesis of normal distribution of the network produced outputs. Figure S2 in Supplementary Material shows the flute plots of inclusion SoS contrast Δc distribution across 100 samples for 8 sample patients, with the results showing normality. For the frame selected based on uncertainty (based on pixel-wise sample SoS variation), Δc is then calculated for the sample mean SoS image. AUC and F1-score of the methods for differentiating CA from FA are shown in Table 1, indicating that

Table 1 Contrast-based lesion classification results of SoS reconstruction approaches based on selecting the first (S1), third (S3), and a random (SR) frame, as well as informed by inclusion uncertainty (SI^{inc}) and relative (inclusion normalized by background) uncertainty (SI^{rel})

Method	AUC (%)				F1 Score (%)			
	LBFGS	VN	BVI	MCD	LBFGS	VN	BVI	MCD
S1	54.81	49.04	54.81	58.65	55.17	61.54	58.82	57.14
S3	59.62	54.81	51.90	51.90	63.59	51.94	51.82	51.82
SR	63.47	45.21	54.42	59.61	64.02	59.28	58.81	63.59
SI^{inc}	–	–	61.54	61.54	–	–	63.16	61.54
SI^{rel}	–	–	**73.08**	**75.96**	–	–	**80.00**	**70.59**

Bold indicates best metric per uncertainty method

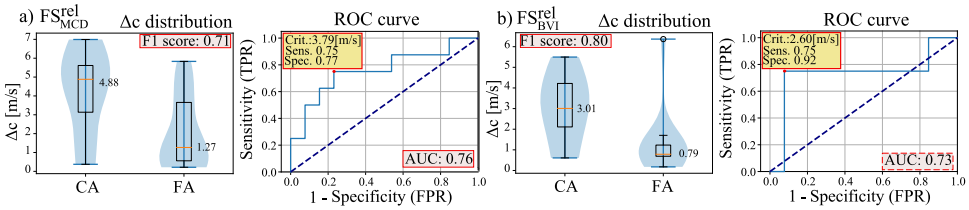

Fig. 4 The Δc distributions and ROC curves for CA vs. FA classification, for the uncertainty estimation approaches selecting the frames using **a** SI^{rel}_{MCD} for MCD-reconstructed frames and **b** SI^{rel}_{BVI} for BVI-reconstructed frames. The ROC curves display the classification criterion, sensitivity, and specificity at the point that maximizes the cumulative sensitivity and specificity

that uncertainty-informed automatic frame selection shows potential for improving differential diagnosis performance compared to using a fixed (S1 or S3) or random (SR) acquisition. These three uninformed selection approaches showing no systematic order of superiority across the four reconstruction methods (columns) show that there is no major bias of classification results based on frame acquisition order.

The Δc distributions and ROC curves of the top two methods (SI^{rel}_{MCD} and SI^{rel}_{BVI}) are presented in Fig. 4, both of which are informed by relative uncertainty metric. Both methods show an impressive improvement compared to the studied baselines, including the analytical method LBFGS and the baseline network architecture VN.

Discussion and conclusion

Although uncertainty estimation has been studied widely in the literature, these methods are yet to find their clinical utility and niche application scenarios. In this work, it is one of the first times we are proposing a use-case of uncertainty for automatic frame selection in scenarios where multiple repeated acquisitions are possible. We have demonstrated this on VN-based SoS reconstructions, for differential breast cancer diagnosis of patients in BI-RADS class 4, which involves suspicious cases that cannot be differentiated by other means such as the US B-mode. Accordingly, the differential diagnosis task involves classifying lesions that are visible and already identified by the radiologist on the US image.

We have adapted and trained VN for a specific VS Tx sequence used in the clinical data collection and have accordingly estimated SoS reconstruction uncertainty for the first time. To that end, we used Monte Carlo and Bayesian techniques. For the Bayesian approach, we have utilized a mathematical reformulation of the conventional loss function, which leads to a tractable implementation and increases training robustness. We have shown that the proposed solutions are able to preserve (and even improves) the reconstruction accuracy of the baseline VN, while enabling an assessment of trust (confidence) to put on a particular reconstruction. In this work, frame selection based on the proposed minimization of inclusion-to-background normalized uncertainty metric (SI^{rel}) proved most effective as a surrogate of trust in reconstructions. This may be due to such metric seeking for consistency in uncertainty between different image regions, i.e., inclusion and background. Different problem settings may require different treatments, but our work highlights the challenge of directly comparing uncertainties across separate images, even of similar scenes.

Note that the VN framework learns the regularizers and priors, which makes it interpretable as opposed to blackbox learning algorithms. Also, VN takes in the known imaging model matrix as input, so it does not have to learn it from data, making this method robust to train and easier to generalize across domain shift to in vivo data, even from only simulated training. Although the forward imaging model is fixed in VN, one could also learn and calibrate the imaging model in a restricted form using data, as shown in [28]. Similarly to the conventional analytical methods (e.g., LBFGS), deep learned VN that we used also relies on the initial assumption of beamforming SoS. This nevertheless can be estimated from displacement data [29], to be incorporated in the future for improved accuracy.

In this manuscript, we propose a practical use-case of uncertainty estimation for selecting optimal frames from

multiple data acquisitions. We study this for breast cancer differential diagnosis based on VN-based US image reconstruction. Such solution can be extended and applied on many other problem settings and image modalities with pixel-wise regression-like outputs, where selecting one out of several possibilities is relevant.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11548-025-03402-4.

Funding Open access funding provided by Uppsala University. Funding was provided partially by Uppsala Medtech Science & Innovation Centre. We acknowledge computation hardware support by an Nvidia grant.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethics approval and Informed Consent Clinical data was collected at Kantonsspital Baden, Switzerland, with informed consent and ethics approval (EKNZ, BASEC 2020-01962).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Houghton SC, Hankinson SE (2021) Cancer progress and priorities: breast cancer. Cancer Epidemiol Biomark Prev 30(5):822–844
2. Guray M, Sahin AA (2006) Benign breast diseases: classification, diagnosis, and management. Oncologist 11(5):435–449
3. Bae MS, Moon WK, Chang JM, Koo HR, Kim WH, Cho N, Yi A, La Yun B, Lee SH, Kim MY (2014) Breast cancer detected with screening US: reasons for nondetection at mammography. Radiology 270(2):369–377
4. Anderson ME, Trahey GE (1998) The direct estimation of sound speed using pulse-echo ultrasound. J Acoust Soc Am 104:3099–106
5. Perrot V, Polichetti M, Varray F, Garcia D (2021) So you think you can DAS? A viewpoint on delay-and-sum beamforming. Ultrasonics 111:106309
6. Jaeger M, Held G, Peeters S, Preisser S, Grünig M, Frenz M (2015) Computed ultrasound tomography in echo mode for imaging speed of sound using pulse-echo sonography: proof of principle. Ultrasound Med Biol 41(1):235–250
7. Sanabria SJ, Ozkan E, Rominger M, Goksel O (2018) Spatial domain reconstruction for imaging speed-of-sound with pulse-echo ultrasound: simulation and in vivo study. Phys Med Biol 63(21):215015
8. Rau R, Schweizer D, Vishnevskiy V, Goksel O (2021) Speed-of-sound imaging using diverging waves. Int J Comput Assist Radiol Surg 16(7):1201–1211
9. Schweizer D, Rau R, Bezek CD, Kubik-Huch RA, Goksel O (2023) Robust imaging of speed-of-sound using virtual source transmission. IEEE Trans UFFC 70(10):1308–1318
10. Bezek CD, Goksel O (2024) Model-based speed-of-sound reconstruction via interpretable pruned priors. In: IEEE UFFC joint symposium
11. Bezek CD, Farkas M, Schweizer D, Kubik-Huch RA, Goksel O (2025) Breast density assessment via quantitative sound-speed measurement using conventional ultrasound transducers. Eur Radiol 35:1490–1501
12. Feigin M, Freedman D, Anthony B (2019) A deep learning framework for single-sided sound speed inversion in medical ultrasound. IEEE Trans Biomed Eng 67:1142–1151
13. Ruby L, Sanabria SJ, Martini K, Dedes KJ, Vorburger D, Oezkan E, Frauenfelder T, Goksel O, Rominger MB (2019) Breast cancer assessment with pulse-echo speed of sound ultrasound from intrinsic tissue reflections: proof-of-concept. Invest Radiol 54(7):419–427
14. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS (2018) Image reconstruction by domain-transform manifold learning. Nature 555(7697):487–492
15. Jin KH, McCann M, Froustey E, Unser M (2017) Deep convolutional neural network for inverse problems in imaging. IEEE Trans Image Process 26(9):4509–4522
16. McCann MT, Unser M (2019) Biomedical image reconstruction: from the foundations to deep neural networks. Found Trends Signal Process 13(3):283–359
17. Monga V, Li Y, Eldar YC (2021) Algorithm unrolling: interpretable, efficient deep learning for signal and image processing. IEEE Signal Process Mag 38(2):18–44
18. Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, Knoll F (2018) Learning a variational network for reconstruction of accelerated MRI data. Magn Reson Med 79(6):3055–3071
19. Vishnevskiy V, Rau R, Goksel O (2019) Deep variational networks with exponential weighting for learning computed tomography. In: Medical image computing and computer assisted intervention (MICCAI), Shenzhen, China, pp 310–318
20. Bernhardt M, Vishnevskiy V, Rau R, Goksel O (2020) Training variational networks with multidomain simulations: speed-of-sound image reconstruction. IEEE Trans UFFC 67(12):2584–2594
21. Schlemper J, Castro DC, Bai W, Qin C, Oktay O, Duan J, Price AN, Hajnal J, Rueckert, D (2018) Bayesian deep learning for accelerated MR image reconstruction. In: International workshop on machine learning for medical image reconstruction, pp 64–71
22. Narnhofer D, Effland A, Kobler E, Hammernik K, Knoll F, Pock T (2021) Bayesian uncertainty estimation of learned variational MRI reconstruction. IEEE Trans Med Imaging 41(2):279–291
23. Ekmekci C, Cetin M (2022) Uncertainty quantification for deep unrolling-based computational imaging. arXiv:2207.00698
24. Gal Y, Ghahramani Z (2016) Dropout as a bayesian approximation: representing model uncertainty in deep learning. In: International conference on machine learning (ICML), pp 1050–1059
25. Broyden CG (1970) The convergence of a class of double-rank minimization algorithms: 1. General considerations. IMA J Appl Math 6(1):76–90
26. Roth S, Black MJ (2009) Fields of experts. Int J Comput Vis (IJCV) 82:205–229
27. Kobler E, Klatzer T, Hammernik K, Pock T (2017) Variational networks: connecting variational methods and deep learning. In: German conference pattern recognition (GCPR), pp 281–293
28. Bezek CD, Haas M, Rau R, Goksel O (2024) Learning the imaging model of speed-of-sound reconstruction via a convolutional for-

mulation. IEEE Trans Med Imaging (in print). https://ieeexplore.ieee.org/document/10716684
29. Bezek CD, Goksel O (2023) Analytical estimation of beamforming speed-of-sound using transmission geometry. Ultrasonics 134:107069

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.