

## RESEARCH ARTICLE OPEN ACCESS

# A Deep Learning-Based Approach to Characterize Skull Physical Properties: A Phantom Study

Deepika Aggrawal<sup>1</sup> | Loïc Saint-Martin<sup>2</sup> | Rayyan Manwar<sup>2</sup>  | Amanda Siegel<sup>2</sup> | Dan Schonfeld<sup>1</sup> | Kamran Avanaki<sup>2,3</sup> 

<sup>1</sup>Department of Electrical and Computer Engineering, University of Illinois, Chicago, Illinois, USA | <sup>2</sup>Department of Biomedical Engineering, University of Illinois, Chicago, Illinois, USA | <sup>3</sup>Department of Dermatology, University of Illinois, Chicago, Illinois, USA

**Correspondence:** Kamran Avanaki ([avanaki@uic.edu](mailto:avanaki@uic.edu))

**Received:** 28 March 2024 | **Revised:** 22 September 2024 | **Accepted:** 9 October 2024

**Funding:** This work was supported by the National Institutes of Health R01EB027769–01 and R01EB028661–01.

**Keywords:** characterization | deep learning | photoacoustic | porosity | skull | thickness

## ABSTRACT

Transcranial ultrasound imaging is a popular method to study cerebral functionality and diagnose brain injuries. However, the detected ultrasound signal is greatly distorted due to the aberration caused by the skull bone. The aberration mechanism mainly depends on thickness and porosity, two important skull physical characteristics. Although skull bone thickness and porosity can be estimated from CT or MRI scans, there is significant value in developing methods for obtaining thickness and porosity information from ultrasound itself. Here, we extracted various features from ultrasound signals using physical skull-mimicking phantoms of a range of thicknesses with embedded porosity-mimicking acoustic mismatches and analyzed them using machine learning (ML) and deep learning (DL) models. The performance evaluation demonstrated that both ML- and DL-trained models could predict the physical characteristics of a variety of skull phantoms with reasonable accuracy. The proposed approach could be expanded upon and utilized for the development of effective skull aberration correction methods.

## 1 | Introduction

Transcranial ultrasound imaging is a potentially valuable noninvasive method to assess cerebral functionality and diagnose possible brain injuries [1]. However, a major limitation in transcranial imaging is the presence of skull bone, which acts as a dispersive barrier and causes distortion to the amplitude and phase of the received acoustic waves [2–4]. This distortion is evident in terms of acoustic attenuation, dispersion, signal broadening, and temporal shift [5, 6]. The degree of distortion caused by these phenomena depends primarily on the acoustic properties of the cranium (i.e., speed of sound, density, absorption, scattering, and acoustic impedance) [7, 8]. However, these acoustic properties can be determined from the mechanical characteristics of the skull bone (i.e., bone type, density, porosity, and thickness). Among these parameters, thickness and

porosity, which vary depending on location, age, and gender, have one of the most significant impacts [5, 9–14]. Since thicker skull bones force the acoustic pressure wave to propagate longer distances with a high absorption rate, the amplitude is greatly impacted in terms of attenuation. Simultaneously, higher porosity increases the frequency of propagating media interfaces (bone-fluid), introduces scattering phenomena [9], and affects the phase information in terms of dispersion [8], broadening (widening of the pulse) [15], and temporal shift [8, 15–20]. Therefore, the estimation of the thickness and porosity of the skull bone is of great interest. Methods to noninvasively estimate these parameters would help recover the original signal and would be a major advancement in brain imaging.

While computed tomography (CT) and magnetic resonance imaging (MRI) using methods such as ultrashort time echo

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Journal of Biophotonics* published by Wiley-VCH GmbH.

(UTE) and zero time echo (ZTE) imaging can predict skull thickness and porosity [21], they are not widely available and require relatively long scan times [21, 22]. Ultrasound is widely available, and innovative technologies could be harnessed to determine thickness and porosity, but few analytical-based approaches have been tested to make such predictions [23, 24]. Some that have been tried include porosity prediction from ultrasound backscatter using multivariate analysis [20]; use of axial transmission velocities [12]; and thickness prediction using Sono Pointer [25]. However, these methods predict either porosity or thickness, require user inputs, and demonstrate results that vary significantly between studies. With the ability to train models from image sets, machine learning (ML) and deep learning (DL) techniques have the potential for better interpretation of imaging data across diverse applications [26–31], offering a solution to challenges that traditional methods encounter [32].

If appropriate features can be extracted from the data, ML models can be trained using a relatively small dataset. Conversely, if no relevant features are known, DL models can explore existing data to identify the most relevant features. Over the years, DL models have shown prominent results in the field of phase recovery and scattering imaging for different modalities including ultrasound imaging [33, 34]. To our knowledge, while there are studies using DL models for extracting skull characteristics from other imaging technologies such as MRI and CT [35], there are few using ultrasound data [36–39]. In [36], the authors estimated the micro-architectural parameters of cortical bone such as pore diameter, pore density, and porosity using an artificial neural network. However, the research was limited by only using simulation data for training and testing. Also, the results were limited to predicting the porosity of cortical layers, where the size of the pores is in the micron range. Porosity in this range does not affect the acoustic wavelengths typically used in transcranial imaging.

According to [5], the diploë layer of skull bone consists of a more complex formation of porous and bone structures that dominates the acoustic scattering, and hence distortion of the ultrasound pressure wave. In addition to the variability of porosity, the thickness of the diploë layer varies with age [11]. Therefore, a good model must predict both porosity and thickness of diploë type skull bone in order to compensate for aberrated acoustic signal [17].

In this paper, our objective was to achieve high accuracy for determining skull bone thickness and porosity in order to improve our understanding of the influence of diploë layer characteristics. Such an objective could enable the development of effective skull aberration correction methods. We have utilized both ML and DL approaches to test their capability to determine the “thickness” and “porosity” of solid skull phantoms, where we varied phantom thickness to mimic changes in skull thickness and mixed varying concentrations of poppy seeds into resin to mimic porosity [40, 41] (200 phantoms with varied thickness and porosity were made). After we explain the method of preparing the skull phantoms, we describe the selected ML and DL approaches. The results for determining thickness and porosity are reported separately.

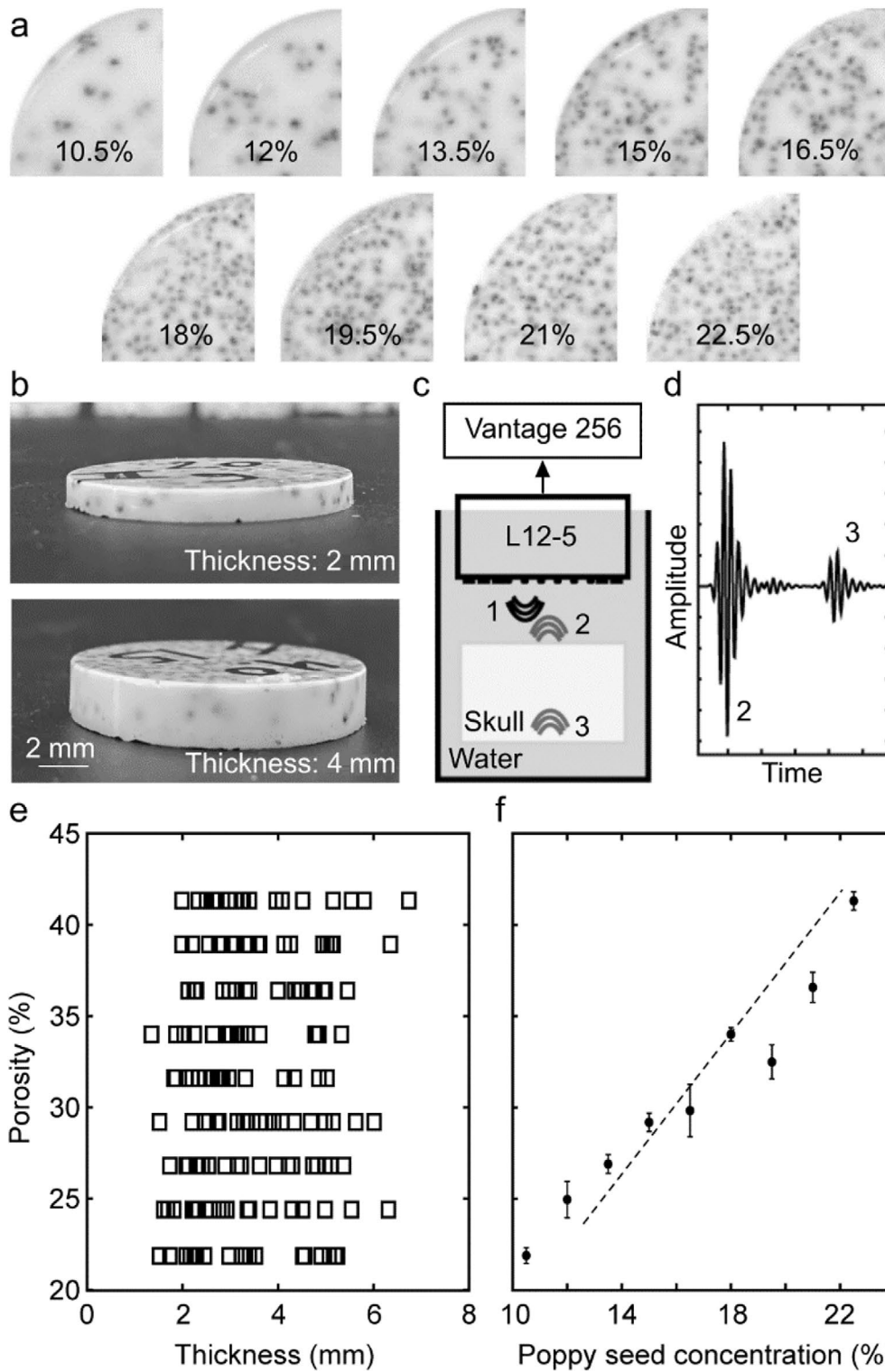
## 2 | Materials And Methods

### 2.1 | Sample Preparation and Experimental Setup

The diploë type skull bone phantom was prepared from a mixture of titanium dioxide (Loudwolf, USA), (1% w/w) and different concentrations of poppy seeds in an epoxy resin solution (Clearcast 7050, The Epoxy Resin Store) (2:1 hardener: epoxy). First, the titanium dioxide is homogenized with the resin hardener with the help of a vortex mixer and a sonicator bath. Then, the resin was added, and the mixture was vortexed and sonicated again before being poured into a Petri dish and placed inside a vacuum chamber for 30 min to eliminate any air bubbles before solidification. A predefined number of poppy seeds are added to the solution inside the Petri dish before placing it in the vacuum chamber. A further detailed explanation of the skull bone phantom preparation method can be found in [42]; top and side views of phantoms showing a range of poppy seed concentrations and thicknesses can be seen in Figure 1a (top view) and Figure 1b (side view).

A total of 180 samples (diameter: 30 mm) were created with different thicknesses ranging between 1 and 7 mm and different concentrations of poppy seeds (range from 10.5% to 22.5% w/w). Addition of poppy seeds (diameter: ~1 mm) was used to represent fluid-filled pockets within the diploë layer of the skull bone [42]. This range of poppy seed concentrations was selected to mimic a range in porosity of 21%–42% (Figure 1f). We calculate porosity by image segmentation of each sample's cross-section, using the ratio between the detected seeds in the image and the total surface. The presence of the seed inside the sample generates similar acoustic behavior as is produced in the real skull porous region due to the impedance mismatch (high changes in density and speed of sound between two environments at their interface), creating phenomena such as reflection, mode conversion, refraction, and signal distortion [43]. For each seed mass concentration, 20 samples were made with variable thickness (ranging from 1 to 7 mm).

Skull phantoms were suspended in a water-filled (coupling medium) bucket that was normal to the ultrasound sensor surface. To detect ultrasound pressure waves generated from acoustic reflection at the phantom-water interface, we used a linear array transducer (L12-5, ATL-Philips, USA). The acquired signals were sampled at 62.5MS/s using a Vantage 256 system (Vera Sonics Inc., USA). We acquired 50 signals from each element of a linear array probe, averaging them to suppress background noise and address the random distribution of seeds in the sample. The width of the probe, with elements spaced 0.2mm apart and positioned parallel to the sample's center, means each element experiences a slightly different acoustic path through the sample. These signals were averaged to represent the characteristics of the sample (Figure 1c). Each averaged signal was decomposed into two parts: the first echo (#2 in Figure 1d) represents the reflection at the top interface between the water and the sample and the second echo (#3 in Figure 1d) represents the reflection at the bottom interface between the sample and the water. To isolate each echo during the signal processing, the envelope of the entire signal was obtained, locating the maximum of each signal and their starting/ending points.



**FIGURE 1** | Skull phantoms and signal acquisition from same. (a) Top view of samples with different concentrations of poppy seeds (10%–22.5%). (b) Side view of samples showing samples of different thicknesses. (c) Schematic of the experimental setup used to obtain the signals. (d) An example of the acoustic signal obtained from the samples. (e) Thickness and calculated porosity of 180 phantoms. (f) Relationship between poppy seed concentration and sample porosity as empirically measured.

## 2.2 | Dataset

We utilized the averaged signals from our 180 phantoms of varying thickness and porosity, shown in Figure 1e, to train the system. For the thickness predictions, 144 samples were

used for training, which, for DL models, includes 129 samples for training and 15 for validation, and then 36 samples were used for testing. For the porosity predictions, we have the same 180 samples representing 9 different porosity values. Among the 180 samples, 171 were used for training, which,

for DL models, includes 153 for training and 18 for validation, and 9 (one from each porosity) were used as a test dataset. The rationale for selecting the size of training/testing/validation datasets can be found in [44].

### 2.3 | Dataset Preprocessing and Feature Extraction

As a first data preprocessing step, we used min–max normalization to the signals to maintain data consistency [45]. This ensures that all signals are scaled to the same range, which is essential for effective model training. Next, we extracted features from the normalized signals. In our previous work [42], we extracted many features from ultrasound signals of skull phantoms and analyzed whether they varied with changes in phantom width and porosity. There, we identified nine features that were the most significant. These features are shown in Table 1. Detailed descriptions of these features can be found in our prior publication [42]. To ensure that features with larger numerical values do not disproportionately influence the results, we normalized them using standard scaling before applying ML and DL techniques.

For our DL experiments, we used either the same nine features or the envelope of the signal.

### 2.4 | Implementation of Three ML Regression Models

We employed ML techniques, where we used the nine features identified above as inputs. ML regression models use combinations of independent variables (in our case signal features) to map input features to an output value (thickness or porosity). Three algorithms of increasing complexity were tested. The first method is multiple linear regression (MLR), which explores linear relationships between the nine features and the target variables [46]. This method allowed us to quantify the influence of each feature on

the target by estimating the regression coefficients. The general form of the MLR equation is

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i \quad (1)$$

where  $\beta_0$  represents intercept,  $\beta_i$  represents the learned coefficient of  $X_i$ ,  $X_i$  represents the  $i$ th feature value, and  $Y$  represents the target value. Ridge linear regression (RLR) is an extension of MLR that includes a regularization term to prevent overfitting and enhance model generalization. By adding a penalty proportional to the squared magnitude of the coefficients [47], RLR reduces the impact of less important features, thereby providing more stable and reliable coefficient estimates. While the equation for RLR is the same as for MLR, RLR utilizes a loss function with an additional penalty component,  $\lambda$ , which demonstrates a regularization parameter (controls the amount of shrinkage applied to coefficients).

Loss Function:

$$\min_{\beta} \left( \sum_{i=1}^n \left( Y_i - \beta_0 - \sum_{j=1}^i \beta_j X_j^2 \right) \right) + \lambda \sum_{j=1}^i \beta_j^2 \quad (2)$$

Polynomial Regression (PNR) is another extension of MLR used to identify nonlinear relationships between the nine features and the dependent variables [48]. The model is defined as

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 \dots + \beta_i X^i \quad (3)$$

where  $Y$  is the dependent variable,  $X$  is one of the features,  $\beta_0, \beta_1, \beta_2, \dots, \beta_i$  are the coefficients, and  $i$  is the degree of the polynomial. For this study, we used a polynomial degree of 2, which incorporates all squared features and all pairwise interaction terms. This degree was chosen because it allows for simpler interpretation by limiting the model to linear and quadratic interactions, while also minimizing the risk of overfitting. For our system with nine original features, this translates to a total of 54 features (9 original features, 9 squared features, and 36 paired interactions).

TABLE 1 | Feature table with corresponding descriptions.

Feature name	Feature description
Delay	Delay between the first and second peak (TD)
Slope	Slope between the first and second peak (TD)
SA-Total-T	Surface area of complete signal (TD)
Phase-Shift-T	Phase shift of complete signal (TD)
FWHM-SP-T	FWHM of second peak (TD)
SA-Total-F	Surface area of complete signal (FD)
SA-SP-F	Surface area of second peak (FD)
FWHM-Total-F	FWHM of complete signal (FD)
FWHM-SP-F	FWHM of second peak (FD)

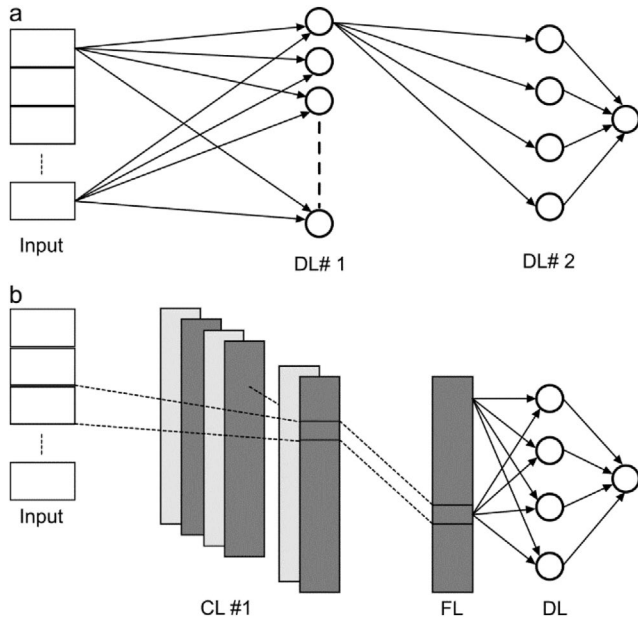
Abbreviations: FD, frequency domain; FWHM, full width half maximum; SA, surface area; SP, second peak; TD, time domain.

### 2.5 | Implementation of Two DL Models

Additionally, we explored two DL techniques with two different input datasets: One using the same nine features as inputs (see Table 1) and another using the envelope as input. Use of the envelope allows the models to autonomously learn the most relevant features. This approach minimizes the risk of overlooking potentially important features that might be missed with manual feature selection. By using the entire signal, the model is able to learn intricate patterns and complexities within the data, thereby identifying the most significant features without relying on prior assumptions.

To reduce model complexity while still capturing intricate features that might have been overlooked during manual selection, we designed two DL models, the first of which is a feed-forward neural network (FFN) consisting of an initial dense layer with 8 units, followed by a dense layer with 4 units, and finally, an output layer for prediction (Figure 2a). We modify the FFN by introducing a convolution layer to build our second DL model,





**FIGURE 2** | A schematic overview of the FFN and CNN architectures used in this study. (a) FFN architecture; DL#1 (8 units) and DL#2 (4 units) are both dense layers. (b) CNN architecture; CL #1 represents a convolutional layer (8 filters), FL represents a Flatten layer, and DL represents a dense layer (4 units). In (b), the dotted lines illustrate how each part of the input is processed by different filters in the convolutional layers, visually representing the learning process and the interaction of various filters with different regions of the input data.

a convolution neural network (CNN) [49, 50] to add complexity, as it allows hierarchical decomposition of input, and can extract high-level features from the input. The CNN design is shown in Figure 2b. The architecture consists of one convolutional hidden layer with a set of filters whose parameters were learned throughout the training process and a feed-forward network. These parameters are utilized to obtain the unique and hidden learning features which are essential for mapping.

Given different ultrasound signals acquired from skull phantoms with varied thickness and porosity values, our neural networks learn the distinctive features from the signal and map them to the actual thickness or porosity value. Over the years, FFN and CNN and their variations have been used in biomedical imaging to solve problems such as seizure prediction [31, 51] and EEG data classification [52]. More complex DL architectures such as long short-term memory (LSTM) and generative adversarial networks (GANs) require large amounts of training datasets, whereas simple neural networks (few layers) have performed well on small datasets [53]. Another reason for selecting simple neural networks is the relatively small size of the inputs: in one set of experiments, we use only nine features, and in another, we employ an envelope of length 200 which is straightforward, lacking complexities such as dips or noise. We aimed to predict continuous values—thickness and porosity—by optimizing filter parameters through the minimization of a loss function using as input either the extracted features (CNN-F or FFN-F) or the extracted envelope (CNN-E or FFN-E). Hyperparameter tuning led us to utilize the Adam optimizer with 2500 epochs and employ the mean absolute error (MAE) as the loss function for porosity and mean squared error (MSE) as the loss function for thickness.

**TABLE 2** | Performance comparison of models for predicting thickness values (mm): MAE, MSE, and RMSE Metrics (models depicted in bold are graphed in Figure 3).

Model	MAE	MSE	RMSE
<b>MLR</b>	<b>0.29</b>	<b>0.13</b>	<b>0.36</b>
PNR	0.30	0.14	0.38
RLR	0.32	0.15	0.39
<b>FFN-F</b>	<b>0.31</b>	<b>0.24</b>	<b>0.49</b>
CNN-F	0.71	0.89	0.94
FFN-E	0.65	0.68	0.82
CNN-E	0.64	0.66	0.81

### 3 | Results And Discussion

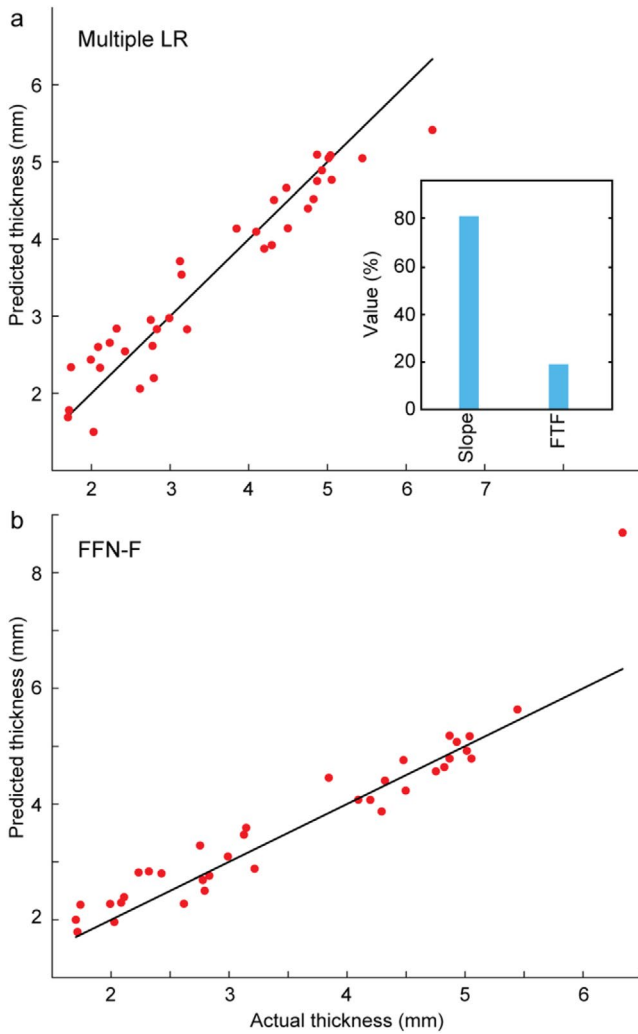
#### 3.1 | Thickness

Table 2 reports the combined results for all methodologies (ML and DL) predicting thickness. The results of the models are summarized as follows: The MLR model achieved the best performance of all the models, but all the ML models were very similar (and good, average  $R^2 = 0.907 \pm 0.004$ ). MLR showed an MAE of 0.29, MSE of 0.13, and RMSE of 0.36. The PNR model yielded an MAE of 0.30, MSE of 0.14, and RMSE of 0.38, while RLR showed an MAE of 0.32, MSE of 0.15, and RMSE of 0.39. Among the DL models, FFN-F had the best performance, with an MAE of 0.31, MSE of 0.24, RMSE of 0.49, and  $R^2 = 0.85$ . The CNN-F model had higher error metrics with an MAE of 0.71, MSE of 0.89, and RMSE of 0.94. When using the envelope as input, the FFN-E model recorded an MAE of 0.65, MSE of 0.68, and RMSE of 0.82, whereas the CNN-E model showed an MAE of 0.64, MSE of 0.66, and RMSE of 0.81. Overall, the ML models slightly outperformed the DL models for determining thickness. Among the DL models, those using features as inputs (FFN-F and CNN-F) performed better than those using envelopes (FFN-E and CNN-E). This suggests that envelope-based inputs may not be as effective as features for predicting thickness.

Figure 3 shows the best-performing models from both ML and DL methodologies for predicting thickness: MLR for ML (Figure 3a) and FFN-F for DL (Figure 3b). The figure highlights the accuracy of the predicted thickness values. Some outliers are observed: we anticipate that increasing the size of the training dataset would help reduce these discrepancies and improve the overall accuracy of the models.

#### 3.2 | Feature Contributions to ML Thickness Model

The relative contributions of the features for the MLR model are shown in Figure 3a. Slope contributes the most to the model, with a weight of 80.19%. The second most significant feature is FWHM-Total-F, contributing 19.21%. The results indicate that thickness can be reasonably well portrayed by linear combinations of the two identified features.



**FIGURE 3** | Evaluation of the performance of thickness prediction on the test data. (a) best ML architecture (MLR: multiple LR) and (b) best DL architecture (FFN-F). Percent contributions of the most significant features are shown for the ML architecture. Multiple LR: multiple linear regression; FFN-F: feed-forward network (features as input).

### 3.3 | Porosity Results

Table 3 contains the combined results for all methodologies predicting porosity using the same performance metrics (MAE, MSE, and RMSE). Among the models, the CNN-E, DL model achieved the best overall performance, with the lowest MAE of 2.98, MSE of 17.94, and RMSE of 4.23. The RLR model achieved the best performance from the ML category, exhibiting a competitive performance with an MAE of 3.42, MSE of 17.01, and RMSE of 4.13 ( $R^2=0.47$ ). While FFN-E also performed well, achieving an MAE of 3.21, MSE of 19.04, and RMSE of 4.36, the feature-based DL models (FFN-F and CNN-F) had higher error metrics. FFN-F recorded an MAE of 4.10, MSE of 23.82, and RMSE of 4.88, while CNN-F had an MAE of 3.18, MSE of 23.79, and RMSE of 4.87. Among all other ML models, PNR showed relatively higher error rates with an MAE of 3.42, a surprising MSE of 94.54, RMSE of 9.72, and  $R^2$  of 0.10. MLR performed nearly as well as RLR, with an MAE of 3.42, MSE of 17.16, RMSE of 4.14 and  $R^2$  of 0.46. The superior performances of the FFN-E and CNN-E

**TABLE 3** | Performance comparison of models for predicting porosity values (%): MAE, MSE, and RMSE Metrics (models depicted in bold are graphed in Figure 4).

Model	MAE	MSE	RMSE
MLR	3.42	17.16	4.14
PNR	3.42	94.54	9.72
<b>RLR</b>	<b>3.42</b>	<b>17.01</b>	<b>4.13</b>
FFN-F	4.10	23.82	4.88
CNN-F	3.18	23.79	4.87
FFN-E	3.21	19.04	4.36
<b>CNN-E</b>	<b>2.98</b>	<b>17.94</b>	<b>4.23</b>

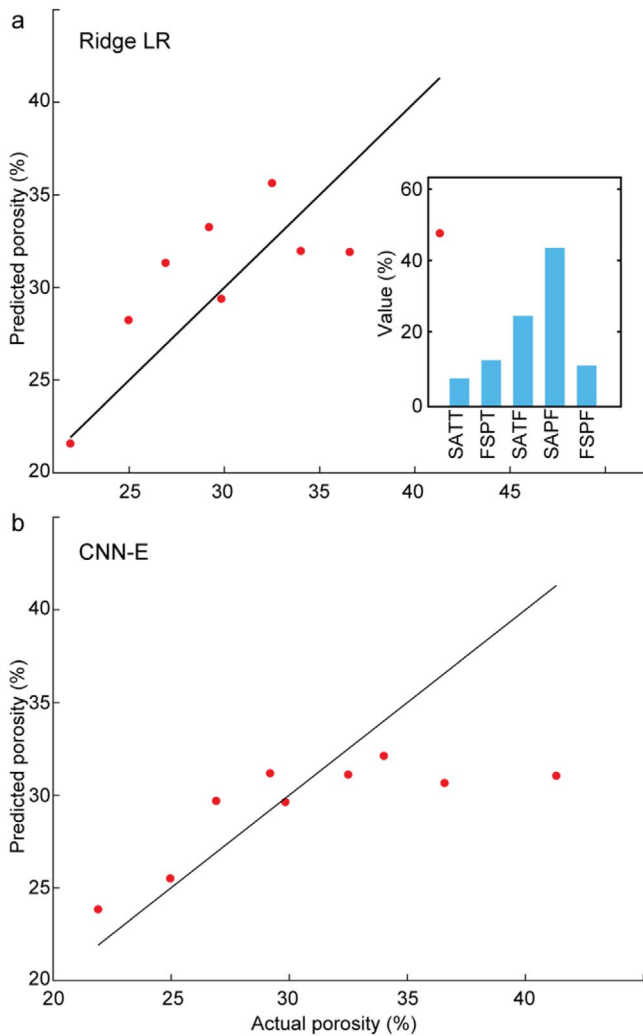
models highlight the advantage of using envelope data for improved porosity prediction accuracy despite potential relationships suggested between features and porosity in [42].

### 3.4 | Feature Contributions to Porosity ML Models

The relative contributions of the RLR features are shown in Figure 4a: SA-SP-F contributes the most to the model, with a weight of 43.15%. The second most significant feature is SA-Total-F, contributing 23.74%, followed by FWHM-SP-T at 12.35%. The other features, FWHM-SP-F and SA-Total-T, contribute 10.25% and 7.57%, respectively. This distribution suggests that surface area (SA) and FWHM, both in the time and FD, are the most influential features in predicting porosity, with some contributions from other features as well.

Reviewing Figure 4, it is clear that while the general trend is captured, the performance at the extremes, particularly on higher values, suggests a need for improvement. This indicates that the models struggle to predict at the upper end of the range, possibly due to the limited data available for these higher values. Our use of poppy seeds simulated a porosity range of 21%–42%. For comparison, the porosity of the parietal bone varies, with the outer and inner tables typically having porosities between 10% and 30%, while the central diploe layer has a much higher porosity, ranging from 50% to 80% [10]. On average, the entire bone likely exhibits a porosity in the range of 20%–41%, which is consistent with the range simulated by our poppy seed concentrations. While our model successfully captures this average porosity, more complex phantoms with multiple layers exhibiting different porosity values would provide a more accurate representation of bone structure.

Up to this point, we have described our results for independent models generated to predict either thickness or porosity from our phantoms, where a range of models for each variable were trained and tested separately. We also designed a model to predict both thickness and porosity simultaneously. To enhance the simultaneous model's accuracy, we introduced additional complexities to capture more detailed lower-level features. However, this approach resulted in a high MSE for both predictions (data not shown). We then sought to analyze the variables sequentially by first finding thickness, and then feeding thickness values into



**FIGURE 4** | Evaluation of the performance of porosity prediction on the test data. (a) best ML architecture (RLR: Ridge LR), and (b) best DL architecture (CNN-E). Percent contributions of the most significant features are shown for the ML architecture. Ridge LR: ridge linear regression. CNN-E: convolutional neural network (envelope as input). SATT: SA-Total-T, FSPT: FWHM-SP-T, SATF: SA-Total-F, SAPF: SA-SP-F, FSPF: FWHM-SP-F.

the porosity model. However, incorporating thickness values into the porosity model did not improve performance. To understand these results, we generated a correlation matrix encompassing the nine identified features as well as thickness and porosity. Analysis of the matrix indicated that thickness and porosity are driven by distinct signal features, with no significant correlation between them. Based on these findings, we decided to develop two separate models, one for each variable, to optimize accuracy for both thickness and porosity separately.

Analyzing the range of models tested for predicting thickness, it is evident that the ML models performed exceptionally well, demonstrating that they were able to effectively leverage the information contained in the features to extract thickness. This may be because the ML models worked with specific features that had previously been demonstrated to be highly correlated with thickness [42]. Two key features that significantly contributed to thickness prediction in ML models were Slope in the

time domain and FWHM of the complete signal in the frequency domain (FWHM-Total-F). Slope combines delay and amplitude, which were previously shown to correlate with thickness prediction [42]. FWHM-Total-F is expected to be broadened by the length of time the signal takes to pass through the phantom, making it a reasonable key indicator for thickness. The DL models also showed promising results, if not quite as accurate as the ML models. The DL-F models (CNN-F and FFN-F) had the same initial inputs as the ML models, and here, the unstructured learning of neural networks showed, principally, that the sample size was perhaps too small to achieve high accuracy. The DL-E models (CNN-E and CNN-F) also performed reasonably well, likely because the envelope somewhat recapitulates aspects of the broadening extracted by the slope and FWHM-Total-F features.

For porosity predictions, the best results were observed using the RLR method in the ML models. We noticed that FWHM and SA features, both in the frequency (FD) and time (TD) domains, made significant contributions. In the FD, acoustic aberrations caused by porosity appear to introduce variations in the frequency components of the signal. Both FWHM and SA effectively capture these changes: FWHM reflects the broadening of the signal in the FD, while SA captures changes in the signal's area. Similarly, in TD, porosity-induced echoes can lead to a reduction in signal peak amplitude. Both SA and FWHM in TD are sensitive to these changes, allowing them to capture the effects of porosity on the signal. The feature-based DL and the ML models performed worse than the envelope models, demonstrating the extracted features dataset was insufficient to fully capture porosity values. The envelope-based methods, particularly the CNN-E model, were able to extract features that are relevant to porosity, though these features were not fully captured due to the limited sample size as demonstrated by the modest  $R^2$  value (0.46).

This limitation suggests that future work in this area, with a larger dataset, could further improve the accuracy of porosity prediction using DL methods. The envelope worked better for porosity prediction because it captures much of the critical information from the ultrasound signal, such as peak position, amplitude, and FWHM. The waveform also simplifies the signal, reducing the complexity compared to the full signal, which contains multiple peaks with amplitude variations. Such variations are often influenced by hardware, making DL applications using full signals as input more challenging than applications using the envelope as input.

Given a larger dataset ( $\gg 180$  samples), the DL models might have been able to capture these relationships more effectively. Our sample size was limited by the requirement to fabricate and test each sample manually. Multiple simulation studies have been reported such as [25], but we were dissatisfied with the capability of that or other studies to capture the complex effect of the acoustic mismatch presented by the presence of diploë in skull. Here, we developed physical phantoms that capture this mismatch to some extent (although not perfectly, as they are of approximately uniform porosity throughout). A limitation of our physical phantoms is that there is some estimation required for characterizing porosity (see, e.g., Figure 1f) due to local/internal variations in poppy seed concentrations, which could affect the accuracy of the training and testing data sets.

## 4 | Conclusion

The cranial bone poses a significant barrier to sound waves in ultrasound-based transcranial imaging techniques, such as focused ultrasound, photoacoustic, thermoacoustic, and x-ray acoustic imaging [3, 4, 34, 54–59]. Standard reconstruction algorithms used in these procedures typically assume that ultrasound waves travel through a homogenous medium [60]. The assumption often leads to distorted images of the target area due to the heterogeneous nature of the cranial bone [60]. To produce more accurate images, it is crucial to incorporate the physical characteristics of the skull into the reconstruction algorithms. Transcranially acquired ultrasounds have aberrations due to the presence of skull bone largely due to variations in skull thickness and bone porosity. Therefore, it is essential to identify appropriate methods for determining these parameters to correct the aberrations. In this study, we evaluated the performance of seven different models—comprising three ML models and four DL models—for predicting thickness and porosity independently. None of the models used the full signal but instead were based on either the envelope of the signal or extracted features. This reduces the effect of extraneous input caused by system noise while still capturing critical information. Among these models, the MLR method, which is an ML approach, achieved the lowest MAE of 0.29 for thickness prediction. For porosity prediction, the CNN-E, a DL technique, demonstrated the best performance with the lowest MAE of 2.98. For practical applications to a human adult cranium, a three-layer model should be considered, including the inner and outer tables of the skull bone around the diploe. Part of our study could be applied to the inner and outer tables, where similar thickness variations occur but with lower porosity levels. Accurate thickness and porosity for each layer should be evaluated within the ranges reported in the literature to examine the impact of these physical characteristics on signal features. This work can help in understanding the influence of the diploe layer on the ultrasound signal and suggests a process for developing an effective skull aberration compensating method.

---

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

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

### References

1. T. S. Riis, T. D. Webb, and J. Kubanek, “Acoustic Properties Across the Human Skull,” *Ultrasonics* 119 (2022): 106591, <https://doi.org/10.1016/j.ultras.2021.106591>.
2. P. J. White, G. T. Clement, and K. Hynynen, “Longitudinal and Shear Mode Ultrasound Propagation in Human Skull Bone,” *Ultrasound in Medicine & Biology* 32 (2006): 1085–1096, <https://doi.org/10.1016/j.ultrasmedbio.2006.03.015>.
3. R. Manwar, M. T. Islam, S. M. Ranjbaran, and K. Avanaki, “Transfontanelle Photoacoustic Imaging: Ultrasound Transducer Selection Analysis,” *Biomedical Optics Express* 13 (2022): 676–693.
4. R. Manwar, L. S. McGuire, M. Islam, et al., “Transfontanelle Photoacoustic Imaging for In-Vivo Cerebral Oxygenation Measurement,” *Scientific Reports* 12 (2022): 1–7.

5. R. Manwar, K. Kratkiewicz, and K. Avanaki, “Investigation of the Effect of the Skull in Transcranial Photoacoustic Imaging: A Preliminary Ex Vivo Study,” *Sensors* 20 (2020): 4189.
6. R. Keshari, M. Vatsa, R. Singh, and A. Noore, “Learning Structure and Strength of CNN Filters for Small Sample Size Training,” In *Proceedings of Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 9349–9358.
7. E. Merčep, J. L. Herraiz, X. L. Deán-Ben, and D. Razansky, “Transmission–Reflection Optoacoustic Ultrasound (TROPUS) Computed Tomography of Small Animals,” *Light: Science & Applications* 8 (2019): 18, <https://doi.org/10.1038/s41377-019-0130-5>.
8. B. Liang, S. Wang, F. Shen, Q. H. Liu, Y. Gong, and J. Yao, “Acoustic Impact of the Human Skull on Transcranial Photoacoustic Imaging,” *Biomedical Optics Express* 12 (2021): 1512–1528, <https://doi.org/10.1364/boe.420084>.
9. H. Zhang, Y. Zhang, M. Xu, et al., “The Effects of the Structural and Acoustic Parameters of the Skull Model on Transcranial Focused Ultrasound,” *Sensors* 21 (2021): 5962, <https://doi.org/10.3390/s21175962>.
10. S. L. Alexander, K. Rafaels, C. A. Gunnarsson, and T. Weerasooriya, “Structural Analysis of the Frontal and Parietal Bones of the Human Skull,” *Journal of the Mechanical Behavior of Biomedical Materials* 90 (2019): 689–701, <https://doi.org/10.1016/j.jmbbm.2018.10.035>.
11. N. Lynnerup, J. G. Astrup, and B. Sejrsen, “Thickness of the Human Cranial Diploe in Relation to Age, Sex and General Body Build,” *Head & Face Medicine* 1 (2005): 13, <https://doi.org/10.1186/1746-160X-1-13>.
12. M. Gräsel, C. C. Glüer, and R. Barkmann, “Characterization of a New Ultrasound Device Designed for Measuring Cortical Porosity at the Human Tibia: A Phantom Study,” *Ultrasonics* 76 (2017): 183–191, <https://doi.org/10.1016/j.ultras.2017.01.001>.
13. L. Mohammadi, H. Behnam, J. Tavakkoli, and K. Avanaki, “Skull Acoustic Aberration Correction in Photoacoustic Microscopy Using a Vector Space Similarity Model: A Proof-Of-Concept Simulation Study,” *Biomedical Optics Express* 11 (2020): 5542–5556.
14. R. Manwar, K. Kratkiewicz, and M. R. N. Avanaki, “Study the Effect of the Skull in Light Illumination and Ultrasound Detection Paths in Transcranial Photoacoustic Imaging,” In *Proceedings of Photons Plus Ultrasound: Imaging and Sensing* 1164 (2021): 380–388.
15. X. L. Deán-Ben, D. Razansky, and V. Ntziachristos, “The Effects of Acoustic Attenuation in Optoacoustic Signals,” *Physics in Medicine and Biology* 56 (2011): 6129–6148, <https://doi.org/10.1088/0031-9155/56/18/021>.
16. G. Pinton, J. F. Aubry, E. Bossy, M. Muller, M. Pernot, and M. Tanter, “Attenuation, Scattering, and Absorption of Ultrasound in the Skull Bone,” *Medical Physics* 39 (2012): 299–307, <https://doi.org/10.1118/1.3668316>.
17. L. Mohammadi, H. Behnam, J. Tavakkoli, and M. R. N. Avanaki, “Skull’s Photoacoustic Attenuation and Dispersion Modeling With Deterministic Ray-Tracing: Towards Real-Time Aberration Correction,” *Sensors* 19 (2019): 345.
18. B. Jing and B. D. Lindsey, “Effect of Skull Porous Trabecular Structure on Transcranial Ultrasound Imaging in the Presence of Elastic Wave Mode Conversion at Varying Incidence Angle,” *Ultrasound in Medicine & Biology* 47 (2021): 2734–2748, <https://doi.org/10.1016/j.ultrasmedbio.2021.05.010>.
19. A. Kyriakou, E. Neufeld, B. Werner, M. M. Paulides, G. Szekely, and N. Kuster, “A Review of Numerical and Experimental Compensation Techniques for Skull-Induced Phase Aberrations in Transcranial Focused Ultrasound,” *International Journal of Hyperthermia* 30 (2014): 36–46, <https://doi.org/10.3109/02656736.2013.861519>.
20. C. T. M. Eneh, I. O. Afara, M. K. H. Malo, J. S. Jurvelin, and J. Töyräs, “Porosity Predicted From Ultrasound Backscatter Using Multivariate Analysis Can Improve Accuracy of Cortical Bone Thickness Assessment,” *Journal of the Acoustical Society of America* 141 (2017): 575–585, <https://doi.org/10.1121/1.4973572>.



21. S. B. Cho, H. J. Baek, K. H. Ryu, et al., "Clinical Feasibility of Zero TE Skull MRI in Patients With Head Trauma in Comparison With CT: A Single-Center Study," *AJNR. American Journal of Neuroradiology* 40 (2019): 109–115, <https://doi.org/10.3174/ajnr.A5916>.
22. W. A. Edelstein, M. Mahesh, and J. A. Carrino, "MRI: Time Is Dose—and Money and Versatility," *Journal of the American College of Radiology* 7 (2010): 650–652, <https://doi.org/10.1016/j.jacr.2010.05.002>.
23. G. Iori, F. Heyer, V. Kilappa, et al., "BMD-Based Assessment of Local Porosity in Human Femoral Cortical Bone," *Bone* 114 (2018): 50–61, <https://doi.org/10.1016/j.bone.2018.05.028>.
24. J. G. Minonzio, N. Bochud, Q. Vallet, et al., "Bone Cortical Thickness and Porosity Assessment Using Ultrasound Guided Waves: An Ex Vivo Validation Study," *Bone* 116 (2018): 111–119, <https://doi.org/10.1016/j.bone.2018.07.018>.
25. S. H. Tretbar, P. K. Plinkert, and P. A. Federspil, "Accuracy of Ultrasound Measurements for Skull Bone Thickness Using Coded Signals," *IEEE Transactions on Biomedical Engineering* 56 (2009): 733–740, <https://doi.org/10.1109/TBME.2008.2011058>.
26. A. Voulodimos, N. Doulamis, G. Bebis, and T. Stathaki, "Recent Developments in Deep Learning for Engineering Applications," *Computational Intelligence and Neuroscience* 2018 (2018): 1–2, <https://doi.org/10.1155/2018/8141259>.
27. R. J. G.v. Sloun, R. Cohen, and Y. C. Eldar, "Deep Learning in Ultrasound Imaging," *Proceedings of the IEEE* 108 (2020): 11–29, <https://doi.org/10.1109/JPROC.2019.2932116>.
28. R. Zemouri, N. Zerhouni, and D. Racoceanu, "Deep Learning in the Biomedical Applications: Recent and Future Status," *Applied Sciences* 9 (2019): 1526.
29. D. Aggrawal, M. Zafar, M. T. Islam, R. Manwar, D. Schonfeld, and K. Avanaki, "E-Unet: A Deep Learning Method for Photoacoustic Signal Enhancement," *In Proceedings of Photons Plus Ultrasound: Imaging and Sensing* 12379 (2023): 189–193.
30. D. Aggrawal, M. Zafar, D. Schonfeld, and K. Avanaki, "Deep Learning-Boosted Photoacoustic Microscopy With an Extremely Low Energy Laser," *In Photons Plus Ultrasound: Imaging and Sensing* 11960 (2022): 256–265.
31. R. Singh, R. Mehta, and N. Rajpal, "Efficient Wavelet Families for ECG Classification Using Neural Classifiers," *Procedia Computer Science* 132 (2018): 11–21, <https://doi.org/10.1016/j.procs.2018.05.054>.
32. N. O'Mahony, S. Campbell, A. Carvalho, et al., "Deep Learning versus Traditional Computer Vision," in *Proceedings of Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC)*, vol. 1, (New York, NY: Springer International Publishing, 2019): 128–144.
33. Y. Gao, W. Xu, Y. Chen, W. Xie, and Q. Cheng, "Deep Learning-Based Photoacoustic Imaging of Vascular Network Through Thick Porous Media," *IEEE Transactions on Medical Imaging* 41 (2022): 2191–2204, <https://doi.org/10.1109/TMI.2022.3158474>.
34. R. Manwar, X. Li, S. Mahmoodkalayeh, E. Asano, D. Zhu, and K. Avanaki, "Deep Learning Protocol for Improved Photoacoustic Brain Imaging," *Journal of Biophotonics* 13 (2020): e202000212.
35. N. Alqahtani, F. Alzubaidi, R. T. Armstrong, P. Swietojanski, and P. Mostaghimi, "Machine Learning for Predicting Properties of Porous Media From 2d X-Ray Images," *Journal of Petroleum Science and Engineering* 184 (2020): 106514, <https://doi.org/10.1016/j.petrol.2019.106514>.
36. K. Mohanty, O. Yousefian, Y. Karbalaiesadegh, M. Ulrich, Q. Grimal, and M. Muller, "Artificial Neural Network to Estimate Micro-Architectural Properties of Cortical Bone Using Ultrasonic Attenuation: A 2-D Numerical Study," *Computers in Biology and Medicine* 114 (2019): 103457, <https://doi.org/10.1016/j.compbiomed.2019.103457>.
37. E. Hosseini-Asl, R. Keynton, and A. El-Baz, "Alzheimer's Disease Diagnostics by Adaptation of 3D Convolutional Network," In *Proceedings of 2016 IEEE International Conference on Image Processing (ICIP)* 25–28 Sept. 2016, 126–130.
38. A. Hasan, H. Jalab, F. Meziane, H. Kahtan, and A. Al-Ahmad, "Combining Deep and Handcrafted Image Features for MRI Brain Scan Classification," *IEEE Access* 7 (2019): 79959–79967, <https://doi.org/10.1109/ACCESS.2019.2922691>.
39. M. Rebsamen, Y. Suter, R. Wiest, M. Reyes, and C. Rummel, "Brain Morphometry Estimation: From Hours to Seconds Using Deep Learning," *Frontiers in Neurology* 11 (2020): 244, <https://doi.org/10.3389/fneur.2020.00244>.
40. A. Wydra, E. Malyarenko, K. Shapoori, and R. G. Maev, "Development of a Practical Ultrasonic Approach for Simultaneous Measurement of the Thickness and the Sound Speed in Human Skull Bones: A Laboratory Phantom Study," *Physics in Medicine and Biology* 58 (2013): 1083–1102, <https://doi.org/10.1088/0031-9155/58/4/1083>.
41. A. Wydra and R. G. Maev, "A Novel Composite Material Specifically Developed for Ultrasound Bone Phantoms: Cortical, Trabecular and Skull," *Physics in Medicine and Biology* 58 (2013): N303–N319, <https://doi.org/10.1088/0031-9155/58/22/n303>.
42. L. Saint-Martin, D. Aggrawal, R. Manwar, D. Schonfeld, and K. Avanaki, "Relationship Between Skull Phantoms Physical Characteristics and Acoustic Wave Signal Features," *In Proceedings of Photons Plus Ultrasound: Imaging and Sensing* 12379 (2023): 161–168.
43. B. Hu, M. Wang, and Y. Du, "The Influence of Impedance Mismatch of the Medium With Gradient Change of Impedance on Acoustic Characteristics," *MATEC Web of Conferences* 283 (2019): 09005, <https://doi.org/10.1051/mateconf/201928309005>.
44. B. Toleva, "The Proportion for Splitting Data Into Training and Test Set for the Bootstrap in Classification Problems," *Business Systems Research Journal* 12 (2021): 228–242, <https://doi.org/10.2478/bsrj-2021-0015>.
45. A. Jain, K. Nandakumar, and A. Ross, "Score Normalization in Multimodal Biometric Systems," *Pattern Recognition* 38 (2005): 2270–2285, <https://doi.org/10.1016/j.patcog.2005.01.012>.
46. J. D. Jobson, "Multiple Linear Regression," in *Applied Multivariate Data Analysis: Regression and Experimental Design*, ed. J. D. Jobson (New York, NY: Springer New York, 1991), 219–398, [https://doi.org/10.1007/978-1-4612-0955-3\\_4](https://doi.org/10.1007/978-1-4612-0955-3_4).
47. D. Marquardt and R. Snee, "Ridge Regression in Practice," *American Statistician* 29 (1975): 3–20, <https://doi.org/10.1080/00031305.1975.10479105>.
48. E. Ostertagová, "Modelling Using Polynomial Regression," *Procedia Engineering* 48 (2012): 500–506, <https://doi.org/10.1016/j.proeng.2012.09.545>.
49. Y. Tanaka, Y. Seino, and K. Hattori, "Measuring Brinell Hardness Indentation by Using a Convolutional Neural Network," *Measurement Science and Technology* 30 (2019): 065012, <https://doi.org/10.1088/1361-6501/ab150f>.
50. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE* 86 (1998): 2278–2324, <https://doi.org/10.1109/5.726791>.
51. N. D. Truong, A. D. Nguyen, L. Kuhlmann, et al., "Convolutional Neural Networks for Seizure Prediction Using Intracranial and Scalp Electroencephalogram," *Neural Networks* 105 (2018): 104–111, <https://doi.org/10.1016/j.neunet.2018.04.018>.
52. W. L. Mao, H. I. K. Fathurrahman, Y. Lee, and T. W. Chang, "EEG Dataset Classification Using CNN Method," *Journal of Physics. Conference Series* 1456 (2020): 12017, <https://doi.org/10.1088/1742-6596/1456/1/012017>.
53. L. Alzubaidi, J. Zhang, A. J. Humaidi, et al., "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future

Directions,” *Journal of Big Data* 8 (2021): 53, <https://doi.org/10.1186/s40537-021-00444-8>.

54. M. T. Graham, J. Huang, F. X. Creighton, and M. A. Lediju Bell, “Simulations and Human Cadaver Head Studies to Identify Optimal Acoustic Receiver Locations for Minimally Invasive Photoacoustic-Guided Neurosurgery,” *Photoacoustics* 19 (2020): 100183, <https://doi.org/10.1016/j.pacs.2020.100183>.

55. R. Manwar, K. Kratkiewicz, S. Mahmoodkalayeh, et al., “Development and Characterization of Transfontanelle Photoacoustic Imaging System for Detection of Intracranial Hemorrhages and Measurement of Brain Oxygenation: Ex-Vivo,” *Photoacoustics* 32 (2023): 100538.

56. R. Manwar, M. T. Islam, A. Shoo, D. A. Pillers, and K. Avanaki, “Development of Ex-Vivo Brain Hemorrhage Phantom for Photoacoustic Imaging,” *Journal of Biophotonics* 16 (2023): e202200313.

57. S. M. Ranjbaran, M. Zafar, L. Saint-Martin, M. T. Islam, and K. Avanaki, “A Practical Solution to Improve the Field of View in Circular Scanning-Based Photoacoustic Tomography,” *Journal of Bio photonics* 17 (2024): e202400125.

58. J. Benavides-Lara, R. Manward, L. S. McGuire, et al., “Trans fontanelle Photoacoustic Imaging of Intraventricular Brain Hemorrhages in Live Sheep,” *Photoacoustic* 33 (2023): 100549.

59. K. Avantika and J. G. Gelovani, “Ultrasound and Multispectral Photoacoustic Systems and Methods for Brain and Spinal Cord Imaging Through Acoustic Windows Google Patents,” 2020.

60. L. Nie, X. Cai, K. Maslov, A. Garcia-Urbe, M. A. Anastasio, and L. V. Wang, “Photoacoustic Tomography Through a Whole Adult Human Skull With a Photon Recycler,” *Journal of Biomedical Optics* 17 (2012): 110506, <https://doi.org/10.1117/1.Jbo.17.11.110506>.