Article | Received 23 March 2025; Accepted 22 July 2025; Published 4 August 2025
https://doi.org/10.55092/neuroelectronics20250008

Domain enriched learning for skull induced phase aberration correction in ultrasound neuromodulation

Haichuan Zhang*, **Rudra Biswas**, **Mehdi Kiani and Vishal Monga**

Department of Electrical Engineering, Pennsylvania State University, University Park, USA

* Correspondence author; E-mail: haichuan@psu.edu.

Highlights:

- A Dual-Branch Skull-Induced Phase Aberration Correction network (DB-SIPAC) is proposed as a machine learning (ML) method for predicting time delay profiles.
- An Iterative Time-Delay Search algorithm (ITDS) is introduced to generate ground-truth training data for DB-SIPAC.
- DB-SIPAC outperforms state-of-the-art alternatives in both accuracy and speed.
- DB-SIPAC can excel even with limited training being domain enriched.
- DB-SIPAC enables precise neuromodulation in real-time.

Abstract: Transcranial focused ultrasound stimulation (tFUS) is a promising noninvasive neuromodulation technique that offers high spatial resolution compared to other non-invasive alternatives. However, its effectiveness is hindered by skull-induced phase aberrations, which distort the ultrasound beam and reduce focal accuracy. Existing solutions are often computationally intensive or fail to account for the effects of ultrasound wave changes outside the expected focal spot. In this paper, we approach phase correction as a model inversion problem and propose the Dual-Branch Skull-Induced Phase Aberration Correction Network (DB-SIPAC), a novel domain-enriched machine learning framework designed to efficiently predict time-delay profiles. The DB-SIPAC model consists of two specialized branches: the Pathway Branch, which focuses on the direct path from the ultrasound transducer to the target, and the Skull Branch, which incorporates full skull structure information to account for reflections and refractions. This dual-branch design enables rapid and accurate time-delay predictions, reducing computational time from minutes to less than a second. To facilitate learning, we generated a training dataset consisting of skull images and ground-truth time-delay profiles across a range of skull shapes and thicknesses. Since existing methods cannot guarantee time-delay profiles that ensure the expected focal spot has the maximum pressure within the brain, we introduce Iterative Time Delay Search (ITDS), a novel numerical algorithm that iteratively refines time-delay profiles generated by state-of-the-art methods to generate the ground truth for DB-SIPAC. Simulation results demonstrate that DB-SIPAC outperforms state-of-the-art alternatives, achieving perfect focal point alignment with the target and maximizing focal pressure, all

while providing real-time inference.

Keywords: ultrasound; phased array; transducer; high resolution; transcranial; neuromodulation; machine learning

1. Introduction

Low-intensity transcranial focused ultrasound stimulation (tFUS), functioning as a noninvasive method for neuromodulation, has proven to be effective in animals and humans, exhibiting superior millimeter-scale spatial resolution when compared to its noninvasive alternatives [1–4] for precise focusing within the brain. Recently, phased ultrasound (US) arrays have been developed to provide both focusing and steering capability using electronic beamforming [5–10]. However, these US transducers only provide precise focusing under ideal conditions with a homogeneous propagation medium. However, a significant challenge encountered by this technology is the aberration caused by the skull, particularly, in large animals and humans with thick skulls, leading to deviations from the ideal scenario. Aberration of phase, introduced by the skull can lead to considerable degradation in the accuracy, resolution, and focal pressure of the beam, thereby impairing its overall performance and effectiveness, defocusing of the ideal beamforming, and degrading the beam shape [11–14].

Effective experimental methods of phase correction, such as the implanted hydrophone by Clement and Hynyen, provide high-quality refocusing. However, the implication of this method requires invasive procedures, and the steering capability is limited [15], leading the authors to adopt the non-invasive beam steering approach [16]. Other effective non-invasive techniques such as Acoustic Stars [17], the Time Reversal (TR) method [18], Hybrid Angular Spectrum (HAS) [19], and Acoustoelectric Time Reversal (AETR) [20] are some examples of effective attempts on non-invasive US neuromodulation. Nevertheless, emerging machine learning (ML) algorithms offer promising solutions to correct phase aberrations and achieve optimal focus in transcranial ultrasound (US), independent of skull-induced distortions.

In recent years, ML has been widely applied across various domains, including medical imaging [21], speech recognition [22], and autonomous systems [23]. By training on large datasets, ML models can uncover complex patterns and make accurate predictions, often outperforming traditional techniques [24]. Deep learning, a subfield of ML, has gained significant attention for its capacity to model intricate data relationships using multi-layered neural networks. Recently, graph-based deep models have been proposed to capture non-local brain-network structures, examples include the Spatio-Temporal Graph Hubness Propagation model for dynamic brain-network classification [25], and the Long-Interval Spatio-Temporal Graph Convolution framework for brain-disease diagnosis [26]. At the same time, Convolutional Neural Networks (CNNs) are particularly well-suited for image-related tasks. CNNs, such as ResNet, have demonstrated exceptional performance in image recognition and classification due to their ability to extract hierarchical features through stacked convolutional layers [27,28]. In the context of transcranial ultrasound (US), deep learning, and CNNs in particular, has been successfully employed to reduce skull-induced artifacts in US imaging, as shown in recent simulation studies [29]. These models can predict and correct distortions caused by skull heterogeneities, thereby enhancing US image quality. Furthermore,

they are capable of generalizing across different skull anatomies, effectively adapting to variations in skull shape and thickness, and thus offer a robust solution for compensating phase aberrations.

Simulating the pressure generated by a US array in a specific skull configuration using a given time delay profile can be easily achieved using tools such as the k-Wave, a MATLAB toolbox (MathWorks R2023b, Natick, MA) [30]. However, the inverse problem—generating an optimal time delay profile based on a desired pressure distribution—is significantly more complex. This challenge positions skull-induced phase aberration correction as a model inversion problem.

A widely used approach for skull phase correction employs time-reversal techniques [18], where a simulated point source is placed at the desired focal point. The pulse emitted by this source propagates through the medium and is recorded by the transducer array. The recorded signals are then time-reversed and re-emitted, causing the wavefront to retrace its original path and refocus at the source, as if the experiment were played backward [31,32]. An alternative but conceptually similar approach is forward propagation, where the acoustic field from each transducer element is individually propagated to the target. By calculating the phase at the target location from these individual fields, the required corrections can be determined and applied to the transducer array [19]. Despite differences in implementation, both methods rely on simulation-based phase recovery, requiring at least one simulation run to achieve precise refocusing. However, this process is computationally expensive and time-consuming, making it unsuitable for real-time applications. Additionally, time-reversal methods only optimize pressure at the expected focal position while neglecting pressure distribution elsewhere in the brain. As a result, unintended regions may experience higher pressure than the intended focus spot, potentially compromising safety and effectiveness.

To address the computational limitations of time-reversal techniques, machine learning-based approaches have been introduced. Naftchi *et al.* [33] proposed an end-to-end deep learning model for simultaneously predicting the pressure field and correcting phase aberrations. In a key innovation, Tian *et al.* [34] developed a deep learning approach leveraging pulse-echo ultrasound to process backscattered US radio-frequency (RF) signals from the skull. Their approach estimates the local skull thickness and speed of sound (SOS) map and then applies the Fast Marching Method (FMM) [35] to generate time delay profiles. However, both methods focus on a small skull region, without considering the impact of the entire skull or cases where the focal spot is located deep within the brain, far from the skull; our work naturally extends these approaches by processing a full skull image.

In this paper, we propose a domain-enriched Dual-Branch Skull-Induced Phase Aberration Correction Network (DB-SIPAC), a machine-learning framework designed to predict time-delay profiles with high accuracy and speed. To generate a ground-truth time-delay profile dataset for training the model, we develop Iterative Time Delay Search (ITDS), a novel numerical iterative search algorithm that enhances time-delay profiles generated by other model-based methods, such as [18,19]. Using ITDS, we create a first-of-its-kind training dataset consisting of CT skull images, desired focal regions, and their corresponding ground-truth time-delay profiles. Our work makes the following contributions:

- A Fast New Domain Knowledge Driven ML Based Phase Aberration Correction Network for tFUS: We proposed DB-SIPAC, a machine learning framework designed to correct skull-induced phase aberrations with dual branch structure. This model integrates skull properties and US pathways to

generate time-delay profiles that incorporate full skull information while specifically focusing on the essential pathway. The Pathway Branch targets the direct path from the US transducers to the focal spot, capturing the primary effects of skull-induced aberrations. The Skull Branch processes the entire skull structure, accounting for reflections and refractions that affect the US beam beyond the direct pathway. This network will be trained on the ground-truth dataset and can predict time-delay profiles with high precision, reducing computation time from minutes to mere seconds, making it suitable for real-time clinical applications.

- Iterative Time Delay Search: A New Iterative approach to generate training ground truth. We introduce ITDS, a novel iterative method designed to enhance time-delay profiles generated by existing approaches, such as the time-reversal method. Initially, it leverages model-based methods such as TR [18] or HAS [19], to generate preliminary time-delay profiles. Then, using an iterative gradient descent algorithm, ITDS optimizes the time-delay profile for each US element through simulations. This ensures precise US wave focusing despite skull-induced distortions, providing an efficient and accurate solution for ground truth beam focusing in heterogeneous skull environments.
- A New Skull Image and Ground-truth Time Delay Profile Dataset: This study also introduces the Skull-Induced Phase Aberration Correction (SIPAC) Dataset. To our knowledge, it is the first to pair real-world skull images with their corresponding ground truth time delay profiles dataset. Generated using the previously mentioned ITDS approach, this dataset captures diverse skull shapes and thicknesses by optimizing the time delay profiles for each unique skull configuration. The dataset enables machine learning models to train and validate on a diverse range of skull shapes, thicknesses, and focal spot positions, ensuring robust generalization across different subjects. The dataset will be available on https://github.com/6zhc/SIPAC.

The rest of the article is organized as follows: In Section 2, we describe the generation of the binary skull file and the parameter settings for the simulation. In Section 3, we present the details of the proposed ML method, DB-SIPAC. In Section 4, we introduce the SIPAC dataset, which is used to train and validate the ML model, along with ITDS, which generates the ground-truth time-delay profiles. In Section 5, we validate the ML method along with other State-of-the-art method through simulation studies, demonstrating the practical advantages of DB-SIPAC. Section 6, concludes the paper and discusses future work.

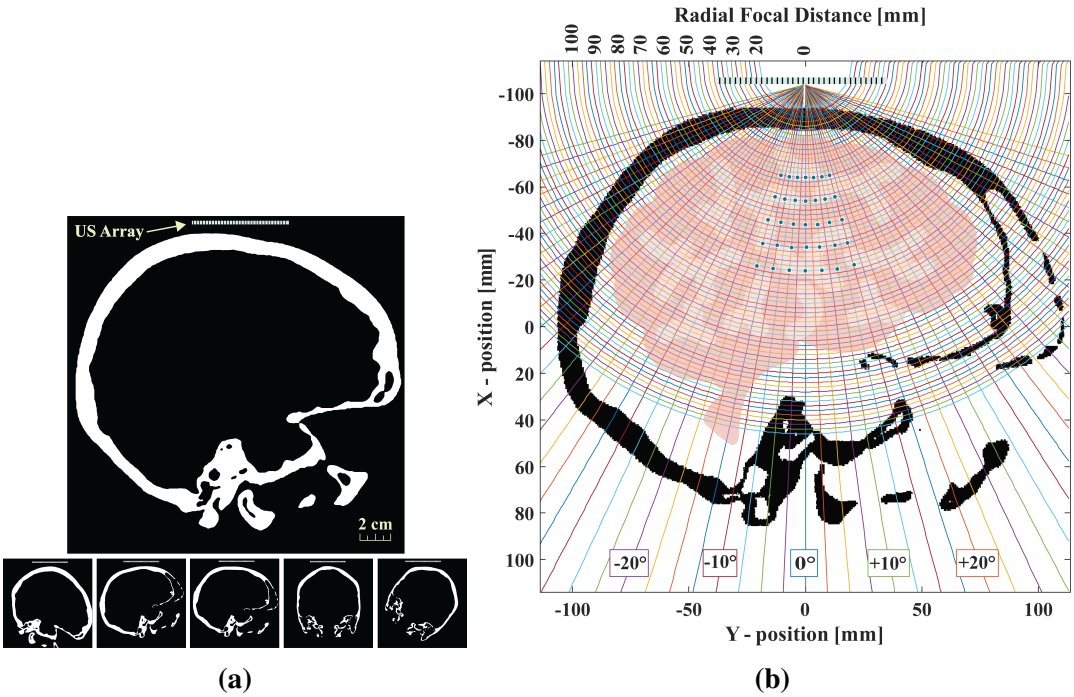
2. Methodology

2.1. Numerical skull model

The aim of this study is to develop a machine learning algorithm for generalized phase correction in a tFUS neuromodulation array, enhancing beamforming performance by accounting for the heterogeneous properties of the skull.

To validate the effectiveness of the method, a large dataset comprising skull images (to learn the effects of skull shapes and thicknesses) and focal points (focal length and steering angle pairs) is required. The base acoustic models employed in this study are developed using datasets from the Visible Human Project® (VHP) and the Center for Magnetic Resonance Research (CMRR) at the University of

Minnesota [36,37]. The dataset comprises CT scans from 12 adult subjects (9 males, 3 females), with each patient is selected for three CT slices, resulting in a total of 36 skull models. The average age of the subjects was 62.2 years with a standard deviation of 14.8 years. These datasets of skull cross-section figures were then converted to binary images using ImageJ, a popular image editing software for revising biomedical images. A total of 36 different binary skulls were extracted. To augment the dataset, existing images were rotated while keeping the array position fixed to target different skull portions. We further introduced systematic variations in the relative positioning between the skull and the ultrasonic device, emulating a clinical configuration in which the transducer is secured to a head-mounted baseplate. Inspired by [38], this strategy is predicated on the expectation that, in real-world applications, patient movement will not substantially alter the array-skull positioning. Figure 1a consists of some sample images from the dataset of different binary skulls.



(a) (b)

Figure 1. Overview of the binary skull model and focal points in polar coordinates. **(a)** The binary skull image used to simulate the effects of skull thickness and shape on the US beam, where white regions represent the skull and black regions denote soft tissue or water. **(b)** The skull in polar coordinates, illustrating focal points (blue dots) marked by the intersections of straight lines (representing steering angles) and arcs (representing focal distances). The US array is positioned at the origin. This figure demonstrates the selection of different focal spots within the skull for analysis.

To denote focal points inside the brain region enclosed by the skull as targets for stimulation, different regions are selected. An example can be seen in Figure 1b, where the straight lines denote the steering angle spaced at 2.5° angles from adjacent and the arcs denote the focal distance from the array, which are at 2 mm spacing from the adjacent. Each crossing of the straight lines and the arcs denotes a focal point. Focal points were uniformly sampled on a polar grid bounded by azimuthal angles from $-15°$ to $+15°$ in 5° increments, and radial distances of 40, 55, 70, 85, and 100 mm (measured from the center of the transducer aperture). These limits were determined during a preliminary sweep: steering

beyond $\pm 15^{\circ}$ or deeper than 110 mm resulted in loss of focal pressure and pronounced grating-lobe artifacts. Each binary-skull slice was therefore rotated to ensure that all grid points lay within the cranial cavity, guaranteeing unbiased, uniform coverage of all regions within the array's effective beamwidth. Approximately 35 focal points were selected from each skull dataset. With 36 skulls in the dataset, this yields a total of 1,260 focal points.

Once the dataset is obtained, the skull model is imported into the k-Wave toolbox. To ensure alignment with real-world measurements, the transducer dimensions specified in [9] are replicated. For the skull measurements, the guidelines in [39] are adhered to, with the Glabella-to-back-of-head distance maintained at 18–21 cm and the Bitragion breadth at 13–16 cm [40].

2.2. Simulation setting

We start with US focusing using phased linear arrays. Based on the previous elaborate works on US arrays for neuromodulation [9,10,41], a 2D numerical model is built using the k-Wave toolbox [30]. The 32-element linear transducer array, which was optimized in [10] for tFUS applications at 500 kHz, was modeled and simulated in k-Wave in this work. Although higher driving frequencies yield progressively greater on-axis pressures, owing to reduced focal spot size, they also incur intensified beam distortion and sidelobe formation using this low frequency array, which broadens the effective beam area. Simultaneously, attenuation in cranial tissue increases more than linearly with frequency, further diminishing net pressure at the focus. These competing phenomena result in a monotonic decline in the FoM as frequency rises above the baseline. Consequently, 500 kHz setting delivers the optimal balance between spatial resolution, transmission efficiency, and sidelobe suppression using this 500 kHz phased array. If a higher frequency is desired, a phased array transducer can be designed for such frequency using our design procedure in [9], based on which the proposed machine learning algorithm can be trained for optimal performance. For the simulation, a grid resolution of $\lambda/16$ is selected, which ensures 16 grid points per wavelength (λ). Such resolution is chosen to balance accuracy and computational efficiency. Consequently, the grid size is set to $(N_x, N_y) = (876, 876)$. The design parameters are listed in the following Table 1.

Table 1. List of simulation parameters in k-Wave.

Parameters	**Optimized US Array**
Sonication Frequency, f (kHz)	500
Number of US Elements, N	32
US Array Aperture, D (mm)	49.7
US Element Width, a (mm)	1.3
US Interelement Spacing, d (mm)	1.56
Kerf (mm)	0.26

In practical applications, the transducer should be positioned in close proximity to the skull, with a wave-guiding medium such as US gel used to fill the intervening space. This approach minimizes acoustic reflections and energy loss. Additionally, it is important to consider that the skull is overlaid by the scalp, and beneath it lies the cerebrospinal fluid (CSF) and brain tissue. The density and speed of sound (c) of these layers are given in Table 2.

Table 2. Acoustic parameters.

Layer	Density (g/cm^3)	Speed of Sound (m/s)
US Gel	0.99–1.02	1510
Skin	0.985	1540
CSF	1.0003–1.0008	1505
Brain Tissue	1.081	1546
Water	1	1500
Skull	~1.9	~2900

The Courant-Friedrichs-Lewy (CFL) condition is satisfied by setting its value to 0.3 [30]. In order to reduce complexity, it is a common approach to model the system replacing the US Gel, Skin, CSF, and Brain Tissue layers with water, as it shares properties very similar to the soft layers [37].

In a uniform medium, optimal beam steering and focusing at the point with azimuthal angle (θ_s) and focus distance (F) requires calculating the excitation time delay $\mathbf{t}[n]$ for the n^{th} element as follows:

$$\mathbf{t}[n] = \left(\frac{F}{c}\right)\left(1 - \sqrt{1 + \left(\frac{nd}{F}\right)^2 - \frac{2nd\sin(\theta_s)}{F}}\right) + t_0 \tag{1}$$

where F is the focal distance, θ_s is the azimuthal angle, c is the ultrasound velocity, and t_0 is a common delay offset [42]. Utilizing the time delay profile calculated for a uniform medium in the presence of a skull can cause the focal spot to deviate from the intended position.

To assess the impact of skull aberration, we first simulated a homogeneous medium composed entirely of water. We analyzed beamforming with a delay profile designed to focus and steer US waves to a target area at a depth (F) of 80 mm with a 0° beam-steering angle (θ_s), as shown in Figure 2a. The beamforming performance was further evaluated using a Full Width at Half Maximum (FWHM) profile, presented in Figure 2c. The results demonstrate ideal beamforming, characterized by minimal focal spread, and maximum pressure at the focal point (Target), as illustrated in Figure 2a,b.

In contrast, when the US arrays were positioned in front of the skull using the same delay profile with the initial uncorrected phase, the beam profile, depicted in Figure 2c, showed significant shifting of the maximum pressure point from the intended focal target. The beam was shifted laterally and axially by 1.23 mm and 17.86 mm, respectively, and the maximum US pressure at the corresponding focal point was reduced by 1.88 times . At the target focal point of $F = 80$ mm, $\theta_s = 0°$, the US pressure is reduced by 2.09 times.

In addition to the non-idealities introduced by the heterogeneous nature of the skull, a significant challenge of inconsistency must be addressed. In practice, the skull shape and thickness will vary among subjects, meaning that the phase correction optimized for one subject is unlikely to be applicable to another. Furthermore, these properties can vary within the same individual; for example, the parietal section of a subject's skull may have different properties compared to the frontal or temporal sections. This variability necessitates individualized phase correction for each subject.

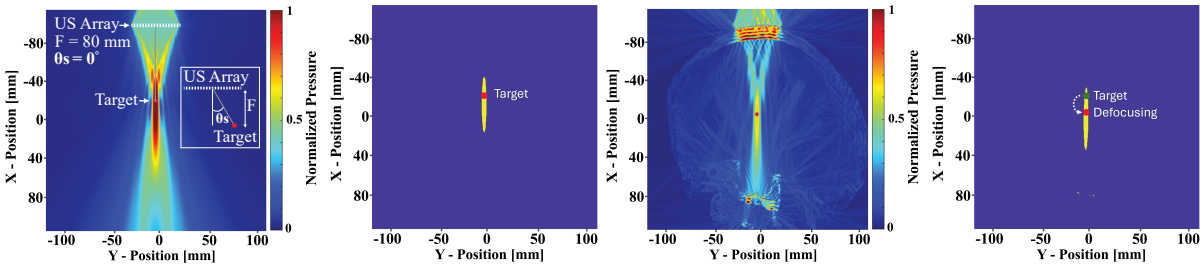


Figure 2. Beamforming and aberration effects in US focusing. **(a)** Ideal beamforming scenario in a homogeneous water medium, where the US beam focuses precisely at the target location ($F = 80$ mm, $\theta_s = 0°$). **(b)** Binary beam profile for ideal focusing, exhibiting sharp focusing with minimal spread and maximum pressure at target focus. **(c)** Beamforming in the presence of skull-induced phase aberrations, resulting in defocusing and a distorted beam shape. **(d)** Binary beam profile for skull-aberrated focusing, demonstrating significant degradation in focus quality & Defocusing due to skull effects.

2.3. Figure of Merit (FoM) for optimizing phased arrays for US neuromodulation

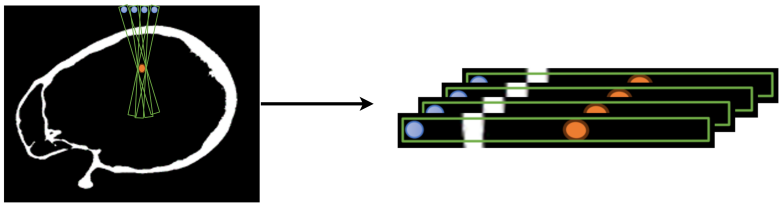
To evaluate the crossing condition for the transcranial focused US system, a metric is necessary for the quantitative comparison of different configurations. As discussed in [9], achieving an appropriate balance between power efficiency and spatial resolution is critical. This ensures sufficient US pressure at the target location while maintaining a small focal spot for better specificity, all with minimal input electrical power. To quantify this trade-off, we employ the FoM defined in [9], given by the following equation:

$$\text{FoM} = \frac{P_{\text{max}}}{\sqrt{A_{\text{beam}}}} \tag{2}$$

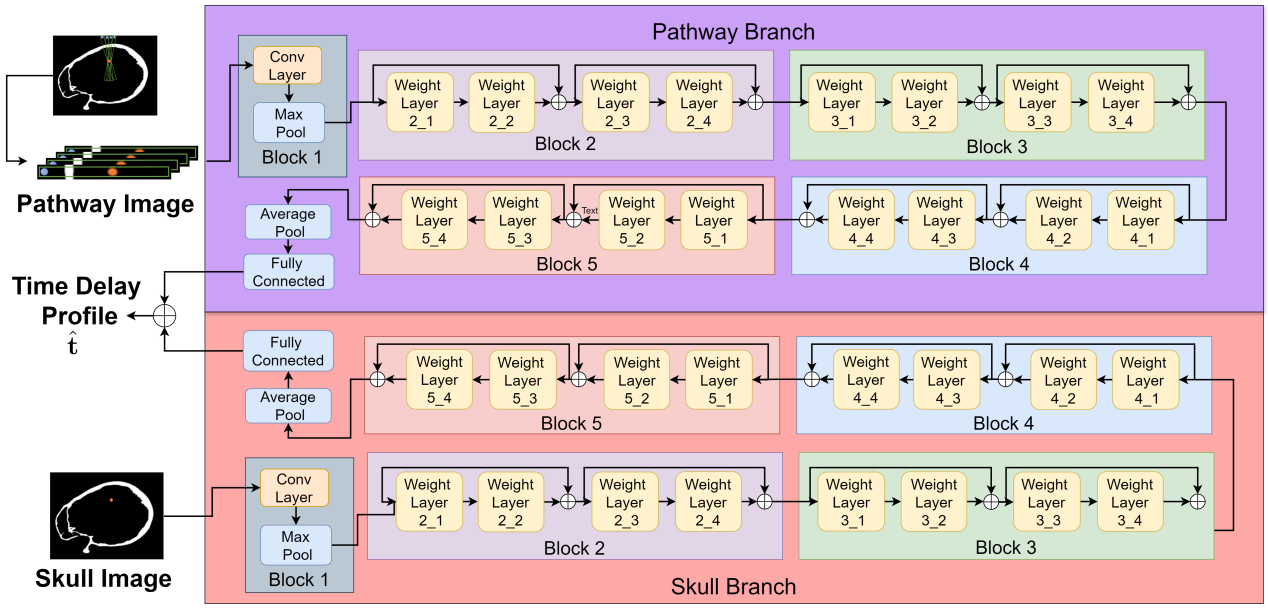
where P_{max} represents the maximum US pressure at the focal spot, and A_{beam} denotes the FWHM beam area. The FoM effectively captures the trade-off between achieving high focal pressure and minimizing the focal spot size, offering a clear metric for optimizing the tFUS system design.

3. Dual-branch skull-induced phase aberration correction network (DB-SIPAC)

Existing methods, such as the Hybrid Angular Spectrum (HAS) method [19], the Time Reversal (TR) method [18], rely on time-consuming simulations to generate time-delay profiles, making them impractical for real-time applications. Additionally, it does not account for pressure distribution across the entire brain, potentially leading to suboptimal time-delay profiles. As a result, neuromodulation treatments using this method may cause unintended side effects. To overcome these limitations, we propose a deep neural network called the Dual-Branch Skull-Induced Phase Aberration Correction Network (DB-SIPAC), whose detailed structure is illustrated in Figure 3. DB-SIPAC is designed to rapidly generate time-delay profiles while considering the entire skull and brain. The network consists of two specialized branches: the Pathway Branch and the Skull Branch.



(a) Pathway Image Extraction Progress



(b) DB-SIPAC Network Architecture

Figure 3. Overview of the DB-SIPAC network for phase aberration correction. **(a)** Pathway image extraction process: The blue points represent the US array elements, and the orange point indicates the expected focal spot. The US pathway in the skull image is bounded by a green box, which is extracted and input into the network. **(b)** Detailed architecture of the Dual-Branch Skull-Induced Phase Aberration Correction Network (DB-SIPAC): This illustration shows the flow of information through the Pathway Branch and Skull Branch. The Pathway Branch processes the direct path between the US elements and the expected focal spot, while the Skull Branch incorporates the entire skull's information to account for reflections and refractions. Both branches collaborate to predict the time delay profile for correcting phase aberrations, ensuring precise US focusing.

3.1. Pathway Branch

Due to the variable relative positions between the expected focal spot and the US array, learning the relationship between the skull image (and the expected position) and the time delay profile is challenging. To address this, all paths that US elements must traverse to reach the expected focal spot are first extracted, thereby reducing the network's computational burden by limiting it to generating time delay profiles corresponding to each path. In the extracted paths, the US elements are fixed on the left side of the image, and the medium information containing the skull along the path from the US element to the expected focal spot is marked as positive, while the medium information in areas to the right of the expected focal spot is marked as negative. This encoding informs the network about the position of the expected focal spot. The pathway extraction procedure is illustrated in Figure 3a, where the blue points represent the US array elements and the orange point signifies the expected focal spot. In the original skull image, the US

pathway is enclosed within a green box. This bounded region is subsequently extracted and input into the network, effectively capturing the geometric relationships between each US array element and the expected focal spot.

The Pathway Branch borrows from the design principles of ResNet [28], which addresses the vanishing gradient problem prevalent in deep networks. This branch includes an input block and four ResBlocks. The input block extracts features from the input, while each ResBlock processes these features sequentially. All ResBlocks share a similar architecture, comprising four weighted layers. Each weighted layer consists of a 3×3 convolutional layer, a batch normalization layer, and a ReLU activation layer. The specific parameters are detailed in Table 3. Additionally, the input features of each block are added to the output features of the second layer, and similarly, the output features of the second layer are added to the output features of the fourth layer. This residual connection allows the network to learn residual functions relative to the layer inputs, enhancing training efficiency and performance.

Table 3. Architecture detail for DB-SIPAC. The "P Output Size" column represents the output size of the block in the Pathway Branch, and the "S Output Size" column represents the output size t of the block in the Skull branch. Downsampling is performed by the first convolutional layer in Block3, Block4, and Block5.

Block Name	**Block Structure**	**P Output Size**	**S Output Size**
	Input	$512 \times 64 \times 32$	$876 \times 876 \times 2$
Block1	7×7 Conv, stride 2 3×3 max pool, stride 2	$256 \times 32 \times 64$	$438 \times 438 \times 64$
Block2	$\begin{bmatrix} 3 \times 3 \text{ Conv, BN, ReLU} \\ 3 \times 3 \text{ Conv, BN, ReLU} \end{bmatrix} \times 2$	$128 \times 16 \times 64$	$219 \times 219 \times 64$
Block3	$\begin{bmatrix} 3 \times 3 \text{ Conv, BN, ReLU} \\ 3 \times 3 \text{ Conv, BN, ReLU} \end{bmatrix} \times 2$	$64 \times 8 \times 128$	$110 \times 110 \times 128$
Block4	$\begin{bmatrix} 3 \times 3 \text{ Conv, BN, ReLU} \\ 3 \times 3 \text{ Conv, BN, ReLU} \end{bmatrix} \times 2$	$32 \times 4 \times 256$	$55 \times 55 \times 256$
Block5	$\begin{bmatrix} 3 \times 3 \text{ Conv, BN, ReLU} \\ 3 \times 3 \text{ Conv, BN, ReLU} \end{bmatrix} \times 2$	$16 \times 2 \times 512$	$28 \times 28 \times 512$
	average pool, 32-d fully conv	$1 \times 1 \times 32$	$1 \times 1 \times 32$

3.2. Skull Branch

Due to the reflection and refraction of ultrasound (US) waves, the entire skull's structure plays a crucial role in accurate predictions. Therefore, in addition to focusing on pathway information, the network must extract relevant details from the global skull image to accurately estimate the time delay profile. Since the

location of the focal spot varies in each case and is critical for information extraction, we added an extra channel to the binary skull image. This additional channel is entirely zeros except for a 7×7 solid circle marked with ones at the expected focal spot.

The structure of the Skull Branch is similar to that of the Pathway Branch, with differences in input dimensions. Like the Pathway Branch, the Skull Branch includes an input block and four ResBlocks, with specific parameters and dimensional differences outlined in Table 3. The output of the Skull Branch is added to the output of the Pathway Branch to obtain the final time delay profile.

In summary, the Pathway Branch primarily focuses on the pathway from the US elements to the expected focal spot, disregarding information that has minimal impact on the result, thereby facilitating easier training of the network. Conversely, the Skull Branch processes the entire skull information, capturing the effects of US reflection and refraction caused by non-pathway regions of the skull on the focal spot, thereby enhancing the generated time delay profile. Through our comparative simulations in Section 5, we found that the Pathway Branch plays a major role in the network, while the Skull Branch helps correct the pathway, resulting in improved outcomes.

4. Skull-induced phase aberration correction dataset

A machine learning approach can significantly accelerate time delay prediction. However, it requires a comprehensive training dataset that includes skull images, expected focal spots, and corresponding time delay profiles. To address this need, we introduce the Skull-Induced Phase Aberration Correction (SIPAC) Dataset. We also propose the Iterative Time Delay Search (ITDS) Algorithm, which improves upon the time delay profile generated by the existing model-based method, such as the Hybrid Angular Spectrum (HAS) method [19] and the Time Reversal (TR) method [18], to get the ground-truth time delay profiles.

Algorithm 1 provides the complete pseudocode for ITDS. The ITDS algorithm begins by selecting an expected focal spot location and computing an initial time delay profile using an existing model-based method, such as HAS [19] or the TR [18]. In this method, each element transmits a signal separately, allowing measurement of the travel time from the element to the expected focal position. Initializing the time delay profile with the existing model-based method results in faster convergence compared to other initialization methods, as it is typically close to the final ground truth time delay. Additionally, this initialization enhances the spatial peak US pressure at the expected focal spot, forming a regular elliptical shape within the FWHM beam region.

Directly comparing time delay profiles through analytical calculations is challenging; therefore, MATLAB's k-Wave is utilized to simulate and obtain the pressure map generated by the US elements using the time delay profile. To evaluate and compare different profiles, we define a comparison function $\mathrm{cmp}(\mathbf{t}, \mathbf{t}')$, where $\mathbf{t}$ denotes the current time delay profile and $\mathbf{t}'$ denotes the profile being compared. The function returns 1 if $\mathbf{t}$ is deemed superior to $\mathbf{t}'$, and -1 otherwise, based on the simulation results from k-Wave. The function is formally defined as follows:

$$\mathrm{cmp}(\mathbf{t},\mathbf{t}') = \begin{cases} 1 & \text{if } NormP(\mathbf{t}) > NormP(\mathbf{t}') \\ 1 & \text{if } NormP(\mathbf{t}) = NormP(\mathbf{t}') \\ & \quad \text{and} \quad P(\mathbf{t}) > P(\mathbf{t}') \\ -1 & \text{otherwise} \end{cases} \tag{3}$$

where $P(\cdot)$ represents the US pressure at the expected focal spot, and NormP$(\cdot)$ denotes the normalized US pressure, defined as:

$$\mathrm{NormP}(\mathbf{t}) = \frac{P(\mathbf{t})}{\mathrm{MaxP}}, \tag{4}$$

with MaxP being the maximum pressure within the skull. This function evaluates two time delay profiles $\mathbf{t}$ and $\mathbf{t}'$ based on their US pressure and normalized US pressure at the expected focal spot. The conditions ensure that $\mathbf{t}'$ is considered better only if it has a higher normalized pressure or, in the case of equal normalized pressure, a higher pressure at the expected focal spot.

Algorithm 1 Iterative Time Delay Search (ITDS) Algorithm

procedure GROUNDTRUTHTIMEDELAY($\mathbf{t}$)
  Pick an expected focal spot location, characterized by focal distance F and azimuthal angle θ_s.
  Calculate initial time delay using a model-based method, such as HAS [19] and TR [18] method.
  Simulate the time delay profile $\mathbf{t}$ with k-Wave.
  while not converged **do**
    for each element $n = 1,\ldots,N$ **do**
      $\mathbf{d} \leftarrow \mathrm{cmp}(\mathbf{t}+\mathbf{t}_n\xi, \mathbf{t}-\mathbf{t}_n\xi)\cdot\mathbf{t}_n$
      if $\mathrm{cmp}(\mathbf{t}+\beta\mathbf{d}, \mathbf{t}) > 0$ **then**
        $\beta \leftarrow \xi$
        while $\mathrm{cmp}(\mathbf{t}+2\beta\mathbf{d}, \mathbf{t}+\beta\mathbf{d}) > 0$ **do**
          $\beta \leftarrow 2\beta$
        end while
        $\mathbf{t} \leftarrow \mathbf{t}+\beta\mathbf{t}_n\mathbf{d}$
      end if
    end for
  end while
end procedure

In each iteration of the ITDS Algorithm, only one element of the time delay profile is optimized while keeping the time delays of all other elements constant. To achieve this selective optimization, an index vector $\mathbf{t}_n$ is defined as:

$$\mathbf{t}_n[i] = \begin{cases} 1 & \text{if } i = n, \\ 0 & \text{otherwise}, \end{cases}$$

where i ranges from 1 to N. This vector effectively isolates the n^{th} element in the time delay profile. Consequently, when updating the time delay profile from $\mathbf{t}$ to $\mathbf{t}+\beta\mathbf{t}_n\mathbf{d}$, only the n^{th} element's time delay

is modified by βd_n, while all other elements remain unchanged. This approach ensures precise and controlled optimization of individual time delays within the profile during each algorithmic iteration.

Since the comparison function between time delay profiles $\mathbf{t}$ and $\mathbf{t}'$ is not continuous, traditional gradient-based methods are not applicable. Instead, the direction $\mathbf{d}$ is defined as:

$$\mathbf{d} = \mathrm{cmp}(\mathbf{t} + \mathbf{t}_n\xi, \mathbf{t} - \mathbf{t}_n\xi) \cdot \mathbf{t}_n,$$

as shown in line 7 of Algorithm 1. This direction points towards the better time delay profile between adding or subtracting a minimal value ξ (which, in our simulations, represents the smallest unit affecting the simulation outcomes, specifically the time step in the k-Wave simulation) to the time delay of the n^{th} element. This strategy initially directs the algorithm towards maximizing the normalized pressure at the expected focal spot, typically approaching 1, which indicates that the pressure at the expected focal spot is the maximum pressure within the brain for that time delay profile. Subsequently, it enhances the pressure at the expected focal spot while maintaining the normalized pressure at 1.

With the direction $\mathbf{d}$ defined, the algorithm then focuses on determining the optimal search step size, as outlined from lines 8 to 12 in Algorithm 1. Unlike traditional optimization methods that employ a fixed learning step size β, we initiate β with ξ and subsequently double it to assess whether performance improves. This exponential growth method for determining the optimal learning step size significantly reduces the search time, saving approximately 75% of the computational time.

The ITDS program was implemented in MATLAB, running on MATLAB R2022a on a system equipped with an Intel(R) Core(TM) i9-7960X CPU @ 2.80GHz, 64 GB of RAM, and an NVIDIA GeForce RTX 4090 GPU.

In the SIPAC Dataset, 36 skull images were selected from the Visible Human Project® (VHP) and the Center for Magnetic Resonance Research (CMRR) project. These images were preprocessed as described in Section 2.1 to generate binary images, where a value of 1 denotes the skull and 0 represents tissue or water. Based on expert recommendations, 7 azimuthal angles and 5 distances from the US array to the focal spot were selected for each skull image, resulting in 35 expected focal spots per image. The selection of these expected focal spots is explained in detail in Section 2.1.

For each expected focal spot in a given skull image, the Iterative Time Delay Search (ITDS) algorithm was employed to compute the ground-truth time delay profile. This process yielded a total of 36 skull images $\times$ 7 azimuthal angles $\times$ 5 distances $=$ 1260 data points, each comprising a ground-truth time delay profile paired with a skull image and its corresponding expected focal spot, thus forming the SIPAC Dataset. Specifically, each data sample is mathematically represented as $\{(\mathbf{I}_j, F_j, \theta_j); \mathbf{t}_j\}_{j=1}^{M=1260}$, where $\mathbf{I}_j$ denotes a skull image, the azimuthal angle θ_j along with the focal distance F_j define the expected focal spot, and $\mathbf{t}_j$ represents the ground-truth time delay profile generated by the ITDS algorithm for that specific configuration.

To validate the effectiveness of the dataset, k-Wave was used to simulate US pressure based on the time-delay profiles. All pressures at the expected focal spot were confined within the FWHM beam area. Moreover, the average ratio of pressure at the expected focal spot to the maximum brain pressure was 1, indicating that the expected focal spot consistently had the highest pressure across all data points. Additionally, the average pressure at the expected focal spot was 1.430, and the average FoM, as defined

in Equation (2), was 0.0299.

5. Simulation results

5.1. Dataset preparation for training and evaluation

To efficiently leverage the limited data, 5-fold cross-validation was employed. First, 6 skull images and their corresponding data points were set aside as the test dataset, which did not participate in model training or validation. Then, the remaining 30 skull images were evenly divided into five parts, using one part for validation and the others for training in each fold, thereby obtaining 5 trained models. These models were tested separately, and the average of performance metrics (along with standard deviation) is reported. The dataset sizes used for each training, validation, and testing were as follows:

- 840 pairs (24 skull images) for training (66.67%),
- 210 pairs (6 skull images) for validation (16.67%),
- 210 pairs (6 skull images) for testing (16.67%).

5.2. Training loss function

L_1 loss, also known as Mean Absolute Error (MAE), is widely used in various domains due to its simplicity and robustness in minimizing the error between predicted and true values [43]. It is beneficial when dealing with data that contains outliers, as L_1 loss is less sensitive to extreme values compared to other loss functions, such as L_2 loss (Mean Squared Error) [44]. Recognizing that adding a constant to all 32 time delays in a profile does not impact the US array's actual performance, the loss calculation was adjusted. The delay profiles were normalized by subtracting the minimum value from each of the 32 time delays in both the ground truth and the network output. Then, the L_1 distance between these normalized values was calculated. This adjustment shifts the focus of the loss calculation to relative timing differences rather than absolute values, enhancing the network's ability to generalize across new datasets. The loss function is defined as follows:

$$\text{Loss}\left(\text{Norm}\left(\mathbf{t}\right),\text{Norm}\left(\hat{\mathbf{t}}\right)\right)=\frac{1}{M}\sum_{i=1}^{M}\left|\text{Norm}\left(\mathbf{t}_i\right)-\text{Norm}\left(\widehat{\mathbf{t}_i}\right)\right| \tag{5}$$

where

$$\text{Norm}\left(\mathbf{t}\right)=\left(t_1-\min\left(\mathbf{t}\right),t_2-\min\left(\mathbf{t}\right),\ldots,t_n-\min\left(\mathbf{t}\right)\right)$$

and $\mathbf{t}$ is the ground truth time delay profile and $\hat{\mathbf{t}}$ is generated time delay profile from the network.

The Adam optimizer was employed to learn the DB-SIPAC network parameters, initialized with a learning rate of 0.001 during the training. The models are implemented using the PyTorch framework [45], and all simulations are performed on two NVIDIA Titan X (12GB) GPUs.

5.3. Competing methods

- TR (Time Reversal) [18]: Time Reversal, also named virtual source, is opposite the target and transducer location. A simulation was run by sending a test pulse from the intended target to the US transducer array and recording the received delay at each element.

- HAS (Hybrid Angular Spectrum) [19]: The HAS method, a TR-based variant, individually propagates ultrasound waves from each transducer element and records the received delay at the target. These recorded delays are then used to construct the time-delay profile for phase correction.
- MedViT [46]: MedViT, a leading SOTA network in medical image classification, was adapted for time delay profile generation. Since no prior work exists on generating time delay profiles from skull images, we modified MedViT by extracting features before its last layer and feeding them into a fully connected layer to generate the time delay profiles. [a] MedViT's effectiveness has been demonstrated across various modalities, including US and CT images, making it a relevant comparison for our task.
- TUSNet [33]: TUSNet is the first end-to-end deep learning approach designed to jointly estimate the pressure field and perform phase aberration correction. We use the pretrained model provided by the authors and freeze the pressure field and absolute pressure decoders while fine-tuning the phase decoder and its associated encoder on our dataset.
- P-SIPAC (Pathway-focused Skull-Induced Phase Aberration Correction Network): An original ablation study introduced in this work. P-SIPAC uses only the Pathway Branch of DB-SIPAC as a standalone network, thereby isolating the impact of direct-path skull information while omitting global structural features. This ablation evaluates the contribution of the pathway prior by comparing its performance against the full dual-branch architecture.
- DB-SIPAC (Dual-Branch Skull-Induced Phase Aberration Correction Network): As described in Section 3 DB-SIPAC is a dual-branch network that, once trained, can takes patient-specific CT skull images as input and directly generate the element time-delay profile in under a second.

5.4. Quantitative performance measures

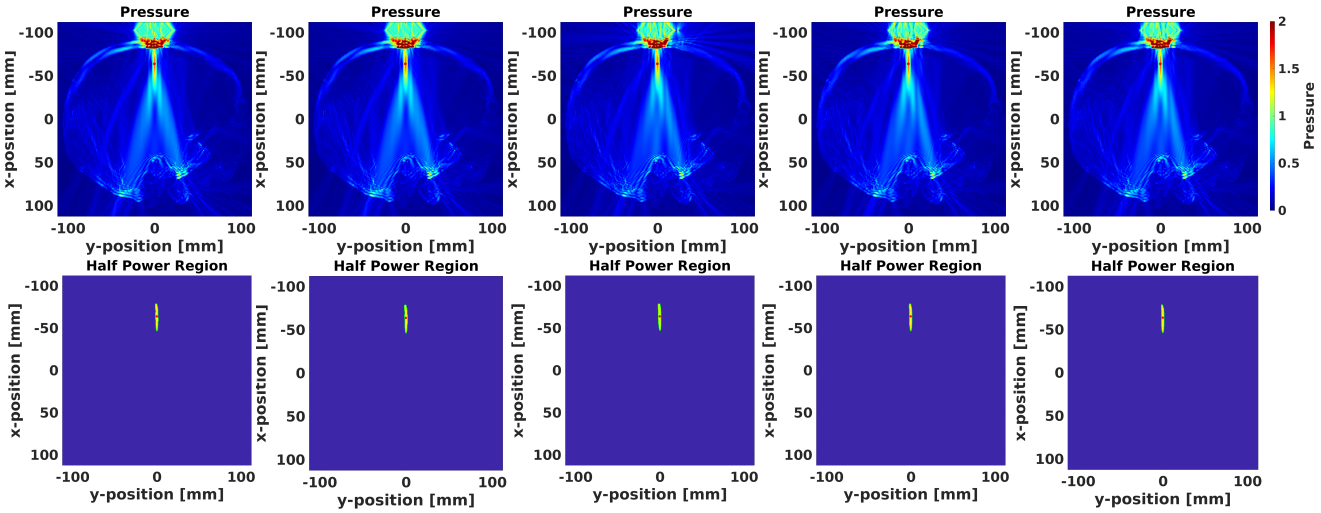
The five methods in Section 5.3 are compared with the following metrics. The first is the pressure at the expected focal spot, where a higher value is preferred. The second is normalized pressure at the expected focal spot, calculated using equation (4), which represents the ratio of pressure at the expected focal spot to the maximum pressure in the brain; a higher value is better, with a maximum of 1. The third metric is the Figure of Merit (FoM), defined in Section 2.3, where a higher FoM is achieved when the pressure at the expected focal spot is maximized while minimizing the FWHM beam area. The final metric is running time, which measures how long each method takes to generate a single time-delay profile for a given skull image and target focal spot.

5.5. Numerical evaluation

Figure 4 presents both the simulated acoustic beam profiles (top row) and their FWHM regions (bottom row), with the red dot marking the intended focus. In the TR and HAS cases, skull-induced aberrations visibly warp the main lobe and shift its half-power contour, such that the red-dot target sits at the periphery of the FWHM region rather than its center, indicating misalignment. MedViT successfully places the expected focal spot within the FWHM beam region; however, the maximum pressure in the brain appears several pixels away from the expected spot. By contrast, both P-SIPAC and DB-SIPAC enable the US

[a] As reviewed in Section 1, [34] have been proposed recently as ML methods for time delay prediction but do not utilize the CT skull image as input but instead process RF signals or an image patch. Therefore, we cannot experimentally compare with [34] but compare with MedViT [46], whose input is the same as that of P-SIPAC and DB-SIPAC.

array to accurately focus on the expected spot, placing the target at the point of highest pressure in the brain and centre it within a compact FWHM region. These qualitative observations agree with the quantitative improvements reported in Table 4.



(a) TR [18] **(b)** HAS [19] **(c)** MedViT [46] **(d)** P-SIPAC **(e)** DB-SIPAC

Figure 4. The upper row shows the simulated acoustic beam profiles generated using different time delay methods: Time Reversal (TR) [18], Hybrid Angular Spectrum (HAS) [19], MedViT [46], P-SIPAC, and DB-SIPAC, respectively. The lower row displays the FWHM beam regions, representing the focal spot for each method. The red point indicates the expected focal spot, and the boundary of each FWHM beam region is highlighted with a green line for clarity.

Table 4. Comparative results for different methods, showing the performance of the time delay profiles generated by each approach. The table compares key metrics, including pressure at the expected focal spot, normalized pressure (the ratio of the pressure at the focal point to the maximum pressure in the brain), Figure of Merit (FoM), and the computational time required. The results highlight the balance between accuracy and computational efficiency achieved by the proposed DB-SIPAC method compared to traditional methods.

	Pressure ↑	**Normalized Pressure** ↑	**Figure of Merit** ↑	**Running time** ↓
TR [18]	1.402	0.970	0.0291	32 secs
HAS [19]	1.397	0.973	0.0295	249 secs
MedViT [46]	1.401 ± 0.0043	0.9775 ± 0.0020	0.0289 ± 0.0004	0.53 secs
TUSNet [33]	1.399 ± 0.0056	0.9655 ± 0.0032	0.0286 ± 0.0005	1.20 secs
P-SIPAC	1.4170 ± 0.0038	0.9865 ± 0.0014	0.0295 ± 0.0002	0.32 secs
DB-SIPAC	1.4250 ± 0.0023	0.9923 ± 0.0006	0.0296 ± 0.0001	0.45 secs

Table 4 shows that the proposed DB-SIPAC outperforms all other methods across all metrics to evaluate the time delay profile. Since TR and HAS are simulation-based methods that do not require training, they consistently generate the same time-delay profiles regardless of how the training and validation datasets are divided. However, without the enhancement of ITDS, these methods do not

account for pressure distribution in other parts of the brain. This can result in a lower normalized pressure compared to other methods. For ML methods trained on datasets, performance can vary depending on the training setup. To fairly evaluate the robustness of each ML method, we trained each method five times using different training/validation splits while keeping the testing dataset unchanged. We then reported the average performance metrics along with their standard deviations. MedViT, a state of the art ML method for processing ultrasound image, struggles to fully capture the complex relationship between skull images, expected focal spots, and optimal time-delay profiles, due to its lack of domain knowledge. As a result, its performance is slightly lower than that of DB-SIPAC and P-SIPAC. Additionally, its standard deviation is larger, indicating potential instability because of the training dataset. TUSNet [33], an end-to-end network originally designed for simultaneous pressure field reconstruction and phase aberration correction, incurs longer runtime than other ML methods focused solely on aberration correction. Because it emphasizes pressure estimation as well, its effectiveness in phase correction is reduced, leading to lower performance in our evaluation. P-SIPAC has the shortest runtime but at the cost of slightly reduced performance compared to DB-SIPAC. DB-SIPAC achieves the highest pressure at the expected focal spot, with a normalized pressure close to 1. This suggests that, in most cases, the pressure achieved by DB-SIPAC at the expected focal spot is the highest. Although DB-SIPAC requires an additional 0.13 s of inference time (0.45 s compared to 0.32 s for P-SIPAC), this latency remains well below the sub-second threshold typically considered "real-time" for closed-loop tFUS workflows. Moreover, it is still nearly two orders of magnitude faster than other existing SOTA methods such as TR and HAS. In contrast, DB-SIPAC achieves approximately a 0.8% increase in peak pressure and a 0.6% improvement in normalized pressure. This suggests that less acoustic energy is distributed to unintended locations when maintaining the same pressure at the target spot, resulting in a safer and less harmful profile for patients. Therefore, while the computational overhead introduced by DB-SIPAC is modest and clinically negligible, the corresponding enhancements in acoustic performance provide meaningful practical benefits.

5.6. Training robustness: generalization benefits of DB-SIPAC

While a key advantage of ML-based methods is their versatility across different physical setups and their ability to enable real-time inference; a major practical concern is their ability to generalize beyond the training dataset. The dataset for training and testing is typically derived from either simulated patient skull CT images or real-world measurements using skull specimens, US arrays, and hydrophones, both of which are expensive and difficult to obtain. As a result, acquiring sufficient data for training machine learning models poses a significant challenge.

To evaluate performance under limited training data, we reduced the dataset size from 840 pairs (24 skull images) to 630 pairs (18 skull images), retaining approximately 75% of the original dataset. As shown in Figure 5, all ML methods experienced a decline in performance due to the dataset reduction. However, DB-SIPAC and P-SIPAC, as domain-enriched networks specifically designed for Skull-Induced Phase Aberration Correction in ultrasound neuromodulation, exhibited greater generalization than other methods, maintaining relatively consistent performance despite the smaller dataset. From a practical deployment standpoint, such training robustness is a highly desirable benefit of DB-SIPAC as ML methods often place unreasonable demands on both quantity and quality of the desirable training dataset.

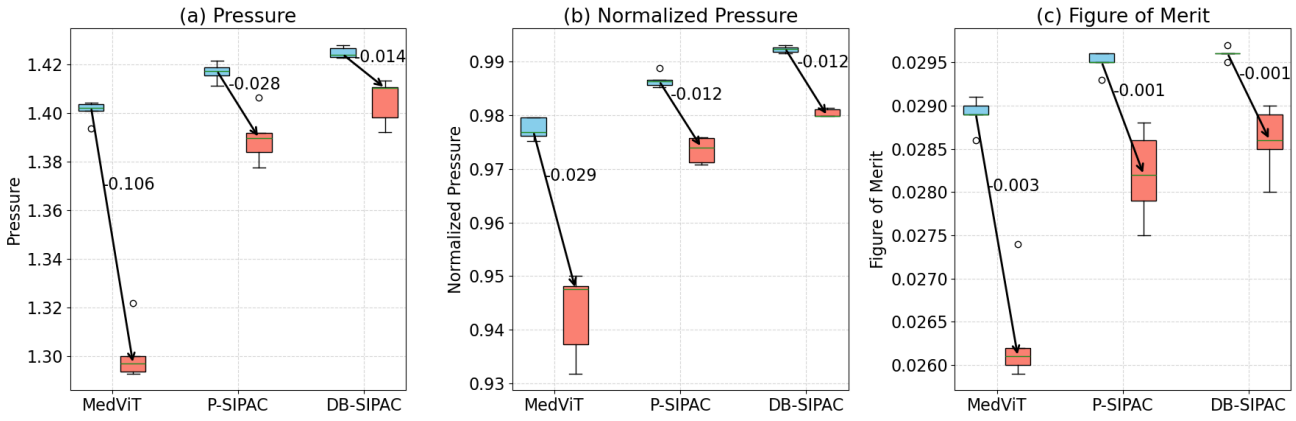


Figure 5. Box plot of merit across methods in fully trained and low-training scenarios. **(a)** Box plot of pressure at the expected focal spot. **(b)** Box plot of normalized pressure, representing the ratio of pressure at the focal spot to the maximum pressure in the brain. **(c)** Box plot of the Figure of Merit (FoM), which quantifies beam focusing efficiency relative to the FWHM beam area. The blue box represents the ML method with full training, while the red box represents the ML method with limited training.

5.7. Model interpretability

For model interpretability, gradient-based saliency maps were computed to determine which skull features most significantly contribute to the predictions of the DB-SIPAC model. The analysis was performed on selected images from the Pathway Branch, shown in Figure 3a. The saliency maps, presented in Figure 6, show that areas of high activation consistently correspond to the bone–water interface and regions of high cortical curvature along the ultrasound propagation path. This provides strong evidence that the network attends to anatomically meaningful structures during the time-delay estimation process.

Figure 6. Model interpretability visualized through gradient-based saliency maps, corresponding to the selected input images in Figure 3a. Brighter pixels represent features most influential for the model's time-delay estimation. These critical features consistently localize to the skull-water interface and sections of high curvature, confirming the network's focus on anatomically relevant structures.

6. Discussion and conclusion

The primary challenge in transcranial focused ultrasound stimulation (tFUS) is overcoming skull-induced phase aberrations, which distort the US beam and reduce both accuracy and effectiveness. Traditional simulation-based methods are computationally intensive and fail to account for US pressure effects across the entire brain, leading to potential inaccuracies. To address these limitations, this study introduces DB-SIPAC, an innovative, noninvasive, machine learning-based approach designed to efficiently correct skull-induced phase aberrations.

DB-SIPAC introduces a dual-branch architecture, comprising a Pathway Branch, which captures the direct US pathway to the target focal spot, and a Skull Branch, which accounts for broader skull

structure details, including reflections and refractions. This comprehensive approach ensures robust correction of phase aberrations, outperforming state of the art model-based methods as well as black-box ML alternatives. The validation on our simulating set-ups show that DB-SIPAC achieves perfect focal point alignment and maximizes pressure at the target location, exceeding the accuracy of competing methods but in a fraction of the time. The ability to generate optimal time delay profiles in real time makes DB-SIPAC particularly suitable for clinical applications, where quick adjustments are crucial.

While DB-SIPAC offers a feasible solution for clinical use with an attractive computational cost-accuracy balance, some limitations remain. We recognize that our SIPAC dataset, comprising 36 skull models from 12 adult donors, represents only a subset of the full spectrum of human cranial anatomy. Consequently, variations in paediatric skull geometry, extremes of bone density, and sex-specific morphological features may be under-represented, potentially limiting the model's generalizability. Expanding the dataset with more diverse skull profiles would enhance the model's generalizability. We also recognize that our uniform grid of 35 focus points per skull, while sufficient for broad coverage, may not be dense enough to resolve highly localized aberration patterns; under-sampling in regions with rapidly varying phase delay could introduce subtle bias. In future work, we will explore increasing the focal-point density to achieve more comprehensive, high-resolution coverage of all relevant intracranial trajectories. Furthermore, DB-SIPAC currently addresses static skull scenarios. Future research should explore dynamic adjustments, accommodating real-time changes in skull properties and patient movement to improve adaptability in clinical environments. Integrating patient-specific data, such as individualized skull characteristics obtained from imaging techniques like CT or MRI, could further refine the model's accuracy. The ultimate goal is to enable precise neuromodulation that tailors the corrections to each patient's unique anatomical features, and increases the overall efficacy of treatments.

Finally, while DB-SIPAC relies on static CT-driven compensation, future work will integrate adaptive, echo-based feedback loops, e.g. real-time time-reversal adjustments or ultrasound backscatter tracking, to correct for patient motion and dynamic skull–brain interface changes. Our model's interpretability offers an initial window into DB-SIPAC's reasoning, showing that it prioritizes physical boundaries and curvature zones—key sources of phase distortion. Building on this, further work could quantify feature importance via integrated gradients or layer-wise relevance propagation. Integrating closed-loop feedback (e.g. from embedded hydrophones or ultrafast imaging) and fusing pre-operative CT with live ultrasound or MRI guidance would enable continuous, real-time phase refinement in clinical settings.

Acknowledgments

This work was supported by the National Science Foundation (NSF) under Grant ECCS-2143557.

Authors' contribution

Conceptualization and formal analysis were led by Rudra Biswas and Mehdi Kiani. The methodology, software, validation, and data curation were developed and performed by Haichuan Zhang and Vishal Monga. Supervision, project administration, resources, and funding acquisition were provided by Mehdi Kiani and Vishal Monga. All authors contributed to the investigation, visualization, and writing of the

manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of interests

The authors declare no conflict of interest.

References

[1] Bystritsky A, Korb AS, Douglas PK, Cohen MS, Melega WP, *et al.* A review of low-intensity focused ultrasound pulsation. *Brain Stimul.* 2011, 4(3):125–136.

[2] Kim T, Park C, Chhatbar PY, Feld J, Mac Grory B, *et al.* Effect of low intensity transcranial ultrasound stimulation on neuromodulation in animals and humans: an updated systematic review. *Front. Neurosci.* 2021, 15:620863.

[3] Baek H, Pahk KJ, Kim H. A review of low-intensity focused ultrasound for neuromodulation. *Biomed. Eng. Lett.* 2017, 7:135–142.

[4] Wagner T, Valero-Cabre A, Pascual-Leone A. Noninvasive human brain stimulation. *Annu. Rev. Biomed. Eng.* 2007, 9(1):527–565.

[5] Tyler WJ, Tufail Y, Finsterwald M, Tauchmann ML, Olson EJ, *et al.* Remote excitation of neuronal circuits using low-intensity, low-frequency ultrasound. *PLoS One* 2008, 3(10):e3511.

[6] Gougheri HS, Kiani M. Microscopic ultrasound stimulation of neural tissue. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Cleveland, USA, October 17–19, 2018, pp. 1–4.

[7] Gougheri HS, Dangi A, Kothapalli SR, Kiani M. A comprehensive study of ultrasound transducer characteristics in microscopic ultrasound neuromodulation. *IEEE Trans. Biomed. Circuits Syst.* 2019, 13(5):835–847.

[8] Seo J, Shin H, Cho S, Lee S, Ryu W, *et al.* A phased array ultrasound system with a robotic arm for neuromodulation. *Med. Eng. Phys.* 2023, 118:104023.

[9] Ilham SJ, Kashani Z, Kiani M. Design and optimization of ultrasound phased arrays for large-scale ultrasound neuromodulation. *IEEE Trans. Biomed. Circuits Syst.* 2021, 15(6):1454–1466.

[10] Ilham SJ, Kiani M. Towards high-resolution ultrasound neuromodulation with crossed-beam phased arrays. *IEEE Trans. Biomed. Circuits Syst.* 2023, 17(3):534–546.

[11] Tsai PC, Gougheri HS, Kiani M. Skull impact on the ultrasound beam profile of transcranial focused ultrasound stimulation. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, July 23–27, 2019, pp. 5188–5191.

[12] Culjat MO, Goldenberg D, Tewari P, Singh RS. A review of tissue substitutes for ultrasound imaging. *Ultrasound Med. Biol.* 2010, 36(6):861–873.

[13] White D, Clark J, Chesebrough J, White M, Campbell J. Effect of the skull in degrading the display of echoencephalographic B and C scans. *J. Acoust. Soc. Am.* 1968, 44(5):1339–1345.

[14] O'Reilly MA, Muller A, Hynynen K. Ultrasound insertion loss of rat parietal bone appears to be proportional to animal mass at submegahertz frequencies. *Ultrasound Med. Biol.* 2011, 37(11):1930–1937.

[15] Seip R, VanBaren P, Ebbini ES. Dynamic focusing in ultrasound hyperthermia treatments

using implantable hydrophone arrays. *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 1994, 41(5):706–713.

[16] Clement GT, Hynynen K. A non-invasive method for focusing ultrasound through the human skull. *Phys. Med. Biol.* 2002, 47(8):1219.

[17] Pernot M, Montaldo G, Tanter M, Fink M. 'Ultrasonic stars' for time-reversal focusing using induced cavitation bubbles. *Appl. Phys. Lett.* 2006, 88(3):034102.

[18] Fink M, Cassereau D, Derode A, Prada C, Roux P, *et al.* Time-reversed acoustics. *Rep. Prog. Phys.* 2000, 63(12):1933.

[19] Leung SA, Moore D, Webb TD, Snell J, Ghanouni P, *et al.* Transcranial focused ultrasound phase correction using the hybrid angular spectrum method. *Sci. Rep.* 2021, 11(1):6532.

[20] Preston C, Alvarez AM, Allard M, Barragan A, Witte RS. Acoustoelectric time-reversal for ultrasound phase-aberration correction. *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 2023, 70(8):854–864.

[21] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, *et al.* A survey on deep learning in medical image analysis. *Med. Image Anal.* 2017, 42:60–88.

[22] Hinton G, Deng L, Yu D, Dahl GE, Mohamed Ar, *et al.* Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process. Mag.* 2012, 29(6):82–97.

[23] Chen C, Seff A, Kornhauser A, Xiao J. Deepdriving: learning affordance for direct perception in autonomous driving. In *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, December 7–13, 2015, pp. 2722–2730.

[24] Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015, 349(6245):255–260.

[25] Zhu Q, Li S, Meng X, Xu Q, Zhang Z, *et al.* Spatio-temporal graph hubness propagation model for dynamic brain network classification. *IEEE Trans. Med. Imag.* 2024, 43(6):2381–2394.

[26] Li S, Zhu Q, Guan D, Shen B, Zhang L, *et al.* Long-interval spatio-temporal graph convolution for brain disease diagnosis. *IEEE Trans. Instrum. Meas.* 2025, 74:4004511.

[27] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015, 521(7553):436–444.

[28] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, June 27–30, 2016, pp. 770–778.

[29] Tang Y, Nekkanti S, Rohera V, Nephew B, King JA, *et al.* Deep-learning-based skull-induced artifact reduction for transcranial ultrasound imaging: simulation study. In *Medical Imaging 2023: Ultrasonic Imaging and Tomography*, San Diego, USA, February 22–23, 2023, pp. 183–189.

[30] Treeby BE, Cox BT. k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. *J. Biomed. Opt.* 2010, 15(2):021314.

[31] Marquet F, Pernot M, Aubry JF, Montaldo G, Marsac L, *et al.* Non-invasive transcranial ultrasound therapy based on a 3D CT scan: protocol validation and in vitro results. *Phys. Med. Biol.* 2009, 54(9):2597.

[32] Aubry JF, Tanter M, Pernot M, Thomas JL, Fink M. Experimental demonstration of noninvasive

transskull adaptive focusing based on prior computed tomography scans. *J. Acoust. Soc. Am.* 2003, 113(1):84–93.

[33] Naftchi-Ardebili K, Singh K, Popelka GR, Pauly KB. A deep-learning model for one-shot transcranial ultrasound simulation and phase aberration correction. *arXiv* 2024, arXiv:2410.19995.

[34] Tian Z, Olmstead M, Jing Y, Han A. Transcranial phase correction using pulse-echo ultrasound and deep learning: a 2D numerical study. *IEEE Trans. Ultrason., Ferroelectr., Freq. Control* 2024, 71(1):117–126.

[35] Sethian JA. Fast marching methods. *SIAM Rev.* 1999, 41(2):199–235.

[36] Ackerman M. The visible human project: a resource for education. *Acad. Med.* 1999, 74(6):667–670.

[37] Mueller JK, Ai L, Bansal P, Legon W. Numerical evaluation of the skull for human neuromodulation with transcranial focused ultrasound. *J. Neural Eng.* 2017, 14(6):066012.

[38] Yang Y, Yuan J, Field RL, Ye D, Hu Z, *et al.* Induction of a torpor-like hypothermic and hypometabolic state in rodents by ultrasound. *Nat. Metab.* 2023, 5(5):789–803.

[39] Lee Jh, Hwang Shin SJ, Istook CL. Analysis of human head shapes in the united states. *Int. J. Hum. Ecol.* 2006, 7(1):77–83.

[40] Bushby K, Cole T, Matthews J, Goodship J. Centiles for adult head circumference. *Arch. Dis. Child.* 1992, 67(10):1286–1287.

[41] Javid A, Ilham S, Kiani M. A review of ultrasound neuromodulation technologies. *IEEE Trans. Biomed. Circuits Syst.* 2023, 17(5):1084–1096.

[42] Von Ramm OT, Smith SW. Beam steering with linear arrays. *IEEE Trans. Biomed. Eng.* 1983, BME-30(8):438–452.

[43] Terven J, Cordova-Esparza DM, Ramirez-Pedraza A, Chavez-Urbiola EA. Loss functions and metrics in deep learning. *arXiv* 2023, arXiv:2307.02694.

[44] Huang X. Evaluating loss functions and learning data pre-processing for climate downscaling deep learning models. *arXiv* 2023, arXiv:2306.11144.

[45] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, *et al.* Pytorch: an imperative style, high-performance deep learning library. *arXiv* 2019, arXiv:1912.01703.

[46] Manzari ON, Ahmadabadi H, Kashiani H, Shokouhi SB, Ayatollahi A. MedViT: a robust vision transformer for generalized medical image classification. *Comput. Biol. Med.* 2023, 157:106791.