Cortical Bone Thickness Assessment from Multi-frequency Ultrasound RF Data using a Convolutional Architecture with Multi-head Attention

Hossam H. Sultan[1], Enrico Grisan[1], Paul Dryburgh[2], Laura Peralta[2], Sevan Harput[1]
[1]South Bank Applied Bioengineering Research, London South Bank University, London, UK
[2]School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK
E-mail: abdelh13@lsbu.ac.uk

***Abstract*— Cortical bone thickness is an important predictor of bone strength and fracture risk, and accurate classification is crucial for the diagnosis and treatment of osteoporosis. The thinning of the cortical layer, indicative of compromised bone microarchitecture due to imbalanced formation and loss, underscores its significance. Nonetheless, quantifying bone thickness is challenging due to the diverse skeletal sites and subject variations in bone structure and properties.**

A potential solution lies in multi-frequency ultrasound assessment of cortical bone, enabling comprehensive property characterization across varying wavelengths and penetration depths. This research strives to establish a robust methodology for evaluating cortical bone thickness by leveraging a convolutional model with an attention mechanism to analyse multi-frequency ultrasound data.

***Keywords*— Bone characterization, Multi-frequency ultrasound, Deep learning, Attention mechanism, Chirp signal.**

I. Introduction

The primary component of the skeleton in humans is made up of mineralized hard bone. It performs a variety of essential tasks, including the generation of red and white blood cells and the storage of minerals [1, 2]. A significant biomarker of bone fragility that indicates a risk of fractures is cortical bone thickness and porosity [3]. Therefore, in order to lower the risk of fracture, it is clinically necessary to evaluate the quality of the cortical bone microstructure.

Over the past several decades, numerous methods utilizing ultrasound have been introduced for the evaluation of both micro and macro bone structures. In the 1990s, ultrasonic technology was first used to assess bones, which was relatively early in the area [4]. Subsequently, various Quantitative Ultrasound (QUS) techniques have been developed, with some methods relying on ultrasound attenuation while traversing through bone. However, the fact that it is only applicable to peripheral sites affects how reliable it is for scanning and evaluating various bone sites. In another study, an ex-vivo research comparing ultrasonic and Micro-CT for porosity and thickness measures has been presented [3]. Nowadays, the most often used techniques to provide a moderately high-resolution ultrasound scan are synthetic aperture and compound imaging. This method requires a lot of computation power and takes a long time to acquire.

The attention mechanism in Convolutional Neural Networks (CNNs) has significance in enhancing the network's capacity to concentrate on important characteristics in an input [5]. By dynamically highlighting significant regions, the attention mechanism allows the CNN to allocate more computational resources to key areas. CNNs frequently utilize approaches like channel attention, which amplifies informative channels in feature maps, and spatial attention, which highlights certain spatial regions.

In this study, a new approach for cortical bone classification by integrating RF data and deep neural networks across three distinct phases is presented. Firstly, employing the Finite-Difference Time-Domain (FDTD) method to simulate the propagation of ultrasound within cortical bone utilizing 8 diverse transmit signals within the frequency range of 1-8 MHz. Secondly, a CNN with 19 layers and 3 parallel paths for multi-head attention was introduced, demonstrating a unique ability for extracting involved features from multidimensional ultrasound RF data. Furthermore, a consensus strategy harmonized RF data from successive ultrasound array elements (64 signal/acquisition) to further eliminate errors. Lastly, the third phase involved real animal bone scanning through single pulse-echo acquisitions.

II. Methods

A. *Simulation environemnt*

Figure 1 illustrates the configuration of the simulation setup. Cortical bone simulation was conducted by employing a 2D FDTD approach [6], including diverse porosity levels, ranging from 0% to 20%, and thickness variations ranging from 1 mm to 8 mm. The cortical bone was represented as a linear elastic 2D plate, and a range of environments were generated to replicate both intact and pathological tissue conditions. The simulation proposed complete immersion of the cortical plate within a soft tissue-like medium, while perfectly matched layers surrounded the model to minimize undesirable boundary reflections. The dynamic propagation of mechanical waves within continuous media, governed by Hooke's law (as expressed in Equation 2), is captured as:

$$\boldsymbol{\rho(x)\frac{\partial v_i}{\partial t}(x,t) = \sum_{j=1}^{d}\frac{\partial T_{ij}}{\partial x_j}(x,t) + f_i(x,t)} \qquad (1)$$

$$\frac{\partial T_{ij}}{\partial t}(x,t) = \sum_{j=1}^{d}\sum_{i=1}^{d} c_{ijkl}(x)\,\frac{\partial v_k}{\partial x_l}(x,t) + \theta_{ij}(x,t) \qquad (2)$$

Where, x and t are the space and time variables, $\rho(x)$ is mass density, $c(x)$ is the fourth order rigidity tensor. $v(x,t)$ are the vector components of the particle velocities. $T(x,t)$ are the components of the stress tensor. f_i and θ_i are source terms, in which f is the vector components of force sources and θ is the tensor components of strain sources.

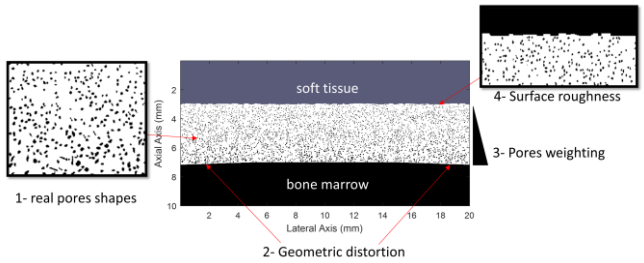

Figure 1: Simulation environment shows the bone model with realistic pores extracted from ex-vivo bone samples, curvature, pores weighting, and surface roughness [7].

B. *Multi-frequency acquisitions*

A typical linear transducer array consisting of 128 elements, featuring multiple centre frequencies, spanning from 1 to 8 MHz, was used for the study. Figure 2 shows a representation of RF data acquired from a single channel. Each frequency reveals a distinct facet of information, significantly influenced by multiple factors including porosity, thickness, and frequency variations. For the presented RF response, certain frequencies (1-3MHz) effectively capture reflections originating from the distal sides of the cortical bone. This phenomenon can be attributed to the complicated balance between frequency and wavelength, offering insights into the complex interactions occurring within the bone medium. This interchange between frequency and depth sensitivity highlights the richness of information embedded within the RF data.

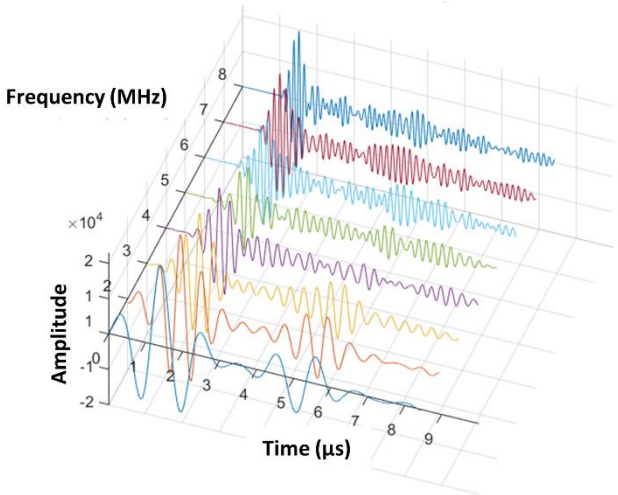

Figure 2: Received RF data from the same channel at different frequencies, 1-8 MHz, show how the frequency affect the echo from the distal bone side.

C. *Convolutional neural networks*

Figure 3 shows a CNN, characterized by a 19-layer architecture and engineered with 3 concurrent pathways for dot product attention followed by a multi-head attention layer [5], for the purpose of discriminating endosteal and periosteal features inherent within multi-dimensional ultrasound RF data. The attention process directs its focus to specific sections of the input by employing weighted multiplication operations [8]. The three parallel paths are used to extract relevant features from the signal, by using different kernel sizes for convolutional operation, allowing the model to focus on different aspects simultaneously.

Subsequently, the obtained output serves as an input for a multi-head self-attention layer. Multi-head self-attention essentially operates as an ensemble of attention mechanisms. This ensemble effect can improve the model's ability to generalize by capturing diverse perspectives of the data, leading to a more robust and accurate classification. The attention mechanism employed in the convolutional model for the input V can be described using the following equations for the query (Q), key (K), and value (V) transformations:

$$\boldsymbol{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \boldsymbol{softmax}\left(\frac{\boldsymbol{QK^T}}{\sqrt{\boldsymbol{d_k}}}\right)\boldsymbol{V} \qquad (3)$$

Where, d_k is the dimension of the key vectors.

Additionally, multi-head self-attention involves linearly projecting the input query, key, and value vectors multiple times and then concatenating the results as a further step. Given h different sets of learned linear projections W_q^i, W_k^i, W_v^i for the i^{th} head, the multi-head self-attention is calculated as follows:

$$\boldsymbol{MultiHead}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \boldsymbol{Concat}(\boldsymbol{Head_1},\dots,\boldsymbol{Head_h}).\boldsymbol{W_o} \quad (4)$$

Where, $Head_i = Attention\ Mechansim(Q.W_q^i, K.W_k^i, V.W_v^i)$, and W_o is the learned output projection and h is the total number of heads.

D. *Consensus mechanism*

A consensus approach has been employed to leverage the collective agreement of the majority within the array of 64 channels present in each test sample. This consensus strategy involves determining the predominant class with 50% or higher agreement threshold among these channels as:

$$\boldsymbol{Consensus}(\boldsymbol{C}) = \frac{\boldsymbol{No.\ of\ Channels\ Agreeing\ on\ Class\ (C)}}{\boldsymbol{Total\ Channels}}\boldsymbol{x100} \quad (5)$$

Where, $Consensus(C)$ is the consensus agreement for a specific class. $No.of\ Channels\ Agreeing\ on\ Class$ is the count of channels agreeing on given class. $Total\ Channels$ is the total number of channels.

The threshold value can be adjusted for accuracy; a higher threshold increases precision but may require more measurements for confidence.

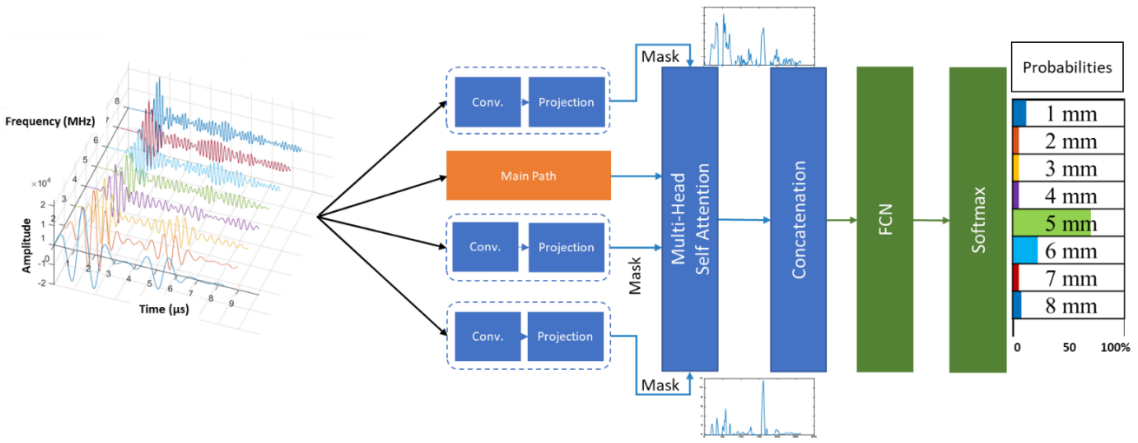

Figure 3: Multi-frequency RF data are fed into a CNN model with attention mechanism to extract features of the echoes from the endosteum and periosteum.

The proposed models are trained on Intel i7-7700HQ CPU @ 2.80GHz, NVIDIA GeForce GTX 1060 (6 GB) GPU, 16.0 GB RAM. All models are validated using a new generated simulations to avoid overfitting with 10,752 and 2560 data points for training and testing respectively. Based on the system convergence, 50 epochs have been trained with a normal distribution to set the initial layers weights, initial learning rate of 0.001, batch size of 16, and ADAM optimizer to update the training weights.

III. RESULTS AND DISCUSSION

A. *Classification and prediciton of bone thickness*

The confusion matrix shown in Figure 4 reveals the classifier's performance in classifying critical bone thicknesses. For low bone thicknesses, 1 and 2 mm, the classifier exhibited a good accuracy, with nearly 100%. However, as the bone thickness increased, the classifier's accuracy became more variable. For example, for a bone thickness of 6 mm, the classifier encountered challenges. While most of instances were correctly classified as 6 mm, some were slightly under or overestimated as 5 or 7 mm respectively. This discrepancy can be attributed to the fundamental principles of porosity impacting the speed of distal side echoes. This phenomenon highlights the complexity of the classification task, with various factors leading to occasional misclassifications. In real-life applications, a variance of 1 mm in the classification of cortical bone thickness can still be considered acceptable, given the inherent measurement uncertainties and biological variability. Overall, the classifier shows promising relevant results, particularly considering the complexities introduced by varying porosity levels.

The scatter plot in Figure 5 illustrates the relationship between the predicted cortical bone thickness values generated by CNN model and the corresponding target thickness. Each point on the plot represents a specific data instance among 320 data points for each thickness, with its horizontal position indicating the actual predicted thickness and its vertical position representing the CNN's target values. The plot shows how well the CNN model's predictions align with the true cortical bone thicknesses. Ideally, the points would form a diagonal line from the bottom left to the upper right, indicating perfect alignment between predictions and targets.

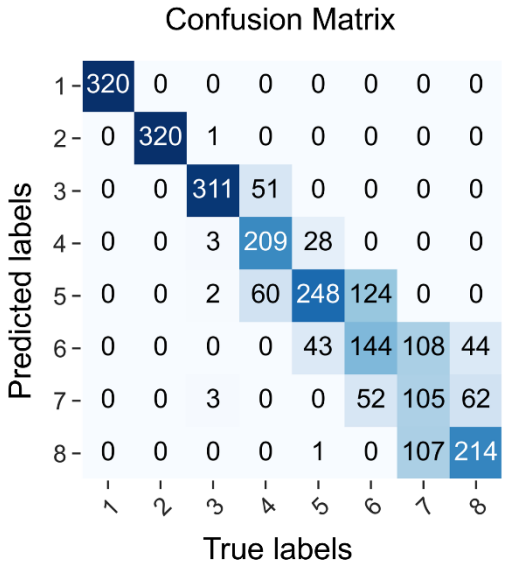

Figure 4: Confusion matrix illustrating the classification results of a cortical bone thickness classifier for thicknesses ranging from 1 to 8 mm.

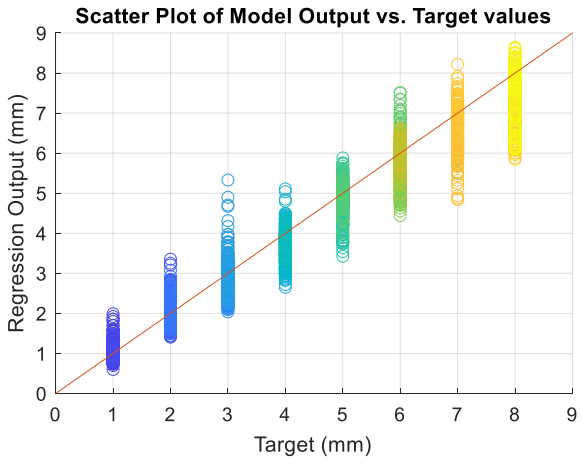

Figure 5: Scatter plot showing the regression output of the CNN model against the actual target cortical bone thicknesses ranging from 1 to 8 mm. The plot shows the alignment between the CNN's predictions and the true thickness values. Each colour on the plot corresponds to a specific label representing a distinct thickness value within this range.

B. *Consensus mechanism*

Consensus mechanism works by harmonizing RF data collected from consecutive array elements and helps to effectively reduce errors and enhance the overall robustness of the model. In Figure 6, the ROC curve on the left shows the classifier's performance without employing the consensus technique, while the curve on the right shows the enhanced performance achieved through the consensus, highlighting the significant improvement in accuracy. Through this method, a considerable improvement has been observed, achieving levels surpassing 80% compared to 72.5%.

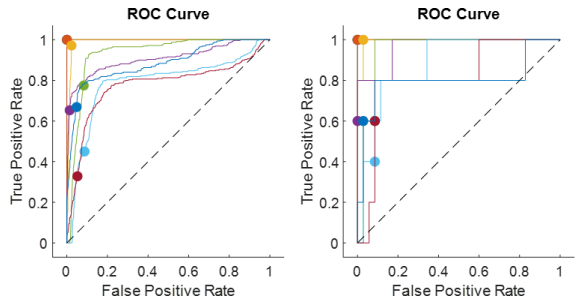

Figure 6: (Left) the ROC curve shows the classifier's performance without employing consensus, (Right) demonstrates the enhanced performance achieved through applying the consensus technique.

C. *Verification of the CNN with GradCAM*

Gradient-weighted Class Activation Mapping (GradCAM) results present a comprehensive visualization of the CNN focus during classification as in Figure 7. This analysis reveals the network's ability to recognise relevant patterns, specifically emphasizing proximal and distal echoes as key factors for accurate thickness classification.

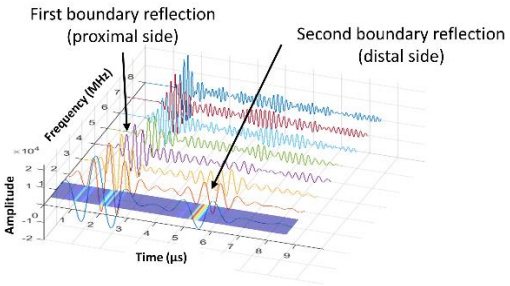

Figure 7: Heatmap generated using GradCAM technique showing regions of interest for cortical bone thickness classification.

D. *Ex-vivo measurements*

For further validation, we conducted an ex-vivo validation study utilizing bovine and swine bone samples with varied porosity levels, 9.23% for swine and 5.45% for bovine. Bone porosity was quantified as the proportion of pore area encompassing structures such as Haversian and Volkmann's canals, as well as lacunae, within the cortical bone region.

Empirical measurements were conducted using a ULA-OP 256 ultrasound system across a sample set comprising 14 tibia bone acquisitions. Chirp signals with spanning frequencies of 3-7 MHz were utilized, with a 3 MHz bandwidth and a 2.5 µs pulse duration. Figure 8 shows the classification results of cortical thickness. Each point on the plot represents a specific acquisition, with its actual measured thickness on the x-axis and the corresponding classified thickness by the CNN model on the y-axis.

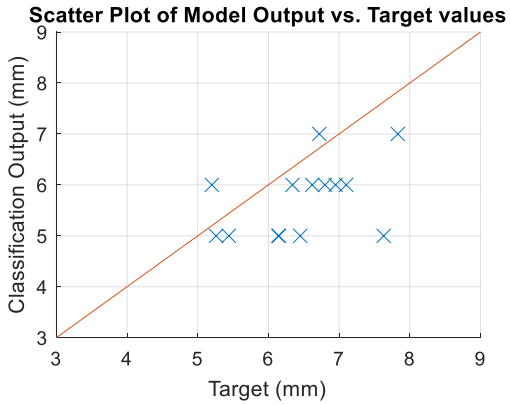

Figure 8: Scatter plot showing the classification output of the model against the actual target cortical bone thicknesses. The plot shows the alignment between the CNN's classification and the true thickness values.

IV. Conclusion

This work presented a numerical study aiming to classify cortical bone thicknesses based on muti-frequency acquisitions. Single pulse-echo approach enabled to acquire large dataset, a total number of 10,752 RF data points, which is a combination of 8 different thicknesses, 21 different porosities, and 8 frequencies. It is demonstrated that CNN can be used for extracting the transient features from the RF data directly. The attention mechanism has a remarkable ability to highlight key zones, particularly the endosteum and periosteum. The outcomes showed the potential of CNN and multi-frequency to be used in cortical bone quantification.

References

[1] P. Augat and S. Schorlemmer, "The role of cortical bone and its microstructure in bone strength," *Age and ageing,* vol. 35, no. suppl_2, pp. ii27-ii31, 2006.

[2] D. Cooper, C. Kawalilak, K. Harrison, B. Johnston, and J. Johnston, "Cortical bone porosity: what is it, why is it important, and how can we detect it?," *Current osteoporosis reports,* vol. 14, pp. 187-198, 2016.

[3] J.-G. Minonzio *et al.*, "Bone cortical thickness and porosity assessment using ultrasound guided waves: An ex vivo validation study," *Bone,* vol. 116, pp. 111-119, 2018.

[4] P. Laugier, B. Fournier, and G. Berger, "Ultrasound parametric imaging of the calcaneus: in vivo results with a new device," *Calcified tissue international,* vol. 58, pp. 326-331, 1996.

[5] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing,* vol. 452, pp. 48-62, 2021.

[6] E. Bossy, M. Talmant, and P. Laugier, "Three-dimensional simulations of ultrasonic axial transmission velocity measurement on cortical bone models," *The Journal of the Acoustical Society of America,* vol. 115, no. 5, pp. 2314-2324, 2004.

[7] H. H. Sultan, E. Grisan, L. Peralta, and S. Harput, "Estimation of Cortical Bone Strength Using CNN-based Regression Model," in *2022 IEEE International Ultrasonics Symposium (IUS)*, 2022, pp. 1-4: IEEE.

[8] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems,* vol. 30, 2017.