Journal of Research High School
Vol. 2025 (11) 904 – 913
ISSN 2688-3651 [online]

Multimodal Deep Learning Model for Predicting Thrombus Thickness and Elasticity using Ultrasound Imaging

Jennifer Choi[1] *

[1]Chadwick International, Incheon, South Korea
*Corresponding Author: jenniferchoi013@gmail.com

Advisor: Dr. Eunjoo Lee, ejlee0902@sogang.ac.kr

Received January 7, 2025; Revised July 6, 2025; Accepted August 15, 2025

Abstract

Blood clots, known as thrombi, can block blood vessels and lead to life-threatening conditions such as stroke and heart attack. Accurate assessment of thrombus thickness and elasticity is critical for early diagnosis and appropriate treatment. However, conventional ultrasound analysis often relies on single data types, limiting prediction accuracy. In this study, a deep learning model that integrates multiple types of ultrasound data is introduced to predict the thickness and elasticity of thrombi. The model utilizes time-domain ultrasound signals, frequency-domain images, and Doppler ultrasound data, each providing unique information about thrombus characteristics. These data streams are processed individually, and their features are combined to enhance prediction performance. Experimental results using a publicly available vascular ultrasound dataset demonstrate that the proposed multimodal approach significantly outperforms single-modality models. This study highlights the potential of multimodal deep learning to support more accurate and reliable detection of thrombus properties, contributing to improved clinical decision-making in cardiovascular care.

Keywords: Multimodal Deep Learning, Ultrasound Imaging, Thrombus Prediction, Convolutional Neural Network, Doppler Ultrasound

1. Introduction

Blood clots, medically known as thrombi, can form within blood vessels and obstruct normal blood flow, leading to life-threatening events such as stroke and heart attack. Precise measurement of thrombus thickness is critical because even submillimeter differences can alter shear stress on vessel walls, influencing the risk of clot propagation or embolization. Similarly, quantifying elasticity provides insight into the mechanical stability of the clot: softer, low-elasticity thrombi are generally more responsive to pharmacological thrombolysis, whereas harder, high-elasticity clots often require mechanical intervention or surgery. Accurate assessment of these parameters therefore directly impacts clinical decision-making—guiding choices between anticoagulant therapy, catheter-directed thrombolysis, or surgical thrombectomy, and enabling personalized treatment plans that can reduce both procedural risks and long-term complications. Ultrasound imaging is a widely used and non-invasive technique that delivers real-time visualization of internal body structures (Avola et al., 2021). Its safety, portability, and cost-effectiveness make it especially valuable in cardiovascular diagnostics, yet conventional ultrasound analysis typically relies on a single data modality and may therefore fail to capture the full complexity of thrombus properties. Recent advances in artificial intelligence have transformed medical image analysis, especially through Convolutional Neural Networks (CNNs), which automatically extract hierarchical features from raw data. While CNN-based methods have achieved impressive results in unimodal tasks, they remain limited when applied to heterogeneous biological structures such as blood clots. Multimodal learning—integrating complementary data streams—offers a promising solution by combining diverse information to achieve a more comprehensive understanding.

Vol. 2025 (11) 904 – 913
ISSN 2688-3651 [online]

In this study, a deep learning framework that unites three ultrasound modalities is proposed: time-domain ultrasound signals that record echo amplitudes over time; frequency-domain images generated via the Fast Fourier Transform (FFT) to reveal spectral content; and Doppler ultrasound data converted into spectrograms using the Short-Time Fourier Transform (STFT) to capture blood flow dynamics. Each modality is analyzed with a specialized CNN architecture to extract features tailored to its unique characteristics, and these features are then fused in fully connected layers to predict both thrombus thickness and elasticity. Our main contributions are threefold. First, we introduce a novel multimodal deep learning approach that synergistically combines time-domain signals, frequency-domain images, and Doppler data for enhanced thrombus characterization. Second, we design and optimize separate CNN architectures for each modality and develop an effective fusion strategy. Third, we validate our framework on a publicly available vascular ultrasound dataset with synthetic thrombi, demonstrating significant performance gains over single-modality models. The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 details the proposed methodology; Section 4 describes the experimental setup; Section 5 presents and discusses the results; and Section 6 concludes the study and suggests directions for future research..

2. Related Works

2.1 Ultrasound Imaging and Thrombosis

Ultrasound imaging is essential in diagnosing and managing thrombosis, offering a non-invasive and real-time method to visualize blood vessels and detect blood clots. By emitting high-frequency sound waves, ultrasound allows clinicians to assess both the structural and functional aspects of the vascular system. Techniques like B-mode imaging provide detailed views of vessel walls and potential thrombi, while Doppler ultrasound measures blood flow velocity and direction, enhancing the detection of abnormalities caused by clots. The combination of these methods enables accurate assessment of thrombus size, location, and impact on circulation, making ultrasound a vital tool in preventing and treating cardiovascular complications associated with thrombosis.

Basic Principles of Ultrasound Technology and Medical Applications

Ultrasound technology is widely used in medical applications due to its non-invasive, real-time imaging capabilities. The basic principle of ultrasound imaging involves the transmission of high-frequency sound waves into the body using a transducer. These sound waves interact with tissues and are reflected back to the transducer, where they are converted into electrical signals and processed to create an image. Ultrasound is particularly advantageous because it is safe for both the patient and the medical staff, as it does not use ionizing radiation like X-rays or CT scans. It is also relatively inexpensive and portable, making it accessible in various medical settings, including resource-limited areas.

Medical applications of ultrasound are vast, with its use spanning from diagnostics to surgical guidance. Ultrasound is frequently employed in obstetrics for fetal imaging, providing crucial information on fetal health and development without exposing the mother or fetus to harmful radiation. Additionally, it is used for diagnosing conditions related to the heart (echocardiography), liver, kidneys, and thyroid, among other organs. In emergency medicine, ultrasound offers immediate feedback in critical situations such as trauma or when guiding procedures like needle biopsies or catheter placements. Despite its strengths, ultrasound imaging has limitations, such as susceptibility to noise and lower image resolution compared to MRI or CT scans, particularly in cases involving deeper or denser tissues.

Measuring Vascular Thickness in Ultrasound Images

Measuring vascular thickness, particularly through the assessment of carotid intima-media thickness (CIMT), is an essential tool in the prediction and prevention of cardiovascular diseases. CIMT measurement is done via ultrasound imaging, where the thickness of the two innermost layers of the carotid artery—the intima and media—is measured. This measurement is a widely recognized marker of atherosclerosis, a condition characterized by the buildup of plaques in the arterial walls, which leads to cardiovascular events such as heart attacks and strokes.

Ultrasound measurement of CIMT is non-invasive and provides real-time imaging, making it a valuable tool for assessing subclinical atherosclerosis in asymptomatic patients. The predictive value of CIMT has been further highlighted by studies that show its ability to enhance traditional risk scores, such as the Framingham Risk Score (Den Ruijter et al., 2012). Despite some debate regarding its routine clinical use, CIMT remains a crucial method in both research and clinical settings for identifying individuals at higher risk for cardiovascular events.

Doppler Ultrasound

Doppler ultrasound is a specialized technique used to measure the velocity and direction of blood flow within the body. It operates based on the Doppler effect, which describes the change in frequency of sound waves as they reflect off moving objects, such as red blood cells. When sound waves emitted by the ultrasound transducer encounter moving blood cells, the frequency of the returning waves changes depending on the speed and direction of blood flow. These frequency shifts are processed to create an image or graph, displaying flow patterns and providing quantitative measurements.

Doppler ultrasound is used extensively in cardiovascular diagnostics, as it enables real-time visualization of blood flow and detection of abnormal patterns that may indicate conditions such as stenosis, thrombosis, or aneurysms. By measuring the velocity of blood flow, Doppler ultrasound can help assess the severity of vascular diseases, identify regions of turbulent flow, and determine the presence of occlusions or blockages. Its non-invasive nature, combined with the ability to provide dynamic information on blood movement, makes Doppler ultrasound a crucial tool in evaluating both arterial and venous systems.

2.1 Artificial Intelligence and Healthcare

Artificial Intelligence (AI) has shown significant promise in advancing medical image analysis, especially through deep learning techniques. Deep learning models, particularly CNNs, have been successfully applied to various medical imaging tasks such as disease diagnosis and image segmentation. For example, CNNs have been utilized in detecting diabetic retinopathy from retinal images and classifying skin lesions, achieving performance on par with human specialists. These models are particularly effective because they automatically extract and learn features from raw image data, bypassing the need for manual feature engineering, which is typically required in traditional image processing approaches. Moreover, AI models have been applied to MRI and CT scans for conditions such as Alzheimer's and breast cancer, demonstrating improved accuracy and efficiency compared to manual diagnosis (Miotto et al., 2018).

2.2 What is Multimodal Data?

Multimodal data refers to the integration of data from different modalities, such as text, images, audio, and video, to extract complementary information for more comprehensive analysis. In healthcare, multimodal data is particularly important because it allows combining various types of clinical and patient data, which can improve the accuracy and performance of applications, such as diagnosis, treatment recommendations, and patient monitoring. Each modality provides unique information that, when fused with other data types, offers a more holistic view of the patient's health condition. For instance, in speech recognition, both the audio signals and visual data related to lip and mouth movements can provide valuable complementary insights. Multimodal data enables healthcare systems to better understand complex relationships between various data points, thus allowing for more effective decision-making. Deep learning methods have become a solution for handling such multimodal data, automatically extracting and learning representations from different modalities without manual feature engineering, which improves the performance of healthcare applications (Tobón et al., 2022).

2.3 Thrombosis and Cardiovascular Diseases

What is Thrombosis?

Thrombosis is the formation of a blood clot within a blood vessel, which can obstruct the flow of blood through

Vol. 2025 (11) 904 – 913
ISSN 2688-3651 [online]

the circulatory system. This process can occur in arteries or veins and is triggered by factors such as injury to the vessel wall, sluggish blood flow, or conditions that increase clotting. Platelets and clotting proteins work together to form clots that can either resolve naturally or lead to serious complications if they block blood flow to vital organs. For example, arterial thrombosis can result in heart attacks or strokes, while venous thrombosis, such as deep vein thrombosis, can cause pain and swelling, and may lead to life-threatening pulmonary embolism if the clot travels to the lungs (Violi et al., 2020).

The Importance of Basic Ultrasound Examinations for Cardiovascular Disease Prevention

The importance of basic ultrasound examinations, particularly carotid ultrasound, for cardiovascular disease prevention lies in its ability to detect subclinical atherosclerosis, which is strongly linked to an increased risk of cardiovascular events. Carotid ultrasound enables the measurement of CIMT and the detection of atherosclerotic plaques. These indicators are valuable in assessing the atherosclerotic burden, especially in patients who appear asymptomatic. Approximately 40-80% of apparently healthy individuals may show increased CIMT or plaque formation, which are associated with a higher likelihood of future cardiovascular complications.

Ultrasound examinations offer a safe, non-invasive, and cost-effective method for early detection of cardiovascular risks, making them a useful tool in preventive strategies. They can particularly benefit patients in the low-to-intermediate cardiovascular risk category, where the presence of carotid abnormalities might prompt more aggressive prevention measures, such as lifestyle changes or pharmacological intervention. Despite their potential, carotid ultrasounds are underutilized in clinical practice, often due to inconsistent guidelines and varying interpretations of CIMT measurements. However, when incorporated with other risk factors, these examinations can provide clinically relevant information that may lead to better patient outcomes (Ray et al., 2015).

3. Methodology

The proposed multimodal deep learning model aims to predict thrombus thickness and elasticity by integrating multiple data modalities: 1D LGFU signals, 2D frequency spectrum data, and Doppler ultrasound data. Each modality provides unique and complementary information—1D signals capture temporal dynamics, 2D spectra offer morphological insights, and Doppler data provides flow-related characteristics. By combining these features, the model aims to improve both classification of thrombus size and regression of elasticity, supporting more accurate and personalized medical decisions. As shown in Figure 1, the overall architecture of the model is illustrated, and Figure 2 presents the pseudocode for the proposed method. Before outlining our specific objectives, several key technical terms are defined for use throughout this work. Laser-Generated Focused Ultrasound (LGFU) refers to a technique in which a pulsed laser induces thermoelastic expansion in a target medium, producing highly focused ultrasound waves that capture fine temporal variations in thrombus microstructure. The Fast Fourier Transform (FFT) is a computational method that converts time-domain ultrasound signals into two-dimensional frequency-domain images, highlighting morphological patterns such as layer boundaries and texture heterogeneities. The Short-Time Fourier Transform (STFT) extends FFT by applying it over short, overlapping time windows to generate spectrograms, which visualize how frequency content evolves over time and are especially informative of flow dynamics in Doppler data. A spectrogram thus represents signal power as a function of both time and frequency, enabling our model to learn features related to blood-flow disturbances and clot elasticity.

3.1 Study Objectives and Hypotheses

The primary objective of this study is to develop and validate a multimodal deep learning framework that integrates 1D ultrasound signals, frequency-domain images, and Doppler data to accurately predict thrombus thickness and elasticity. It is hypothesized that: (1) Temporal features derived from raw 1D ultrasound signals will lead to improved accuracy in classifying thrombus thickness. (2) Spectral characteristics obtained from FFT-generated images will enhance the morphological characterization of thrombi. (3) Hemodynamic information extracted from STFT-based Doppler spectrograms will yield more precise elasticity predictions.

Vol. 2025 (11) 904 – 913
ISSN 2688-3651 [online]

3.2 Data Description

The three ultrasound modalities were selected to capture complementary aspects of thrombus characteristics. First, time-domain ultrasound signals record raw echo amplitudes over time, providing high-resolution temporal information about tissue interfaces; these signals are sensitive to microstructural variations that directly relate to thrombus thickness. Second, frequency-domain images, obtained via the Fast Fourier Transform (FFT), reveal spectral patterns and harmonic content that emphasize morphological features of the clot, such as layer boundaries and compositional heterogeneities. Third, Doppler ultrasound data, converted into time–frequency spectrograms using the Short-Time Fourier Transform (STFT), characterize blood flow dynamics around the thrombus, supplying critical information about mechanical elasticity and flow obstruction. By integrating these three data streams, the model leverages temporal, spectral, and hemodynamic signatures to improve both thickness classification and elasticity regression.

Laser-Generated Focused Ultrasound (LGFU)

LGFU is an advanced ultrasound technique that utilizes laser-induced thermoelastic expansion to generate highly focused ultrasound waves. This method involves directing a high-energy laser pulse onto a small, targeted area of a medium, typically a tissue-mimicking phantom or biological tissue. The rapid absorption of laser energy leads to localized heating and expansion, producing an acoustic wave that propagates through the medium. LGFU is unique in its ability to generate high-frequency ultrasound pulses with extremely narrow beam widths, allowing for precise imaging and measurement at microscopic scales.

In the context of thrombus characterization, LGFU offers several advantages over conventional ultrasound methods. Its high spatial resolution enables the detection of minute variations in thrombus structure and thickness, which may be critical for accurately predicting clot elasticity and stability. Moreover, LGFU can be used to produce tailored ultrasound pulses with specific frequencies, making it possible to selectively target and analyze different tissue layers within the thrombus. By integrating LGFU-generated signals with other ultrasound modalities, the proposed multimodal deep learning model can achieve enhanced performance in predicting thrombus properties, offering a more comprehensive assessment of clot formation and risk factors.

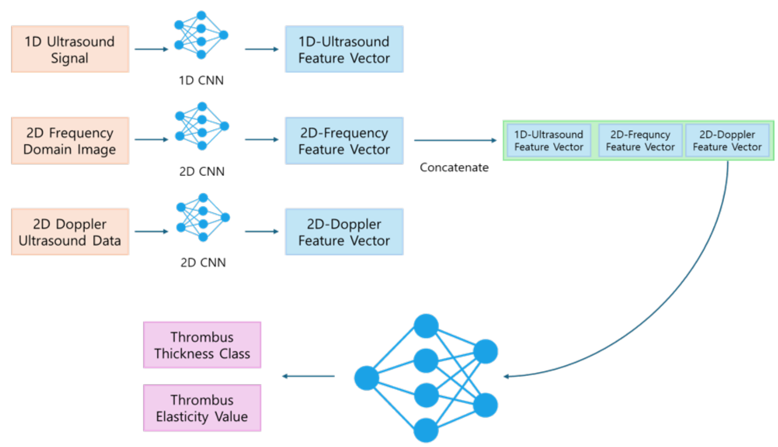

Figure 1. Schematic Diagram for Proposed Method

1-D Ultrasound Signal

In ultrasound imaging, a 1-D ultrasound signal, also known as the raw signal, represents the amplitude of sound waves that are reflected back from various tissues and structures within the body. This signal is typically a time-domain representation, showing the variations in echo intensity as a function of time. In the context of thrombus prediction, analyzing the 1-D signal provides valuable information about the acoustic properties of the tissue layers within the blood vessels, including reflections caused by thrombi.

For the proposed model, the 1-D ultrasound signal will serve as the initial input for the deep learning architecture. The signal will undergo preprocessing steps, including noise reduction and normalization, to ensure that it is clean and ready for feature extraction.

2-D Frequency-Domain Image

The 2-D frequency-domain image, obtained by performing a Fourier transform on the 1-D signal, represents the

Vol. 2025 (11) 904 – 913
ISSN 2688-3651 [online]

frequency components of the ultrasound signal. This transformation shifts the data from the time domain to the frequency domain, where it becomes easier to identify periodic patterns and structures that may be obscured in the raw signal. In the case of thrombus detection, specific frequency signatures might correspond to different tissue characteristics or thrombus formations.

Algorithm 1 Multi-modal CNN for Thrombus Prediction

Input: $X_{1D} \in \mathbb{R}^{n_1}$, $X_{2D} \in \mathbb{R}^{h \times w}$, $X_{\text{Doppler}} \in \mathbb{R}^{t \times v}$
Output: Predicted Thickness $\hat{y}_{\text{thickness}}$, Predicted Hardness $\hat{y}_{\text{elasticity}}$

// Step 1: Data Preprocessing
$X'_{1D} = \text{preprocess}(X_{1D})$
$X'_{2D} = \text{preprocess}(X_{2D})$
$X'_{\text{Doppler}} = \text{preprocess}(X_{\text{Doppler}})$

// Step 2: Feature Extraction
$\boldsymbol{F}_{1D} = \text{CNN}_{1D}(X'_{1D}) \quad \in \mathbb{R}^{k_1}$
$\boldsymbol{F}_{2D} = \text{CNN}_{2D}(X'_{2D}) \quad \in \mathbb{R}^{k_2}$
$\boldsymbol{F}_{\text{Doppler}} = \text{CNN}_{\text{Doppler}}(X'_{\text{Doppler}}) \quad \in \mathbb{R}^{k_3}$

// Step 3: Concatenation of Feature Vectors
$\boldsymbol{F}_{\text{concat}} = \text{Concat}(\boldsymbol{F}_{1D}, \boldsymbol{F}_{2D}, \boldsymbol{F}_{\text{Doppler}}) \quad \in \mathbb{R}^{k_1+k_2+k_3}$

// Step 4: Final 1D CNN and Prediction
$\boldsymbol{F}_{\text{final}} = \text{CNN}_{1D}(\boldsymbol{F}_{\text{concat}}) \quad \in \mathbb{R}^{m}$

// Fully Connected Layer for Predictions
$\hat{y}_{\text{thickness}} = \text{Softmax}(\boldsymbol{W}_{\text{thickness}}\boldsymbol{F}_{\text{final}} + \boldsymbol{b}_{\text{thickness}})$
$\hat{y}_{\text{hardness}} = \boldsymbol{W}_{\text{hardness}}\boldsymbol{F}_{\text{final}} + \boldsymbol{b}_{\text{elasticity}}$
return $\hat{y}_{\text{thickness}}, \hat{y}_{\text{elasticity}}$

Figure 2. Pseudocode for Proposed Method

In the proposed model, the 2-D frequency-domain images will be treated as another modality in the multimodal framework. Convolutional neural networks (CNNs) will be employed to extract spatial features from these images. The CNN architecture will detect patterns and features in the frequency domain, such as variations in harmonic content or shifts in dominant frequencies, which can be linked to changes in thrombus structure or thickness. These extracted features will then be integrated with the temporal features from the 1-D ultrasound signal in a multimodal learning setup to improve the overall accuracy of thrombus thickness prediction.

Doppler Ultrasound Data

Doppler ultrasound data provide critical insights into blood flow dynamics, including speed, direction, and turbulence. For this study, Doppler data will be incorporated into the multimodal learning framework to predict thrombus elasticity, as it reflects the mechanical properties of the clot and surrounding blood vessels. Doppler data are represented as a series of time-frequency maps, capturing changes in blood flow over time and highlighting regions of altered hemodynamics caused by thrombi.

To preprocess the Doppler ultrasound data, the raw Doppler signals are first converted into spectrograms using short-time Fourier transforms (STFT). The resulting spectrograms are then filtered to remove noise and enhance key flow patterns related to thrombus presence and elasticity. The processed Doppler data will be used as an additional modality in the proposed deep learning architecture, complementing the temporal and spatial features extracted from the 1-D and 2-D ultrasound inputs. By incorporating Doppler information, the model aims to improve its ability to differentiate between stable and unstable thrombin, providing a more accurate prediction of thrombus behavior in various clinical scenarios.

3.3 Data Source and Processing

The ultrasound data used in this study were derived from the Mus-V: Multimodal Ultrasound Vascular Segmentation dataset(Geng et al., 2024) available on Kaggle. This dataset provides multimodal ultrasound images, including vascular structures, with detailed annotations. For this study, synthetic blood clot data were simulated based on the vascular ultrasound images in the dataset. Blood clots were categorized into varying thickness ranges, from under 100 µm to over 1000 µm, to mimic realistic clinical scenarios. Ultrasound signals and images from the dataset were processed to serve as input for the LGFU-based analysis.

- 1D Ultrasound Signal Data: The raw time-domain ultrasound signal was processed to remove noise and enhance key features related to blood clot thickness. The processed signals were then used as input for the deep learning model.
- 2D Frequency Spectrum Data: A FFT was applied to convert the 1D signal into a 2D frequency spectrum, which provided additional information about the signal's spectral characteristics. This transformed data was

Vol. 2025 (11) 904 – 913
ISSN 2688-3651 [online]

used as input for the 2D portion of the multi-modal model.

- Doppler ultrasound data : Signals were collected to capture information about blood flow velocity and direction, which are affected by the presence and elasticity of thrombi. The raw Doppler data underwent preprocessing steps, including noise reduction using filters and normalization to ensure consistency across samples. A
- STFT was applied to generate spectrograms representing the Doppler data in the time-frequency domain. These spectrograms highlighted dynamic flow patterns and were used as input for the Doppler-specific CNN component of the model.

3.4 Model Architecture

The proposed multimodal model consists of three parallel convolutional neural network (CNN) branches—one for each ultrasound modality—followed by a feature-fusion stage and final prediction layers. By combining temporal cues from the 1D branch, morphological insights from the 2D FFT branch, and hemodynamic signatures from the Doppler branch, the model delivers superior performance to any single-modality counterpart.

- 1D Signal Processing: A 1D CNN processes raw time-domain ultrasound signals to extract temporal features related to thrombus structure. This branch comprises four convolutional layers with ReLU activations, each followed by average pooling to progressively reduce dimensionality. The resulting feature map is flattened into a vector of size 64×7.
- 2D Frequency-Domain Image Processing: After applying the Fast Fourier Transform (FFT) to convert the 1D signals into 2D frequency-domain images, a 2D CNN extracts spatial and spectral patterns. This network includes three convolutional layers (each with ReLU activation and pooling), producing a 321-dimensional feature vector.
- Doppler Ultrasound Processing: Doppler data are first converted into time–frequency spectrograms via the Short-Time Fourier Transform (STFT). A specialized 2D CNN branch—with three convolutional layers (ReLU + max-pooling)—then captures dynamic flow features indicative of clot elasticity. The final Doppler branch outputs a 1,024-dimensional feature vector.
- Multi-Modal Fusion and Prediction: The three feature vectors are concatenated and passed through fully connected layers with Batch Normalization and Dropout to prevent overfitting. For thickness classification, a softmax output layer maps to the predefined size categories. For elasticity prediction, a separate regression head outputs a continuous Young's modulus estimate.

3.5 Evaluation Metrics

Accuracy and F1-Score for Thickness Prediction

High-frequency vascular ultrasound systems (12–15 MHz) achieve an axial resolution on the order of one-tenth of a millimeter, so thrombi smaller than 0.1 mm approach the physical limits of reliable detection and are most appropriately classified as "small" for monitoring only. Thrombi between 0.1 mm and 1 mm represent a clinically significant intermediate range in which shear stress on the vessel wall may increase abruptly, elevating the risk of symptom onset; these "medium" thrombi typically prompt anticoagulant therapy or additional imaging follow-up. Lesions exceeding 1 mm are large enough to pose a substantial embolic and flow-obstruction risk, and thus warrant consideration of catheter-directed thrombolysis or surgical thrombectomy rather than conservative management.

The prediction of thrombus thickness is treated as a classification problem. This approach aligns with how thrombus size is interpreted in clinical practice, where size is categorized into distinct ranges such as small, medium, or large. These discrete categories simplify treatment decisions, enabling medical professionals to act quickly and effectively based on well-established thresholds. For example:

- Small thrombus(~0.1mm): May only require monitoring.
- Medium thrombus(0.1mm~1mm): Could call for medication or further diagnostic imaging.
- Large thrombus(1mm~): Likely requires surgical intervention.

Since thrombus size does not require continuous measurements for clinical use, a classification model ensures predictions are aligned with actionable medical outcomes. This method is preferred because it reduces ambiguity in treatment protocols by mapping predictions to predefined size categories.

Root Mean Square (RMS) for Elasticity Prediction

The prediction of thrombus elasticity is handled as a regression problem. Unlike size, which can be discretized, elasticity represents a continuous variable that reflects the thrombus's mechanical properties, such as Young's Modulus or Shear Modulus. These properties vary along a spectrum and are crucial for determining the appropriate treatment approach. For example:

- Lower elasticity (soft thrombus): More likely to dissolve with medication.
- Higher elasticity (hard thrombus): Less responsive to medication, may require surgical removal.

Predicting elasticity through regression provides precise values, ensuring that the medical decisions are tailored to the specific mechanical state of the thrombus. This precision aids in personalized treatment planning and offers better outcomes than classification would, as it avoids oversimplification. By using both classification for size and regression for elasticity, the model ensures that each aspect of thrombus prediction is optimized for clinical relevance and decision-making accuracy.

4. Result

In this section, we present the analysis of the results for thrombus thickness (classification) and elasticity (regression) predictions using the multi-modal deep learning model. We compare the proposed model's performance with baseline models to demonstrate the effectiveness of multi-modal integration.

4.1 Thickness Prediction

Proposed Model Performance

As shown in Table 1, the proposed multi-modal CNN, which integrates 1D LGFU signals, 2D frequency spectrum, and Doppler data, achieves the highest performance with an accuracy of 0.853 and an F1-score of 0.828. This result highlights the effectiveness of combining multiple modalities for capturing more comprehensive thrombus features.

Table 1. Thrombus Thickness Prediction Results

Model	Accuracy	F1-Score	Recall	Precision
1D CNN(LGFU only)	0.700	0.682	0.660	0.710
2D CNN(Spectrum only)	0.780	0.738	0.750	0.730
Doppler-only CNN	0.652	0.664	0.670	0.655
Proposed Multi-modal CNN	**0.853**	**0.828**	**0.840**	**0.821**

Baseline Comparison

Among the individual models, the 2D spectrum model achieves the highest performance, suggesting that spectral images provide more morphological information. In contrast, the Doppler-only model achieves lower performance, as Doppler data primarily captures flow dynamics, which may not be sufficient for accurate thickness classification.

4.2. Elasticity Prediction

Proposed Model Performance

As shown in Table 2, the multi-modal CNN achieves the best RMSE of 0.854 kPa, demonstrating the value of combining multiple modalities to predict elasticity more accurately. Integrating LGFU, spectral, and Doppler data enables the model to leverage both morphological and flow-related features for precise predictions.

Table 2. Thrombus Elasticity Prediction Results

Model	RMSE(kPa)
1D CNN(LGFU only)	1.250
2D CNN(Spectrum only)	1.153
Doppler-only CNN	1.102
Proposed Multi-modal CNN	**0.854**

Vol. 2025 (11) 904 – 913
ISSN 2688-3651 [online]

Baseline Comparison

The Doppler-only model achieves a better RMSE than the LGFU-only and Spectrum-only models, as Doppler ultrasound provides direct insight into blood flow and tissue elasticity. However, the multi-modal model outperforms all individual models, highlighting the benefit of combining complementary information from multiple sources. As shown in Figure 3, the ROC curves confirm that the multimodal approach yields the highest area under the curve across all thrombus size classes. The multi-modal model achieves an accuracy of 0.853 and an F1-score of 0.828, outperforming the individual models, and also achieves the best RMSE of 0.854 kPa, indicating its ability to predict thrombus elasticity with high precision. These results demonstrate the advantage of integrating 1D LGFU signals, 2D spectral data, and Doppler ultrasound data, providing better performance in both classification and regression tasks.

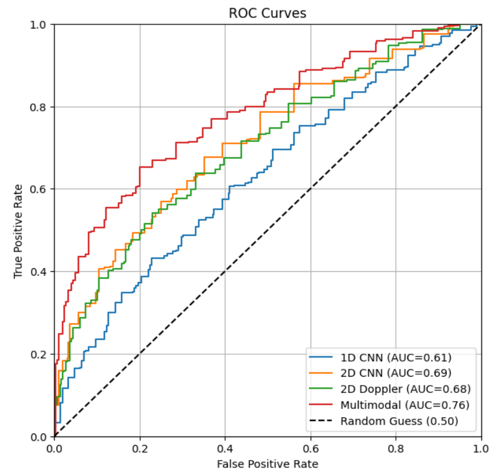

Figure 3. Comparison of ROC Curves for 1D CNN, 2D CNN, 2D Doppler, and a Multimodal Model

5. Discussion

The results confirm the effectiveness of the proposed multimodal deep learning model. By integrating temporal, spectral, and hemodynamic features, the model achieved an accuracy of 0.853 and an F1-score of 0.828 for thickness classification, as well as an RMSE of 0.854 kPa for elasticity regression—each a statistically significant improvement over single-modality baseline. Clinically, these improvements translate into more reliable decision support. The enhanced thickness classification can help clinicians distinguish small, medium, and large thrombi with greater confidence, guiding whether to monitor, administer anticoagulants, or pursue interventional therapy. Likewise, the refined elasticity predictions enable better assessment of clot stability: softer clots (low predicted elasticity) can be managed pharmacologically, while harder clots (high predicted elasticity) may be flagged for more aggressive mechanical removal. Modality-specific insights explain these gains. Frequency-domain features (FFT images) yielded the largest single-modality boost in thickness classification (accuracy 0.780) because they emphasize morphological boundaries that correlate strongly with clot size. Doppler data performed best among single modalities for elasticity (RMSE = 1.102 kPa), reflecting its sensitivity to flow disturbances caused by stiff versus soft thrombi. The 1D ultrasound signals, while less discriminative on their own (accuracy 0.700, RMSE = 1.250 kPa), contribute fine temporal resolution that, when fused, sharpen both classification and regression outputs. Taken together, the multimodal fusion leverages each data stream's strengths: FFT-derived morphology anchors size predictions, Doppler-derived hemodynamics anchor elasticity predictions, and raw time-domain signals fill in the intermediate temporal patterns. This complementary synergy leads to robust performance across clinically relevant tasks. Future work should statistically validate these findings on patient-derived clinical datasets and investigate how additional modalities—such as patient metadata or alternative imaging techniques—could further improve predictive power and generalizability.

6. Conclusion

A multi-modal deep learning model was developed that integrates 1D LGFU signals, 2D frequency spectrum data, and Doppler ultrasound data to predict thrombus thickness and elasticity. Each data modality provides unique and complementary information that, when combined, gives a more complete understanding of thrombus characteristics. Specifically, 1D signals capture temporal dynamics, allowing for insights into periodic and frequency-based attributes, while 2D spectra offer morphological insights, capturing shape and structure critical for understanding thrombus size and composition. Doppler data contributes flow-related characteristics, revealing changes in blood flow around the thrombus, which may be indicative of specific thrombus elasticity and blockage levels.

By merging these distinct types of data, enhanced prediction accuracy was achieved for both classification and regression tasks, supporting more accurate and tailored clinical decisions. Furthermore, FFT and STFT techniques were employed to process each modality, allowing frequency information to be captured in both static and dynamic contexts. FFT was applied to the 1D signals to analyze overall frequency content, while STFT was used with Doppler data to monitor changes over time, enabling a more comprehensive view of thrombus characteristics.

The results demonstrate that combining these data modalities significantly improves prediction performance compared to single-modality models. This outcome highlights the clear advantage of multimodal integration in complex medical prediction tasks, as the combined modalities provide a richer set of features that enhance accuracy and robustness. This approach underscores the potential of multimodal deep learning in medical diagnostics, particularly in situations where reliance on a single data type may yield an incomplete understanding.

In summary, the findings from this study support the use of multi-modal models for improving thrombus prediction accuracy. The success of this approach offers promising directions for future research, where integrating additional data sources, such as clinical metadata or patient history, could further strengthen prediction capabilities and contribute to the development of personalized, data-driven clinical applications.

Future work could focus on incorporating clinical metadata alongside multi-modal ultrasound data to enable more personalized thrombus predictions, tailoring the model to individual patient profiles. Additionally, integrating advanced techniques like transformers could enhance the model's ability to capture complex interactions between different modalities, improving prediction accuracy and robustness. Expanding the model's application to include other vascular conditions, such as plaque formation or arterial stiffness, could further broaden its clinical utility and impact.

References

Avola, D., et al. (2021). Ultrasound medical imaging techniques: a survey. *ACM Computing Surveys (CSUR)*, 54(3), 1-38. https://doi.org/10.1145/3447243

Den Ruijter, H. M., et al. (2012). Common carotid intima-media thickness measurements in cardiovascular risk prediction: a meta-analysis. *JAMA*, 308(8), 796-803. doi:10.1001/jama.2012.9630

Geng, Y., et al. (2024). Force Sensing Guided Artery-Vein Segmentation via Sequential Ultrasound Images. In: Linguraru, M.G., *et al. Medical Image Computing and Computer Assisted Intervention* – MICCAI 2024. MICCAI 2024. Lecture Notes in Computer Science, vol 15004. Springer, Cham. https://doi.org/10.1007/978-3-031-72083-3_61

Miotto, R., et al. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6), 1236-1246. https://doi.org/10.1093/bib/bbx044

Tobon, D. P., et al. (2022). Deep learning in multimedia healthcare applications: a review. *Multimedia systems*, 28(4), 1465-1479. https://doi.org/10.1007/s00530-022-00948-0

Violi, F., et al. (2020). Nutrition, thrombosis, and cardiovascular disease. *Circulation research*, 126(10), 1415-1442. https://doi.org/10.1161/CIRCRESAHA.120.315892