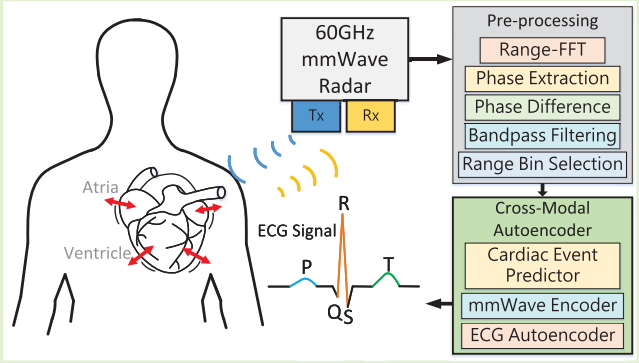
# A Cross-Modal Autoencoder for Contactless Electrocardiography Monitoring Using Frequency-Modulated Continuous Wave Radar

Kai-Chun Liu, *Member, IEEE*, Sheng-Yu Peng, *Senior Member, IEEE*, Yu Tsao, *Senior Member, IEEE*, Che-Yu Liu, Zhu-An Chen, Zong Han Han, Wen-Chi Chen, Po-Quan Hsieh, You-Jin Li, Yu-Juei Hsu, and Shun-Neng Hsu

***Abstract*—While traditional electrocardiogram (ECG) monitoring provides vital clinical information, its electrode-based setup restricts patient movement. To address this limitation, contactless ECG monitoring using frequency-modulated continuous-wave (FMCW) radar and deep learning has been developed. However, such approaches face challenges owing to the limited availability of training data and inherent discrepancies between radar and ECG signals. This article introduces a novel approach to transforming high-fidelity ECG signals from millimeter-wave (mmWave) radar signals reflecting cardiac mechanical activity. The proposed method uses a cascade framework with a cross-modal autoencoder trained using joint waveforms, spectrograms, and deep feature losses. This strategy enables the model to leverage a pretrained ECG-to-ECG autoencoder and a cardiac event (CE) predictor for learning general ECG representations while simultaneously capturing time- and frequency-domain features from limited data. We evaluated the effectiveness of the proposed autoencoder model in terms of signal quality and CE integrity using ablation studies on data from 20 healthy participants. The model achieved high transformation accuracy with a cross correlation of 0.914 and average timing errors below 31 ms for critical ECG features. These findings demonstrate the feasibility and effectiveness of the proposed FMCW radar-based contactless ECG monitoring method, particularly in overcoming the limitations imposed by small datasets and domain discrepancies.**




***Index Terms*—Cross-modal autoencoder, electrocardiogram (ECG) monitoring, frequency-modulated continuous-wave radar, neural networks (NNs).**

## I. Introduction

ELECTROCARDIOGRAPHY (ECG) has been widely employed to monitor the heart's electrical activities in medical centers [1] and homes [2] for healthcare. Recorded ECGs provide vital physical health information and psychological status in clinical assessments, supporting heart disease diagnosis from minor to life-threatening disorders [3], [4]. Previous studies have shown that continuous ECG monitoring and analysis is beneficial for the diagnosis, evaluation, and prevention of cardiovascular diseases (CVDs) [5], [6].

Typically, ECG measurements require electrodes attached to the skin to obtain electrical signals from the heart. However, this direct contact-based approach faces several technical challenges in medical practice. The attachment of electrodes is uncomfortable for humans with skin problems such as rashes or burns [7]. In addition, long-term ECG monitoring is inconvenient because it may lead to skin irritation after prolonged usage [8]. These issues limit the feasibility of continuous health monitoring and willingness to use ECG monitoring devices.

Received 26 August 2024; accepted 9 October 2024. Date of publication 30 October 2024; date of current version 13 December 2024. This work was supported by the National Taiwan University of Science and Technology–Tri-Service General Hospital Joint Research Program (NTUST-TSGH Joint Research Program) under Project TSGH-NTUST-111-02. The associate editor coordinating the review of this article and approving it for publication was Dr. Yuyong Xiong. *(Corresponding author: Sheng-Yu Peng.)*

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee at the National Taiwan University under Application No. NTU-REC 202203EM027 and performed in line with the protocol entitled as "60GHz Millimeter-Wave Radar Physiological Signal Acquisition System Based on Deep Learning."

Please see the Acknowledgment section of this article for the author affiliations.

Digital Object Identifier 10.1109/JSEN.2024.3486154

Several contact-free ECG measurement approaches have recently been proposed to address these issues [9]. A capacitive ECG (cECG) detects the electrical activity of the heart through capacitive coupling between surface electrodes without direct skin contact [10]. However, it is noise-sensitive and typically delivers lower-quality signals as the sensing distance increases beyond 30 cm [11]. These disadvantages limit the use of cECG for diagnosing cardiac disorders.

The growing demand for unobtrusive health monitoring has led to significant interest in contactless ECG monitoring using frequency-modulated continuous-wave (FMCW) radars that detect cardiac mechanical activities remotely, offering an alternative for continuous cardiac health assessment [9]. Deep learning techniques can be used to synthesize ECG representations from detected nonlinear and noisy millimeter-wave (mmWave) signals to support long-distance, unobtrusive, and high-resolution ECG measurements [12], [13], [14], [15], [16]. However, several issues regarding its reliability and effectiveness still need to be addressed. First, ECG patterns vary among individuals depending on their age, gender, and health status. Building a generalized model to reconstruct an ECG signal from an unknown individual is exceptionally challenging. Second, neural networks (NNs) trained with small-scale data are noise-sensitive. Limited training datasets often lead to technical difficulties in overfitting or vanishing gradients. Moreover, the waveform characteristics of the signals received from mmWave radars and ECG monitors differ. It is not feasible to train simple and conventional end-to-end NNs for ECG reconstruction.

This article proposes a novel cross-modal autoencoder model for ECG waveform transformation from FMCW radar signals. The proposed model transforms received FMCW radar signals representing cardiac mechanical activity into ECG waveforms. The received mmWave radar signals are initially processed through fast Fourier transform (FFT), phase extraction, phase difference, bandpass filtering, and range-bin selection to extract chest vibration activities. A cascade framework is then utilized for ECG event detection and transformation. An autoencoder that learns the embeddings from a pretrained ECG-to-ECG encoder–decoder pair with deep loss is adopted to efficiently enhance the generalization performance of the reconstruction model. Furthermore, joint learning tasks, including both ECG waveform and spectrogram reconstruction, are employed in the proposed model to obtain both time- and frequency-domain representations for ECG reconstruction.

The main contributions of this article are as follows.

1) We propose a novel cross-modal autoencoder to transform FMCW radar signals that detect cardiac mechanical activities into ECG waveforms. The experimental results showed that the proposed model achieves a cross correlation of 0.914 with an average timing error of critical ECG events/features less than 31 ms, satisfying the standards set by the Association for the Advancement of Medical Instrumentation (AAMI) [17], [18].
2) We demonstrate the feasibility of the cascade framework for ECG event detection and mmWave-to-ECG transformation in which the ECG event detector provides auxiliary and crucial information for ECG waveform reconstruction.
3) We show that the proposed cross-modal autoencoder enables knowledge acquisition from a pretrained ECG autoencoder and a cardiac event (CE) predictor. Domain knowledge transfer notably enhances the model generalization.
4) Comprehensive evaluation and ablation studies validated the effectiveness and feasibility of the proposed cross-modal autoencoder for ECG waveform transformation.

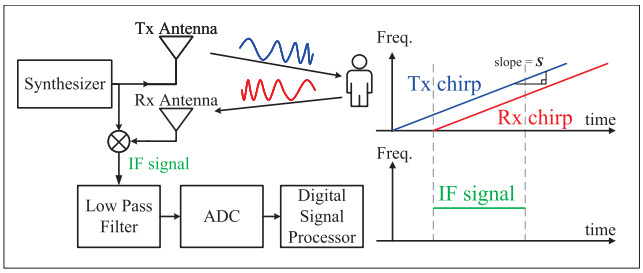



Fig. 1. Block diagram of an FMCW radar for object detection. The radar transmits a linear frequency-modulated chirp signal from the transmitter antenna (Tx). It receives the reflected chirp signal through the receiver antenna (Rx), resulting in an IF signal. The beat frequency of the IF signal is proportional to the distance of the detected object from the radar sensor.

The remainder of this article is organized as follows. Section II introduces prior work relevant to FMCW radar-based ECG monitoring and cross-modal knowledge transfer. Section III details the experimental protocols for data collection and FMCW radar signal preprocessing. Section IV describes the implementation of the proposed cross-modal autoencoder. The evaluation results of the proposed approach and a discussion are presented in Section V. Finally, Section VI provides the conclusions.

## II. Related Works

### A. FMCW Radars for Detection and Monitoring

An FMCW radar utilizes a frequency synthesizer to generate chirp signals with linearly increasing frequency over time that are transmitted through a transmitter (Tx) antenna. Upon reflection from an object or the human body, the received chirp signals are captured by a receiver (Rx) antenna. The received mmWave signals are mixed with the original chirp signals to generate an intermediate frequency (IF) signal. As shown in Fig. 1, the frequency of the IF signal directly corresponds to the distance between the object and the FMCW radar sensor. Consequently, object information within a specific range can be revealed by analyzing the spectrum of the received IF signal obtained through an FFT, known as a range-FFT. The relationship between the distance of the detected object from the FMCW radar and the IF can be expressed as

$$\mathrm{f_{IF}} = \frac{S \cdot 2R}{c} \tag{1}$$

where $\mathrm{f_{IF}}$ is the frequency of the IF signal, $S$ is the chirp rate in Hz/sec, $R$ is the object distance, and $c$ is the light speed.

Although the IF frequency provides information on the distance between the radar sensor and the detected object, the chirp bandwidth limits the range resolution. However,

the phase of the IF signal, denoted as $\phi_b$, offers valuable insights into minute object displacements with the relationship expressed as

$$\phi_b = \frac{4\pi R}{\lambda} \tag{2}$$

where $\lambda$ represents the wavelength of the radar signal at the corresponding range frequency. The sensitivity of the detected IF signal's phase change to minute object displacements makes FMCW radars suitable for subtle vibration detection. Consequently, the object displacement, $\Delta R$, can be expressed as

$$\Delta R = \frac{\lambda \Delta \phi_b}{4\pi}. \tag{3}$$

Therefore, the velocities or vibrations associated with respiration and cardiac activity can be detected by extracting the phase difference of the IF signal. Since the FMCW radar emits relatively higher power through linear frequency-modulated signals and exhibits a higher signal-to-noise ratio than other radar technologies, such as ultrawideband (UWB), leading to better signal quality for ECG signal reconstruction. Recent research has explored the application of FMCW radars in conjunction with NNs for contactless fall detection [19], [20], [21], vital sign monitoring [22], [23], [24], [25], [26], [27], and ECG monitoring [12], [13].

Specifically, a three-transmitter, four-receiver (3T4R) FMCW radar sensor was employed to detect spatial and temporal cardiac mechanical activities in [12]. Feature extraction and fusion were achieved using a 1-D convolutional neural network (CNN) and transformer blocks. Subsequently, a temporal convolutional network (TCN) was utilized for ECG signal reconstruction. In another study [13], a single-transmitter, single-receiver FMCW radar was used for cardiac activity detection. This approach employed a long short-term memory (LSTM) model with an attention mechanism to extract temporal features from mmWave radar signals. In addition, a wavelet-based loss function was implemented to facilitate accurate heartbeat detection by focusing on learning the peaks and valleys of the ECG signals.

### B. Conformer

The Conformer is an NN architecture that combines a convolutional NN and a Transformer, first introduced by Gulati et al. [28]. Initially, the input sequence is processed by a stack of convolutional layers that extract local features. These features are then passed through subsequent Transformer blocks, consisting of multihead self-attention modules, which enable the model to attend to different parts of the input sequence simultaneously and perform self-attention computation separately for each head. As a result, the model can capture long-term dependencies and learn the complex relationships between tokens in the sequence. An ensuing convolution module is interposed between two feed-forward modules in the Conformer. The convolution module, comprising several layers, performs operations of layer normalization, pointwise convolution, gated linear unit (GLU), depth-wise convolution, batch normalization, Swish activation function, and dropout on the input sequence. This convolution module captures the basic patterns and relationships while introducing regularization to prevent overfitting. The feed-forward modules consist of a fully connected NN layer of linear transformation followed by a Swish activation function and another fully connected layer of linear transformation. These feed-forward modules perform nonlinear transformations on the input features, allowing the network to model complex relationships and reduce the output dimensionality. The convolution and feed-forward modules can be stacked multiple times to form a deep NN architecture. Integrating the convolution and Transformer modules results in a size-efficient NN structure with high computational and memory efficiency.

### C. Cross-Modal Autoencoder

A cross-modal autoencoder can be trained to relate information from different data sources and formats. The encoder transforms input data from one or multiple modalities into a corresponding latent representation. The decoder then converts this latent representation into synthesized target modalities. By learning these latent representations, a cross-modal autoencoder can translate different data sources and understand the connections between input and output modalities. A better representation of cross-modal embedding can improve different downstream tasks, such as classification and prediction [29], and the decoders have the potential to mitigate discrepancies between the source and target modalities [30]. This approach has been widely applied in the fields of speech synthesis [30], [31], [32], computer vision [33], and natural language processing [34], achieving improved comprehensibility.

### D. Deep Feature Loss

Deep feature loss, also known as deep perceptual loss [35], measures the dissimilarity among high-level features, or embeddings, from different NN layers instead of comparing raw output data directly. This approach results in a trained network that achieves a better perceptual quality, which is more important than pixel-level accuracy. By accurately capturing perceptual characteristics from embeddings for generative tasks, deep feature loss has been shown in numerous studies to enhance model performance without increasing network complexity [35], [36].

## III. Data Collection and Preprocessing

### A. Data Collection

The developed contactless ECG monitoring system adopts a TI AWR6843 60-to-64-GHz FMCW radar sensor and a DCA1000 real-time data acquisition board. The IF waveform data collected from the radar sensor were preprocessed to extract the cardiac activities from the chest vibration responses. A TI ADS1298ECG-FE performance demonstration kit was employed as an ECG monitor to record the ground-truth ECG data while the FMCW radar signals were collected simultaneously. The sampling rates of both the FMCW radar sensor and ECG monitor were chosen to be 500 samples per second and triggered by a data-gathering laptop for synchronization. The cross correlation is employed

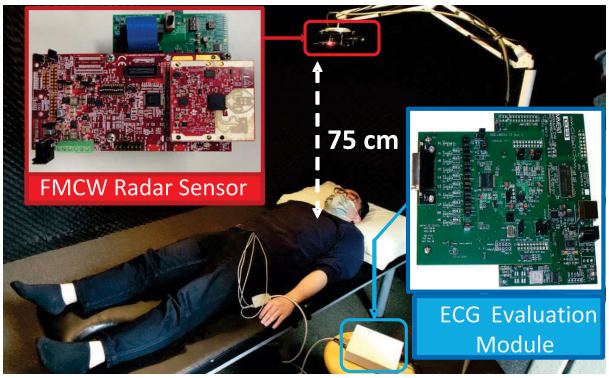



Fig. 2. Experimental setup for FMCW radar and ECG data collection. The FMCW radar is positioned facing the subject's chest cavity at a distance of 75 cm. Standard ECG electrode placements are used on the torso for acquiring reference ECG signals.

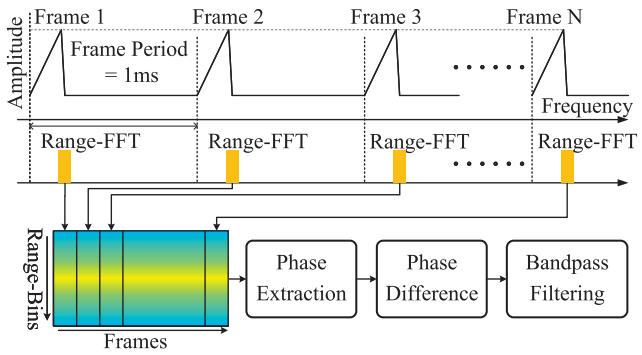



Fig. 3. FMCW radar signal preprocessing for ECG transformation. The IF signal undergoes range-FFT to reveal distance-dependent information. The phase difference is calculated to extract chest motion information. Finally, a bandpass filter isolates the frequency band containing relevant cardiac activities.

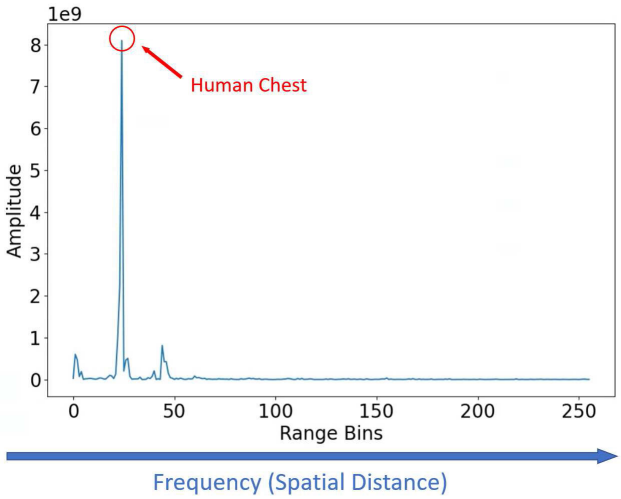



Fig. 4. Example of the FMCW radar IF signal after range-FFT processing.

to align the frames recorded by the ECG monitor and the radar sensor. Similar synchronization and alignment techniques have been implemented and validated in other studies [13].

The study was conducted under a protocol approved by the Institutional Review Board of the National Taiwan University (IRB No. 202203EM027). Twenty participants, consisting of ten male and ten female individuals with ages ranging from 19 to 30, were recruited for mmWave-to-ECG database establishment, which was employed to validate the proposed cross-modal autoencoder for contactless ECG monitoring. All participants provided written informed consent. During the experiment, participants were instructed to lie supine on a bed and remain stationary. Given the employed FMCW radar sensor's azimuth and elevation fields of view of $\pm 60°$ and $\pm 15°$, respectively, the radar sensor was positioned approximately 75 cm above each participant and aligned toward the participant's chest to effectively enclose the participant's torso. The ECG monitor and the FMCW radar signals were collected simultaneously for 15 min during each trial. The experimental protocols were designed under natural lie-on bed conditions without further restrictions or instructions. These experimental setups emulated natural and common physiological conditions. They facilitated the assessment of the feasibility of the proposed remote contactless ECG monitoring system for capturing cardiac activities in daily living. The experimental settings are illustrated in Fig. 2.

### B. Range-FFT and Vibration Extraction

The FMCW radar transmitter emitted chirps at 500 frames per second, and a range-FFT was conducted on the IF signals in each frame to obtain a range–amplitude representation with a size of 256 bins as shown in Fig. 3. The range–amplitude response from a single frame after range-FFT processing is exemplified in Fig. 4, where the $x$-axis represents the frequency or spatial distance and the $y$-axis represents the amplitude. Because the human chest is the object closest to the radar sensor, a strong corresponding peak can be observed in a specific frequency bin, representing the spatial distance between the chest and the radar sensor. Other peaks corresponding to the dc noise and the floor also appeared in the range–amplitude response.

After the range-FFT process, a set of $m$ bins centered around the peak was chosen to capture essential cardiac activities from the chest movement, where we chose $m = 10$ in this study. Since the phase shift of the IF signals is sensitive to the micromovements of the detected object, it enables the detection of subtle and fine-grained vibrations caused by cardiac activities [37]. The phases of the selected range bins were later calculated, unwrapped, and differenced to reconstruct the velocities of the detected chest movements. Subsequently, the phase changes of the FMCW radar IF signal corresponds to the chest vibration signals, $x(m, nT_s)$, which were recorded over time as the input data for the proposed cross-modal autoencoder. The recorded vibration signal, $x(m, nT_s)$, can be expressed as

$$x(m, nT_s) = \frac{\lambda}{4\pi} \Delta\phi_b \tag{4}$$

where $m$ is the selected range bins, $n$ is the chirp index, $T_s$ is the time between consecutive frames, $\lambda$ is the wavelength of the radar radio frequency signal, and $\Delta\phi_b$ is the phase change of the received IF signals. Finally, bandpass filtering between 0.75 and 15 Hz was applied to the chest vibration signals to eliminate respiratory noises and other high-frequency noise and environmental interference amplified by the phase

difference operation while ensuring that the critical QRS complex frequency components were preserved [38], [39].

### C. Sliding-Window and Range-Bin Selection

The sliding-window technique was applied to the recorded vibration signals. Typically, a resting heart rate is approximately 60 beats per minute (bpm). However, a heart rate between 40 and 60 bpm can be normal during sleep. To ensure that at least one complete ECG period was captured, a window size of 2 s was chosen with an overlap of 0.5 s. Similar to that of ECG data collected in clinical settings [40], quality checks were routinely performed to eliminate low-quality data before diagnostic analysis, and noisy frames typically caused by motion artifacts were identified and removed. Ten bins centered around the chest position were used to capture key cardiac activities. The correlation coefficients between each range bin and all the other bins were calculated to obtain the average correlation coefficient for each bin. Slight body movements can cause significant fluctuations in the collected bin signals, resulting in low correlation coefficients between the bins within a frame. Channels with an average correlation coefficient below an empirical threshold value of 0.6 were considered outliers and removed from the input bins. Frames with fewer than five correlated bins were considered corrupted by motion artifacts and discarded. This crucial step ensures that only the relevant and representative data is used for ECG reconstruction. After outlier removal, min–max normalization was applied to the selected bins, scaling them to a range from $-1$ to 1. The averaged vector of the selected range bins with dimensions of $1 \times 1000$, named the mmWave signal, was the input into the proposed cross-modal autoencoder for ECG waveform reconstruction. Such range-bin selection mitigates the influence of uncorrelated bins on ECG transformation models.

## IV. Proposed Cross-Modal Autoencoder

### A. Architecture

Previous studies have demonstrated that the autoencoder architecture can effectively discard less salient features, such as noise or artifacts by compressing input signals into a lower-dimensional space [41]. Moreover, joint learning tasks involving both ECG waveform and spectrogram reconstruction can aid in extracting the underlying structure of the data while suppressing noise. To mitigate the impact of motion artifacts on CE detection, we propose a cross-modal mmWave-to-ECG autoencoder. This model employs a pretraining strategy on an ECG autoencoder and a CE predictor to address the challenges associated with the direct training of the mmWave-to-ECG autoencoder owing to the significant discrepancies between the signal modalities and the inherent noise in FMCW radar data. The pretrained ECG autoencoder and CE predictor generate informative and perceptually meaningful ECG and mmWave embeddings, which enhance initialization and feature representation during the final ECG reconstruction stage.

The architecture of the proposed mmWave-to-ECG autoencoder is shown in Fig. 5. Initially, an ECG autoencoder and a CE predictor were pretrained to obtain the shared ECG decoder and the shared CE predictor, respectively. Next, the mmWave encoder took the mmWave signals and the shared CE predictor outputs as input features to generate mmWave embeddings, which were fed into the shared ECG decoder for ECG waveform reconstruction. The short-time Fourier transform (STFT) was applied to the synthesized ECG waveforms to obtain the synthesized ECG spectrograms. The mmWave encoder was optimized by training with multiple ECG features, including embeddings, ECG waveforms, and ECG spectrograms. This pretrained framework allowed the proposed cross-modal autoencoder to capture the perceptual features of the input mmWave signals and learn better feature representations for ECG waveform reconstruction.

During the testing phase, the meticulously cascaded shared CE predictor, mmWave encoder, and shared ECG decoder were used to transform the mmWave data into ECG waveforms. Finally, the synthesized ECG waveforms were comprehensively assessed using various metrics, including signal quality, CE integrity, and an ablation study. This rigorous process was used to evaluate the performance of the proposed cross-modal autoencoder. The details of the employed models and the evaluation process are provided in Sections IV-B–IV-E.

### B. Pretrained ECG Autoencoder

The detailed model of the pretrained ECG autoencoder is illustrated in Fig. 6. The pretrained ECG encoder, denoted as $E_{\text{pt}}$, processes the ECG signal, $w_{\text{tr}}$, through a sequence of three convolutional subsampling blocks, a linear dropout layer, and two repeated Conformer blocks. Each convolutional subsampling block utilizes three 1-D convolutional layers with a kernel size of three. The first layer employs 16 filters, and the remaining two layers use 32 filters each. Batch normalization rectified linear unit (ReLU) activation, and max-pooling with a pool size of three were applied after each convolutional layer. A linear dropout block with a dropout rate of 0.5 was employed to prevent overfitting on the output of the convolutional subsampling block. Finally, the two repeated Conformer blocks were cascaded after the linear dropout layer. Each Conformer block comprises two feedforward modules, a 64-head self-attention module, a convolutional layer, and a normalization layer. The feedforward module consists of two linear layers: the first expands the feature dimensions by a factor of four, and the second downscales them back to the original size. This encoder generated an embedding with dimensions of $120 \times 64$, representing 64 attention heads and 120 extracted features. The corresponding ECG decoder, $D_{\text{pt}}$, utilizes two repeated Conformer blocks followed by a linear layer to synthesize the ECG signal, denoted as $\widehat{w}_{\text{tr}}$, from the encoded embedding representation.

The loss function during pretraining is the mean squared error (MSE), denoted as $L^{\text{pt}}_{\text{ECG}}$, which incorporates two components: waveform loss, $L^{\text{pt}}_{\text{wave}}$, and spectrogram loss, $L^{\text{pt}}_{\text{spec}}$. The waveform loss ($L^{\text{pt}}_{\text{wave}}$) quantifies the discrepancies between the synthesized and recorded ground-truth ECG waveforms. In contrast, the spectrogram loss ($L^{\text{pt}}_{\text{spec}}$) measures the differences in their spectrograms. These loss functions can be

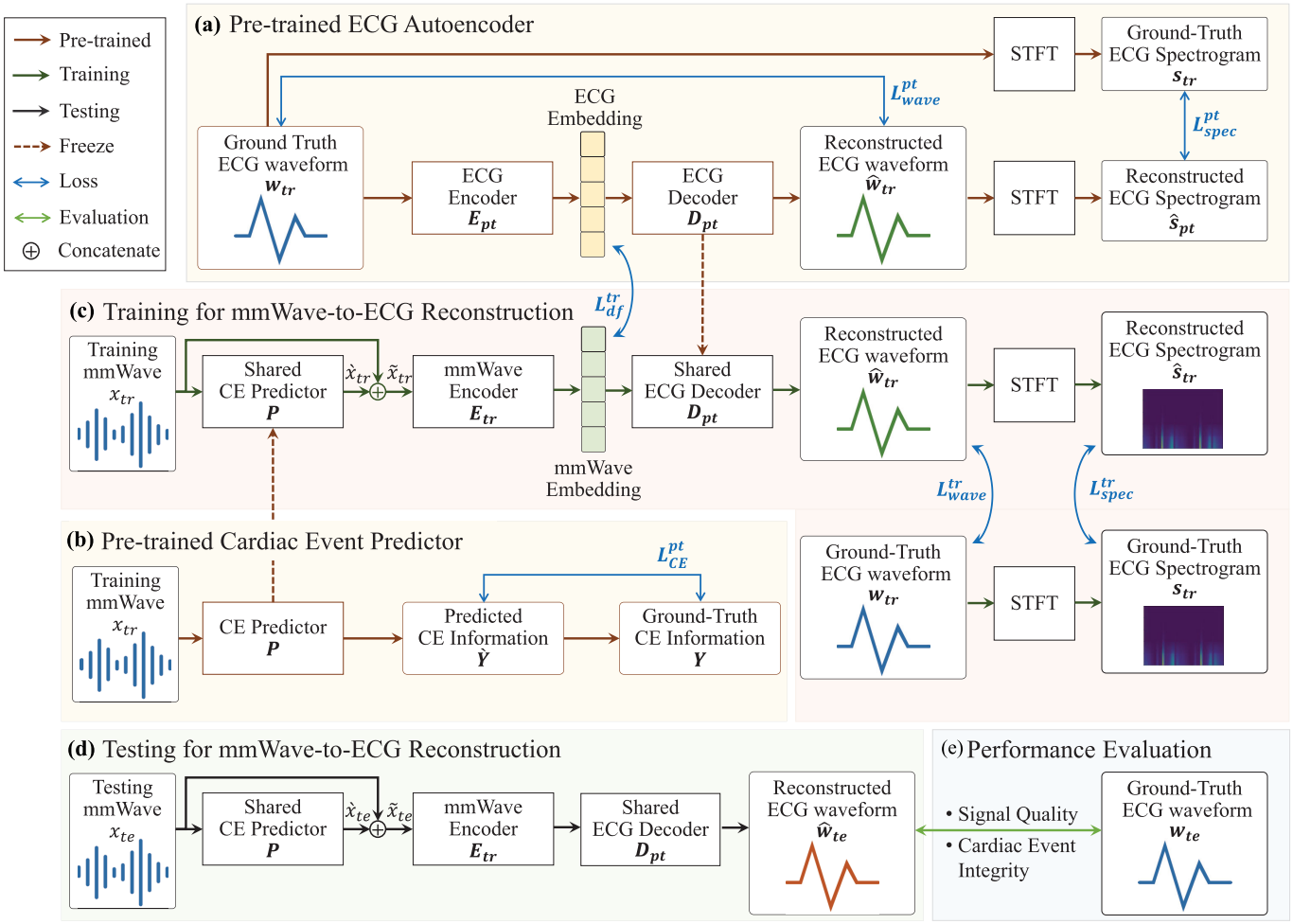



Fig. 5. Framework and training stages of the proposed cross-modal autoencoder for ECG transformation. (a) Pretraining the ECG autoencoder on ECG data for learning general ECG embeddings. (b) Pretraining a CE predictor to identify timing information of P, Q, R, S, and T waves. (c) Training the cross-modal autoencoder to synthesize ECG signals from mmWave radar data. (d) Testing the training the cross-modal autoencoder model on new mmWave data for ECG reconstruction. (e) Evaluating and quantifying the proposed cross-modal autoencoder in signal quality and CE integrity.

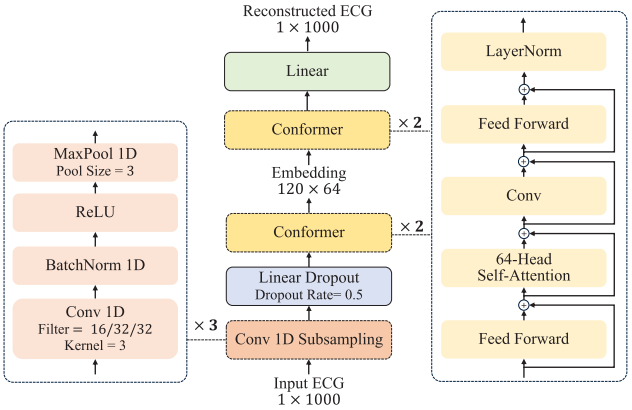



Fig. 6. Architecture of the pretrained ECG autoencoder for learning general ECG latent embeddings. The trained ECG decoder is shared for ECG reconstruction during the training [Fig. 5(c)] and testing [Fig. 5(d)] stages.

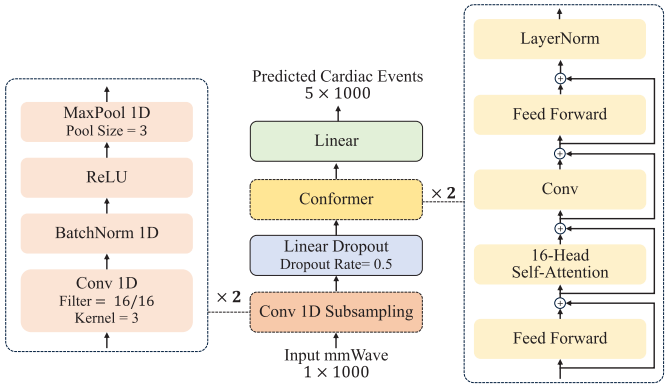



Fig. 7. Architecture of the pretrained CE predictor designed to learn P, Q, R, S, and T wave timing information. The trained CE predictor is shared for ECG reconstruction during the training [Fig. 5(c)] and testing [Fig. 5(d)] stages.

expressed as

$$L^{\text{pt}}_{\text{ECG}} = L^{\text{pt}}_{\text{wave}} + L^{\text{pt}}_{\text{spec}} \tag{5}$$

$$L^{\text{pt}}_{\text{wave}} = \text{MSE}\left(w_{\text{pt}}, \widehat{w}_{\text{pt}}\right) \tag{6}$$

$$L^{\text{pt}}_{\text{spec}} = \text{MSE}\left(s_{\text{pt}}, \widehat{s}_{\text{pt}}\right). \tag{7}$$

### C. Pretrained CE Predictor

The primary purpose of pretraining the CE predictor, denoted as $P$, is to identify the temporal locations of the PQRST waves within the ECG signal. These crucial CE timings, obtained from the pretrained CE predictor,

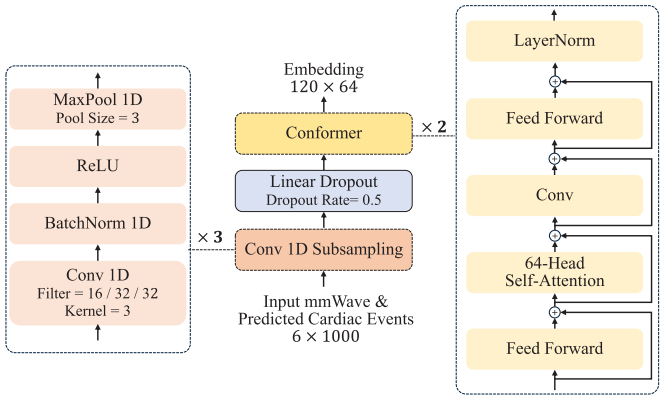



Fig. 8. Architecture of the mmWave encoder for learning general ECG latent embeddings from the mmWave data.

provide auxiliary positional information for the subsequent ECG reconstruction process. This approach leverages the relative simplicity of PQRST wave detection compared with the more complicated task of directly reconstructing the entire ECG waveform pixel-by-pixel.

The architecture of the CE predictor shown in Fig. 7 is similar to that of the pretrained ECG autoencoder. Combining convolutional subsampling, linear dropout, and Conformer blocks enables the model to effectively capture local and global features, leading to improved performance. The CE predictor model utilizes two convolutional subsampling blocks, each consisting of a 1-D convolutional layer with 16 filters and a kernel size of 3. Following each 1-D convolutional layer, batch normalization and ReLU activation are applied. Finally, a maximum pooling layer with a pool size of three is implemented. The output from these blocks is then channeled through a linear dropout layer with a dropout rate of 0.5 to mitigate overfitting, followed by two Conformer blocks. Subsequently, a linear layer is applied to transform and condense the information extracted by the previous layers into a format suitable for further processing, aligning with the final output dimensions required for CE prediction. The loss function used for the CE predictor training can be expressed as

$$L_{\mathrm{CE}}^{\mathrm{pt}} = \mathrm{MSE}\left(Y, Y'\right) \tag{8}$$

where $Y$ is the ground-truth CE information and $Y'$ is the predicted information.

### D. mmWave Encoder

The mmWave encoder, denoted as $E_{\mathrm{tr}}$, takes the concatenation of the mmWave signals and the predicted PQRST wave timings as input. This combined input with dimensions of $6 \times 1000$ leverages information from both sources. The mmWave encoder adopts a similar architecture, size parameters, and output size as that of the pretrained ECG encoder, as illustrated in Fig. 8. This shared architecture allows the mmWave encoder to learn more robust feature representations for the ECG waveform transformation.

The mmWave encoder was trained through a joint deep feature loss approach to extract effective representations from various information sources. The optimization of the mmWave encoder is achieved by minimizing the total mmWave loss, $L_{\mathrm{mmWave}}^{\mathrm{tr}}$, which encompasses three components: the embedding loss, $L_{\mathrm{df}}^{\mathrm{tr}}$, the synthesized ECG waveform loss, $L_{\mathrm{spec}}^{\mathrm{tr}}$, and the synthesized ECG spectrogram loss, $L_{\mathrm{spec}}^{\mathrm{tr}}$, as expressed as follows:

$$L_{\mathrm{mmWave}}^{\mathrm{tr}} = L_{\mathrm{df}}^{\mathrm{tr}} + L_{\mathrm{wave}}^{\mathrm{tr}} + L_{\mathrm{spec}}^{\mathrm{tr}}. \tag{9}$$

The embedding loss, $L_{\mathrm{df}}^{\mathrm{tr}}$, quantifies the dissimilarity between the ECG embeddings generated by the pretrained ECG encoder, $E_{\mathrm{pt}}$, and the mmWave embeddings produced by $E_{\mathrm{tr}}$. A minimized value of $L_{\mathrm{df}}^{\mathrm{tr}}$ indicates closer alignment between these embeddings, which is advantageous for subsequent ECG waveform reconstruction. Furthermore, the STFT was employed to calculate $L_{\mathrm{spec}}^{\mathrm{tr}}$, enabling the analysis of ECG signals in both the time and frequency domains. This dual-domain analysis facilitates a more comprehensive assessment of ECG waveform transformation quality. The overall mmWave loss function combines these individual loss components, guiding the mmWave encoder to extract informative features for ECG reconstruction with similar embeddings generated from the pretrained ECG encoder. These loss functions can be expressed individually as

$$L_{\mathrm{df}}^{\mathrm{tr}} = \mathrm{MAE}\left(E_{\mathrm{tr}}\left(\tilde{x}_{\mathrm{tr}}\right), E_{\mathrm{pt}}(w_{\mathrm{tr}})\right) \tag{10}$$

$$L_{\mathrm{wave}}^{\mathrm{tr}} = \mathrm{MSE}(w_{\mathrm{tr}}, \widehat{w}_{\mathrm{tr}}) \tag{11}$$

$$L_{\mathrm{spec}}^{\mathrm{tr}} = \mathrm{MAE}\left(s_{\mathrm{tr}}, \hat{s}_{\mathrm{tr}}\right) \tag{12}$$

where $w_{\mathrm{tr}}$ is the ground-truth ECG waveform, $\tilde{x}_{\mathrm{tr}}$ is the concatenated mmWave encoder input, $\widehat{w}_{\mathrm{tr}}$ is the synthesized ECG waveform, $s_{\mathrm{tr}}$ is the ground-truth ECG spectrogram, and $\hat{s}_{\mathrm{tr}}$ is the synthesized ECG spectrogram.

### E. Testing and Performance Evaluation

During the test phase, the proposed cross-modal autoencoder facilitates ECG waveform reconstruction. The pretrained CE predictor ($P$) receives the mmWave signal ($x_{\mathrm{te}}$) as input. The output of $P$, represented by $\dot{x}_{\mathrm{te}}$, signifies the predicted timing information for PQRST events. This predicted information is concatenated with the original mmWave signal ($x_{\mathrm{te}}$) to form combined input features, $\tilde{x}_{\mathrm{te}}$. These combined features are subsequently fed into the mmWave encoder ($E_{\mathrm{pt}}$), followed by the shared ECG decoder ($D_{\mathrm{pt}}$) to synthesize the final ECG waveform, $\hat{w}_{\mathrm{te}}$. The synthesized waveform is then evaluated for its signal quality and ability to represent cardiac integrity.

This study randomly selected 16 participants, used their data from 19 873 frames to train the model, and tested the performance of the proposed model on the remaining four participants with 6935 frames. To quantitatively assess the signal quality and resemblance of the synthesized ECG waveform to the ground-truth recordings, we employed three complementary signal-based metrics: cross correlation (XCorr) [42], MSE [43], and root MSE (RMSE) [12]. These metrics offer distinct insights into transformation fidelity. XCorr provides an overall similarity score, whereas MSE and RMSE quantify the average magnitude of the deviations in the amplitude. Utilizing all three metrics facilitates a more comprehensive understanding of how closely the synthesized ECG waveform replicates the recorded ground truth.

TABLE I
ESSENTIAL CE FEATURES

| No. | Features | Description |
|---|---|---|
| $f_1$ | $P_{timing}$ | The P peak timing |
| $f_2$ | $R_{timing}$ | The R peak timing |
| $f_3$ | $T_{timing}$ | The T peak timing |
| $f_4$ | $QRS_{dur}$ | From the start to the end of QRS complex |
| $f_5$ | $RR_{int}$ | Between two successive R peaks (RR interval) |
| $f_6$ | $QT^e_{int}$ | From the start of QRS complex to the end of T wave |
| $f_7$ | $QT^p_{int}$ | From the start of QRS complex to the T peak |
| $f_8$ | $QRS^+_{amp}$ | The maximum positive amplitude of QRS complex |
| $f_9$ | $QRS^-_{amp}$ | The minimum negative amplitude of QRS complex |
| $f_{10}$ | $P_{amp}$ | The amplitude of P wave |
| $f_{11}$ | $T_{amp}$ | The amplitude of T wave |

Accurate transformation of essential CEs, including the P, Q, R, S, and T waves, is paramount for ECG monitoring. These waves provide valuable clinical information through their characteristic features, such as peak amplitude, duration, and the interval between them [44]. This study analyzed 11 key features derived from these essential CEs, as detailed in Table I. We adopted both the absolute error, $\epsilon^{f_i}_{abs}$, and the relative error, $\epsilon^{f_i}_{rel}$ to assess the accuracy of the transformed cardiac features. The absolute error, expressed in milliseconds for timing features, quantifies the raw magnitude of the difference between the estimated value ($c^{f_i}_{est}$) and the ground-truth value ($c^{f_i}_{GT}$) of the $i$th feature and can be expressed as

$$\epsilon^{f_i}_{abs} = \left| c^{f_i}_{est} - c^{f_i}_{GT} \right|. \tag{13}$$

The relative error, expressed as a percentage, provides a normalized measure of this difference relative to the ground-truth value and can be expressed as

$$\epsilon^{f_i}_{rel} = \frac{\left| c^{f_i}_{est} - c^{f_i}_{GT} \right|}{c^{f_i}_{GT}}. \tag{14}$$

Utilizing both metrics allows for a more comprehensive evaluation; the absolute error highlights the actual magnitude of deviations, whereas the relative error accounts for potential variations in the ground-truth values themselves.

Absolute errors quantify deviations in the predicted timing and duration of CEs for the first seven features listed in Table I. Conversely, relative errors quantify event amplitude deviations for the last four features. By analyzing these errors, we can evaluate the effectiveness of the proposed cross-modal autoencoder model in preserving the clinically relevant characteristics of CEs during transformation.

### F. Model Implementation

The proposed model was implemented using PyTorch 1.8.0 and ran on a workstation with 64-bit Ubuntu 20.04.2 LTS (GNU/Linux 5.15.0-91-generic x86_64) and an Intel Xeon CPU E3-1285 v4 at 3.50 GHz. Training and testing were conducted on an NVIDIA RTX 2080 Ti with 11 GB of dedicated memory. The Adam optimizer [45] was employed to minimize the loss functions and iteratively optimize the network parameters. The learning rate and mini-batch size were set to 0.0001 and 64, respectively.

TABLE II
COMPARISON OF SYNTHESIZED WAVEFORM QUALITY

| Model | XCorr | MSE ($\times 10^{-3}$) | RMSE ($\times 10^{-3}$) | FLOPS ($\times 10^{6}$) |
|---|---|---|---|---|
| CNN | .831 ± .101 | 6.5 ± 3.7 | 78.3 ± 21.2 | 16.0 |
| LSTM | .878 ± .099 | 5.1 ± 4.9 | 66 ± 27.2 | 119.1 |
| BLSTM | .912 ± .067 | 3.6 ± 2.8 | 57.3 ± 18.7 | 86.3 |
| Conformer | .905 ± .079 | 4.9 ± 4.9 | 60.8 ± 25.8 | **51.2** |
| **This work** | **.914 ± .067** | **3.5 ± 3.2** | **55.4 ± 21.8** | 59.3 |

**XCorr**: cross correlation, **MSE**: mean square error, **RMSE**: root mean square errors, **FLOPS**: floating-point operations per second.

TABLE III
PERFORMANCE OF TRANSFORMED CE INTEGRITY

| No. | Features | $\epsilon^{f_i}_{abs}$(msec) | $\epsilon^{f_i}_{rel}$ (%) |
|---|---|---|---|
| $f_1$ | $P_{timing}$ | 30.8 ± 1.52 | - |
| $f_2$ | $R_{timing}$ | 4.2 ± 0.81 | - |
| $f_3$ | $T_{timing}$ | 15.9 ± 0.97 | - |
| $f_4$ | $QRS_{dur}$ | 19.5 ± 0.47 | - |
| $f_5$ | $RR_{int}$ | 1.7 ± 0.27 | - |
| $f_6$ | $QT^e_{int}$ | 23.3 ± 0.58 | - |
| $f_7$ | $QT^p_{int}$ | 25.9 ± 0.21 | - |
| $f_8$ | $QRS^+_{int}$ | - | 3.4 ± 0.056 |
| $f_9$ | $QRS^-_{int}$ | - | 3.9 ± 0.044 |
| $f_{10}$ | $P_{amp}$ | - | 2.7 ± 0.043 |
| $f_{11}$ | $T_{amp}$ | - | 5.4 ± 0.126 |

$\epsilon^{f_i}_{abs}$: absolute errors, $\epsilon^{f_i}_{rel}$: relative errors.

## V. RESULTS AND DISCUSSION

### A. Baseline Models and Performance Evaluation

To establish a benchmark, we implemented several deep learning architectures for comparison: CNN, LSTM, bi-directional LSTM (BLSTM), and Conformer. The CNN comprised five 1-D convolutional layers with a kernel size of three and varying filter numbers (40, 20, 20, 20, and 40). Both LSTM and BLSTM models utilized four stacked layers, each with 32 filters. All baseline models culminated in a fully connected layer with 64 filters. The baseline Conformer model utilized the same architecture as the ECG autoencoder with mmWave signals as the input.

Table II compares the synthesized ECG waveform quality metrics (XCorr, MSE, and RMSE) for these baseline models (CNN, LSTM, BLSTM, and Conformer) and the proposed cross-modal autoencoder. Our approach achieved superior performance across all metrics, with an XCorr of 0.914, MSE of $3.5 \times 10^{-3}$, and RMSE of $55.4 \times 10^{-3}$. Notably, the BLSTM model exhibited a comparable transformation accuracy but incurred a higher computational cost (86.3 MFLOPs) than that of our model (59.3 MFLOPs), signifying a 30% reduction in computational resources with the proposed method.

### B. Transformation Quality of CEs

Table III provides a comprehensive analysis of the integrity of the transformed CEs achieved using the proposed method. All average absolute errors, including the P, R, and T peak locations and QRS, RR, and QT interval durations, fell well below the allowable tolerance of 150 ms set by the

TABLE IV
ABLATION STUDY OF THE PROPOSED CROSS-MODAL AUTOENCODER WITH JOINT DEEP FEATURE LOSS

| mmWave Encoder | mmWave Decoder | Waveform Loss | CE Predictor | Shared ECG Decoder | Deep Feature Loss | Spectro Loss | XCorr. | MSE ($\times 10^{-3}$) | RMSE ($\times 10^{-3}$) |
|---|---|---|---|---|---|---|---|---|---|
| ✓ | ✓ | ✓ | | | | | .905 ± .079 | 4.4 ± 4.9 | 60.8 ± 25.8 |
| ✓ | ✓ | ✓ | ✓ | | | | .912 ± .071 | 3.9 ± 3.6 | 58.0 ± 22.5 |
| ✓ | | ✓ | ✓ | ✓ | | | .913 ± .072 | 3.7 ± 3.7 | 56.0 ± 23.6 |
| ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | .914 ± .067 | 3.5 ± 3.2 | 55.4 ± 21.8 |

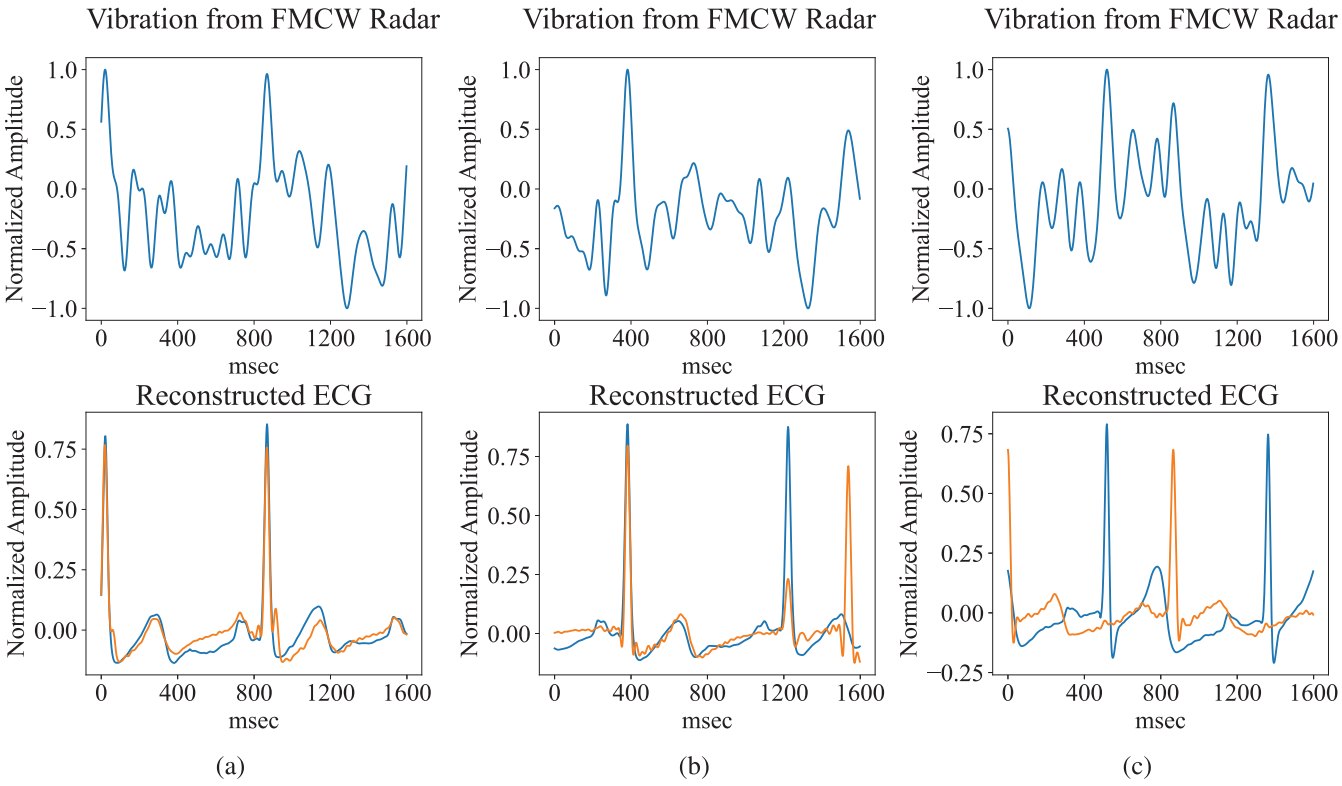



Fig. 9. Three examples of corresponding mmWave (top in blue), ground-truth (GT) ECG (bottom in blue), and synthesized ECG waveforms from the proposed cross-modal autoencoder (bottom in orange), with their cross correlation values of (a) 0.950, (b) 0.678, and (c) 0.666, respectively.

AAMI [18]. In addition, the transformed P- and T-wave amplitudes exhibited minimal errors of 2.7% and 5.4%, respectively. In summary, the proposed model achieved high-fidelity ECG reconstruction with amplitude and time errors of less than 10% and 5%, respectively, satisfying the diagnostic accuracy standards set by the AAMI [17], further substantiating the reliability of our results.

An ablation study assessed the contribution of pretrained modules to the overall transformation performance of CE detection. The results in Table IV reveal that incorporating pretrained event information improves transformation accuracy with the cross correlation increasing from 0.905 to 0.912, MSE decreasing from 4.4 to 3.9, and RMSE decreasing from 60.8 to 50.8. Notably, utilizing a shared ECG decoder and a joint loss function further enhanced performance with a cross correlation of 0.914, MSE of $3.5 \times 10^{-3}$, and RMSE of $55.4 \times 10^{-3}$, as reported in Table II.

### C. Discussion

Our study introduces a novel cross-modal autoencoder for contactless ECG monitoring using signals from an FMCW radar sensor. Experimental results validate the effectiveness of cross-modal learning with deep and joint loss functions in overcoming the limitations of small datasets and domain discrepancies between the ECG and radar signals. The proposed model achieved high-fidelity ECG reconstruction and satisfied the diagnostic accuracy standards set by the AAMI [17], [18]. These findings hold promise for applying contactless FMCW radars in long-term, noninvasive ECG monitoring, potentially enabling earlier diagnoses of various cardiac conditions like arrhythmias and hyperkalemia [5], [46].

Fig. 9 showcases examples of paired raw mmWave signals (top in blue) and their corresponding synthesized ECG waveforms (bottom in orange). Fig. 9(a) exemplifies the proposed model's ability to accurately capture critical ECG events and waveform details, achieving a high correlation coefficient of 0.950 between the synthesized and recorded ECG signals. This high correlation indicates that the mmWave signals from the FMCW radar contain sufficient information for reliable ECG waveform transformation. Conversely, Fig. 9(b) depicts a negative example in which missing features in the raw mmWave signal hindered accurate ECG waveform

reconstruction. Fig. 9(c) further highlights the impact of misalignment between the synthesized and actual P-peak on overall performance. These observations underscore the importance of high-quality mmWave signals for robust ECG transformations.

Several previous works have explored ECG reconstruction using FMCW radar signals [12], [13]. Our study achieved comparable performance, with correlation coefficients exceeding 90%. However, direct comparisons were hindered by variations in the experimental setups, sensing ranges, and receiver/transmitter configurations employed in these studies. While prior research has focused on efficient FMCW radar signal preprocessing and feature engineering, this study delved into advanced deep-learning approaches to tackle the limitations of limited data availability and inherent domain dissonance between FMCW radar signals and ECG.

Although the dataset collected in this study was sufficient to demonstrate the proof-of-concept of the prototype cross-modal autoencoder for ECG monitoring using the FMCW radar, training the ECG reconstruction model with this limited ECG dataset was challenging. The primary reason for this is that ECG patterns and features vary significantly among individuals; therefore, training an end-to-end ECG reconstruction model with a limited dataset may fail to capture the critical features of ECGs from unseen subjects. We observed that the baseline models often failed in the ECG reconstruction of subjects with distinctive ECG patterns. Future efforts will focus on expanding the validation by recruiting healthy subjects across different age groups and patients with various cardiac conditions to further assess the reliability and usability of the proposed approach. This enabled a more comprehensive assessment of the reliability and usability of the proposed approach in real-world scenarios. Advanced signal enhancement and preprocessing techniques, such as the wavelet transform [47] and empirical mode decomposition [48], will be explored to improve the quality of mmWave signals, potentially leading to enhanced transformation accuracy. In addition, we intend to investigate the influence of extrinsic factors on the system reliability, including the sensing range, background noise levels, and device position. By addressing these aspects, we aim to significantly improve the robustness and practical applicability of FMCW radar-based ECG monitoring systems.

## VI. Conclusion

A novel cascade framework that incorporates a cross-modal autoencoder to transform FMCW radar signals into ECG signals was introduced in this article. This approach addresses the challenges associated with limited training data and inherent differences between the two signal domains. The proposed mmWave-to-ECG transformation system achieved high accuracy with a cross correlation of 0.914 and low average timing errors (under 31 ms) for critical ECG features. These findings demonstrate the feasibility of FMCW radar with deep learning technologies for long-term, contactless ECG monitoring, offering significant advantages over traditional methods. This paves the way for advancements in continuous remote patient care for the monitoring of chronic heart conditions.

## Acknowledgment

Kai-Chun Liu is with the College of Information and Computer Sciences, University of Massachusetts, Amherst, MA 01003 USA (e-mail: kaichunliu@umass.edu).

Sheng-Yu Peng, Che-Yu Liu, Zhu-An Chen, and Zong Han Han are with the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan (e-mail: sypeng@mail.ntust.edu.tw).

Yu Tsao, Wen-Chi Chen, Po-Quan Hsieh, and You-Jin Li are with the Research Center for Information Technology Innovation, Academia Sinica, Taipei 115, Taiwan (e-mail: yu.tsao@sinica.edu.tw).

Yu-Juei Hsu and Shun-Neng Hsu are with the Division of Nephrology, Department of Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei 114, Taiwan.

## References

[1] I. Zagan, V. G. Gaitan, N. Iuga, and A. Brezulianu, "M-GreenCARDIO embedded system designed for out-of-hospital cardiac patients," in *Proc. Int. Conf. Develop. Appl. Syst. (DAS)*, May 2018, pp. 11–17.

[2] S. R. Steinhubl et al., "Effect of a home-based wearable continuous ECG monitoring patch on detection of undiagnosed atrial fibrillation: The mSToPS randomized clinical trial," *JAMA*, vol. 320, pp. 146–155, Jul. 2018.

[3] J. J. Gierałtowski, K. Ciuchcski, I. Grzegorczyk, K. Ka, M. Solki, and P. Podziemski, "Heart rate variability discovery: Algorithm for detection of heart rate from noisy, multimodal recordings," in *Proc. Comput. Cardiol.*, Sep. 2014, pp. 253–256.

[4] U. Satija, B. Ramkumar, and M. S. Manikandan, "Automated ECG noise detection and classification system for unsupervised healthcare monitoring," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 3, pp. 722–732, May 2018.

[5] T.-M. Chen et al., "SRECG: ECG signal super-resolution framework for portable/wearable devices in cardiac arrhythmias classification," *IEEE Trans. Consum. Electron.*, vol. 69, no.3, pp. 250–260, Jan. 2023.

[6] S. M. P. Dinakarrao, A. Jantsch, and M. Shafique, "Computer-aided arrhythmia diagnosis with bio-signal processing: A survey of trends and techniques," *ACM Comput. Surveys*, vol. 52, no. 2, pp. 1–37, Mar. 2019.

[7] S. S. Sofos, H. Tehrani, K. Shokrollahi, and M. I. James, "Surgical staple as a transcutaneous transducer for ECG electrodes in burnt skin: Safe surgical monitoring in major burns," *Burns*, vol. 39, no. 4, pp. 818–819, Jun. 2013.

[8] A. Searle and L. Kirkup, "A direct comparison of wet, dry and insulating bioelectric recording electrodes," *Physiolog. Meas.*, vol. 21, no. 2, pp. 271–283, May 2000.

[9] Y. Zhang, R. Yang, Y. Yue, E. G. Lim, and Z. Wang, "An overview of algorithms for contactless cardiac feature extraction from radar signals: Advances and challenges," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–20, 2023.

[10] D. Uguz et al., "Car seats with capacitive ECG electrodes can detect cardiac pacemaker spikes," *Sensors*, vol. 20, no. 21, p. 6288, Nov. 2020.

[11] X. Tang, W. Chen, S. Mandal, K. Bi, and T. Özdemir, "High-sensitivity electric potential sensors for non-contact monitoring of physiological signals," *IEEE Access*, vol. 10, pp. 19096–19111, 2022.

[12] J. Chen, D. Zhang, Z. Wu, F. Zhou, Q. Sun, and Y. Chen, "Contactless electrocardiogram monitoring with millimeter wave radar," *IEEE Trans. Mobile Comput.*, vol. 23, no. 1, pp. 270–285, Jan. 2024.

[13] C. Xu et al., "CardiacWave: A mmWave-based scheme of non-contact and high-definition heart activity computing," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 3, pp. 1–26, Sep. 2021.

[14] Z. Chen, T. Zheng, C. Cai, and J. Luo, "MoVi-Fi: Motion-robust vital signs waveform recovery via deep interpreted RF sensing," in *Proc. 27th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, New York, NY, USA, 2021, pp. 392–405.

[15] S. Wu et al., "Person-specific heart rate estimation with ultra-wideband radar using convolutional neural networks," *IEEE Access*, vol. 7, pp. 168484–168494, 2019.

[16] K. Yamamoto, R. Hiromatsu, and T. Ohtsuki, "ECG signal reconstruction via Doppler sensor by hybrid deep learning model with CNN and LSTM," *IEEE Access*, vol. 8, pp. 130551–130560, 2020.

[17] *Diagnostic Electrocardiographic Devices*, document ANSI/AAMI EC11:1991 (R2007), AAMI, 2007.

[18] *AAMI, Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*, document ANSI/AAMI/ISO EC57:1998 (R2008) (ANSI/AAMI/ISO EC 57:1998 (R2008), 2008.

[19] M. Gu, Z. Chen, K. Chen, and H. Pan, "IR-ST: A lightweight transformer network for human fall detection based on FMCW radar," *IEEE Sensors J.*, vol. 23, no. 20, pp. 25128–25135, Oct. 2023.
[20] W.-L. Hsu, J.-X. Liu, C.-C. Yang, and J.-S. Leu, "A fall detection system based on FMCW radar range-Doppler image and bi-LSTM deep learning," *IEEE Sensors J.*, vol. 23, no. 18, pp. 22031–22039, Nov. 2023.
[21] Y. Yao et al., "Unsupervised-learning-based unobtrusive fall detection using FMCW radar," *IEEE Internet Things J.*, vol. 11, no. 3, pp. 5078–5089, Feb. 2024.
[22] L. Cao, R. Wei, Z. Zhao, D. Wang, and C. Fu, "A novel frequency-tracking algorithm for noncontact vital sign monitoring," *IEEE Sensors J.*, vol. 23, no. 19, pp. 23044–23057, Oct. 2023.
[23] J.-H. Choi, K.-B. Kang, and K.-T. Kim, "RF-vital: Radio-based contactless respiration monitoring for a moving individual," *IEEE Internet Things J.*, vol. 11, no. 8, pp. 13137–13151, Apr. 2024.
[24] W. Kang, Y. Li, and W. Wu, "In-vehicle multiple passengers respiration monitoring based on surface-circuit metasurface tags using time-division FMCW radar," *IEEE Internet Things J.*, vol. 11, no. 5, pp. 7756–7771, Mar. 2024.
[25] L. Liu, J. Zhang, Y. Qu, S. Zhang, and W. Xiao, "MmRH: Noncontact vital sign detection with an FMCW mm-wave radar," *IEEE Sensors J.*, vol. 23, no. 8, pp. 8856–8866, Apr. 2023.
[26] L. Qu, C. Liu, T. Yang, and Y. Sun, "Vital sign detection of FMCW radar based on improved adaptive parameter variational mode decomposition," *IEEE Sensors J.*, vol. 23, no. 20, pp. 25048–25060, Oct. 2023.
[27] Y. Wang et al., "A novel non-contact respiration and heartbeat detection method using frequency-modulated continuous wave radar," *IEEE Sensors J.*, vol. 24, no. 7, pp. 10434–10446, Apr. 2024.
[28] A. Gulati et al., "Conformer: Convolution-augmented transformer for speech recognition," 2020, *arXiv:2005.08100*.
[29] A. Radhakrishnan et al., "Cross-modal autoencoder framework learns holistic representations of cardiovascular state," *Nature Commun.*, vol. 14, no. 1, p. 2436, Apr. 2023.
[30] Y.-W. Chen et al., "EMA2S: An end-to-end multimodal articulatory-to-speech system," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2021, pp. 1–5.
[31] X. Tan, T. Qin, F. Soong, and T.-Y. Liu, "A survey on neural speech synthesis," 2021, *arXiv:2106.15561*.
[32] K.-C. Wang, K.-C. Liu, H.-M. Wang, and Y. Tsao, "EMGSE: Acoustic/EMG fusion for multimodal speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 1116–1120.
[33] L. Ruan et al., "MM-diffusion: Learning multi-modal diffusion models for joint audio and video generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 10219–10228.
[34] D. Wang et al., "End-to-end voice conversion via cross-modal knowledge distillation for dysarthric speech reconstruction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 7744–7748.
[35] F. G. Germain, Q. Chen, and V. Koltun, "Speech denoising with deep feature losses," 2018, *arXiv:1806.10522*.
[36] L. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.
[37] A. Pearce, J. A. Zhang, R. Xu, and K. Wu, "Multi-object tracking with mmWave radar: A review," *Electronics*, vol. 12, no. 2, p. 308, Jan. 2023.
[38] M. Alizadeh, G. Shaker, J. C. M. D. Almeida, P. P. Morita, and S. Safavi-Naeini, "Remote monitoring of human vital signs using mm-wave FMCW radar," *IEEE Access*, vol. 7, pp. 54958–54968, 2019.
[39] J. Fraden and M. R. Neuman, "QRS wave detection," *Med. Biol. Eng. Comput.*, vol. 18, no. 2, pp. 125–132, Mar. 1980.
[40] K. van der Bijl, M. Elgendi, and C. Menon, "Automatic ECG quality assessment techniques: A systematic review," *Diagnostics*, vol. 12, no. 11, p. 2578, Oct. 2022.
[41] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[42] M. S. Manikandan and S. Dandapat, "Wavelet energy based diagnostic distortion measure for ECG," *Biomed. Signal Process. Control*, vol. 2, no. 2, pp. 80–96, Apr. 2007.
[43] N. Tulyakova and O. Trofymchuk, "Real-time filtering adaptive algorithms for non-stationary noise in electrocardiograms," *Biomed. Signal Process. Control*, vol. 72, Feb. 2022, Art. no. 103308.
[44] Y. Zigel, A. Cohen, and A. Katz, "The weighted diagnostic distortion (WDD) measure for ECG signal compression," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 11, pp. 1422–1430, Nov. 2000.
[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[46] C.-S. Lin et al., "A deep-learning algorithm (ECG12Net) for detecting hypokalemia and hyperkalemia by electrocardiography: Algorithm development," *JMIR Med. Informat.*, vol. 8, no. 3, Mar. 2020, Art. no. e15931.
[47] S.-S. Wang, P. Lin, Y. Tsao, J.-W. Hung, and B. Su, "Suppression by selecting wavelets for feature compression in distributed speech recognition," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 3, pp. 564–579, Mar. 2018.
[48] Y.-D. Lin, Y. K. Tan, and B. Tian, "A novel approach for decomposition of biomedical signals in different applications based on data-adaptive Gaussian average filtering," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103104.

**Kai-Chun Liu** (Member, IEEE) received the M.S. and Ph.D. degrees in biomedical engineering from National Yang-Ming University, Taipei, Taiwan, in 2015 and 2019, respectively.

From 2020 to 2023, he was a Postdoctoral Scholar with the Research Center for Information Technology Innovation, Academia Sinica, Taipei. He is currently a Postdoctoral Research Associate with the College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, USA. His research interests include pervasive healthcare, wearable computing, machine learning, and biosignal processing.

**Sheng-Yu Peng** (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1995 and 1997, respectively, the M.Sc. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2004, and the Ph.D. degree in electrical and computer engineering from Georgia Institute of Technology, Atlanta, GA, USA, in 2008.

He joined the National Taiwan University of Science and Technology (NTUST), Taipei, in 2011. He is currently a Professor with the Department of Electrical Engineering. His research interests include interface circuits for sensors and biomedical applications, reconfigurable analog circuits and systems, power-efficient analog signal processing, and low-power machine learning algorithms.

Dr. Peng received the NTUST 2022, 2021, and 2013 Excellent Teaching Awards, the NTUST 2022 Excellent Research Award, the IEEE Taipei Section 2018 Best Master Thesis Advisor Award, and the IEEE Taipei Section 2018 Best Ph.D. Dissertation Advisor Award. He also received the Best Student Paper Award at the 2016 IEEE International Ultrasonics Symposium.

**Yu Tsao** (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1999 and 2001, respectively, and the Ph.D. degree in electrical and computer engineering from Georgia Institute of Technology, Atlanta, GA, USA, in 2008.

From 2009 to 2011, he was a Researcher with the National Institute of Information and Communications Technology, Tokyo, Japan, where he engaged in research and product development in automatic speech recognition for multilingual speech-to-speech translation. He is currently a Research Fellow (Professor) and the Deputy Director with the Research Center for Information Technology Innovation, Academia Sinica, Taipei. He is also a Jointly Appointed Professor with the Department of Electrical Engineering, Chung Yuan Christian University, Taoyuan, Taiwan. His research interests include assistive oral communication technologies, audio coding, and biosignal processing.

Dr. Tsao was a recipient of the Academia Sinica Career Development Award in 2017, national innovation awards from 2018 to 2021, Future Tech Breakthrough Award 2019, Outstanding Elite Award, Chung Hwa Rotary Educational Foundation from 2019 to 2020, NSTC FutureTech Award 2022, and NSTC Outstanding Research Award 2023. He is the corresponding author of a paper that received the 2021 IEEE SIGNAL PROCESSING SOCIETY (SPS), Young Author, Best Paper Award. He is currently an Associate Editor for the IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and IEEE SIGNAL PROCESSING LETTERS.

**Che-Yu Liu** received the B.S. degree in electrical engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 2022, where he is currently pursuing the M.S. degree in electrical engineering.

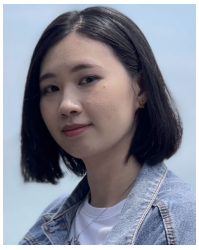
His researches are mainly in the field of deep learning and contactless vital sign monitoring.

**Zhu-An Chen** received the B.S. degree in electronic engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 2023, where he is currently pursuing the M.S. degree in electric engineering.

His research interests include contactless physiological measurements, deep learning, and health monitoring systems.

**Zong Han Han** received the B.S. degree in electrical engineering from Tunghai University, Taichung, Taiwan, in 2020, and the M.S. degree in electrical engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 2023.

His research interests include deep learning, physiological signal processing, and vital sign detection.



**Wen-Chi Chen** received the B.S. and M.S. degrees from the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, in 2017 and 2022, respectively.

From 2021 to 2022, she was a Research Assistant with the Research Center for Information Technology Innovation, Academia Sinica, Taipei. She is currently a Senior Engineer with StreamTeck, Taipei. Her current research interests include radar AI in the medical field.

**Po-Quan Hsieh** received the B.S. degree in electrical engineering from Fu Jen Catholic University (FJU), Taipei, Taiwan, in 2018, and the M.S. degree from National Taiwan University of Science and Technology (NTUST), Taipei, in 2021.

From 2021 to 2023, he was a Research Assistant with the Research Center for Information Technology Innovation, Academia Sinica, Taipei. He is currently a Senior Engineer with StreamTeck, where he is working on the Millimeter Wave Business Unit. His current research interests include antennas at mmWave frequencies, radar applications, hardware design, and measurements.

**You-Jin Li** received the B.S. degree from the Department of Electronic Engineering, National Ilan University, Ilan, Taiwan, in 2014, and the M.S. degree from the Department of Electrical Engineering with communications from the National Ilan University, in 2016. He is currently pursuing the Ph.D. degree with the Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan.

His research interests cover signal processing, speech enhancement, beamforming, deep learning, and multichannel compression.

**Yu-Juei Hsu** received the M.D. degree from the National Defense Medical Center, Taipei, Taiwan, in 1997, and the Ph.D. degree in medical science from Radboud University Nijmegen, Nijmegen, The Netherlands, in 2009.

He was the Chief Resident of the Nephrology Division of Tri-Service General Hospital (TSGH), Taipei, from 2003 to 2004. In 2004, he became a Faculty Member at the nephrology division. He is currently the Deputy Superintendent of TSGH, a Professor at the National Defense Medical Center, and a Member of the IRB board. His research interests include mineral and bone disorders, as well as cardiovascular disorders caused by chronic kidney disease.

Dr. Hsu was awarded as the Best Intern and Resident during his medical training. He also received awards for Excellent Research from Taiwan Society of Nephrology in 2003 and 2016, Young Investigator from the National Defense Medical Center in 2008, Excellent Military Physician of the Medical Affairs Bureau in 2014 and 2021, and Excellent Teacher of the National Defense Medical Center in 2018. In 2021 and 2022, he has been listed among the world's top 2% scientists, published by Stanford University and Elsevier.

**Shun-Neng Hsu** received the M.D. degree from the National Defense Medical Center (NDMC), Taipei, Taiwan, in 2008, and the Ph.D. degree in medical science from the Roslin Institute, University of Edinburgh, Edinburgh, U.K., in 2022.

He was the Chief Resident of the Nephrology Division of Tri-Service General Hospital (TSGH), Taipei, from 2013 to 2014. In 2022, he became a Faculty Member at the Nephrology Division. He is currently the Chief of the Dialysis Center, Division of Nephrology of TSGH, and an Assistant Professor at the NDMC. His research specializes in diagnosing and managing chronic kidney disease-mineral bone disorders (CKD-MBD), collaborating extensively with renowned experts and laboratories dedicated to renal bone diseases.

Dr. Hsu received awards for Excellent Research from the Taiwan Society of Nephrology in 2022 and 2023 and the New Investigator Award from the Bone Research Society (BRS) in 2021.