



國立臺灣大學電機資訊學院暨中央研究院

資料科學學位學程

博士論文

Data Science Degree Program

College of Electrical Engineering and Computer Science

National Taiwan University and Academia Sinica

Doctoral Dissertation

人工智慧在心電圖的心律不整分類

Artificial Intelligence in Electrocardiography

for Cardiac Arrhythmia Classification

陳在民

Tsai-Min Chen

指導教授：曹昱 博士、沈俊嚴 博士

Advisor: Yu Tsao, Ph.D., Chun-Yen Shen, Ph.D.

中華民國 113 年 11 月

November, 2024



國立臺灣大學博士學位論文 口試委員會審定書

PhD DISSERTATION ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

人工智慧在心電圖的心律不整分類

Artificial Intelligence in Electrocardiography for Cardiac Arrhythmia Classification

本論文係 陳在民 (D08946007) 在國立臺灣大學資料科學學位學程完成之博士學位論文，於民國 112 年 7 月 10 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Data Science Degree Program on 10 July 2023 have examined a PhD dissertation entitled above presented by CHEN, TSAI-MIN (D08946007) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

曹星

(指導教授 Advisor)

Chun-Yu Shen

(指導教授 Advisor)

郭弘

王曉嵐

謝伯廷

王榮德

學程主任 Director:

謝宏明

誌謝



在完成這篇論文的過程中，我特別感謝指導教授曹昱博士、沈俊嚴博士，感謝他們給予我最大的研究自由與空間。同時，衷心感謝中央研究院資訊科技創新研究中心的曾煥鑫博士、劉凱鈞博士，感謝他們提供的寶貴經驗，使我在研究過程中少走了許多彎路。此外，中央研究院生物醫學科學研究所的黃明經博士、胡瑜峰博士、施宣誠博士、吳致勳博士、陳思豪博士、方偉泉博士、王宣凱博士、吳育樂博士、Sirapop Nuannimnoi 博士、孫慶妹、黃之瀚、郭宜欣、陳致余，感謝你們的支持與鼓勵。

也感謝台灣人工智慧學校基金會的陳昇瑋博士、蔡明順博士、蔡源鴻博士、林德全博士、李國淵和湯沂達，感謝他們在技術、硬體與經濟等方面提供的寶貴資源。特別感謝亞東紀念醫院人工智慧中心的王榮德博士、李念庭，以及元智大學的方士豪博士，感謝他們提供臨床驗證的機會。最後，感謝 Google PhD Fellowship Program 2023 的無償贊助，與 Google Health Clinical Lead 的 Joe Ledsam 博士的支持。

中文摘要



人工智慧在提升心電圖於心律不整分類中的應用已變得至關重要。本研究利用 2018 年中國生理信號挑戰賽提供的包含 6,877 筆記錄的大型 12 導聯心電圖資料集，結合卷積神經網路和循環神經網路(CNN+RNN)開發出一個人工智慧模型。該模型在 2,954 筆心電圖隱藏測試集上的九類心律不整分類中，達到了 0.84 的整體 F1-score 中位數，展現了其在心律不整檢測中的卓越效能。進一步分析顯示，該模型能充分預測具有多種心律不整診斷的患者。值得注意的是，使用單導聯資料時，模型性能僅輕微下降，其中 aVR 和 V1 導聯最具資訊價值。

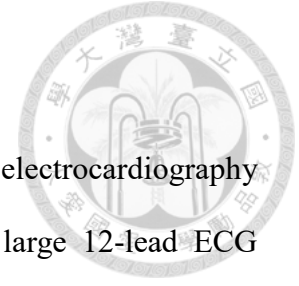
為解決便攜式/穿戴式裝置在長期心電圖監測中，因電池壽命和傳輸頻寬限制所使用的低解析度信號帶來的準確性下降，我提出了一個基於深度學習的心電圖信號超解析框架(SRECG)。SRECG 透過在應用於高解析度訊號多分類器(HMC)的心律不整分類時提升低解析度心電圖信號的準確性。實驗結果顯示，與傳統的插值方法相比，SRECG 顯著提高了 HMC 的心律不整分類準確性，證實了 SRECG 在增強來自便攜式/穿戴式裝置的低解析度心電圖信號以改善基於雲端的 HMC 性能方面的可行性。

此外，我研究了最初為心電圖心律不整分類設計的模型在其他生理信號中的多模態應用潛力，尤其是結合語音信號、人口統計資料和結構化醫療記錄以檢測聲門腫瘤。在從良性聲音障礙檢測聲門腫瘤的應用中，我的模型取得了顯著的準確性，展現了人工智慧方法學在跨不同生理信號診斷上的潛力。

總結而言，人工智慧技術，尤其是深度學習模型，在心電圖分析和心律不整分類方面取得了重大進展。強大的 CNN+RNN 模型、SRECG 等信號增強技術，以及人工智慧模型在其他生理信號中的多模態適應性，展示了人工智慧在廣泛的醫學診斷領域中的應用潛力。

關鍵字：人工智慧、心電圖、心律不整、深度學習、信號增強、穿戴式裝置、多模態

ABSTRACT



Artificial intelligence (AI) has become crucial in enhancing electrocardiography (ECG) classification for cardiac arrhythmia (CA). Leveraging a large 12-lead ECG dataset with 6,877 records from the 2018 China Physiological Signal Challenge (CPSC2018), I developed an AI model combining convolutional neural network and recurrent neural network (CNN+RNN). This model achieved a median overall F1-score of 0.84 across nine CA categories on a hidden test set of 2,954 ECG data, demonstrating high efficiency in CA detection. Further analysis showed that the model effectively predicted coexisting CAs in patients with multiple CA diagnoses. Notably, performance only slightly decreased when using single-lead data compared to the full 12-lead dataset, with the aVR and V1 leads providing the most informative signals.

To address challenges in long-term ECG monitoring with portable/wearable (P/W) devices, such as battery constraints and limited transmission bandwidth resulting in low-resolution signals, I proposed a deep learning (DL)-based ECG super-resolution framework (SRECG). SRECG enhances low-resolution ECG signals, focusing on accuracy for CA classification when applied to high-resolution multiclass classifiers (HMC). Experimental results showed that SRECG significantly improved HMC accuracy over traditional interpolation methods, confirming the feasibility of using SRECG to enhance low-resolution ECG signals from P/W devices, improving cloud-based HMC performance.

Additionally, I investigated the adaptability of the model, initially designed for ECG CA classification, to other physiological signals. Specifically, I explored its multimodal potential to detect glottic tumors in benign vocal disorders by integrating voice signals, demographic data, and structured medical records. My model achieved remarkable

accuracy, highlighting the potential of AI methodologies to transition across different physiological signals for diagnostic purposes.

In summary, AI, particularly DL models, has achieved significant advances in ECG analysis and CA classification. The integration of powerful CNN+RNN models, signal enhancement techniques such as SRECG, and adaptability to other physiological signals underscores the transformative impact of AI's potential for broader medical diagnostics.

Key words: artificial intelligence, electrocardiography, cardiac arrhythmia, deep learning, signal enhancement, wearable device, multimodality

CONTENTS



口試委員會審定書	i
誌謝	ii
中文摘要	iii
ABSTRACT	iv
CONTENTS	vi
LIST OF FIGURES	viii
LIST OF TABLES	ix
Chapter 1 ECG Signal Recognition	1
1.1 Introduction	1
1.2 Methods	4
Data acquisition.....	4
Proposed ECG model architecture	5
Performance evaluation	7
Ensemble and optimization.....	10
1.3 Results.....	11
Best validation models on 10-fold tests and ensemble model on	
hidden test	11
Concurrent CA types	13
Model performances with single-lead ECG	16
Comparison between human assessments and AI	20
1.4 Discussion	22
1.5 Conclusion	23
Chapter 2 ECG Signal Reconstruction.....	25
2.1 Introduction	25
2.2 Methods	31
Data acquisition.....	31
Benchmark ECG model architecture	32
Performance evaluation	34

Traditional spline interpolation for up-sampling	34
Proposed SRECG framework	35
2.3 Results.....	40
Effects of sampling frequency on CA classification.....	40
Effects of traditional spline interpolation methods	41
Effects of joint training losses in SRECG.....	44
2.4 Discussion	47
2.5 Conclusion	50
Chapter 3 Multimodality Adaptation	51
3.1 Introduction	51
3.2 Methods	53
Data acquisition.....	53
Proposed multimodal model architecture	57
Performance evaluation	59
3.3 Results.....	61
Internal validation on FVD.....	61
External validation on SVD	65
Intrinsic validation on CPSC2018 ECG training set.....	67
Comparison between human assessments and AI	68
3.4 Discussion	70
3.5 Conclusion	74
REFERENCE.....	75

LIST OF FIGURES



Figure 1 Architecture of proposed DL network for 12-lead ECG	6
Figure 2 Probabilities output on 10-fold tests by the best validation models	14
Figure 3 The ranked F1-score results of single-lead models	18
Figure 4 Comparison of DL workflows for a P/W device.....	30
Figure 5 Architecture of benchmark DL network for single-lead ECG	33
Figure 6 Architecture of proposed SRECG network for enhancing ECG	38
Figure 7 Framework for training proposed SRECG network.....	39
Figure 8 Predictive power changes of CA classifiers on downsampled ECG data.....	41
Figure 9 Enhanced ECG signals visualization of different spline interpolation degrees.....	43
Figure 10 Enhanced ECG signals visualization of different SRECG losses	46
Figure 11 Architecture of proposed multimodal DL network.....	58
Figure 12 Performance of proposed multimodal model with only voice signals on FVD	62
Figure 13 Performance of proposed multimodal model with voice and demographics on FVD	63
Figure 14 Performance of proposed multimodal model with voice, demographics, and medical records on FVD.....	64
Figure 15 Performance of proposed multimodal model with only voice signals on SVD	66
Figure 16 The performance of proposed multimodal model with voice signals and demographics on SVD.....	67

LIST OF TABLES



Table 1 Details of recordings in the CPSC2018 ECG database.....	5
Table 2 Numbers and sizes of hyperparameters and trainable parameters of proposed DL network.....	7
Table 3 The performances of models trained with multi-labeled data	12
Table 4 CPSC2018's top 10 models and results.....	12
Table 5 Label count statistics of the 476 multi-labeled subjects.....	15
Table 6 The performances of models trained without multi-labeled data.....	15
Table 7 Successfully predicted CA types for the 476 multi-labeled subjects	16
Table 8 The Bayes factors of each lead's performance (F1-score) in log scale.....	19
Table 9 Leads in the top-performing group (Bayes factor < 3.0)	19
Table 10 Leads in the top-performing group (Bayes factor < 0.33)	20
Table 11 Comparisons of CA diagnosis between conference-assigned label, AI model prediction, and expert cardiologists	21
Table 12 Median overall F1-scores (%) of HMCs trained at 250 Hz with different leads and spline interpolation degrees	42
Table 13 Median overall F1-scores (%) of HMCs trained at 250 Hz with different leads and SRECG losses	45
Table 14 Numbers of voice disorders in the FVD.....	55
Table 15 Descriptions of structured medical records in the FVD	56
Table 16 Confusion matrix for FVD (voice + medical records) and SVD (voice + demographics)	64
Table 17 Comparison of glottic neoplasm diagnosis between AI model prediction and pathologists	68
Table 18 Prediction results between 2 independent raters.....	69

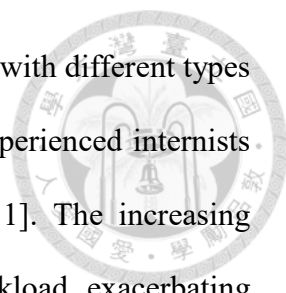
Chapter 1 ECG Signal Recognition



1.1 Introduction

Cardiac arrhythmia (CA) is an indicator of cardiovascular diseases and is associated with an increased risk of mortality [1]. It is typically diagnosed using electrocardiography (ECG), which are non-invasive, cost-effective, and widely employed clinical tools for monitoring heart function. Diagnosis hinges on analyzing specific waveform features of the ECG, such as the P wave, QRS complex, and T wave. A standard ECG captures signals from twelve leads: six limb leads (I, II, III, aVR, aVL, aVF) and six chest leads (V1 to V6). Each lead records the heart's electrical activity from a unique angle, encompassing both the frontal plane (limb leads) and the transverse plane (chest leads) [2, 3].

Each ECG lead highlights particular signal characteristics linked to specific types of CA. For example, atrial fibrillation (AF) is identified by fibrillatory atrial waves and irregular QRS conduction [4]. Left bundle branch block (LBBB) presents distinctive QRS morphologies in leads I, aVL, V1, V2, V5, and V6, while right bundle branch block (RBBB) is recognized by an rsR' pattern in leads V1 and V2 [5]. First-degree atrioventricular block (I-AVB) is characterized by a consistent PR interval exceeding 0.2 seconds [6]. Premature atrial contractions (PAC) and premature ventricular contractions (PVC) signify ectopic electrical impulses, with the P wave or QRS morphology deviating from normal heartbeats [7, 8]. An ST segment is deemed abnormal if the elevation (STE) or depression (STD) surpasses 0.1 mV [9].



Accurately identifying these complex ECG features associated with different types of CA requires extensive training. Studies have shown that even experienced internists and cardiologists may occasionally misdiagnose CA types [10, 11]. The increasing volume of ECG examinations adds to healthcare professionals' workload, exacerbating this issue. Developing computer algorithms that can provide precise and automatic diagnoses may help alleviate the burden. Despite challenges posed by the significant variability in the geometrical and physiological features of ECG signals [12], substantial progress has been made in recent years [13].

As artificial intelligence (AI) prospers most recently, there are two main strategies for developing automatic CA diagnostic tools. The first involves segmenting ECG signals into individual heartbeats or characteristic waveform cycles. This beat-based approach can generate large datasets for machine learning models, even with a limited number of subjects. However, extracting precise morphological features from ECG signals is often challenging [13]. While some studies have reported prediction accuracies above 99% using this method [13], results may be confounded if training and test data originate from the same individuals. For instance, when test beats were sourced from patients not included in the training set, the classification accuracy for six CA types dropped from 99.7% to 81.5% [14].

Most previous CA prediction studies have utilized publicly available data from the MIT-BIH Arrhythmia Database (MIT-BIH AD) [15, 16] and the UCI Machine Learning Repository's Arrhythmia Data Set (UCIAD) [17], containing 48 and 452 subjects, respectively. Small sample sizes can lead to overfitting, particularly with neural network algorithms [18]. Overfitting may also result from unbalanced data if certain CA types are overrepresented. These issues can bias analyses using MIT-BIH AD and UCIAD [19, 20].

For example, a study using UCIAD achieved a 92% classification accuracy with an 80-20 training-test split, but accuracy dropped to 60% with a 50-50 split [19]. Additionally, MIT-BIH AD includes ECG data from only two leads, and UCIAD provides only extracted features rather than raw 12-lead ECG data.

The second strategy offers an end-to-end solution, avoiding the main challenges of the beat-based approach. This method requires a large ECG database and the development of suitable deep learning (DL) neural networks. Recent advancements have made this approach increasingly viable. For example, the PhysioNet/Computing in Cardiology Challenge 2017 (CinC2017) released single-lead ECG data from 8,528 subjects across four heart rhythm categories to promote open-source research [21]. Utilizing a convolutional neural network combined with long short-term memory layers, Xiong et al. achieved top performance in CinC2017 with an F1-score of 0.82 on a hidden test set of 3,658 subjects.

Building on this, the China Physiological Signal Challenge 2018 (CPSC2018), hosted by the 7th International Conference on Biomedical Engineering and Biotechnology [22], provided a large 12-lead ECG database and a hidden test set for global participants. Unlike CinC2017, CPSC2018 included subjects categorized into normal rhythm and eight types of CA: AF, I-AVB, LBBB, RBBB, PAC, PVC, STD, and STE. This represents the largest publicly available 12-lead ECG database with the most labeled CA types to date. In this study, I present a DL neural network model developed using the CPSC2018 ECG data and discuss the results that secured first place in the competition (<http://2018.icbeb.org/Challenge.html>).

1.2 Methods

Data acquisition



As detailed by Liu et al. [22], the CPSC2018 ECG database comprises 9,831 recordings collected from 9,458 individuals across 11 hospitals in China (see Table 1). All data from these 11 hospitals were combined to constitute the challenge database. This database consists of 12-lead ECG recordings sampled at 500 Hz and standardized using Z-score normalization, with durations ranging from a few seconds to over a minute, and an outlier lasting up to 144 seconds. Each recording is labeled as normal or as one of eight abnormal CA types previously mentioned. The dataset was randomly split into 70% for training (training set) and 30% for testing (hidden test set), but only the training set was made publicly available. The distributions of sex, age, hospital sources, and CA types were fairly balanced between the training set and hidden test set [22]. Among the 6,877 recordings in the training set, 470 had two CA-type labels, and six had three.

Table 1 Details of recordings in the CPSC2018 ECG database

Challenge set	Type	# recording of each hospital											# total recording
		A	B	C	D	E	F	G	H	I	J	K	
Training	Normal	73	90	74	73	70	90	77	76	78	77	140	918
	AF	45	184	28	26	138	155	280	104	27	47	64	1098
	I-AVB	39	71	0	8	150	129	212	19	15	27	34	704
	LBBB	7	12	7	2	41	30	51	13	9	18	17	207
	RBBB	79	272	29	30	423	205	401	140	26	30	60	1695
	PAC	57	147	4	8	105	24	111	50	2	57	9	574
	PVC	76	111	3	21	150	80	132	13	18	32	17	653
	STD	117		22	32	525	36	52	1	26		15	826
	STE	3	90	6	7	38	12	16	0	22	0	8	202
Total/Average		490	980	169	202	1645	762	1338	416	223	288	364	6877
Hidden Test	Normal	31	38	32	32	30	38	33	33	34	33	60	394
	AF	19	79	12	11	59	66	120	44	11	20	28	469
	I-AVB	17	30	0	3	64	56	91	8	6	11	15	301
	LBBB	3	5	3	1	18	13	22	5	4	8	8	90
	RBBB	34	116	13	13	181	88	173	61	12	13	25	729
	PAC	23	65	1	3	45	12	48	22	3	24	4	250
	PVC	30	49	2	10	65	33	56	6	8	15	7	281
	STD	50	0	9	14	225	16	23	0	11	0	6	354
	STE	2	38	2	3	16	5	7	0	10	0	3	86
Total/Average		209	420	74	90	703	327	573	179	99	124	156	2954

For privacy reasons, the names of the 11 hospitals are represented by A-K (11 letters).

Proposed ECG model architecture

My model features an architecture composed of five convolutional neural network (CNN) blocks, followed by a bidirectional gated recurrent unit (GRU) [23], an attention layer [24], and a final dense (fully connected) layer (see Figure 1). The selection of five CNN blocks and specific neural network layers resulted from a limited trial-and-error process; I did not extensively explore other potential architectures. Each CNN block includes two convolutional layers followed by a pooling layer to reduce parameters, computational load, and to mitigate overfitting [25]. Additionally, I applied a 20% dropout rate between CNN blocks and other independent layers, including between the last CNN block and the bidirectional GRU layer, to further prevent overfitting. I chose CNNs and recurrent neural network (RNN) due to their proven effectiveness in handling noisy signals and time-series data [26, 27]. The GRU, a newer variant of RNN, may offer

advantages over long short-term memory networks in terms of training efficiency and iteration requirements [28, 29]. Batch normalization was used to adjust and scale inputs from the attention layer [30], which follows a form proposed by Yang et al. that assigns importance weights [24], to the dense layer. I employed the LeakyReLU activation function, a variant of the rectified linear unit, in all layers except the dense layer, where I used the Sigmoid activation function [31]. Table 2 provides details on the hyperparameters, including their types, sizes, and the number of trainable parameters for each network layer.

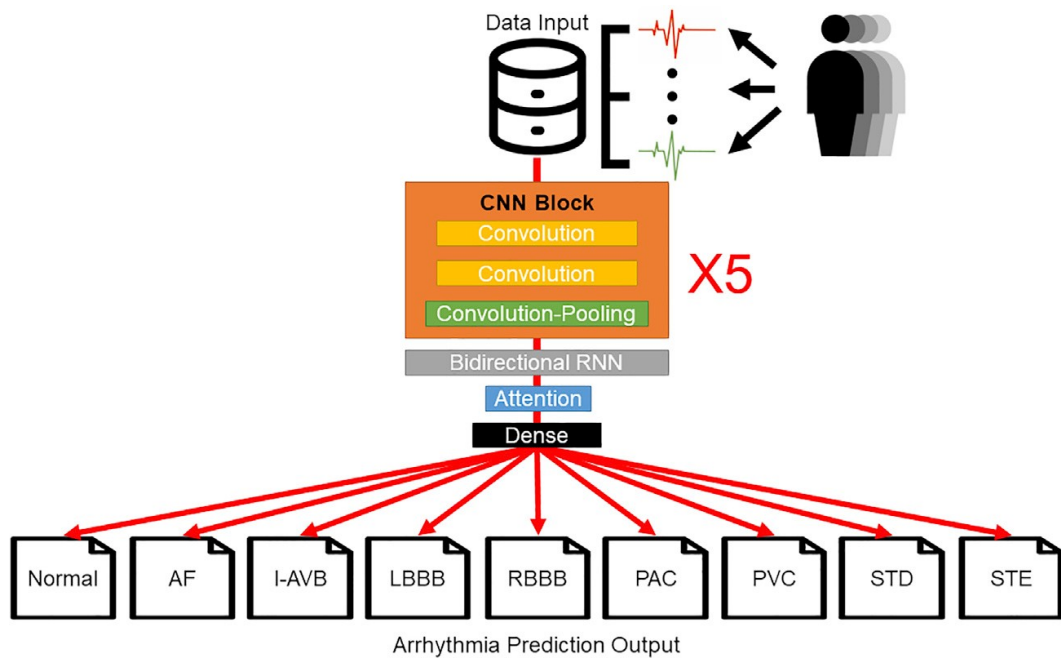


Figure 1 Architecture of proposed DL network for 12-lead ECG

The architecture consists of layers and blocks displayed in rectangular shapes. The label “X5” denotes that five CNN blocks are sequentially connected prior to linking with the bidirectional RNN layer, which is a GRU. The output layer at the base presents the model's predicted probabilities for the nine different CA classes. The prediction corresponds to the class with the highest probability for the given ECG input.



Table 2 Numbers and sizes of hyperparameters and trainable parameters of proposed DL network

CNN block & other network layers	CNN layer	Kernel	Kernel size	Padding	Stride	Parameters*
1	1	12	3	1	1	$12 \times (12 \times 3 + 1) = 444$
	2	12	3	1	1	$12 \times (12 \times 3 + 1) = 444$
	3 (pooling)	12	24	1	2	$12 \times (12 \times 24 + 1) = 3,468$
2	4	12	3	1	1	$12 \times (12 \times 3 + 1) = 444$
	5	12	3	1	1	$12 \times (12 \times 3 + 1) = 444$
	6 (pooling)	12	24	1	2	$12 \times (12 \times 24 + 1) = 3,468$
3	7	12	3	1	1	$12 \times (12 \times 3 + 1) = 444$
	8	12	3	1	1	$12 \times (12 \times 3 + 1) = 444$
	9 (pooling)	12	24	1	2	$12 \times (12 \times 24 + 1) = 3,468$
4	10	12	3	1	1	$12 \times (12 \times 3 + 1) = 444$
	11	12	3	1	1	$12 \times (12 \times 3 + 1) = 444$
	12 (pooling)	12	24	1	2	$12 \times (12 \times 24 + 1) = 3,468$
5	13	12	3	1	1	$12 \times (12 \times 3 + 1) = 444$
	14	12	3	1	1	$12 \times (12 \times 3 + 1) = 444$
	15 (pooling)	12	48	1	2	$12 \times (12 \times 48 + 1) = 6,924$
Bi-GRU	-	-	-	-	-	$2 \times 12 \times 3 \times (12 + 12 + 2) = 1,872$
Attention	-	-	-	-	-	$12 \times 2 \times (12 \times 2 + 2) = 624$
Batch-normalization	-	-	-	-	-	$12 \times 2 \times 4 = 96$
Dense	-	-	-	-	-	$9 \times (12 \times 2 + 1) = 225$
Total						28,035

* Calculated as follows:

CNN layer: $\#lead \times (\#kernels \times kernel\ size + bias)$

Bi-GRU: $\#directions \times \#cells \times \#gates \times (\#leads + \#leads + bias)$

Attention: $\#leads \times 2 \times (\#leads \times 2 + bias)$

Batch normalization: $\#leads \times 2 \times 4$

Dense: $\#output \times (\#leads \times 2 + bias)$

$\#directions = 2$ (bi-directional)

$\#gates = 3$ (reset, update, and current memory gates)

$\#leads = 12$ (12 ECG leads)

$\#output = 9$ (9 rhythm classifications).

Performance evaluation

The F1-score is an evaluation metric that considers both precision and recall, defined as:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \quad (1)$$



where precision and recall are calculated by:

$$Precision = \frac{TP}{TP + FP}, \quad (2)$$

$$Recall = \frac{TP}{TP + FN}, \quad (3)$$

and here, TP (true positive), FP (false positive), TN (true negative), and FN (false negative), are defined as:

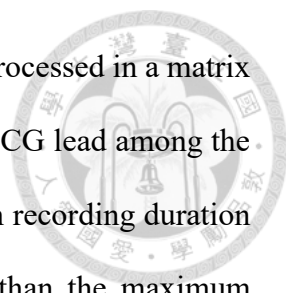
$$TP_j = |\{(x_i, y_i) | y_i = f_j(x_i) = 1, 1 \leq i \leq N\}|, \quad (4)$$

$$FP_j = |\{(x_i, y_i) | y_i = 0 \text{ and } f_j(x_i) = 1, 1 \leq i \leq N\}|, \quad (5)$$

$$TN_j = |\{(x_i, y_i) | y_i = f_j(x_i) = 0, 1 \leq i \leq N\}|, \quad (6)$$

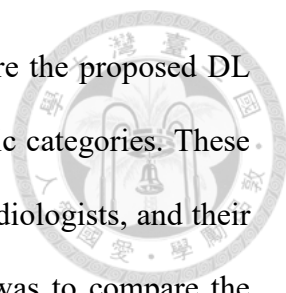
$$FN_j = |\{(x_i, y_i) | y_i = 1 \text{ and } f_j(x_i) = 0, 1 \leq i \leq N\}|, \quad (7)$$

and in these equations, (x_i, y_i) represents the input and label of the i -th sample out of N samples, and $f = (f_1, \dots, f_j, \dots, f_C) \in \{0, 1\}^C$ is a classifier predicting C -categories, where each f_j is binary-valued. The notation $|\cdot|$ denotes the cardinality of a set. The overall F1-score is obtained by averaging the F1-scores across all categories.



In my implementation, the CPSC2018 ECG training set were processed in a matrix format consisting of three components: the subject ID, the specific ECG lead among the 12 leads, and the 72,000 ECG values corresponding to the maximum recording duration (144 seconds) at a sampling rate of 500 Hz. Recordings shorter than the maximum duration were prefixed with zeros for alignment. I extracted the 476 subjects with multiple labels and randomly divided the remaining 6,401 subjects into 10 equal parts to establish an 8-1-1 training, validation, and testing scheme. To address label imbalance, I employed stratified sampling to maintain proportional representation of each CA type. The multi-labeled subjects were then reintegrated into the training part. Classification training was conducted using the categorical cross-entropy loss function and the ADAM optimizer, utilizing the GPU version of TensorFlow through the Keras package [32]. Models were trained for 100 epochs, each epoch representing one full pass through the training part, and evaluated based on performance on the validation part. The model with the lowest validation loss was further assessed by calculating its F1-score on the test part. This process was repeated 10 times to complete a 10-fold training, validation, and testing cycle, resulting in 10 optimal validation models. The median F1-score for each CA label, including the normal type, across the 10 test parts was computed using the F1-score function from Scikit-learn [33].

I also investigated the performance using data from only a single ECG lead. For each lead, I set the ECG values of the other 11 leads to zero and applied the same network architecture and 10-fold cross-validation and testing procedure as before. This resulted in 120 optimal single-lead validation models and a median F1-score for each of the 12 leads across the nine CA labels.



In my sub-analysis, I examined nine false positive cases where the proposed DL model predicted a high probability for CA across different diagnostic categories. These cases were then subjected to independent review by three expert cardiologists, and their assessments were compared to the model's predictions. The goal was to compare the model's predictions with the consensus of the human experts and the original conference-assigned labels, which served as the “ground truth” for training the model and considered as the performance of general cardiologists.

Ensemble and optimization

For the CPSC2018 competition, I combined the 130 best validation models (10 from full-lead training and 120 from single-lead training) into an ensemble model. The output probabilities from the 130 models for each CA type were averaged and adjusted using a weight vector to obtain the final probability for that CA type. The weight vector, consisting of nine weights corresponding to the nine CA types, was optimized using a genetic algorithm (GA) [34] to achieve the best overall median F1-score across the 10 test parts. In the GA optimization, I established a mating system with 40 genes, each with a DNA length of 9, and a population size of 100. The optimization process proceeded for 100 generations, using a crossover probability of 0.5 and a mutation probability of 0.5. For a given ECG recording input, the CA type with the highest probability from the ensemble model was selected as the predicted CA type. This ensemble model was my submission to CPSC2018.

1.3 Results

Best validation models on 10-fold tests and ensemble model on hidden test



Table 3 presents a comparison for each CA type between the median accuracy, area under the receiver operating characteristic curve (AUC), and F1-score derived from the best validation models across ten 10-fold tests, and those obtained from the ensemble model. Additionally, it includes the ensemble model's F1-score on the hidden test set from CPSC2018. The results indicate that the ensemble model slightly outperformed the individual best validation models, which is expected since it combines and optimizes their outputs to enhance the 10-fold test outcomes. Furthermore, the ensemble model demonstrated consistent performance across all CA types when transitioning from the publicly available data to the hidden test set, reflecting the similar compositions of these datasets.

Table 3 also highlights varying levels of difficulty in predicting different CA types. Specifically, prediction accuracy decreased progressively from AF to bundle branch blocks, premature contractions, and finally ST abnormalities, with the normal rhythm being among the most challenging to predict. The model's predictions for STE had the lowest F1-score, ranging from 0.5 to 0.6, which could be partly attributed to differing medical opinions on diagnosing STE [35]. This trend, including the difficulty in predicting the normal type, was consistent across all other top-performing models in the CPSC2018 competition (see Table 4). In fact, nearly all the leading models achieved very high F1-scores (>0.9) for AF and bundle branch blocks. My model, however, provided significantly better predictions than others for several CA types, PAC, PVC, STD, and STE, which explains my superior performance in the competition (Table 4).



Table 3 The performances of models trained with multi-labeled data

CA Typ	Best Validation Models			Ensemble Model			
	Median Accuracy	Median AUC (95% CI)	Median F1-Score	Median Accuracy	Median AUC (95% CI)	Median F1-Score	Hidden Set F1-Score
Normal	0.940	0.890 (0.810–0.942)	0.795	0.949	0.867 (0.832–0.973)	0.808	0.801
AF	0.969	0.928 (0.902–0.985)	0.897	0.983	0.963 (0.914–0.993)	0.944	0.933
I-AVB	0.972	0.899 (0.864–0.988)	0.865	0.977	0.950 (0.875–0.990)	0.899	0.875
LBBS	0.990	0.914 (0.748–1.000)	0.821	0.995	0.942 (0.763–1.000)	0.899	0.884
RBSB	0.955	0.956 (0.887–0.988)	0.911	0.952	0.946 (0.871–0.976)	0.903	0.910
PAC	0.957	0.867 (0.749–0.955)	0.734	0.963	0.920 (0.779–0.981)	0.797	0.826
PVC	0.970	0.928 (0.841–0.988)	0.852	0.977	0.932 (0.864–0.996)	0.874	0.869
STD	0.951	0.878 (0.797–0.972)	0.788	0.959	0.906 (0.815–0.970)	0.834	0.811
STE	0.976	0.707 (0.558–0.995)	0.509	0.977	0.773 (0.603–0.993)	0.600	0.624

The results shown are based on the top-performing validation models and the ensemble model across the ten 10-fold tests. The figures in the final column (in bold) represent the median F1-scores of the ensemble model for the CPSC2018 hidden test set.

Table 4 CPSC2018's top 10 models and results

Rank	Overall F1	F _{af}	F _{block}	F _{pc}	F _{st}
1	0.837	0.933	0.899	0.847	0.779
2	0.830	0.931	0.912	0.817	0.761
3	0.806	0.914	0.879	0.801	0.742
4	0.802	0.918	0.89	0.789	0.718
5	0.791	0.924	0.882	0.779	0.709
6	0.783	0.905	0.902	0.722	0.708
7	0.782	0.911	0.891	0.775	0.670
8	0.778	0.921	0.858	0.797	0.676
9	0.776	0.906	0.876	0.773	0.711
10	0.766	0.894	0.857	0.733	0.683

My model's results, ranked first, are shown in bold, with the top scores in each sub-competition highlighted in red. The overall F1-score represents the average of the F1-scores across all classification types. F_{af} refers to the F1-score for AF; F_{block} corresponds to the F1-scores for I-AVB, LBBS, and RBSB; F_{pc} represents the F1-scores for PAC and PVC; and F_{st} includes the F1-scores for STD and STE.



Concurrent CA types

One factor contributing to reduced accuracy for certain CA types is the prediction of multiple CA types with nearly equal probabilities in some patients. Figure 2 illustrates the probability outputs from the best validation models for ECG subjects when they were part of the test fold in the 10-fold tests. It is evident that normal rhythm, STD, and STE lack distinctive probability scores to differentiate them from the other eight types, aligning with the performance results shown in Table 3. Further analysis of the model probabilities revealed that many patients with AF also had a high probability of RBBB, and RBBB frequently coincided with PAC and PVC, in addition to AF (Figure 2). These concurrent CA probability results correspond well with the statistics of the 476 multi-labeled subjects; specifically, the three most common multi-labeled occurrences were AF/RBBB, RBBB/PAC, and RBBB/PVC (Table 5). An ensemble model that excluded these 476 multi-labeled subjects from the CPSC2018 ECG training set still performed well in predicting these multiple CA types (Tables 6 and 7), indicating the model's capacity to capture ECG features of concurrent CAs. These findings are generally consistent with clinical observations that rate-dependent (phase 3) block during ectopic atrial beats or AF can lead to RBBB [36, 37]. However, a larger dataset of multi-labeled subjects is necessary to thoroughly assess my model's performance in diagnosing concurrent CAs.

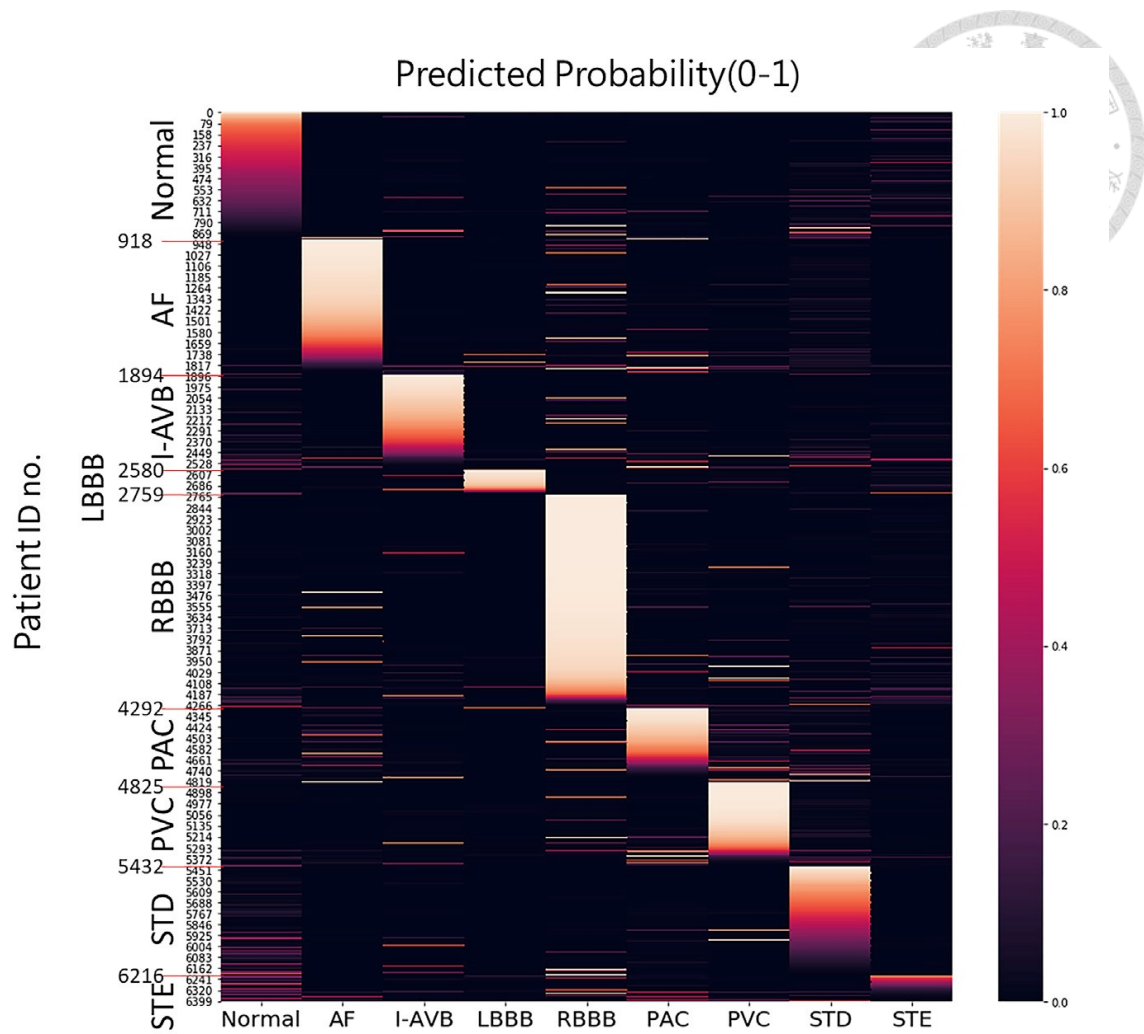


Figure 2 Probabilities output on 10-fold tests by the best validation models
The probability scale is represented on the right, with color coding to indicate the likelihood for each of the nine CA classifications. The class with the highest probability is the model's predicted outcome for the corresponding ECG recording.

Table 5 Label count statistics of the 476 multi-labeled subjects

	AF	I-AVB	LBBB	RBBB	PAC	PVC	STD	STE
AF	0	0	29	172	4	8	33	2
I-AVB		0	8	10	3	5	6	4
LBBB			0	0	10	6	3	4
RBBB				0	55	51	20	19
PAC					2	3	6	5
PVC						0	18	2
STD							0	2
STE								0

Only the upper triangular section of the symmetrical concurrent CA label count matrix is displayed. The three highest counts are highlighted in bold.

Table 6 The performances of models trained without multi-labeled data

CA Type	Best Validation Models			Ensemble Model		
	Median Accuracy	Median AUC (95% CI)	Median F1-score	Median Accuracy	Median AUC (95% CI)	Median F1-score
Normal	0.940	0.908 (0.791-0.916)	0.807	0.937	0.901 (0.808-0.932)	0.794
AF	0.974	0.949 (0.885-0.992)	0.915	0.980	0.955 (0.929-0.995)	0.935
I-AVB	0.973	0.918 (0.852-0.991)	0.876	0.976	0.912 (0.900-0.996)	0.879
LBBB	0.993	0.927 (0.748-1.000)	0.870	0.993	0.913 (0.770-1.000)	0.862
RBBB	0.961	0.954 (0.895-0.981)	0.922	0.944	0.925 (0.880-0.972)	0.885
PAC	0.957	0.852 (0.747-0.961)	0.747	0.965	0.889 (0.798-0.984)	0.796
PVC	0.973	0.926 (0.832-0.989)	0.858	0.974	0.950 (0.820-0.992)	0.870
STD	0.950	0.869 (0.800-0.966)	0.786	0.956	0.914 (0.790-0.950)	0.821
STE	0.974	0.667 (0.491-0.995)	0.394	0.975	0.663 (0.491-0.995)	0.444

The 476 multi-labeled recordings were excluded from the training set for these models. The results shown are from the top-performing validation models and the ensemble model across the ten 10-fold tests. These outcomes are comparable to those shown in Table 3.

Table 7 Successfully predicted CA types for the 476 multi-labeled subjects

	AF	I-AVB	LBBB	RBBB	PAC	PVC	STD	STE
AF	0/0	0/0	17/29	154/172	0/4	6/8	4/33	0/2
I-AVB		0/0	2/8	8/10	0/3	2/5	0/6	0/4
LBBB			0/0	0/0	2/10	4/6	0/3	0/4
RBBB				0/0	34/55	49/51	16/20	5/19
PAC					0/0	3/3	4/6	2/5
PVC						0/0	6/18	0/2
STD							0/0	2/2
STE								0/0

Displayed is only the upper triangular portion of the symmetrical concurrent CA label counts. The values represent the number of correctly predicted subjects versus the total number of multi-labeled subjects for each CA type. The predicted concurrent CA types correspond to the two CA types with the highest and second highest probabilities. The three most frequent concurrent CA labels are highlighted in bold.

Model performances with single-lead ECG

Figure 3 presents the median F1-scores from the 10-fold tests for models using data from individual ECG leads. The performance observed in the best validation models using 12-lead data (Table 3) was largely mirrored when only single-lead data were used. In most instances, there were only slight variations in F1-scores for the classification of specific CA types between the analyses of 12-lead and single-lead ECG data. The results indicate that lead aVR was among the best-performing single leads, ranking first in overall average performance and in three individual CA types (normal rhythm, AF, and STD), and placed within the top three for all CA types except STE and PAC. Another high-performing single lead was lead V1, which ranked first in three CA types (first-degree atrioventricular block I-AVB, RBBB, and PAC), but performed worse than most other leads for other CA types. In contrast, lead I, which is used by the Apple Watch [38],

did not perform as notably in my tests. Lead II, commonly favored by physicians for a quick assessment of an ECG due to its clear signal [39], ranked fifth in overall average performance but was not statistically different from the top-performing leads (p-value from paired t-test < 0.05). These findings are further supported by a Bayesian factor analysis [40] that rigorously evaluated statistical differences among these leads (see Tables 8, 9, and 10).

These performance rankings suggest that my model effectively captured the lead-specific morphological features of various CA types. For instance, the presence of deep and broad S-waves in lead V1 and broad, clumsy R-waves in lead V6 are indicative of LBBB [41], and both V1 and V6 were identified as top-performing single leads. Similarly, the diagnostic criteria for RBBB include the rSR' pattern in leads V1 and V2 [42], which were also among the highest-performing single leads.

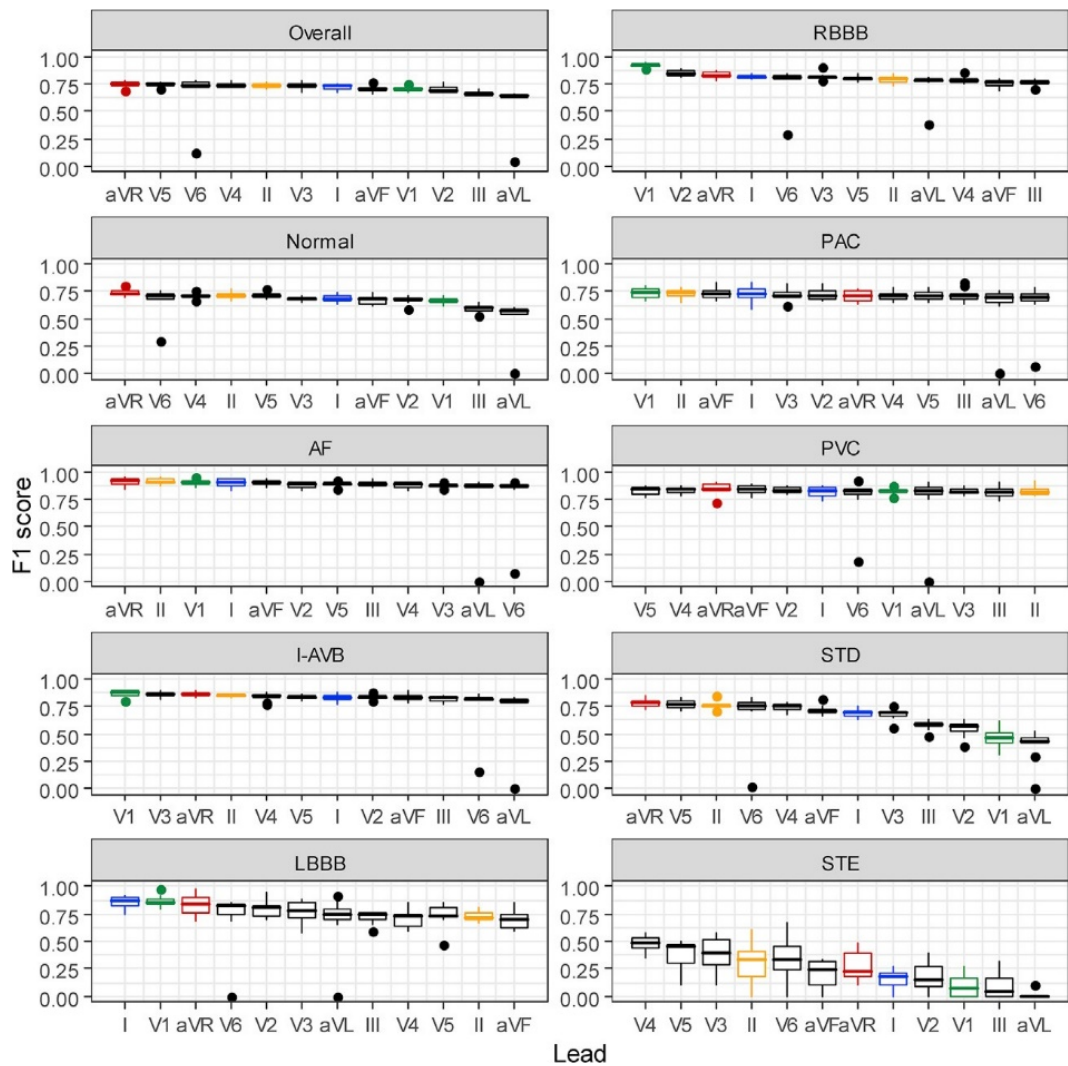


Figure 3 The ranked F1-score results of single-lead models

Displayed on the y-axis are the F1-scores from single-lead models tested in the 10-fold validation. Lead aVR is highlighted in red, V1 in green, I in blue, and II in orange.

Table 8 The Bayes factors of each lead's performance (F1-score) in log scale

	I	II	III	aVR	aVL	aVF	V1	V2	V3	V4	V5	V6
Normal	3.4	0.3	14.1	-0.9	3.1	4.1	7.3	5.3	5.0	0.9	0.4	-0.1
AF	-0.8	-0.8	-0.6	-0.9	-0.3	-0.7	-0.9	0.2	1.1	0.8	-0.3	-0.2
I-AVB	1.2	0.1	3.7	-0.4	0.1	0.8	-0.9	1.3	-0.5	0.5	1.1	-0.1
LBBB	-0.9	6.1	6.4	-0.8	0.4	6.1	-0.9	0.2	1.0	3.8	1.9	-0.1
RBBB	18.3	15.5	22.2	11.4	4.1	18.8	-0.9	9.9	12.9	16.0	18.6	1.8
PAC	-0.9	-0.9	-0.6	-0.3	-0.1	-0.9	-0.9	-0.8	-0.7	-0.2	-0.4	-0.2
PVC	-0.8	-0.9	-0.8	-0.8	-0.6	-0.9	-0.9	-0.9	-0.9	-0.9	-0.9	-0.6
STD	6.0	-0.9	15.4	-0.9	10.2	2.4	14.9	11.9	5.6	0.6	-0.8	-0.5
STE	11.9	1.7	11.3	3.7	23.1	6.7	12.2	7.1	0.3	-0.9	0.4	0.8

Calculated using the BayesFactor function from the R package.

Table 9 Leads in the top-performing group (Bayes factor < 3.0)

	I	II	III	aVR	aVL	aVF	V1	V2	V3	V4	V5	V6
Normal	0	1	0	1	0	0	0	0	0	0	1	1
AF	1	1	1	1	1	1	1	1	0	0	1	1
I-AVB	0	1	0	1	1	0	1	0	1	0	0	1
LBBB	1	0	0	1	1	0	1	1	0	0	0	1
RBBB	0	0	0	0	0	0	1	0	0	0	0	0
PAC	1	1	1	1	1	1	1	1	1	1	1	1
PVC	1	1	1	1	1	1	1	1	1	1	1	1
STD	0	1	0	1	0	0	0	0	0	0	1	1
STE	0	0	0	0	0	0	0	0	1	1	1	0
Total	4	6	3	7	5	3	6	4	4	3	6	7

Leads in the top-performing group are marked with a 1 (those excluded are marked with a 0) for each CA type. Statistically, these leads performed similarly based on a Bayes factor threshold of < 3.0, indicating no significant difference from the leading lead. According to this threshold, leads aVR and V6 had the highest number of top-performing group counts, followed by leads II, V1, and V5.

Table 10 Leads in the top-performing group (Bayes factor < 0.33)

	I	II	III	aVR	aVL	aVF	V1	V2	V3	V4	V5	V6
Normal	0	0	0	1	0	0	0	0	0	0	0	0
AF	1	1	1	1	0	1	1	0	0	0	0	0
I-AVB	0	0	0	0	0	0	1	0	1	0	0	0
LBBB	1	0	0	1	0	0	1	0	0	0	0	0
RBBB	0	0	0	0	0	0	1	0	0	0	0	0
PAC	1	1	1	0	0	1	1	1	1	0	0	0
PVC	1	1	1	1	1	1	1	1	1	1	1	1
STD	0	1	0	1	0	0	0	0	0	0	1	1
STE	0	0	0	0	0	0	0	0	0	1	0	0
Total	4	4	3	5	1	3	6	2	3	2	2	2

Leads in the top-performing group are marked with a 1 (and with a 0 if excluded) for each CA type, based on a Bayes factor threshold of < 0.03, indicating that the null hypothesis of no significant difference from the leading lead holds. According to this threshold, lead V1 had the highest count of top-performing group classifications, followed closely by lead aVR.

Comparison between human assessments and AI

In sub-analyses comparing the AI model's predictions with human evaluations (see Table 11), I observe a mix of outcomes. In some cases, such as LBBB and RBBB, the proposed AI model's predictions aligned with the expert cardiologists' consensus, which is highlighted in red. For instance, case 6380 was labeled as LBBB both by the model and the expert cardiologists, showing the model's high predictive accuracy in certain scenarios. Similarly, case 4237 was identified as I-AVB consistently across the model and the expert cardiologists' assessments. However, in other cases, the model's predictions diverged from both the conference label and the expert cardiologists' opinions. For example, in case 504, the conference labeled a normal SR, which was validated by the expert cardiologists, while the model prediction was set as STE.

These discrepancies highlight both the strengths and weaknesses of model in medical diagnostics. While the model is highly accurate, even better than the performance of general cardiologists, it may still require expert oversight in more ambiguous or complex cases. This comparison demonstrates the potential of DL model to assist cardiologists, but also underscores the need for continuous expert involvement, especially in refining edge cases where DL model predictions may lack the necessary clinical context or nuance.

Table 11 Comparisons of CA diagnosis between conference-assigned label, AI model prediction, and expert cardiologists

PatientID	Conf.	Model	Cardiologist 1	Cardiologist 2	Cardiologist 3	Consensus
6268	STE	Normal	Normal	SR (normal)	SR (normal)	Normal
6215	STD	AF	Narrow-QRS tachycardia, SVT, RBBB, RAD	PSVT	PSVT	PSVT
4237	RBBB	I-AVB	I-AVB	I-AVB, SR, TWI (V1-V4), rSR' (V1)	I-AVB,SR	I-AVB
6380	STE	LBBB	LBBB	LBBB,SR,LAE	LBBB,SR	LBBB
1452	AF	RBBB	AF, MVR, RBBB	AF, RBBB, Q wave, STE with reciprocal change, w/o old MI	AF, RBBB	RBBB
5398	PVC	PAC	PAC	PAC	PAC,SR	PAC
5963	STD	PVC	PVC	PVC,SR,STD(V3-V6)	PVC,SR	PVC
4278	RBBB	STD	Normal	SR, minimal STTC(II, III, aVF)	STD-like, SR	Normal
504	Normal	STE	STE-like	SR, early repolarization	SR	Normal

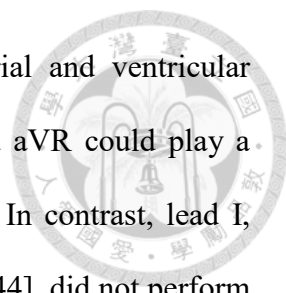
Conference assignments (considered as the “ground truth” for training model) or proposed model's predictions that align with the consensus of the three expert cardiologists are marked in red. Abbreviations: SVT: supraventricular tachycardia; RAD: right axis deviation; MVR: mitral valve replacement; SR: sinus rhythm; PSVT: paroxysmal supraventricular tachycardia; LAE: left atrial enlargement; MI: myocardial infarction; STTC: ST-T change.

1.4 Discussion



The field of AI, particularly DL, has recently made significant advancements in medical diagnostics [43]. Interestingly, recent studies, including mine, have achieved overall F1-scores ranging from 0.82 to 0.84 in predicting CA types. Although these AI-based ECG diagnostic tools have not been extensively tested in real-world clinical settings, they have demonstrated significant improvements in diagnostic accuracy compared to general cardiologists [10, 11]. Consequently, these AI models hold promise for reducing diagnostic errors and easing the workload on medical professionals. While these results are promising, it's important to note that further enhancements in model accuracy may be limited until the majority of "ground truth" CA diagnoses used for training AI models are provided by expert cardiologists (see Table 11).

My analysis indicates that models utilizing single-lead ECG data can predict CA types with performance levels similar to those using all 12 leads. Clinical diagnostic criteria for CA types are often specific to certain leads. In my model, the highest-ranking single leads for diagnosing RBBB or LBBB corresponded with the leads specified in clinical diagnostic guidelines [5], reinforcing the validity of my AI diagnostic approach. The notable performance of lead aVR, a lead often overlooked in clinical practice, in my model is particularly intriguing and merits further attention. Traditionally, leads I, II, and V1 are used in continuous monitoring or portable/wearable (P/W) ECG devices [38]. However, in my AI model, lead aVR outperformed these conventional leads in predicting several CA types. Unlike leads that consider reciprocal information from the left lateral side of the heart, lead aVR focuses on capturing specific information from the right upper side, including the outflow tract of the right ventricle and the basal part of the septum. Its vector alignment is parallel to the anatomical and electrical axis from the atrial base to



the ventricular apex, potentially maximizing the detection of atrial and ventricular depolarization signals. These characteristics may explain why lead aVR could play a unique role in diagnosing CAs, potentially surpassing other leads. In contrast, lead I, which is utilized by devices like the Apple Watch for detecting AF [44], did not perform as well in my analysis. My findings suggest that the most effective single lead for predicting different CA types may vary in clinical applications. This insight could stimulate future research into the use of lead aVR in diagnosing various CA types and its implementation in P/W ECG devices.

CAs are complex conditions, and the occurrence of concurrent CA types is not uncommon, especially among those with related electrophysiological characteristics. While existing ECG-based CA diagnostic models have primarily focused on predicting a single CA type, my analysis demonstrates that AI is capable of diagnosing multiple CA types simultaneously. Future research should explore the detection and classification of concurrent CAs, and my model serves as an initial step in this direction. Beyond CA, ECG data have been effective in detecting other health issues, such as asymptomatic left ventricular dysfunction [45] and non-invasive monitoring of potassium levels [46]. As AI and machine learning techniques continue to evolve and become more accessible to non-specialists, I anticipate that ECG will be increasingly leveraged for diagnosing a broader range of diseases and clinical conditions.

1.5 Conclusion

In this study, I demonstrated the efficacy of my AI-driven approach for the accurate diagnosis of CAs using ECG data collected from 11 hospitals. My DL model, designed

to process both 12-lead and single-lead ECG inputs, not only excels in identifying singular CA events but also shows remarkable capability in diagnosing concurrent CAs with precision. By leveraging an ensemble of 12-lead and single-lead models, my solution achieved the top ranking in the CPSC2018 competition, highlighting its robustness across different lead configurations.

Notably, the performance analysis identified aVR and V1 as the best-performing single leads, providing critical insights into lead selection for efficient CA detection. These findings underscore the potential of focused lead analysis in enhancing diagnostic accuracy while simplifying the complexity of ECG data input.

My model sets a new benchmark in CA diagnosis, with implications for more accurate, scalable, and reliable applications in clinical settings. The results from this research reinforce the promise of AI in revolutionizing cardiac care and support the integration of DL models in everyday diagnostic workflows.

Chapter 2 ECG Signal Reconstruction

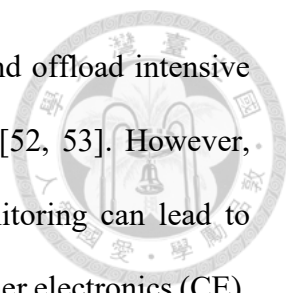


2.1 Introduction

Consumer technologies have significantly impacted the world, particularly in healthcare, prompting traditional medicine to evolve into telemedicine, connected health, e-health, mobile health, and smart health systems [47]. The expansion of portable /wearable (P/W) devices has empowered smart healthcare systems, enabling physicians to collect patient data without geographical limitations and to administer immediate remote treatments [48]. Furthermore, these P/W devices facilitate continuous monitoring of patient health conditions and the collection of relevant data. This information is crucial for alerting medical professionals to critical situations, initiating timely interventions, and providing valuable insights for future medical research.

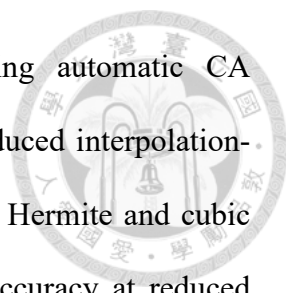
Cardiac arrhythmia (CA) is an early indicator of cardiovascular diseases that pose significant risks to human life [1]. Electrocardiography (ECG), which records the heart's electrical activity, is the primary tool used for assessment. As a non-invasive and cost-effective method, ECG is widely employed in clinical settings to monitor heart function [49, 50, 51]. Diagnosing CA involves analyzing specific waveform features in ECG signals, such as the P wave, QRS complex, and T wave. Physicians typically examine these features manually to detect the presence of CA. However, reviewing long-term ECG recordings is time-consuming and labor-intensive. Therefore, developing an automatic CA classification system is essential to support clinical diagnosis.

Researchers have proposed cloud-based CA monitoring systems that integrate cloud services with advanced deep learning (DL) models within smart healthcare



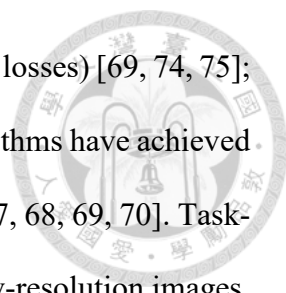
frameworks. These systems use P/W devices to gather ECG data and offload intensive processing tasks to cloud-based DL engines for CA classification [52, 53]. However, continuous sampling and analysis required for long-term CA monitoring can lead to excessive power consumption in P/W devices integrated with consumer electronics (CE). Factors influencing power efficiency include the data source, sampling rate, electronic components, and transmission protocols. To mitigate power consumption, common strategies involve reducing the number of monitoring sources and lowering the sampling frequency. The latter is particularly significant, as high-frequency data sampling can account for up to 90% of a health monitoring system's power usage [54, 55, 56].

To enable long-term health monitoring services, various strategies have been employed to enhance power efficiency in automatic CA classification systems [55, 56, 57]. Despite these advancements, reducing the number of monitoring sources remains a straightforward approach to conserving energy, for example, limiting ECG recordings from a full 12-lead setup to a single lead [38]. In prior research, I conducted an empirical analysis of 12-lead ECG data to identify the most suitable single-lead source for DL-based CA classification [58]. Another effective strategy is lowering the sampling rate of the P/W device, which significantly cuts power consumption. However, decreasing the sampling frequency often leads to information loss, resulting in lower-resolution signals. Such low-resolution data can negatively impact CA classification performance because essential ECG features may be diminished, hindering the monitoring system's ability to automatically detect CA. While some studies have examined ECG sampling frequencies, most focus on heart rate variability, a non-invasive marker of autonomic nervous system activity [59, 60].



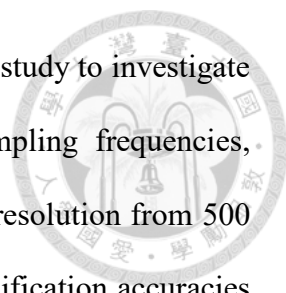
A limited number of studies have focused on developing automatic CA classification methods for low-resolution signals. Sidek et al. introduced interpolation-based signal enhancement techniques, specifically, piecewise cubic Hermite and cubic spline interpolation, to improve signal quality and classification accuracy at reduced sampling frequencies [61]. Their approach achieved CA detection accuracies of approximately 99%, compared to about 97% without signal enhancement [61]. Mathews et al. proposed a DL-based CA classification framework employing Restricted Boltzmann Machines and deep belief networks to enhance classification performance at lower sampling rates, exploring frequencies from 72 Hz to 360 Hz [62]. Their findings confirmed the effectiveness of DL approaches for low-resolution CA classification. However, these studies did not address automatic CA classification at extremely low sampling rates, even though P/W-based monitoring systems with very low sampling frequencies have been developed for other long-term healthcare applications like daily activity recognition and fall detection [55, 63].

To address information loss at very low sampling frequencies, image super-resolution (SR) techniques may offer a viable solution for enhancing low-resolution ECG signals in automatic CA classification. Recently, image SR has gained significant attention and study [64, 65, 66]. The core concept involves reconstructing high-quality information from low-resolution data to facilitate better resolution for subsequent tasks. Specifically, image SR has been pivotal in image processing within computer vision, aiming to generate high-resolution images from low-resolution inputs, a challenging endeavor due to its mathematically ill-posed nature [67, 68, 69, 70]. The rapid advancement of DL techniques has spurred researchers to tackle image SR challenges. Generally, DL-based SR algorithms consist of three components: network architectures (e.g., convolutional layers, residual connections, recursive layers, upsampling layers) [71,



72, 73]; loss function designs (e.g., pixel-wise, perceptual, adversarial losses) [69, 74, 75]; and learning principles and strategies [76, 77, 78, 79]. Some SR algorithms have achieved state-of-the-art performance [80, 81], benefiting other vision tasks [67, 68, 69, 70]. Task-driven SR can significantly improve object detection accuracy on low-resolution images, positively impacting visual recognition [68]. Similarly, in signal processing, audio SR involves increasing the sampling rate of low-resolution audio signals. Inspired by advancements in learning-based algorithms for speech recognition [82, 83, 84], music generation [85, 86], and other fields [87], substantial progress has been made in audio SR through DL, allowing direct modeling of raw signals in the time domain and effectively capturing long-term dependencies [88].

Despite the success of SR techniques in visual tasks, to my knowledge, no prior research has applied SR to ECG signals for automatic CA classification. In this study, I propose and examine a novel DL-based ECG signal SR framework (SRECG) functioning as a processor. This framework aims to address the challenges of low-resolution ECG data from P/W devices, enhancing classification accuracy. The proposed SRECG improves low-resolution ECG signals using composite training loss in the cloud, which enhances the performance of the DL-based high-resolution multiclass classifier (HMC) for CAs. As illustrated in Figure 4, incorporating SRECG as a computing processor allows ECG signals to be sampled at extremely low frequencies, reducing both energy consumption and data transmission size compared to standard smart healthcare workflows. Since both the processor and HMC reside in the cloud, their computational demands are separated from those of the P/W device. Additionally, I conducted a comprehensive analysis of the impact of sampling frequencies on DL-based classification for CAs across 12 different ECG lead sources.



This work makes three primary contributions: (1) it is the first study to investigate ECG signals in CA classification across different leads and sampling frequencies, demonstrating the information loss associated with reducing signal resolution from 500 Hz to 1 Hz; (2) the proposed SRECG framework enhances the classification accuracies of the HMC when processing low-resolution ECG signals from P/W devices at low sampling frequencies, addressing energy consumption and data transmission issues in CE; (3) training SRECG with a composite loss function yields better recovery of HMC classification accuracies compared to using a single loss function. I believe that the proposed SRECG framework can be readily adapted for commercial ECG diagnostic applications, where data storage is a significant concern.

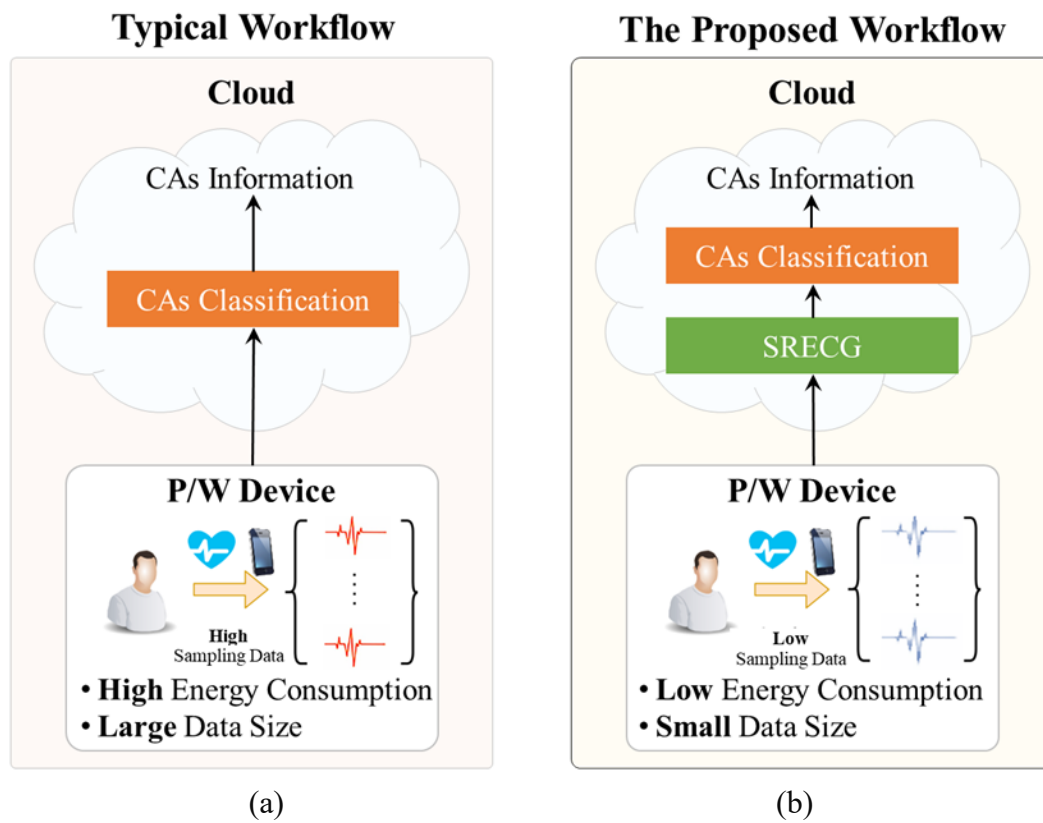


Figure 4 Comparison of DL workflows for a P/W device

(a) Traditional workflow. (b) Proposed workflow with the integration of the SRECG framework.

2.2 Methods

Data acquisition



In this study, my goal was to enhance ECG signals from 12 different lead sources at reduced sampling frequencies. To accomplish this, I utilized the open-source training dataset, benchmark ECG model architecture, and performance evaluation metrics provided by the China Physiological Signal Challenge 2018 (CPSC2018), an international competition focused on automatic ECG classification for CAs [22]. Detailed information about the CPSC2018 ECG training set is available in Chapter 1 (see Table 1). Unlike previous CA prediction studies that often relied on publicly available ECG databases with only 2-lead signals and limited recordings, such as the MIT-BIH Arrhythmia Database with 48 recordings [15, 16], the CPSC2018 ECG training set comprises 6,877 recordings of 12-lead ECG signals with durations ranging from seconds to minutes. Each recording is standardized using Z-score normalization and labeled into one of nine categories: a normal rhythm or one of eight abnormal CAs, including atrial fibrillation (AF), left bundle branch block (LBBB), right bundle branch block (RBBB), first-degree atrioventricular block (I-AVB), premature atrial contraction (PAC), premature ventricular contraction (PVC), ST-segment elevation (STE), and ST-segment depression (STD). Among these recordings, 476 have been assigned more than one CA-type label.

To simulate ECG measurements from different lead sources at lower sampling frequencies, I downsampled the CPSC2018 ECG training set, specifically the 6,401 recordings with a single CA label, from 500 Hz to various lower frequencies: 250, 125, 100, 50, 25, 10, 5, 2, and 1 Hz. The signals from each of the 12 leads were extracted separately for analysis. The low-resolution signals were obtained by applying an integer

factor n , which represents the number of time units skipped when collecting the signal. In this study, I used $n = 1, 3, 4, 9, 19, 49, 99, 249$, and 499 to achieve the aforementioned sampling frequencies. To address recordings of unequal lengths, I applied zero-padding at the beginning of any recording shorter than the maximum duration in the dataset. No additional preprocessing steps were performed on the recordings.

Benchmark ECG model architecture

My proposed ECG model architecture in Chapter 1, which secured the first place in the CPSC2018 competition, was chosen as the reference model to create my HMCs and to assess CA classification performance at different sampling frequencies from different lead sources (see Figure 5). The model contains 28,035 trainable parameters, ensuring it is sufficiently compact to explore statistical relationships using limited computing resources [58].

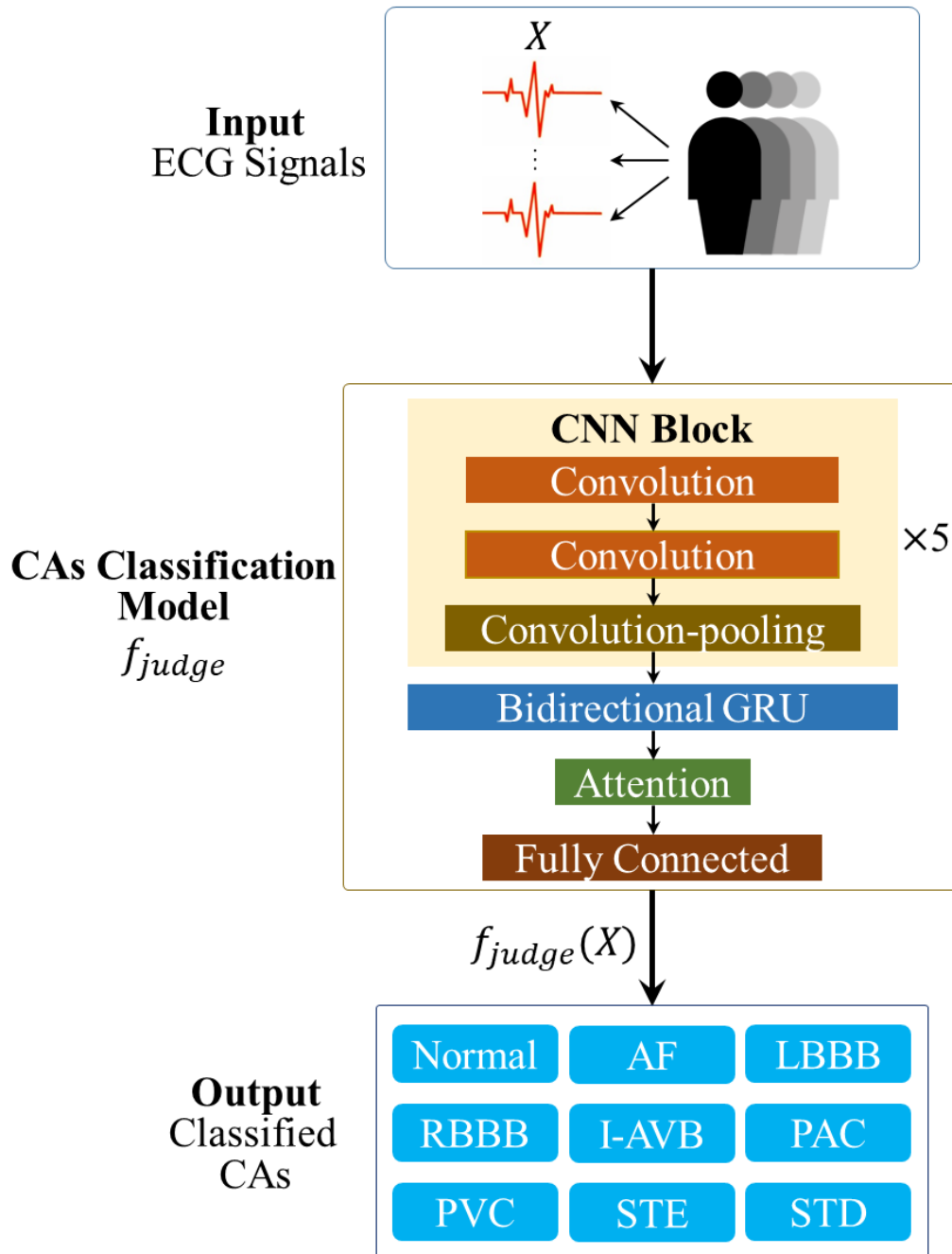


Figure 5 Architecture of benchmark DL network for single-lead ECG

The architecture includes multiple layers and blocks, shown within rectangular boxes. The label “×5” indicates that five convolutional neural network (CNN) blocks are sequentially connected before linking to the bidirectional gated recurrent unit (GRU) layer. The output layer at the bottom presents the model's predicted probabilities for nine different CAs. The CA with the highest probability is the model's predicted classification for the given ECG signal.

Performance evaluation

Similar to the method mentioned in Chapter 1, the dataset was randomly divided into 10 equal parts to establish an 8-1-1 scheme for training, validation, and testing in machine learning. To address label imbalance, I employed stratified sampling to maintain proportional representation of each CA type. With this data split, the model was trained for 100 epochs (each epoch being one complete cycle through the training dataset), generating 100 models (one per epoch). The model that performed best on the validation set was selected as the optimal model for this training process. This best model was then used to compute the F1-score on the test set. This procedure was repeated 10 times to complete the 10-fold training, resulting in 10 best models for each fold. The median F1-score for the overall performance, and for each CA label, including the normal type, across the 10 test sets was calculated using the Scikit-learn package [89]. The overall performance (overall F1-score) was obtained by averaging the F1-scores across all CA labels. This training utilized the ADAM optimizer [32].

Traditional spline interpolation for up-sampling

Spline interpolation is a valuable technique for enhancing low-resolution ECG signals obtained at low sampling frequencies to achieve optimal CA classification. Originally, a spline referred to a flexible strip used to draw smooth curves through a series of points.

In numerical analysis, spline interpolation is realized by fitting piecewise low-degree polynomials between local data pairs, rather than fitting a single high-degree polynomial to all data points. For example, fitting polynomials between each pair of points reduces oscillations at the edges of intervals compared to using a single cubic

polynomial. The spline interpolation curves are defined by a set of polynomials $\{q_i\}$ satisfying $y = q_i(x)$ for data pairs (x_{i-1}, y_{i-1}) and (x_i, y_i) , where $i = 1, 2, \dots, n$. There are n polynomials and $n + 1$ knots in this case, with the first polynomial ending at (x_0, y_0) , and the last at (x_n, y_n) .

Proposed SRECG framework

Instead of using traditional spline interpolation, I propose a SR-based method called SRECG, illustrated in Figure 6. The SRECG architecture is adapted from the well-known SRResNet [77], modified to include one convolutional layer and 16 residual blocks [90]. Each residual block contains two convolutional layers with residual connections, followed by four convolutional layers for up-sampling and signal reconstruction.

All convolutional layers have a kernel size of 16 and a kernel number of 64, except for the last two up-sampling convolutional layers, which have kernel numbers of 320 (for five times up-sampling) and 128 (for two times up-sampling), respectively. Within the residual blocks, batch normalization is applied to adjust and rescale the inputs from the two convolutional layers (no downsampling layers are used). LeakyReLU activation functions are employed between convolutional layers before connecting to residual blocks and within up-sampling layers. Finally, batch normalization is used in the convolutional layer following the residual blocks. This architecture was implemented using the Keras package supported by TensorFlow on GPUs [33, 91].

As depicted in Figure 7, low-resolution ECG data are input into this architecture for training using the same 10-fold cross-validation procedure. During training, the model aims to reconstruct an enhanced ECG signal from the low-resolution input. Unlike

conventional super-resolution tasks that focus solely on increasing image resolution by minimizing mean square error (MSE) differences, I intended to integrate the CA classification results from another DL-based model, i.e. HMC, to assist in the process.

To achieve this, I designed a customized joint loss function to simultaneously optimize ECG signal reconstruction and predictability for CA classification. This ensures that the final enhanced output can be subsequently fed into the HMC for CA judgment. It is important to note that the HMC was independently pre-trained with high-resolution ECG signals using the same 10-fold cross-validation scheme.

The customized joint loss L_J for the SRECG is defined as:

$$L_J(\hat{y}, y, \hat{z}, z) = \gamma L_R(\hat{y}, y) + (1 - \gamma) L_C(\hat{z}, z), \gamma \in [0, 1], \quad (8)$$

where L_R denotes the MSE regression loss between the high-resolution ECG signals y and the enhanced signals $\hat{y}$:

$$L_R(\hat{y}, y) = \frac{1}{M} \sum_{i=1}^M (\hat{y}_i - y_i)^2, \quad (9)$$

where M denotes the total number of time steps of y and $\hat{y}$. The categorical cross-entropy (CCE) loss L_C measures the difference between the true labels from the HMC $z = (z_1, \dots, z_C) \in [0, 1]^C$ and the predicted labels $\hat{z} = (\hat{z}_1, \dots, \hat{z}_C) \in (0, 1]^C$:

$$L_C(\hat{z}, z) = -\sum_{k=1}^C z_k \log(\hat{z}_k), \quad (10)$$



and, in my model, I set $z = f_{judge}(y)$ and $\hat{z} = f_{judge}(\hat{y})$, where f_{judge} is the fixed HMC model providing CA probabilities. Thus, the joint loss becomes:

$$L_J(\hat{y}, y) = L_J(\hat{y}, y, f_{judge}(\hat{y}), f_{judge}(y)), \quad (11)$$

and this loss function is determined once the enhanced ECG signals $\hat{y}$ are generated. It is also implied that all the above losses are averaged over the number of samples for normalization during actual implementation.

By integrating signal reconstruction and classification performance into a unified loss function, the proposed SRECG framework aims to enhance low-resolution ECG signals effectively, thereby improving the accuracy of automatic CA classification in smart healthcare systems utilizing P/W devices.

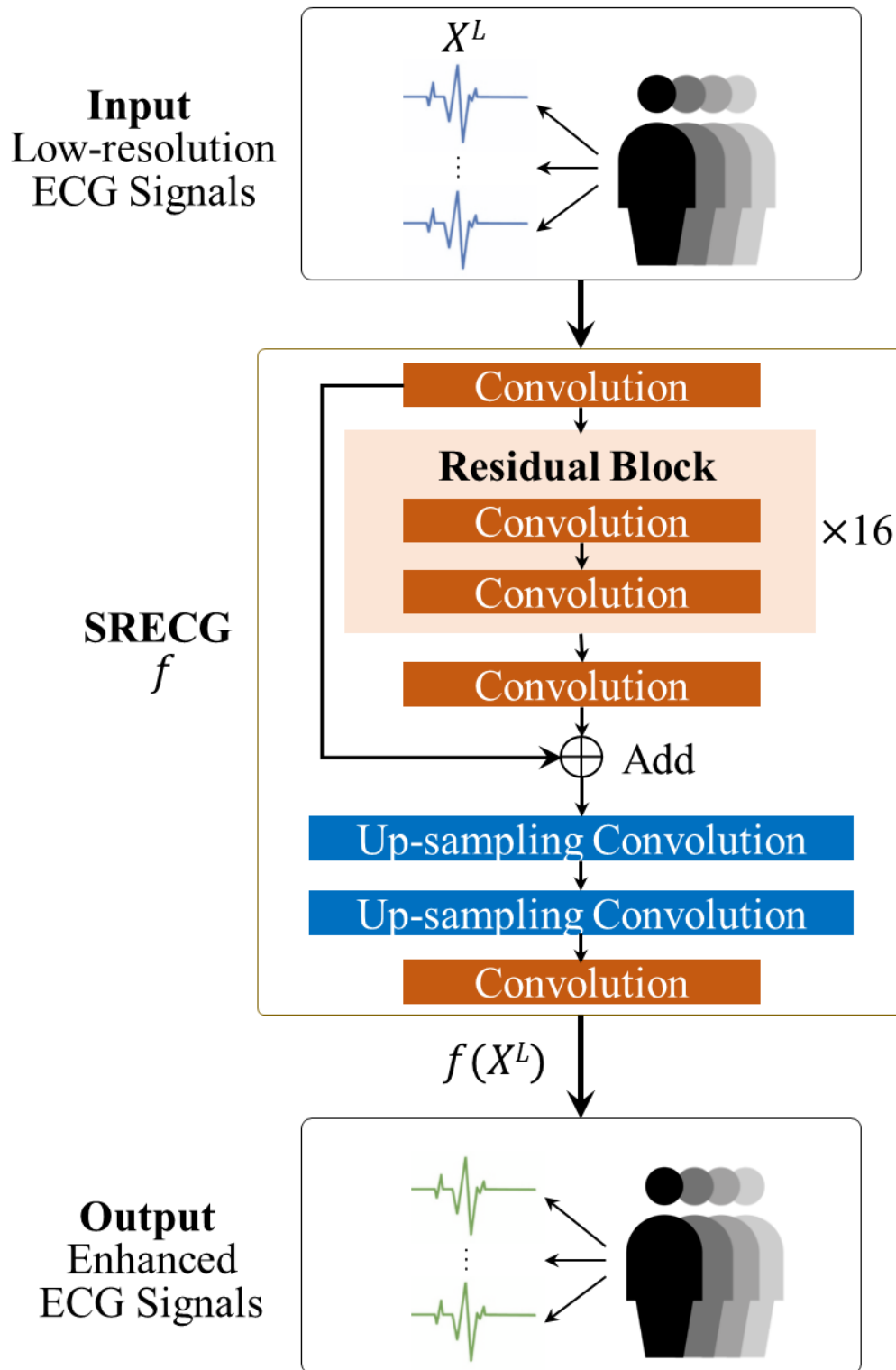


Figure 6 Architecture of proposed SRECG network for enhancing ECG

The input, represented as X^L , refers to the low-resolution ECG signals, while the output, $f(X^L)$, corresponds to the enhanced ECG signals. Layers and blocks are depicted as rectangles. The " $\times 16$ " notation indicates that 16 residual blocks are sequentially connected before being merged with the convolution layer.

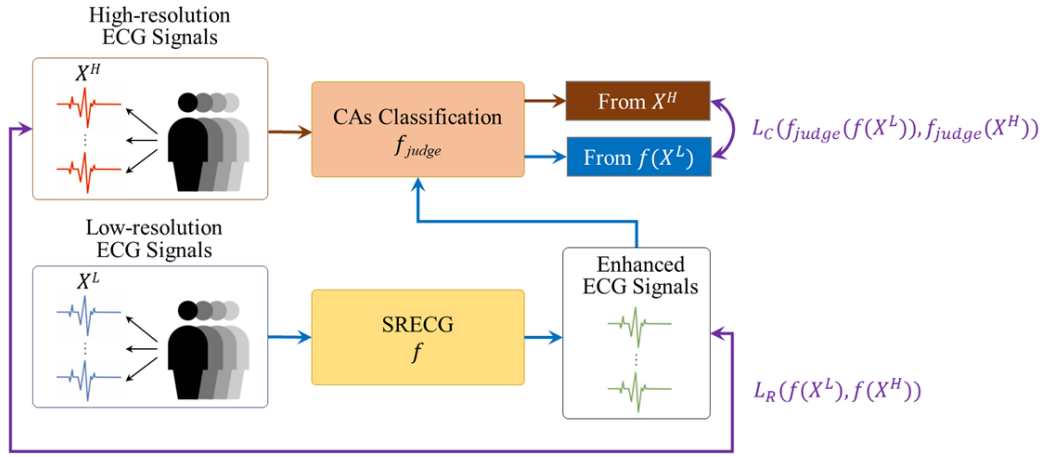


Figure 7 Framework for training proposed SRECG network

The SRECG network, denoted as f , takes low-resolution ECG signals (X^L) as input and generates enhanced ECG signals ($f(X^L)$) as output. This process uses a combined loss function (L_J), which integrates the regression loss (L_R) and the CA classification loss (L_C) as described in Equation 8. The DL-based CA classification and SRECG network architectures are shown in Figures 5 and 6, respectively.

2.3 Results

Effects of sampling frequency on CA classification



I examined how varying sampling frequencies, ranging from 500 Hz down to 1 Hz, influence the performance of a DL-based multiclass CA classifier using ECG signals from 12 distinct lead sources. This analysis was conducted under a 10-fold cross-validation procedure, as illustrated in Figure 8. My preliminary findings revealed that at a sampling frequency of 25 Hz, there was significant information loss, which adversely affected overall CA classification accuracy. In contrast, a sampling frequency of 250 Hz retained most of the essential information necessary for accurate classification. Based on these observations, I employed traditional spline interpolation methods and my proposed SRECG model to enhance the 25 Hz input data, aiming to improve the overall F1-score of HMC trained at 250 Hz.

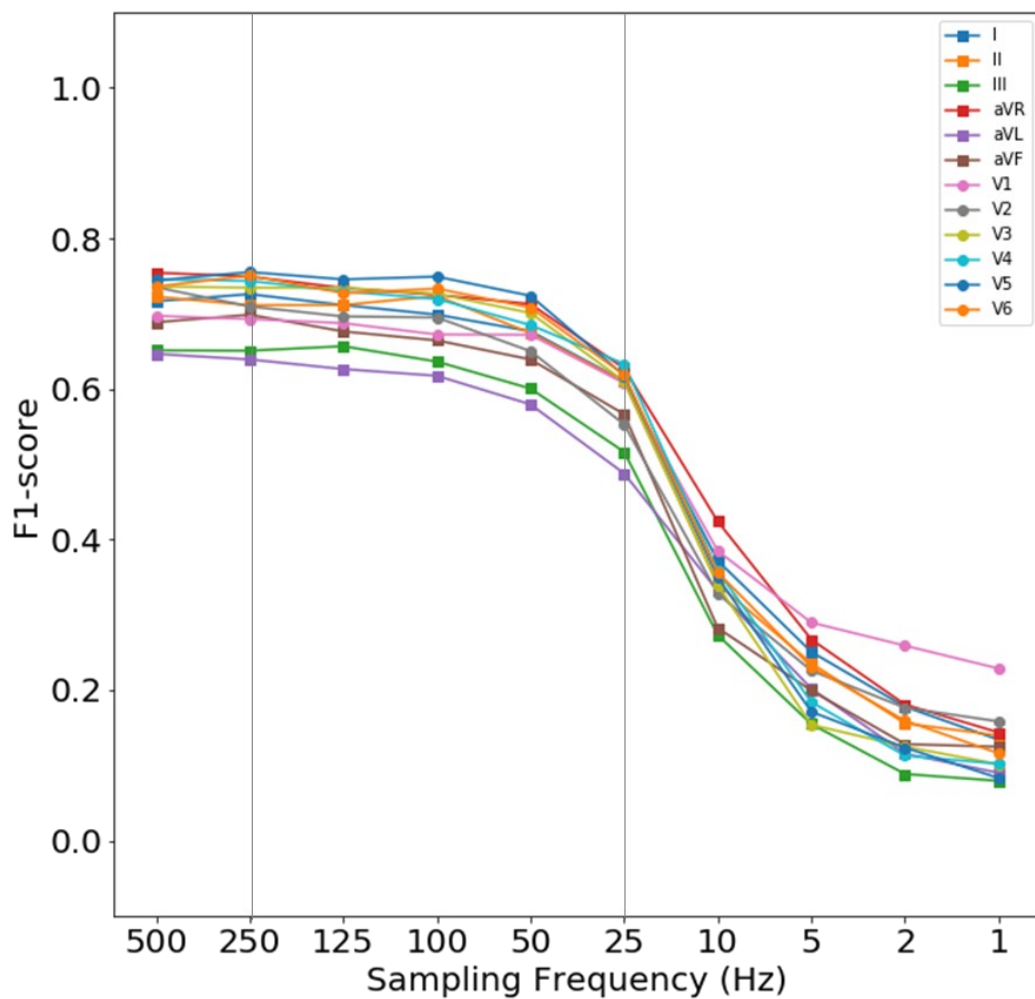


Figure 8 Predictive power changes of CA classifiers on downsampled ECG data

The plot shows the median overall F1-scores of the CA classification models, represented by 12 colored lines. Each line reflects the model's performance across 10-fold tests, where the model was trained on ECG data from various leads and at different sampling frequencies (500, 250, 125, 100, 50, 25, 10, 5, 2, and 1 Hz).

Effects of traditional spline interpolation methods

I explored four standard spline interpolation techniques for one-dimensional signals: using polynomials of degree zero (I_0), one (I_1), two (I_2), and three (I_3). Table 12 presents a comparison of the median overall F1-scores for CA classification using models C_N (trained at 25 Hz), C_P (trained at 250 Hz), and HMC (trained on 250 Hz ECG data) with

25 Hz input enhanced via spline interpolation. Across different ECG lead sources, I observed that none of the interpolation methods resulted in F1-scores exceeding those of C_N . Figure 9 illustrates the signal reconstruction performance. Specifically, the ECG waveforms of a lead II heartbeat showed that I_1 provided superior reconstruction, achieving the lowest MSE of 1.11% and the highest Pearson correlation coefficient (R) of 72.12%, compared to I_0 (MSE: 3.71%; R: 24.55%), I_2 (MSE: 1.26%; R: 71.74%), and I_3 (MSE: 1.32%; R: 70.73%).

From Table 12 and Figure 9, I concluded that although traditional spline interpolation methods can effectively reconstruct ECG signals, they do not sufficiently preserve CA classification performance (as measured by F1-scores). This suggests that traditional spline interpolation fails to recover enough CA-related information during ECG signal enhancement.

Table 12 Median overall F1-scores (%) of HMCs trained at 250 Hz with different leads and spline interpolation degrees

Leads / degrees	C_N	I_0	I_1	I_2	I_3	C_P
<i>I</i>	61	56	47	43	39	73
<i>II</i>	61	57	47	29	24	71
<i>III</i>	52	50	45	36	31	65
<i>aVR</i>	63	56	41	38	36	75
<i>aVL</i>	49	49	42	38	36	64
<i>aVF</i>	57	56	52	35	29	70
<i>V1</i>	61	59	50	45	42	69
<i>V2</i>	55	34	32	23	18	71
<i>V3</i>	61	33	22	18	16	73
<i>V4</i>	63	39	20	13	12	74
<i>V5</i>	62	45	18	14	12	76
<i>V6</i>	62	53	31	26	22	75

The scores shown in bold are closest to their corresponding C_P values.

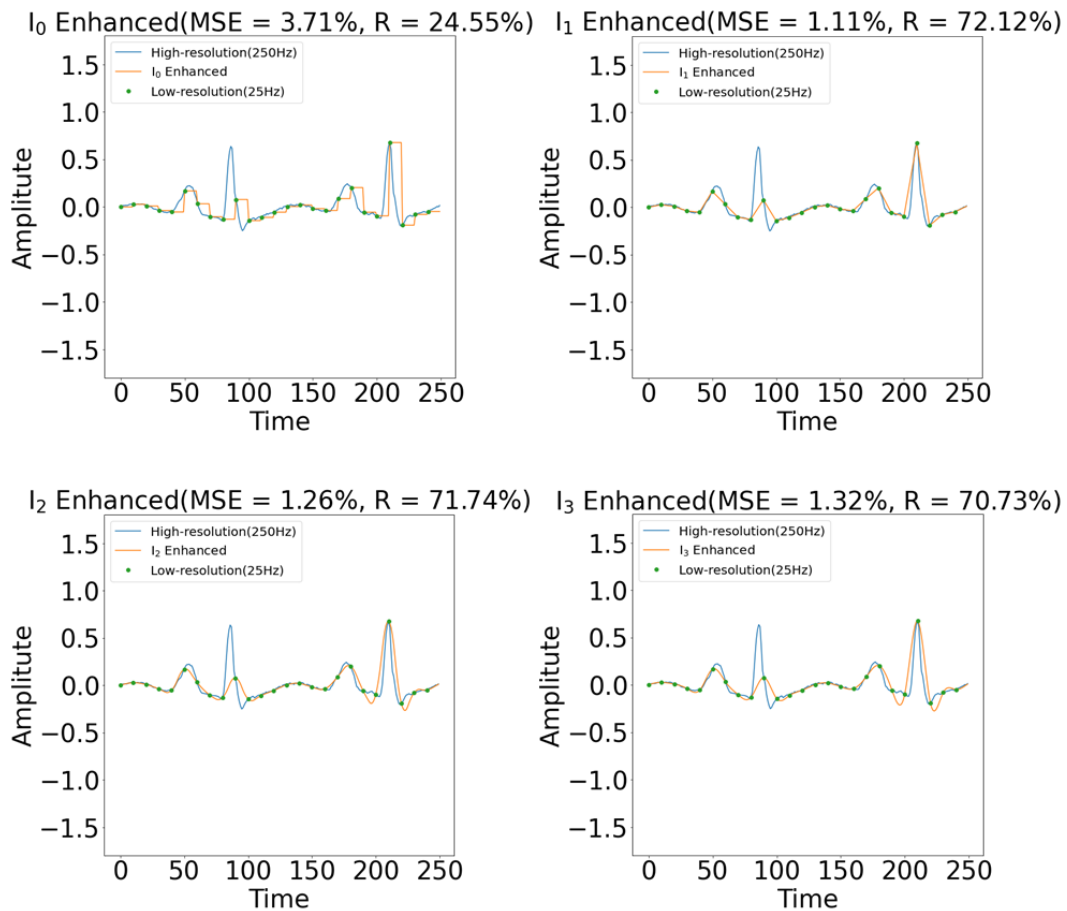
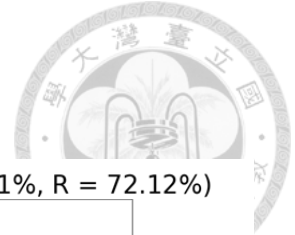
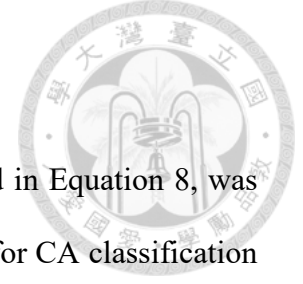


Figure 9 Enhanced ECG signals visualization of different spline interpolation degrees

ECG waveforms for a lead II heartbeat, over a 1-second interval, are shown using four spline interpolation degrees: I_0 , I_1 , I_2 , and I_3 . The x-axis represents time, while the y-axis indicates ECG amplitude. The subtitles provide the MSEs and R values, which reflect the reconstruction quality for each 1-second period.

Effects of joint training losses in SRECG



The composite loss function of the SRECG model, as defined in Equation 8, was designed to optimize two separate objectives simultaneously: CCE for CA classification and MSE for signal reconstruction. The balance between these two losses is controlled by a weighting constant γ within the range $[0,1]$. I illustrate three specific loss conditions in Equation 8: pure regression loss (L_R) when $\gamma = 1$, pure classification loss (L_C) when $\gamma = 0$, and a balanced joint loss (L_J) when $\gamma = 0.5$.

Table 13 compares the median overall F1-scores for models C_N (trained at 25 Hz), C_P (trained at 250 Hz), and HMC (trained on 250 Hz ECG data) when the 25 Hz input is enhanced by SRECG using different loss configurations. Across various ECG lead sources, the highest F1-scores in the enhanced CA classification tasks were generally achieved under the L_J condition, except for leads V1, V2, and V3. In Figure 10, the ECG waveforms of a lead II heartbeat demonstrate that: L_R provides superior reconstruction accuracy, yielding the lowest MSE (0.14%) and highest R (97.10%); L_C may not show clear reconstruction capability (MSE: 35.05%; R: -2.72%) nor resemble typical ECG signals, yet these enhancements result in higher F1-scores than those obtained with L_R (as shown in Table 13), suggesting that machine learning models may capture patterns beyond human intuition; L_J exhibits characteristics of both L_R and L_C enhancements (MSE: 5.40%; R: 50.49%). Incorporating L_C acts as a constraint that both penalizes and incentivizes the model, preventing overfitting and focusing on essential latent features.

Table 13 Median overall F1-scores (%) of HMCs trained at 250 Hz with different leads and SRECG losses

Leads / losses	C_N	L_R	L_C	L_J	C_P
<i>I</i>	61	60	65	66	73
<i>II</i>	61	64	64	66	71
<i>III</i>	52	51	56	58	65
<i>aVR</i>	63	68	71	71	75
<i>aVL</i>	49	48	55	56	64
<i>aVF</i>	57	59	61	62	70
<i>V1</i>	61	62	65	64	69
<i>V2</i>	55	62	64	61	71
<i>V3</i>	61	65	67	65	73
<i>V4</i>	63	65	65	68	74
<i>V5</i>	62	65	70	70	76
<i>V6</i>	62	57	69	69	75

The scores shown in bold are closest to their corresponding C_P values.

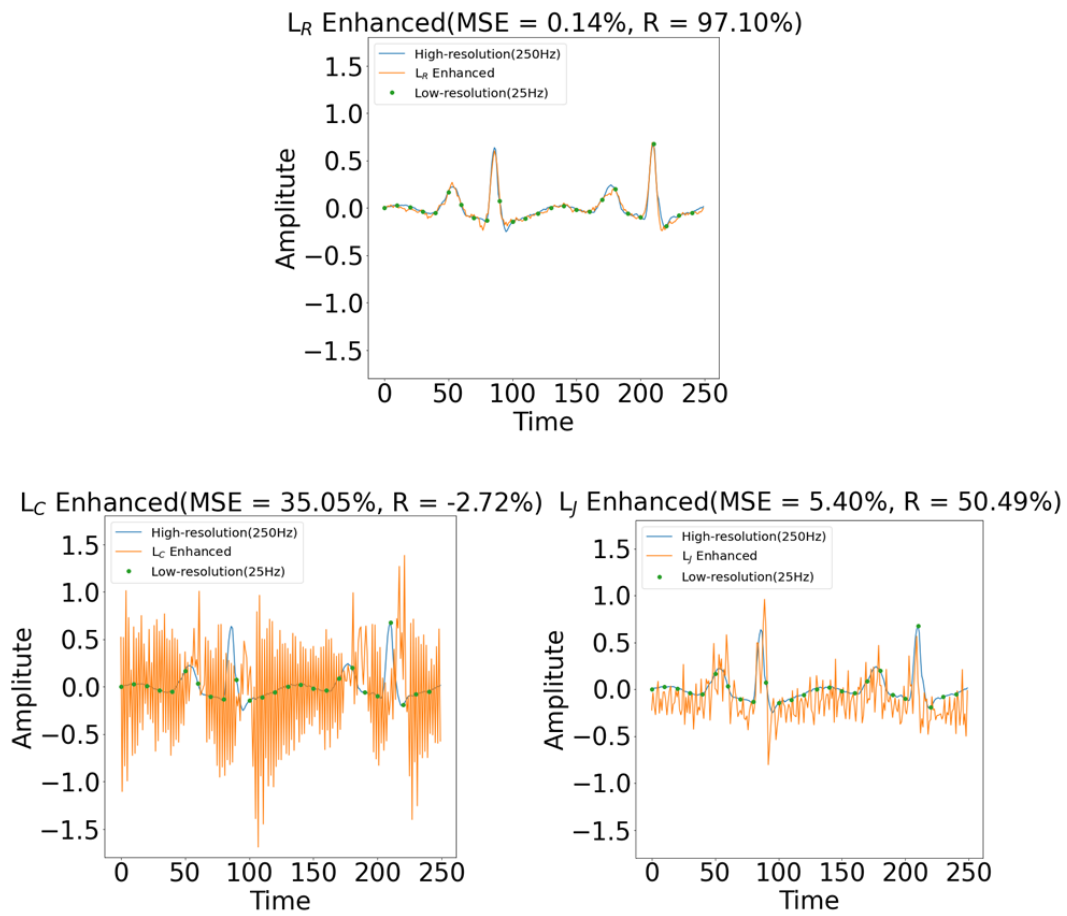


Figure 10 Enhanced ECG signals visualization of different SRECG losses

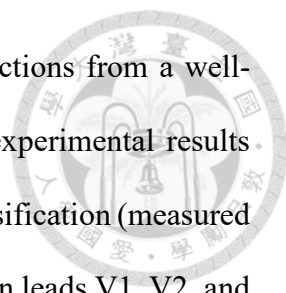
ECG waveforms for a lead II heartbeat, over a 1-second interval, are shown using three different loss functions: L_R , L_C , and L_J . The x-axis represents time, while the y-axis indicates ECG amplitude. The subtitles provide the MSEs and R values, which reflect the reconstruction quality for each 1-second period.

2.4 Discussion

CE play a pivotal role in enhancing the quality of life through smart healthcare. P/W devices integrated with CE for medical purposes have provided straightforward methods for individuals to monitor their health in daily life, such as through bracelets and wrist gadgets [92]. Various studies on smart healthcare systems have utilized these devices to assist visually impaired individuals and to prevent falls among the elderly [93, 94, 95]. Additionally, these devices enable long-term monitoring of physiological signals for clinical diagnosis and evaluation, including ECG, electroencephalography [96], and near-infrared spectroscopy [97].

The integration of cloud-based DL algorithms with P/W devices is becoming increasingly prevalent in smart healthcare systems. However, limitations associated with current hardware capabilities, such as battery life and computational power, pose challenges for long-term ECG monitoring systems implemented in CE. Higher frequencies used to record ECG signals demand more power, which presents an energy consumption constraint. To address this issue, my study aimed to restore the predictive power of the HMC using low-sampling-frequency inputs through my proposed DL-based SRECG framework, a novel SR method. The SRECG is tailored for P/W devices within a cloud-based smart healthcare workflow (see Figure 4).

In traditional SR tasks, the loss function typically involves only the L_R . My experiments revealed that using L_R alone was insufficient to enhance the CA classification accuracies of the HMCs. Therefore, I introduced a teacher–student learning



approach by incorporating the categorical CCE loss, utilizing predictions from a well-trained HMC as the judge model for CA classification [98]. The experimental results indicated that my joint loss L_J achieved better performance in CA classification (measured via F1-scores) across most ECG lead sources, with some exceptions in leads V1, V2, and V3 (Table 13). Interestingly, the best-reconstructed ECG signals (those with the lowest MSEs and highest R_s) did not correspond to the highest F1-scores when applied to HMCs for CA classification. This may be because my proposed model, which combines reconstruction and classification tasks, focuses more on reconstructing features or signals that aid in ECG classification. Consequently, it can achieve higher model F1-scores without necessarily having the lowest MSEs and highest R_s . A similar phenomenon is observed in automatic speech recognition (ASR) in noisy environments [99, 100, 101], where removing background noise through speech enhancement models improves signal quality but often degrades ASR performance due to distortions introduced by the enhancement process.

Clinically, DL-based ECG diagnosis of CAs has demonstrated its potential and importance in enhancing diagnostic accuracy compared to general physicians and cardiologists [10, 11]. Different ECG leads offer distinct features that aid medical professionals in recognizing specific CA types [6, 7, 8, 9, 102]. With growing concerns about medical data privacy [103], it is essential to develop encryption methods to render data unrecognizable, complying with new privacy-preservation regulations such as the General Data Protection Regulation (GDPR) [104]. Interestingly, as observed in Figure 10, I found that ECG signals enhanced using the L_C can produce indistinct ECG

waveforms while simultaneously improving predictive power for CA classification (Table 13). This finding may offer an alternative approach to encrypt ECG data.

As machine learning and DL techniques continue to evolve and become more accessible through P/W devices, ECG will increasingly serve as repositories of physiological information. DL models can be employed to extract information for overlooked or unintended predictions, such as determining age, sex and mortality [105, 106]. Therefore, it is promising to anticipate that ECG will offer greater diagnostic capabilities in numerous forthcoming applications on P/W devices.

2.5 Conclusion

To facilitate automatic clinical diagnosis and long-term smart healthcare services on P/W devices integrated with CE, this study introduces a novel DL-based SRECG as a cloud processor. This processor enhances low-resolution ECG signals sampled at low frequencies for CA classification using the HMC, achieving improved accuracy. In future research, I plan to further investigate the effects of training the SRECG with joint loss functions using different lead sources and their combinations, as well as exploring various preprocessing methods for the SRECG. Given that DL-based ECG signal diagnosis holds great potential for deciphering other physiological conditions, it is worthwhile to apply the SRECG for P/W devices in other healthcare diagnostic criteria, such as cognitive impairment and heart failure, in addition to CA.



Chapter 3 Multimodality Adaptation



3.1 Introduction

Despite electrocardiography (ECG) signals provide profound information for medical diagnosis, voice signals are still the most potential biomarkers for various health conditions. Early signs of diseases like Parkinson's can manifest as changes in voice [107], and artificial intelligence (AI) techniques, including deep learning (DL) algorithms, can enhance the accuracy of detecting such changes [108]. AI is also employed in identifying mood disorders [109] and assessing voice quality [110]. Historically, AI algorithms have been used to automatically detect glottic pathologies using traditional classifiers like hidden Markov models (HMM) [111], support vector machines (SVM) [112], Gaussian mixture models (GMM) [113], artificial neural networks (ANN) [114], and more recently, DL methods [115].

When it comes to clinical concerns for patients with voice disorders, there are two primary considerations: first, determining whether the patient's voice is normal or abnormal; second, if the voice is abnormal, assessing whether this indicates a glottic neoplasm. Most existing AI models focus on the binary classification of voice data as either normal or abnormal [116]. Only a handful of studies have attempted to differentiate between various voice disorders, and these have been limited to benign lesions [117], cases following neoplastic surgery [118], studies with small sample sizes [119], or comparisons between cancer patients and healthy controls [120]. It was only recently that a pilot study showed deep neural networks could detect glottic neoplasms and

preneoplastic conditions in patients matched for age and sex with benign voice disorders, achieving accuracies between 72% and 86% depending on the used thresholds for detection [121]. Building on this initial proof of concept, the current study utilizes two extensive public voice databases to validate the effectiveness of AI models in detecting glottic neoplasms.

3.2 Methods

Data acquisition

I sourced voice samples from the publicly accessible Far Eastern Memorial Hospital Voice Database (FVD) (<http://www.femh-voice-database.org>). As detailed by Wang et al. [122], this database comprises audio recordings from 2,000 individuals diagnosed with various voice disorders. The research protocol received approval from the Research Ethics Review Committee of Far Eastern Memorial Hospital (FEMH-IRB Nos. 111158-F and 112035-E). Within the FVD, there are recordings from 60 patients with glottic neoplasms, including squamous cell carcinoma, carcinoma in situ, and dysplasia, and 1,940 patients with benign voice disorders (see Table 14). Each voice sample is a 2- to 3-second uncompressed waveform audio file format (WAV) file capturing the sustained vowel sound "ah." Recordings from patients with glottic neoplasms were made before any medical interventions, with malignancies confirmed through pathological examination. The FVD also includes structured medical records containing information such as symptom onset and presentation, occupational voice demands, smoking habits, and alcohol consumption (see Table 15).

For external validation, I utilized voice samples from the Saarbrücken Voice Database (SVD) (<https://stimmdb.coli.uni-saarland.de>) [123]. This online public repository includes data from 23 patients with glottic neoplasms and 1,351 patients with other benign voice disorders. The SVD provides only age and sex information for each individual.



To assess intrinsic adaptivity to ECG signals, I employed the CPSC2018 ECG training set, which consists of 6,877 recordings sampled at 500 Hz and standardized using Z-score normalization. Details of this dataset are outlined in Chapter 1 (see Table 1). Based on my prior analysis of ECG leads, I selected the best-performing single leads, aVR and V1, along with age and sex data, to evaluate the performance of binary classification for cardiac arrhythmia (CA).

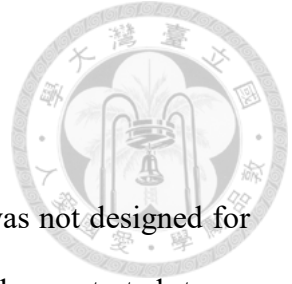
Table 14 Numbers of voice disorders in the FVD

No	Diagnosis	Numbers
1	Polyp	202
2	Nodules	200
3	Atrophy	189
4	Cyst	175
5	Vocal paresis	164
6	Vocal palsy	164
7	Reinke's edema	144
8	Sulcus	127
9	Muscle tension dysphonia	117
10	Presbyphonia	115
11	Vocal process granuloma	86
12	Fibrous mass	84
13	Ulcer	64
14	Spasmodic dysphonia	34
15	Dysplasia	31
16	Laryngeal cancer	29
17	Papilloma	28
18	Tremor	18
19	Scar	15
20	Varix	14



Table 15 Descriptions of structured medical records in the FVD

Variable	Coding	Notions
Sex	1/2	Male/Female
Age	numbers	Years
Narrow pitch range	0/1	No/Yes
Decreased volume	0/1	No/Yes
Fatigue	0/1	No/Yes
Dryness	0/1	No/Yes
Lumping	0/1	No/Yes
Heartburn	0/1	No/Yes
Choking	0/1	No/Yes
Eye dryness	0/1	No/Yes
PND	0/1	No/Yes
Diabetes	0/1	No/Yes
Hypertension	0/1	No/Yes
CAD	0/1	No/Yes
Head and neck cancer	0/1	No/Yes
Head injury	0/1	No/Yes
CVA	0/1	No/Yes
Smoking (status)	0/1/2/3	Never/past/active/e-cigarette
PPD	numbers	pack (of cigarettes) per day
Alcohol	0/1/2	Never/past/active
Frequency (of alcohol consumption)	1/2/3/4	Not/occasionally/weekly/daily
Onset of dysphonia	1/2/3/4/5	Sudden/gradually/on and off/since childhood/other
Noise at work	1/2/3	Not/a little/noisy
Diurnal pattern	1/2/3/4	Worse in AM/worse in PM/similar all day/fluctuating
Occupational vocal demand	1/2/3/4	Always/frequent/occasional/minimal
Voice handicap index - 10	0 to 40	



Proposed multimodal model architecture

Although my proposed ECG model architecture in Chapter 1 was not designed for interpreting and predicting voice signals [58], it could be adapted and demonstrated strong performance in the CODA TB DREAM Challenge (<https://www.synapse.org/Synapse:syn31472953>) to detect pulmonary tuberculosis status based on voice signals standardized by Z-score normalization [124].

To further accommodate the multimodality of FVD and SVD, I developed an adjunct model that incorporates additional demographic data and structured medical records. Categorical variables (e.g., sex) were processed using one-hot encoding, while continuous variables (e.g., age) were standardized using Z-score normalization. The output from the batch-normalized attention layer of the core model, after applying LeakyReLU, was concatenated with the additional demographic and medical record data. This combined data then passed through a series of operations: a fully connected layer, batch normalization, LeakyReLU activation, and another fully connected layer with a hyperbolic tangent activation function to generate adjustments. The final output of the adjunct model was obtained by adding these adjustments to the output of the core model's network (see Figure 11).

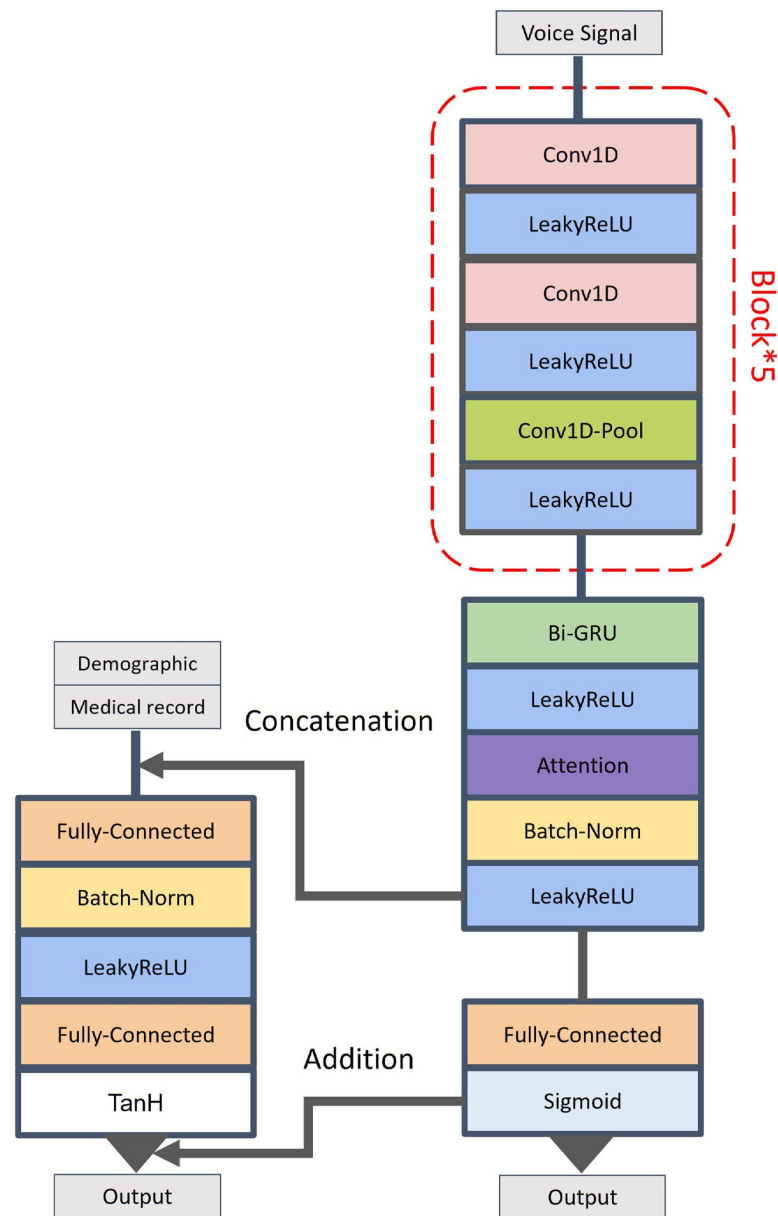


Figure 11 Architecture of proposed multimodal DL network

The diagram illustrates the architecture of the proposed multimodal model, showing multiple layers and connections. The right part processes voice signals through five sequential blocks, each consisting of convolution, LeakyReLU, and convolution-pooling layers. The architecture incorporates bidirectional gated recurrent unit (GRU), attention, and batch normalization. The left part further integrates demographic and medical record data using concatenation and addition operations before outputting predictions.

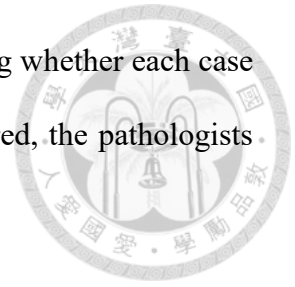


Performance evaluation

Similar to the method mentioned in Chapter 1, I implemented 10-fold cross-validation for training and evaluating the model. The dataset was randomly divided into 10 equal parts to create an 8:1:1 split for training, validation, and testing, respectively. To address class imbalance, I employed stratified sampling to maintain proportional representation of each class (e.g., neoplasm and benign voice disorders). The model was trained over 100 epochs, producing 100 models (one per epoch). The model demonstrating the best performance on the validation set was selected as the optimal model. This optimal model was then applied to the testing data, and its performance was assessed using the area under the receiver operating characteristic (ROC) curve score (AUC). This procedure was repeated 10 times using the ADAM optimizer with the mean square error (MSE) loss function [125], resulting in 10 optimal models for each fold. I aggregated the prediction results from these models on their respective testing data to obtain an overall confusion matrix (i.e., predicted values vs. true labels). For external validation on the SVD dataset, I averaged the outputs of the 10 optimal models trained on the FVD dataset. The model's performance on ECG signals in CA classification was evaluated using the CPSC2018 ECG training set.

Additionally, I conducted a sub-analysis to compare the AI model's predictions with human analysis. I randomly selected 300 benign voice samples from the FVD and 200 from the SVD, including all 60 and 23 cases of neoplasm from the FVD and SVD, respectively. These samples were randomly stratified into three sets: voice only, voice plus demographics, and voice plus medical records. Two speech pathologists

independently assessed these samples in a blinded manner, predicting whether each case was neoplasm or benign. In instances where their predictions differed, the pathologists discussed the case to reach a consensus.



3.3 Results

Internal validation on FVD

My proposed multimodal model achieved a mean AUC of 0.631 using only voice recordings (see Figure 12). When I sequentially incorporated demographic data (illustrated in Figure 13) and medical records (shown in Figure 14), the model's performance improved significantly, reaching AUCs of 0.807 and 0.878, respectively. I then analyzed the confusion matrix derived from the 2,000 cases in the FVD, as presented in Table 16 (left panel). Among the 60 patients diagnosed with glottic neoplasm, the model, which utilized both voice recordings and medical records, correctly identified 47 cases, resulting in a sensitivity of 78.3%. For the 1,940 cases of benign glottic disorders, the model accurately classified 1,583 cases, yielding a specificity of 81.6%. The overall accuracy of the model was calculated to be 81.5%.



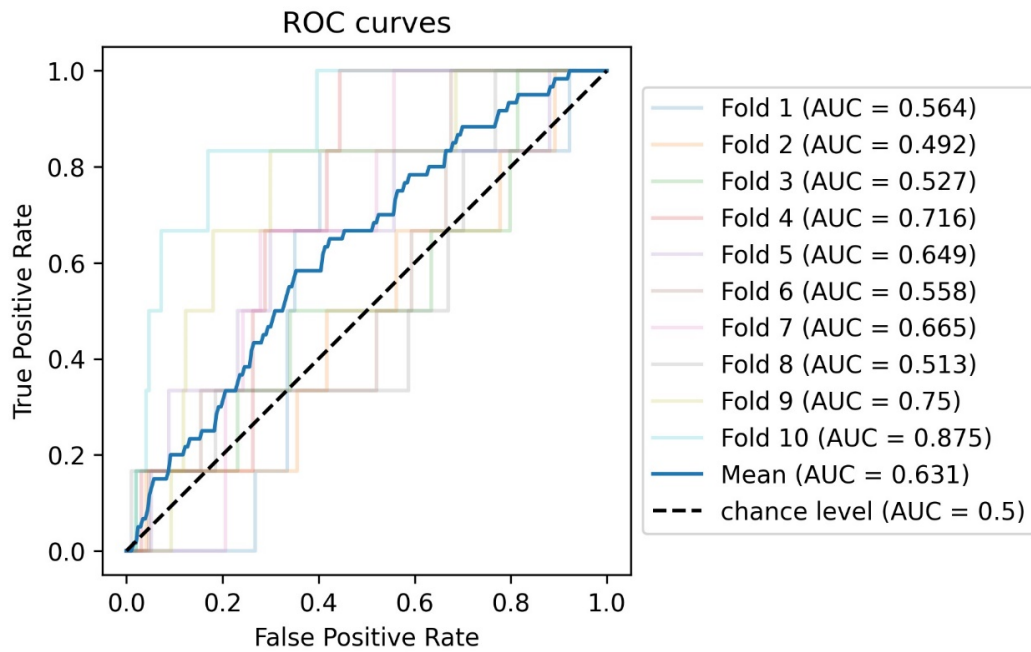


Figure 12 Performance of proposed multimodal model with only voice signals on FVD

The ROC curves illustrate the model's performance across 10 validation folds, with AUCs ranging from 0.492 to 0.875. The mean AUC across all folds is 0.631, indicating poor classification performance. The dashed line represents the chance level (AUC = 0.5).

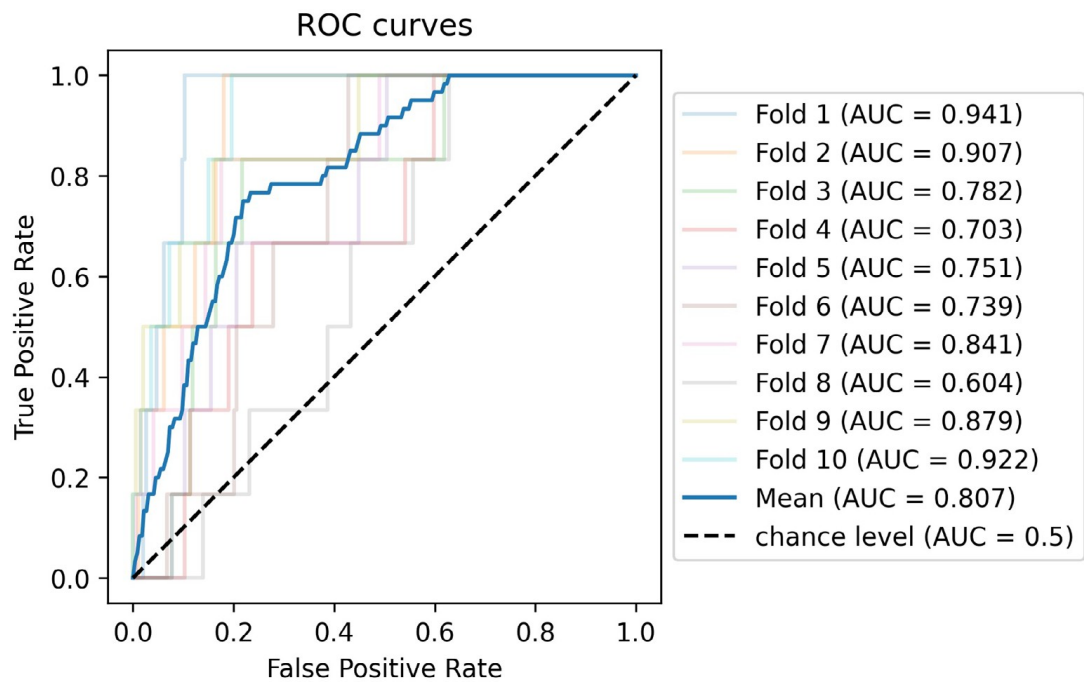


Figure 13 Performance of proposed multimodal model with voice and demographics on FVD

The ROC curves illustrate the model's performance across 10 validation folds, with AUCs ranging from 0.604 to 0.941. The mean AUC across all folds is 0.807, indicating moderate classification performance. The dashed line represents the chance level (AUC = 0.5).

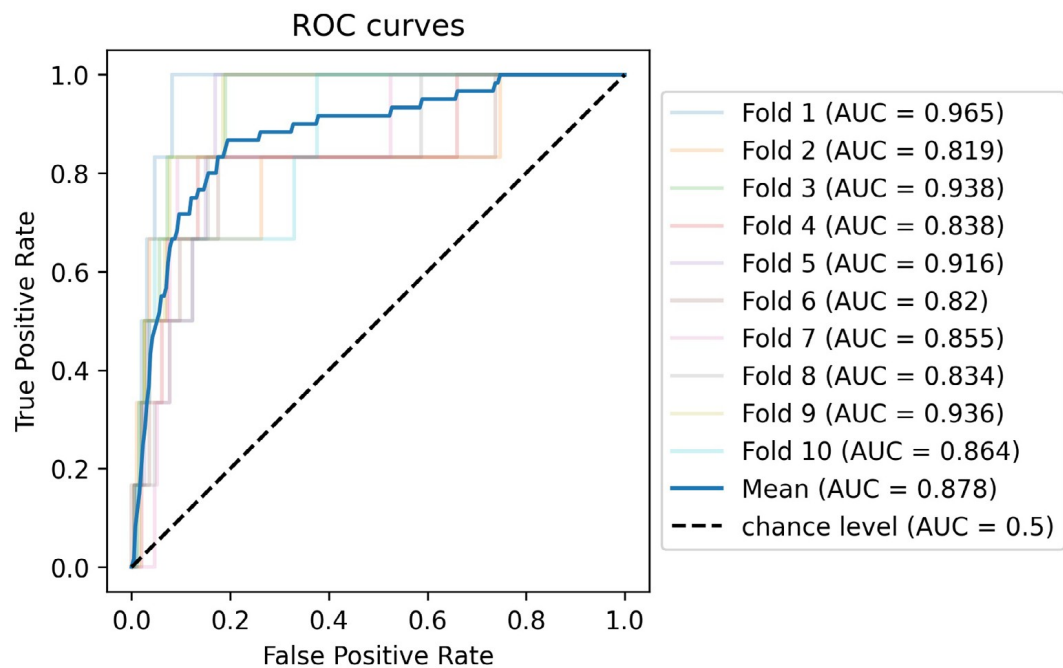


Figure 14 Performance of proposed multimodal model with voice, demographics, and medical records on FVD

The ROC curves illustrate the model's performance across 10 validation folds, with AUCs ranging from 0.819 to 0.965. The mean AUC across all folds is 0.878, indicating strong classification performance. The dashed line represents the chance level (AUC = 0.5).

Table 16 Confusion matrix for FVD (voice + medical records) and SVD (voice + demographics)

		Predicted				Predicted					
		Neoplasm	Benign	Total		SVD	Neoplasm	Benign	Total		
FVD	True label	Neoplasm	47	13	60	True label	Neoplasm	17	6	23	
		Benign	357	1583	1940			Benign	339	992	1331
		Total	404	1596	2000			Total	356	998	1354

External validation on SVD

To validate the model externally, I employed data from the SVD. Using only voice signals, the proposed multimodal model achieved an AUC of 0.470 (refer to Figure 15). When demographic information was added to the model, the AUC increased to 0.785, as depicted in Figure 16. According to the confusion matrix (Table 16, right panel), the model correctly predicted 17 out of 23 cases of glottic neoplasm, corresponding to a sensitivity of 73.9%. Out of 1,331 benign cases, 992 were accurately identified, resulting in a specificity of 74.5%. The overall accuracy of the model in this external validation was 74.5%.

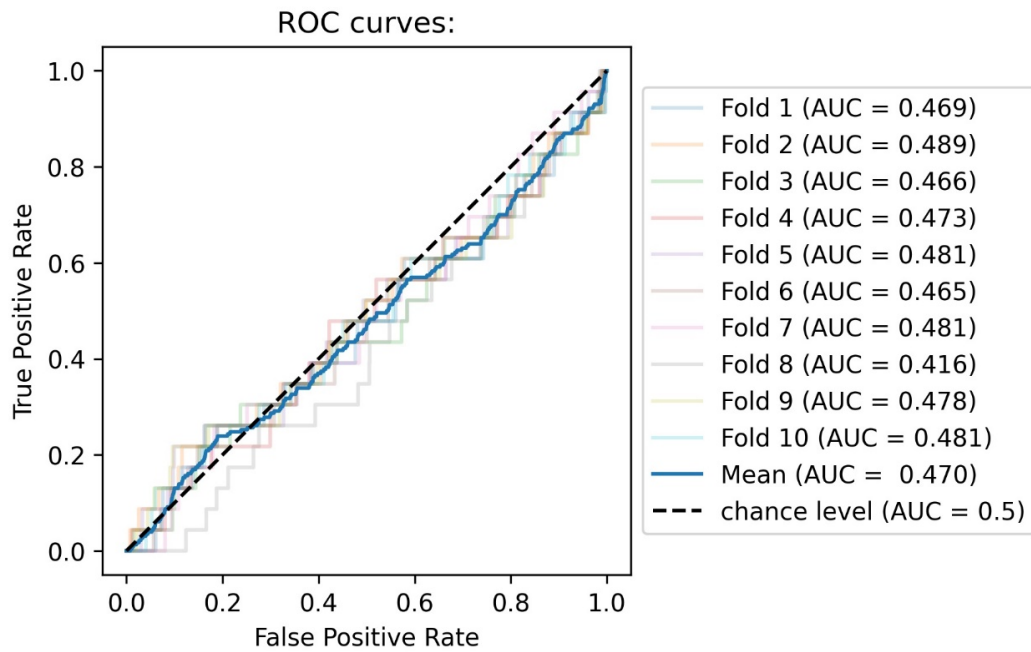


Figure 15 Performance of proposed multimodal model with only voice signals on SVD

The ROC curves illustrate the model's performance across 10 validation folds, with AUCs ranging from 0.416 to 0.489. The mean AUC across all folds is 0.470, indicating poor classification performance. The dashed line represents the chance level (AUC = 0.5).

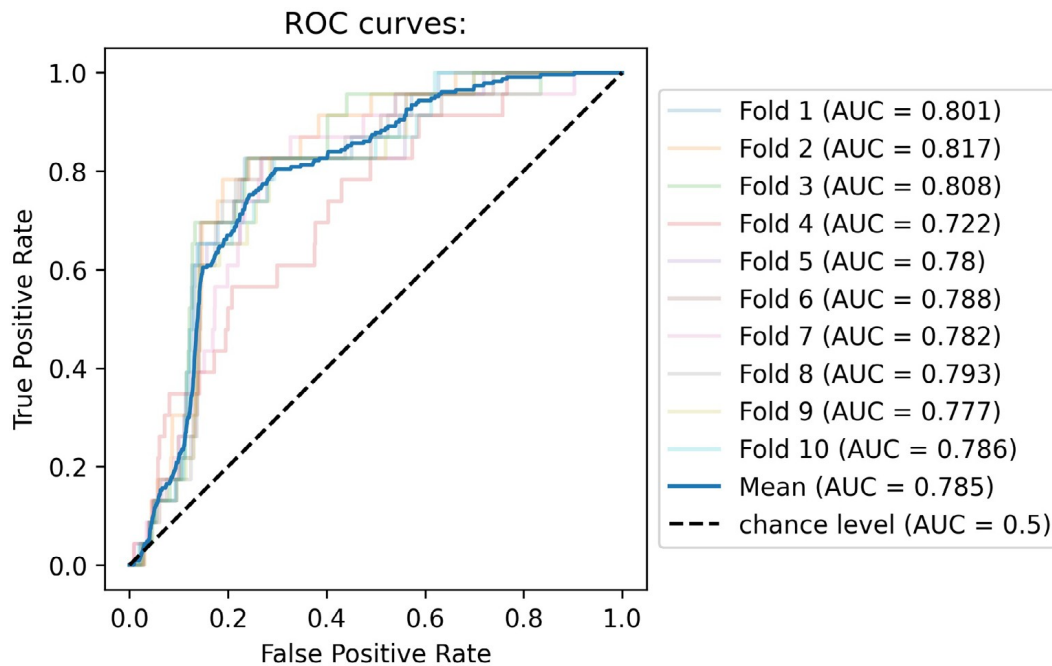


Figure 16 The performance of proposed multimodal model with voice signals and demographics on SVD

The ROC curves illustrate the model's performance across 10 validation folds, with AUCs ranging from 0.722 to 0.817. The mean AUC across all folds is 0.785, indicating moderate classification performance. The dashed line represents the chance level (AUC = 0.5).

Intrinsic validation on CPSC2018 ECG training set

I also validated the intrinsic adaptivity of my multimodal model using ECG data from the CPSC2018 ECG training set. When the model utilized only ECG signals, it achieved AUCs of 0.936 for lead aVR and 0.917 for lead V1. After incorporating demographic information (age and sex) into the model, the AUCs slightly changed to 0.939 for lead aVR and 0.909 for lead V1. These differences were not statistically significant (p-value from paired t-test < 0.05).

Comparison between human assessments and AI

In sub-analyses comparing the AI model's predictions with human evaluations, I observed that the model generally exhibited higher sensitivity, ranging from 48% to 78%, compared to human assessments, which ranged from 9% to 42% ($p < 0.01$, based on two-sample proportion tests; see Table 17). Conversely, human assessments demonstrated higher specificity, between 86% and 98%, compared to the model's specificity of 43% to 82% ($p < 0.01$). When considering only voice data or voice data combined with demographics, human predictions typically showed better overall accuracy than the model ($p < 0.01$). However, with the inclusion of medical records, the model's accuracy increased to 82%, which was comparable to the human assessment accuracy of 83% ($p = 0.78$). Further analysis revealed an agreement rate of 85% between two independent human raters (see Table 18).

Table 17 Comparison of glottic neoplasm diagnosis between AI model prediction and pathologists

	FVD			SVD		
	Human	AI	<i>p</i> -Value*	Human	AI	<i>p</i> -Value*
Voice only						
Sensitivity	10%	55%	<0.01	42%	48%	0.73
Specificity	86%	54%	<0.01	91%	43%	<0.01
Accuracy	73%	54%	<0.01	86%	43%	<0.01
Voice + demographics						
Sensitivity	25%	68%	<0.01	9%	74%	<0.01
Specificity	98%	70%	<0.01	90%	75%	<0.01
Accuracy	86%	70%	<0.01	82%	75%	0.08
Voice + medical records						
Sensitivity	30%	78%	<0.01	-	-	-
Specificity	93%	82%	<0.01	-	-	-
Accuracy	83%	82%	0.78	-	-	-

Bold values indicate statistical significance ($p < 0.05$).

Table 18 Prediction results between 2 independent raters

		Rater 1		Total
		Neoplasm	Benign	
Rater 2	Neoplasm	31	51	82
	Benign	39	462	501
	Total	70	513	583

3.4 Discussion

Over the past decade, AI has increasingly integrated into nearly every aspect of modern life, including medical practice. Numerous review articles have outlined various applications of AI in the field of otorhinolaryngology [126, 127]. A recent systematic review specifically focused on AI applications in laryngeal cancer using different data modalities [128]. In this discussion, I will mainly concentrate on signals beyond ECG, i.e. acoustics and clinical data.

Previous studies have utilized machine learning algorithms to differentiate voice recordings between laryngeal cancer patients and healthy controls. An early small-scale study reported detection rates of 88.7% for vocal cord cancer ($N = 20$) and 92.8% for healthy individuals ($N = 10$) using various spectral features HMM [120]. A more recent study achieved a discrimination accuracy of 96.9% between laryngeal cancer patients ($N = 101$) and normal subjects ($N = 100$) based on glottal flow parameters ANN [119]. Another investigation utilized Mel-frequency cepstral coefficients (MFCCs) as acoustic features, reporting an AI performance with an AUC of 0.85 among 50 laryngeal cancer cases and 45 healthy controls [129].

Despite the high accuracy reported, these studies were limited by small sample sizes and an unbalanced proportion of glottic neoplasm cases [130]. More importantly, these results have limited clinical relevance because an ideal AI application should distinguish glottic neoplasms from other benign voice disorders rather than from healthy individuals [128]. Subsequent efforts to classify voice disorders into distinct disease categories, including glottic neoplasm, using machine learning algorithms have not been as successful. One study demonstrated significant overlap in acoustic features among

voice recordings from glottic neoplasm, phonotrauma, and vocal palsy [131]. My study also found that voice signals alone are insufficient for accurately differentiating glottic neoplasms from other non-malignant voice disorders (Figures 12 and 15).

In this research, I utilized the newly published FVD to develop a DL-based AI model for detecting glottic neoplasms. A key advantage of the FVD over other databases is its inclusion of multiple data modalities, such as voice recordings, demographic information, and structured medical records (Table 15). To my knowledge, only one other small voice database (N = 208) includes lifestyle information like smoking and alcohol consumption in addition to voice recordings [132].

The potential benefits of incorporating demographic data into AI models are supported by previous epidemiological studies [133]. For instance, female patients are more prone to phonotraumatic voice disorders due to factors like higher fundamental frequency, a greater prevalence in voice-dependent occupations [134], and differences in the extracellular matrix composition of vocal folds compared to males [135, 136]. Research has also indicated that the age of patients with voice disorders varies significantly between diagnostic groups. Younger individuals are more likely to develop phonotraumatic lesions such as nodules and polyps, whereas older patients tend to present with neoplastic lesions [137]. A publication observed that patients with glottic neoplasms were generally older and predominantly male compared to those with vocal palsy or phonotraumatic conditions [138]. In the current study, adding demographic data significantly enhanced model performance for both the FVD and SVD (Figures 14 and 16). I found that incorporating medical records into the AI model further increased performance to an AUC of 0.878 (Figure 15). This suggests that a multimodal approach

integrating voice, demographic, and medical data is more effective for AI-assisted detection of voice disorders. Although similar phenomena could not be observed in CA classification with ECG signals, it may implicate that demographic factors like age and sex do not provide additional predictive value for CA classification beyond what is already captured by the signals themselves.

My study employed an end-to-end network structure to predict glottic neoplasms using voice recordings of the sustained vowel sound "ah." While vowel sounds contain less acoustic information than continuous speech, conversations, or paragraph readings, their use allows for the inclusion of voice data from patients speaking different languages. For example, the SVD includes recordings of patients counting from 1 to 10 in German, which cannot be analyzed using an AI model based on English or Mandarin. Similarly, AI models trained on voice samples of reading the Rainbow Passage in English are not applicable to other languages. My results clearly demonstrate the advantage of using vowel sounds as the source of acoustic features. The AI model using voice and demographics from the FVD achieved an AUC of 0.807, and external validation with the SVD data resulted in an AUC of 0.785, indicating that vowel sounds are ideal for building robust, cross-language AI models.

Another benefit of my model is its ability to directly analyze raw voice waveforms without the need for additional extraction of acoustic features such as MFCCs or wavelet parameters commonly used in previous studies [113, 139, 140]. This model was originally designed for ECG and CA classification [58], showcasing the robustness of the AI super-resolution technique and its potential for integration into portable/wearable devices [141]. Future research could explore smartphone-based voice signal acquisition and AI

interpretation, increasing accessibility for both the medical industry and the general public.

My study found that the AI model had higher sensitivity but lower specificity compared to human assessment (Table 17), consistent with recent literature [129]. Considering that glottic neoplasms constitute only a small fraction of cases among patients with voice disorders [130], the high negative predictive value (99%) of the AI model makes it suitable for first-line screening to "rule out" benign cases. In practical application, this model could potentially reduce medical workload by 80%, as only 404 cases suspected of glottic neoplasm would require prompt referral for laryngeal endoscopy and pathology (Table 16). Human experts could collaborate by assessing voice samples to "rule in" possible neoplastic cases, as human assessment showed greater positive predictive value than the AI model (Table 17). It remains important, however, to advise patients with negative AI screening results to remain vigilant in monitoring their voice conditions. Progressive infiltration of glottic neoplasms in false-negative cases may lead to gradual changes in voice patterns, which could be detected in subsequent AI screenings.

As indicated in Table 16, the AI model produced some false positives, misclassifying benign lesions as neoplasms. Further investigation revealed that these false positives were more likely in cases of vocal atrophy and presbyphonia. Since my AI model incorporates demographic data, one possible explanation is that these cases share an older age distribution similar to neoplastic cases. Additionally, the AI model may miss glottic neoplasms that do not originate from the vibrating edge of the vocal folds (e.g., lesions in the ventricle).

3.5 Conclusion

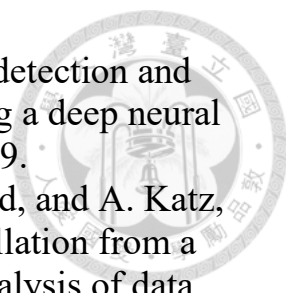
This study presents a new voice database comprising 2,000 cases that combine voice signals, demographic information, and structured medical records for AI analysis. My findings indicate that using voice signals alone may be insufficient for distinguishing glottic neoplasms from other benign voice disorders. Incorporating demographics and medical records into the AI model significantly improved performance, bringing its predictive accuracy closer to that of human assessments.

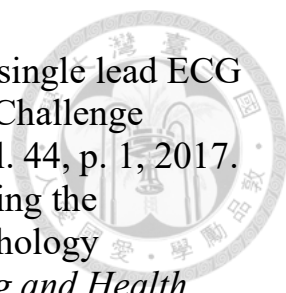


REFERENCE



- [1] A. S. Kibos, B. P. Knight, V. Essebag, S. B. Fishberger, M. Slevin, and I. C. Țintoiu, *Cardiac Arrhythmias: From Basic Mechanism to State-of-the-Art Management*. Springer London, 2013.
- [2] P. Malmivuo, J. Malmivuo, and R. Plonsey, *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA, 1995.
- [3] F. N. Wilson *et al.*, "Recommendations for standardization of electrocardiographic and vectorcardiographic leads," *Circulation*, vol. 10, no. 4, pp. 564-573, 1954.
- [4] A. Bayes de Luna *et al.*, "Interatrial conduction block and retrograde activation of the left atrium and paroxysmal supraventricular tachyarrhythmia," *European heart journal*, vol. 9, no. 10, pp. 1112-1118, 1988.
- [5] B. Surawicz, R. Childers, B. J. Deal, and L. S. Gettes, "AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part III: intraventricular conduction disturbances a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology," *Journal of the American College of Cardiology*, vol. 53, no. 11, pp. 976-981, 2009.
- [6] K. Wesley, *Huszar's ECG and 12-Lead Interpretation - E-Book*. Elsevier Health Sciences, 2016.
- [7] Y. Kobayashi, "Idiopathic Ventricular Premature Contraction and Ventricular Tachycardia: Distribution of the Origin, Diagnostic Algorithm, and Catheter Ablation," *Journal of Nippon Medical School*, vol. 85, no. 2, pp. 87-94, 2018.
- [8] T. Garcia and G. Miller, *Arrhythmia Recognition: The Art of Interpretation*. Jones & Bartlett Learning, 2004.
- [9] E. B. Hanna and D. L. Glancy, "ST-segment depression and T-wave inversion: classification, differential diagnosis, and caveats," *Cleveland Clinic journal of medicine*, vol. 78, no. 6, p. 404, 2011.

- 
- [10] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine*, vol. 25, no. 1, p. 65, 2019.
 - [11] A. Shiyovich, A. Wolak, L. Yacobovich, A. Grosbard, and A. Katz, "Accuracy of diagnosing atrial flutter and atrial fibrillation from a surface electrocardiogram by hospital physicians: analysis of data from internal medicine departments," *The American journal of the medical sciences*, vol. 340, no. 4, pp. 271-275, 2010.
 - [12] R. Hoekema, G. J. Uijen, and A. Van Oosterom, "Geometrical aspects of the interindividual variability of multilead ECG recordings," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 5, pp. 551-559, 2001.
 - [13] A. Lyon, A. Mincholé, J. P. Martínez, P. Laguna, and B. Rodriguez, "Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances," *Journal of the Royal Society Interface*, vol. 15, no. 138, p. 20170821, 2018.
 - [14] Q. Qin, J. Li, L. Zhang, Y. Yue, and C. Liu, "Combining Low-dimensional Wavelet Features and Support Vector Machine for Arrhythmia Beat Classification," *Scientific reports*, vol. 7, no. 1, p. 6067, 2017.
 - [15] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215-e220, 2000.
 - [16] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45-50, 2001.
 - [17] H. A. Guvenir, B. Acar, G. Demiroz, and A. Cekin, "Supervised machine learning algorithm for arrhythmia analysis," *Computers in cardiology*, pp. 433-436, 1997.
 - [18] R. Begg, *Neural Networks in Healthcare: Potential and Challenges: Potential and Challenges*. IGI Global, 2006.
 - [19] A. Mustaqeem, S. M. Anwar, and M. Majid, "Multiclass Classification of Cardiac Arrhythmia Using Improved Feature Selection and SVM Invariants," *Computational and mathematical methods in medicine*, vol. 2018, 2018.
 - [20] C. G. NAYAK, G. Seshikala, U. Desai, and S. G. Nayak, "Identification of arrhythmia classes using machine-learning techniques," *International Journal of Biology and Biomedicine*, vol. 1, pp. 48-53, 2016.

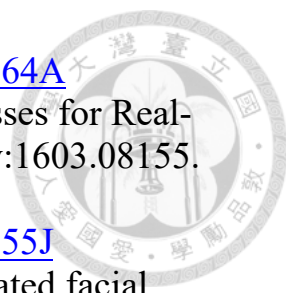
- 
- [21] G. D. Clifford *et al.*, "AF classification from a short single lead ECG recording: The Physionet Computing in Cardiology Challenge 2017," *Proceedings of Computing in Cardiology*, vol. 44, p. 1, 2017.
 - [22] F. Liu *et al.*, "An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1368-1373, 2018.
 - [23] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," (in English), *Ieee T Signal Proces*, vol. 45, no. 11, pp. 2673-2681, Nov 1997, doi: Doi 10.1109/78.650093.
 - [24] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical Attention Networks for Document Classification," San Diego, California, June 2016: Association for Computational Linguistics, in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1480-1489, doi: 10.18653/v1/N16-1174. [Online]. Available: <https://aclanthology.org/N16-1174>
<https://doi.org/10.18653/v1/N16-1174>
 - [25] J. Hearty, *Advanced Machine Learning with Python*. Packt Publishing, 2016.
 - [26] L. Rutkowski, *Computational Intelligence: Methods and Techniques*. Springer Berlin Heidelberg, 2008.
 - [27] A. Pal and P. Prakash, *Practical Time Series Analysis: Master Time Series Data Processing, Visualization, and Modeling using Python*. Packt Publishing, 2017.
 - [28] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.
 - [29] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.
 - [30] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
 - [31] A. L. Maas, "Rectifier Nonlinearities Improve Neural Network Acoustic Models," 2013.
 - [32] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *ArXiv e-prints*, vol. 1412. [Online]. Available:

- <http://adsabs.harvard.edu/abs/2014arXiv1412.6980K>
- [33] M. Abadi *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.
- [34] D. E. Goldberg, "Genetic Algorithms in Search Optimization and Machine Learning," 1989.
- [35] J. M. McCabe *et al.*, "Physician accuracy in interpreting potential ST-segment elevation myocardial infarction electrocardiograms," *Journal of the American Heart Association*, vol. 2, no. 5, p. e000268, 2013.
- [36] J. B. Nielsen, M. S. Olesen, M. Tangø, S. Haunsø, A. G. Holst, and J. H. Svendsen, "Incomplete right bundle branch block: a novel electrocardiographic marker for lone atrial fibrillation," *Europace*, vol. 13, no. 2, pp. 182-187, 2010.
- [37] M. Gertsch, *The ECG Manual: An Evidence-Based Approach*. Springer London, 2016.
- [38] Apple. "Taking an ECG with the ECG app on Apple Watch Series 4." <https://support.apple.com/hr-hr/HT208955> (accessed 4/20, 2019).
- [39] R. Beebe and J. Myers, *Professional Paramedic, Volume I: Foundations of Paramedic Care*. Cengage Learning, 2012.
- [40] S. N. Goodman, "Toward evidence-based medical statistics. 2: The Bayes factor," *Annals of internal medicine*, vol. 130, no. 12, pp. 1005-1013, 1999.
- [41] P. Podrid, M. D. M. S. Rajeev Malhotra, R. Kakkar, and P. A. Noseworthy, *Podrid's Real-World ECGs: Volume 4B, Arrhythmias [Practice Cases]: A Master's Approach to the Art and Practice of Clinical ECG Interpretation*. Cardiotext Publishing, 2015.
- [42] S. Chugh, *Textbook of Clinical Electrocardiography*. Jaypee Brothers, 2014.
- [43] A. Esteva *et al.*, "A guide to deep learning in healthcare," *Nature medicine*, vol. 25, no. 1, p. 24, 2019.
- [44] A. C. Krueger. "FDA document of Electrocardiograph software for over-the-counter use." https://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180044.pdf (accessed October 17, 2018).
- [45] Z. I. Attia *et al.*, "Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram," *Nature medicine*, vol. 25, no. 1, p. 70, 2019.
- [46] Z. I. Attia *et al.*, "Novel Bloodless Potassium Determination Using a

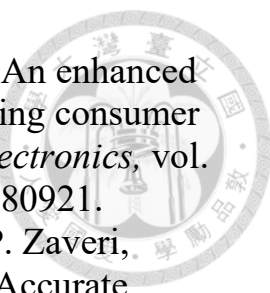
- Signal-Processed Single-Lead ECG," *Journal of the American Heart Association*, vol. 5, no. 1, p. e002746, 2016.
- [47] H. Zhu *et al.*, "Smart Healthcare in the Era of Internet-of-Things," *IEEE Consumer Electronics Magazine*, vol. 8, no. 5, pp. 26-30, 2019, doi: 10.1109/MCE.2019.2923929.
- [48] S. P. Mohanty, U. Choppali, and E. Kougianos, "Everything you wanted to know about smart cities: The Internet of things is the backbone," *IEEE Consumer Electronics Magazine*, vol. 5, no. 3, pp. 60-70, 2016, doi: 10.1109/MCE.2016.2556879.
- [49] M. Shao, Z. Zhou, G. Bin, Y. Bai, and S. Wu, "A Wearable Electrocardiogram Telemonitoring System for Atrial Fibrillation Detection," *Sensors (Basel)*, vol. 20, no. 3, Jan 22 2020, doi: 10.3390/s20030606.
- [50] C. T. Lin *et al.*, "An intelligent telecardiology system using a wearable and wireless ECG to detect atrial fibrillation," *IEEE Trans Inf Technol Biomed*, vol. 14, no. 3, pp. 726-33, May 2010, doi: 10.1109/TITB.2010.2047401.
- [51] S. Wattal, S. K. Spear, M. H. Imtiaz, and E. Sazonov, "A polypyrrole-coated textile electrode and connector for wearable ECG monitoring," in *2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 4-7 March 2018 2018, pp. 54-57, doi: 10.1109/BSN.2018.8329657.
- [52] S. Hong *et al.*, "CardioLearn: A Cloud Deep Learning Service for Cardiac Disease Detection from Electrocardiogram," presented at the Companion Proceedings of the Web Conference 2020, Taipei, Taiwan, 2020. [Online]. Available: <https://doi.org/10.1145/3366424.3383529>.
- [53] X. Wang, Q. Gui, B. Liu, Z. Jin, and Y. Chen, "Enabling Smart Personalized Healthcare: A Hybrid Mobile-Cloud Approach for ECG Telemonitoring," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 3, pp. 739-745, 2014, doi: 10.1109/JBHI.2013.2286157.
- [54] J. Kim and C. Chu, "Analysis of energy consumption for wearable ECG devices," in *SENSORS, 2014 IEEE*, 2-5 Nov. 2014 2014, pp. 962-965, doi: 10.1109/ICSENS.2014.6985162.
- [55] K. C. Liu, K. H. Hung, C. Y. Hsieh, H. Y. Huang, C. T. Chan, and Y. Tsao, "Deep Learning Based Signal Enhancement of Low-Resolution Accelerometer for Fall Detection Systems," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1-1,

- 2021, doi: 10.1109/TCDS.2021.3116228.
- [56] X. Fafoutis, L. Marchegiani, A. Elsts, J. Pope, R. Piechocki, and I. Craddock, "Extending the battery lifetime of wearable sensors with embedded machine learning," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, 5-8 Feb. 2018 2018, pp. 269-274, doi: 10.1109/WF-IoT.2018.8355116.
- [57] N. Petrellis, I.-E. Kosmadakis, M. Vardakas, F. Gioulekas, M. Birbas, and A. Lalos, "Compressing and Filtering Medical Data in a Low Cost Health Monitoring System," in *Proceedings of the 21st Pan-Hellenic Conference on Informatics*, 2017, pp. 1-5.
- [58] T.-M. Chen, C.-H. Huang, E. S. Shih, Y.-F. Hu, and M.-J. Hwang, "Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model," *Iscience*, vol. 23, no. 3, p. 100886, 2020.
- [59] S. Mahdiani, V. Jeyhani, M. Peltokangas, and A. Vehkaoja, "Is 50 Hz high enough ECG sampling frequency for accurate HRV analysis?," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2015, pp. 5948-51, 2015, doi: 10.1109/EMBC.2015.7319746.
- [60] O. Kwon *et al.*, "Electrocardiogram Sampling Frequency Range Acceptable for Heart Rate Variability Analysis," *Healthc Inform Res*, vol. 24, no. 3, pp. 198-206, Jul 2018, doi: 10.4258/hir.2018.24.3.198.
- [61] K. A. Sidek and I. Khalil, "Enhancement of low sampling frequency recordings for ECG biometric matching using interpolation," *Computer Methods and Programs in Biomedicine*, vol. 109, no. 1, pp. 13-25, 2013/01/01/ 2013, doi: <https://doi.org/10.1016/j.cmpb.2012.08.015>.
- [62] S. M. Mathews, C. Kambhamettu, and K. E. Barner, "A novel application of deep learning for single-lead ECG classification," (in eng), *Comput Biol Med*, vol. 99, pp. 53-62, Aug 1 2018, doi: 10.1016/j.compbimed.2018.05.013.
- [63] Y. L. Coelho, F. d. A. S. d. Santos, A. Frizera-Neto, and T. F. Bastos-Filho, "A Lightweight Framework for Human Activity Recognition on Wearable Devices," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24471-24481, 2021, doi: 10.1109/JSEN.2021.3113908.
- [64] Z. Long, T. Wang, C. You, Z. Yang, K. Wang, and J. Liu, "Terahertz image super-resolution based on a deep convolutional neural network," *Appl Opt*, vol. 58, no. 10, pp. 2731-2735, Apr 1 2019, doi: 10.1364/AO.58.002731.
- [65] M. Kwon, S. Han, K. Kim, and S. C. Jun, "Super-Resolution for

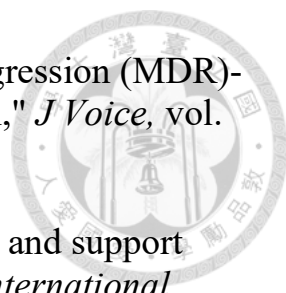
- Improving EEG Spatial Resolution using Deep Convolutional Neural Network-Feasibility Study," *Sensors (Basel)*, vol. 19, no. 23, Dec 3 2019, doi: 10.3390/s19235317.
- [66] K. Umehara, J. Ota, and T. Ishida, "Application of Super-Resolution Convolutional Neural Network for Enhancing Image Resolution in Chest CT," *J Digit Imaging*, vol. 31, no. 4, pp. 441-450, Aug 2018, doi: 10.1007/s10278-017-0033-z.
- [67] D. Dai, Y. Wang, Y. Chen, and L. Van Gool, "Is Image Super-resolution Helpful for Other Vision Tasks?," p. arXiv:1509.07009. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2015arXiv150907009D>
- [68] M. Haris, G. Shakhnarovich, and N. Ukita, "Task-Driven Super Resolution: Object Detection in Low-resolution Images," p. arXiv:1803.11316. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2018arXiv180311316H>
- [69] M. S. M. Sajjadi, B. Schölkopf, and M. Hirsch, "EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis," p. arXiv:1612.07919. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2016arXiv161207919S>
- [70] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "SOD-MTGAN: Small Object Detection via Multi-Task Generative Adversarial Network," in *Computer Vision – ECCV 2018*, Cham, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., 2018// 2018: Springer International Publishing, pp. 210-226.
- [71] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution," p. arXiv:1704.03915. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2017arXiv170403915L>
- [72] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," Cham, 2014: Springer International Publishing, in *Computer Vision – ECCV 2014*, pp. 184-199.
- [73] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," p. arXiv:1511.04587. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2015arXiv151104587K>
- [74] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, Accurate, and Lightweight Super-Resolution with Cascading Residual Network," p. arXiv:1803.08664. [Online]. Available:

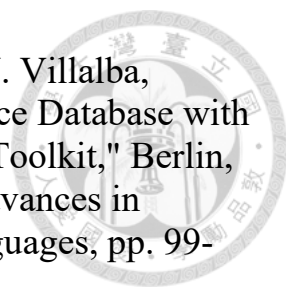
- 
- [75] <https://ui.adsabs.harvard.edu/abs/2018arXiv180308664A>
J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," p. arXiv:1603.08155. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2016arXiv160308155J>
- [76] A. Bulat and G. Tzimiropoulos, "Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs," p. arXiv:1712.02765. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2017arXiv171202765B>
- [77] C. Ledig *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," p. arXiv:1609.04802. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2016arXiv160904802L>
- [78] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," p. arXiv:1707.02921. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2017arXiv170702921L>
- [79] Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, and C. Schroers, "A Fully Progressive Approach to Single-Image Super-Resolution," p. arXiv:1804.02900. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2018arXiv180402900W>
- [80] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," p. arXiv:1807.02758. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2018arXiv180702758Z>
- [81] T. Dai, J. Cai, Y. Zhang, S. Xia, and L. Zhang, "Second-Order Attention Network for Single Image Super-Resolution," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 15-20 June 2019 2019, pp. 11057-11066, doi: 10.1109/CVPR.2019.01132.
- [82] G. Hinton *et al.*, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82-97, 2012, doi: 10.1109/MSP.2012.2205597.
- [83] D. Jurafsky and J. H. Martin, "Speech and language processing," (in English), 2014. [Online]. Available: <http://www.dawsonera.com/depp/reader/protected/external/AbstractView/S9781292037936>.

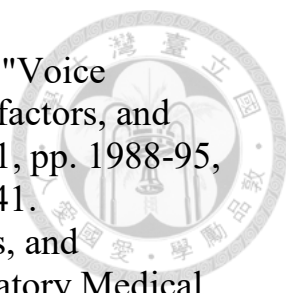
- 
- [84] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising autoencoder," in *INTERSPEECH*, 2013.
 - [85] A. van den Oord *et al.*, "WaveNet: A Generative Model for Raw Audio," p. arXiv:1609.03499. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2016arXiv160903499V>
 - [86] J.-P. Briot, G. Hadjeres, and F.-D. Pachet, "Deep Learning Techniques for Music Generation -- A Survey," p. arXiv:1709.01620. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2017arXiv170901620B>
 - [87] M. A. Acevedo, C. J. Corrada-Bravo, H. Corrada-Bravo, L. J. Villanueva-Rivera, and T. M. Aide, "Automated classification of bird and amphibian calls using machine learning: A comparison of methods," *Ecological Informatics*, vol. 4, no. 4, pp. 206-214, 2009, doi: 10.1016/j.ecoinf.2009.06.005.
 - [88] S. Birnbaum, V. Kuleshov, Z. Enam, P. W. Koh, and S. Ermon, "Temporal FiLM: Capturing Long-Range Sequence Dependencies with Feature-Wise Modulations," p. arXiv:1909.06628. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2019arXiv190906628B>
 - [89] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. null, pp. 2825–2830, 2011.
 - [90] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27-30 June 2016 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
 - [91] P. W. D. Charles, "Project Title," *GitHub repository*, vol. <https://github.com/charlespwd/project-title>, 2013.
 - [92] H. Thapliyal, V. Khalus, and C. Labrado, "Stress Detection and Management: A Survey of Wearable Smart Health Devices," *IEEE Consumer Electronics Magazine*, vol. 6, no. 4, pp. 64-69, 2017, doi: 10.1109/MCE.2017.2715578.
 - [93] S. H. Chae, M. C. Kang, J. Y. Sun, B. S. Kim, and S. J. Ko, "Collision detection method using image segmentation for the visually impaired," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 4, pp. 392-400, 2017, doi: 10.1109/TCE.2017.015101.
 - [94] C. W. Lee, P. Chondro, S. J. Ruan, O. Christen, and E. Naroska, "Improving Mobility for the Visually Impaired: A Wearable Indoor Positioning System Based on Visual Markers," *IEEE Consumer Electronics Magazine*, vol. 7, no. 3, pp. 12-20, 2018, doi: 10.1109/MCE.2018.2797741.

- 
- [95] J. Wang, Z. Zhang, B. Li, S. Lee, and R. S. Sherratt, "An enhanced fall detection system for elderly person monitoring using consumer home networks," *IEEE Transactions on Consumer Electronics*, vol. 60, no. 1, pp. 23-29, 2014, doi: 10.1109/TCE.2014.6780921.
 - [96] M. A. Sayeed, S. P. Mohanty, E. Kougianos, and H. P. Zaveri, "Neuro-Detect: A Machine Learning-Based Fast and Accurate Seizure Detection System in the IoMT," *IEEE Transactions on Consumer Electronics*, vol. 65, no. 3, pp. 359-368, 2019, doi: 10.1109/TCE.2019.2917895.
 - [97] A. M. Joshi, P. Jain, and S. P. Mohanty, "iGLU 3.0: A Secure Noninvasive Glucometer and Automatic Insulin Delivery System in IoMT," *IEEE Transactions on Consumer Electronics*, vol. 68, no. 1, pp. 14-22, 2022, doi: 10.1109/TCE.2022.3145055.
 - [98] J. Li, M. L. Seltzer, X. Wang, R. Zhao, and Y. Gong, "Large-Scale Domain Adaptation via Teacher-Student Learning," p. arXiv:1708.05466. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2017arXiv170805466L>
 - [99] S. O. Sadjadi and J. H. Hansen, "Assessment of single-channel speech enhancement techniques for speaker identification under mismatched conditions," in *Eleventh Annual Conference of the International Speech Communication Association*, 2010.
 - [100] M. Fujimoto and H. Kawai, "One-Pass Single-Channel Noisy Speech Recognition Using a Combination of Noisy and Enhanced Features," in *INTERSPEECH*, 2019, pp. 486-490.
 - [101] H. Sato, T. Ochiai, M. Delcroix, K. Kinoshita, N. Kamo, and T. Moriya, "Learning to Enhance or Not: Neural Network-Based Switching of Enhanced and Observed Signals for Overlapping Speech Recognition," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022: IEEE, pp. 6287-6291.
 - [102] A. B. de Luna, *Clinical Electrocardiography, Enhanced Edition: A Textbook*. Wiley, 2012.
 - [103] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," p. arXiv:1902.04885. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2019arXiv190204885Y>
 - [104] F. Fatehi, F. Hassandoust, R. K. L. Ko, and S. Akhlaghpour, "General Data Protection Regulation (GDPR) in Healthcare: Hot Topics and Research Fronts," *Stud Health Technol Inform*, vol. 270,

- pp. 1118-1122, Jun 16 2020, doi: 10.3233/SHTI200336.
- [105] Z. I. Attia *et al.*, "Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs," *Circ Arrhythm Electrophysiol*, vol. 12, no. 9, p. e007284, Sep 2019, doi: 10.1161/CIRCEP.119.007284.
- [106] C.-M. Liu *et al.*, "Artificial Intelligence-Enabled Electrocardiogram Improves the Diagnosis and Prediction of Mortality in Patients With Pulmonary Hypertension," *JACC: Asia*, vol. 2, no. 3, Part 1, pp. 258-270, 2022/06/01/ 2022, doi: <https://doi.org/10.1016/j.jacasi.2022.02.008>.
- [107] J. Rusz *et al.*, "Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task," *The Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 2171-2181, 2013.
- [108] M. Wodzinski, A. Skalski, D. Hemmerling, J. R. Orozco-Arroyave, and E. Nöth, "Deep learning approach to Parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification," in *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2019: IEEE, pp. 717-720.
- [109] A. Z. Antosik-Wójcińska *et al.*, "Smartphone as a monitoring tool for bipolar disorder: a systematic review including data analysis, machine learning algorithms and predictive modelling," *International journal of medical informatics*, vol. 138, p. 104131, 2020.
- [110] J.-Y. Han *et al.*, "Enhancing the Performance of Pathological Voice Quality Assessment System Through the Attention-Mechanism Based Neural Network," *Journal of Voice*, 2023.
- [111] J. D. Arias-Londono, J. I. Godino-Llorente, N. Saenz-Lechon, V. Osma-Ruiz, and G. Castellanos-Dominguez, "An improved method for voice pathology detection by means of a HMM-based feature space transformation.," *Pattern Recognition.*, vol. 43, no. 9, pp. 3100-3112, 2010.
- [112] M. K. Arjmandi, M. Pooyan, M. Mikaili, M. Vali, and A. Moqarehzadeh, "Identification of voice disorders using long-time features and support vector machine with different feature reduction methods," *J Voice*, vol. 25, no. 6, pp. e275-89, Nov 2011, doi: 10.1016/j.jvoice.2010.08.003.
- [113] G. Muhammad, T. A. Mesallam, K. H. Malki, M. Farahat, A.

- 
- Mahmood, and M. Alsulaiman, "Multidirectional regression (MDR)-based features for automatic voice disorder detection," *J Voice*, vol. 26, no. 6, pp. 817 e19-27, Nov 2012, doi: 10.1016/j.jvoice.2012.05.002.
- [114] N. Souissi and A. Cherif, "Artificial neural networks and support vector machine for voice disorders identification," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, 2016.
- [115] S. H. Fang *et al.*, "Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach," *J Voice*, vol. 33, no. 5, pp. 634-641, Sep 2019, doi: 10.1016/j.jvoice.2018.02.003.
- [116] S. Hegde, S. Shetty, S. Rai, and T. Dodderi, "A Survey on Machine Learning Approaches for Automatic Detection of Voice Disorders," *J Voice*, vol. 33, no. 6, pp. 947 e11-947 e33, Nov 2019, doi: 10.1016/j.jvoice.2018.07.014.
- [117] A. Al-Nasheri *et al.*, "An Investigation of Multidimensional Voice Program Parameters in Three Different Databases for Voice Pathology Detection and Classification," *J Voice*, vol. 31, no. 1, pp. 113 e9-113 e18, Jan 2017, doi: 10.1016/j.jvoice.2016.03.019.
- [118] M. Chaiani, S. A. Selouani, M. Boudraa, and M. Sidi Yakoub, "Voice disorder classification using speech enhancement and deep learning models," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 2, pp. 463-480, 2022, doi: 10.1016/j.bbe.2022.03.002.
- [119] A. Aicha and K. Ezzine, "Cancer larynx detection using glottal flow parameters and statistical tools," presented at the 2016 International Symposium on Signal, Image, Video and Communications (ISIVC), 2016.
- [120] L. Gavidia-Ceballos and J. H. Hansen, "Direct speech feature estimation using an iterative EM algorithm for vocal fold pathology detection," *IEEE Trans Biomed Eng*, vol. 43, no. 4, pp. 373-83, Apr 1996, doi: 10.1109/10.486257.
- [121] C.-T. Wang, Z.-Y. Chuang, C.-H. Hung, Y. Tsao, and S.-H. Fang, "Detection of Glottic Neoplasm Based on Voice Signals Using Deep Neural Networks," *IEEE Sensors Letters*, vol. 6, no. 3, pp. 1-4, 2022, doi: 10.1109/lens.2022.3152738.
- [122] C. T. Wang, T. M. Chen, N. T. Lee, and S. H. Fang, "AI Detection of Glottic Neoplasm Using Voice Signals, Demographics, and Structured Medical Records," (in English), *Laryngoscope*, vol. 134, no. 11, pp. 4585-4592, Nov 2024, doi: 10.1002/lary.31563.

- 
- [123] D. Martínez, E. Lleida, A. Ortega, A. Miguel, and J. Villalba, "Voice Pathology Detection on the Saarbrücken Voice Database with Calibration and Fusion of Scores Using MultiFocal Toolkit," Berlin, Heidelberg, 2012: Springer Berlin Heidelberg, in *Advances in Speech and Language Technologies for Iberian Languages*, pp. 99-109.
 - [124] D. Jaganath *et al.*, "Accelerating cough-based algorithms for pulmonary tuberculosis screening: Results from the CODA TB DREAM Challenge," *medRxiv*, p. 2024.05.13.24306584, 2024, doi: 10.1101/2024.05.13.24306584.
 - [125] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
 - [126] A. M. Bur, M. Shew, and J. New, "Artificial Intelligence for the Otolaryngologist: A State of the Art Review," *Otolaryngol Head Neck Surg*, vol. 160, no. 4, pp. 603-611, Apr 2019, doi: 10.1177/0194599819827507.
 - [127] M. G. Crowson *et al.*, "A contemporary review of machine learning in otolaryngology-head and neck surgery," *Laryngoscope*, Feb 1 2019, doi: 10.1002/lary.27850.
 - [128] Y. Bensoussan, E. B. Vanstrum, M. M. Johns, 3rd, and A. Rameau, "Artificial Intelligence and Laryngeal Cancer: From Screening to Prognosis: A State of the Art Review," *Otolaryngol Head Neck Surg*, vol. 168, no. 3, pp. 319-329, Mar 2023, doi: 10.1177/01945998221110839.
 - [129] H. Kim *et al.*, "Convolutional neural network classifies pathological voice change in laryngeal cancer with high accuracy," *Journal of Clinical Medicine*, vol. 9, no. 11, p. 3415, 2020.
 - [130] S. M. Cohen, J. Kim, N. Roy, C. Asche, and M. Courey, "Prevalence and causes of dysphonia in a large treatment-seeking population," *Laryngoscope*, vol. 122, no. 2, pp. 343-8, Feb 2012, doi: 10.1002/lary.22426.
 - [131] S.-H. Fang, C.-T. Wang, J.-Y. Chen, Y. Tsao, and F.-C. Lin, "Combining acoustic signals and medical records to improve pathological voice classification," *APSIPA Transactions on Signal and Information Processing*, vol. 8, 2019, doi: 10.1017/atsip.2019.7.
 - [132] U. Cesari, G. De Pietro, E. Marciano, C. Niri, G. Sannino, and L. Verde, "A new database of healthy and pathological voices," *Computers & Electrical Engineering*, vol. 68, pp. 310-321, 2018, doi: 10.1016/j.compeleceng.2018.04.008.

- 
- [133] N. Roy, R. M. Merrill, S. D. Gray, and E. M. Smith, "Voice disorders in the general population: prevalence, risk factors, and occupational impact," *Laryngoscope*, vol. 115, no. 11, pp. 1988-95, Nov 2005, doi: 10.1097/01.mlg.0000179174.32345.41.
 - [134] S. R. Best and C. Fakhry, "The prevalence, diagnosis, and management of voice disorders in a National Ambulatory Medical Care Survey (NAMCS) cohort," *Laryngoscope*, vol. 121, no. 1, pp. 150-7, Jan 2011, doi: 10.1002/lary.21169.
 - [135] J. E. Butler, T. H. Hammond, and S. D. Gray, "Gender-related differences of hyaluronic acid distribution in the human vocal fold," (in eng), *Laryngoscope*, vol. 111, no. 5, pp. 907-11, May 2001, doi: 10.1097/00005537-200105000-00029.
 - [136] P. D. Ward, S. L. Thibeault, and S. D. Gray, "Hyaluronic acid: its role in voice," *J Voice*, vol. 16, no. 3, pp. 303-9, Sep 2002. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12395982>.
 - [137] J. C. Stemple, N. Roy, and B. K. Klaben, *Clinical Voice Pathology Theory and Management*. Plural Publishing, 2014.
 - [138] S. Y. Tsui, Y. Tsao, C. W. Lin, S. H. Fang, F. C. Lin, and C. T. Wang, "Demographic and Symptomatic Features of Voice Disorders and Their Potential Application in Classification Using Machine Learning Algorithms," *Folia Phoniatr Logop*, vol. 70, no. 3-4, pp. 174-182, 2018, doi: 10.1159/000492327.
 - [139] J. I. Godino-Llorente, P. Gomez-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters," *IEEE Trans Biomed Eng*, vol. 53, no. 10, pp. 1943-53, Oct 2006, doi: 10.1109/TBME.2006.871883.
 - [140] M. K. Arjmandi and M. Pooyan, "An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine.," *Biomedical Signal Processing and Control*, vol. 7, no. 1, pp. 3-19, 2012.
 - [141] T.-M. Chen *et al.*, "SRECG: ECG Signal Super-resolution Framework for Portable/Wearable Devices in Cardiac Arrhythmias Classification," *IEEE Transactions on Consumer Electronics*, pp. 1-1, 2023, doi: 10.1109/tce.2023.3237715.