Contents

A. Dataset

I. Task 1: CVD Dataset

Population selection

Fig. A1 illustrates the population selection process for the CVD dataset used in Task 1. Patients were included if they had at least one hospitalization occurring within 90 days after an ECG measurement. Patients were excluded if they were under 18 years of age or over 89 years of age, or if their hospital stays were shorter than 24 hours. Records with admissions containing CVD diagnoses but without CVD as the primary diagnosis or operation were excluded, as the cause of admission could not be confirmed to be CVD-related.

After applying these criteria, the dataset comprised 44,790 CVD-related ECG samples from 13,289 patients, further categorized into coronary heart disease (CHD, N=18,445), stroke (N=4,927), and heart failure (HF, N=21,418). In addition, 125,987 non-CVD ECG samples from 52,388 patients were identified. The ICD code definitions for each CVD category are summarized in **Table A1**.

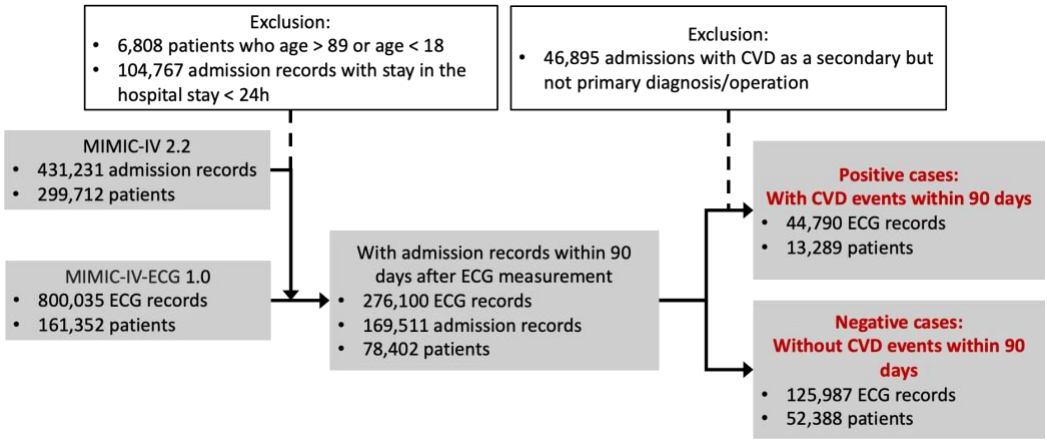

Fig. A1. Population selection process for Task 1 (CVD dataset). Focusing on patients with at least one hospitalization occurring within 90 days after an ECG measurement. The final dataset included 44,790 CVD-related and 125,987 non-CVD ECG samples.

Data description

1) **EHR Static Data:** Patient demographics included gender and age, extracted from the ***patient*** table. Latest outpatient measurements (systolic/diastolic blood pressure, weight, and height) were taken from the ***omr*** table. Medical history was defined based on ten CVD-related conditions reported in the literature [ref], with detailed definitions listed in **Table A1**. Medication history was derived from the ***prescriptions*** table, in which drugs were mapped to Anatomical Therapeutic Chemical (ATC) codes using the RxNorm API [1]. We selected ATC first-level class C (cardiovascular system) and grouped drugs by their third-level categories (e.g., C01A).
   Detailed variable summaries are provided in *DataDescriptions/CVD_Static.csv*, and the codebook is available at *DataDescriptions/Codebook.md*.
2) **Laboratory Results:** Eight laboratory tests relevant to CVD were selected according to prior studies [ref], including estimated glomerular filtration rate (eGFR), troponin T, creatine kinase,

creatine kinase-MB, serum creatinine, HDL cholesterol, LDL cholesterol, and total cholesterol. These values were obtained from the ***labevents*** table.

Detailed variable summaries are provided in *DataDescriptions/CVD_Labs.csv*, with the corresponding codebook in *DataDescriptions/Codebook.md*.

3) **ECG Signals, Text, and Features:** Raw 12-lead ECG recordings (500 Hz, 10 s) underwent the following preprocessing steps: (1) interpolation of missing values, (2) down-sampling to 125 Hz, (3) removal of noise and baseline wander, (4) application of a third-order Butterworth band-pass filter (0.5–40 Hz), and (5) segmentation into 5-second windows.

Besides raw signals, nine time-domain features (e.g., heart rate, PR interval) were got from machine-generated ECG reports. Detailed variable summaries are provided in *DataDescriptions/CVD_ECG.csv*.

Machine-generated ECG reports were further preprocessed and mapped to 143 SNOMED CT clinical terms [ref], providing structured and interpretable diagnostic judgments. Examples of the mapping are shown in **Fig. A2**, and the distribution of mapped samples is summarized in *DataDescriptions/ECG_Notes.csv*.

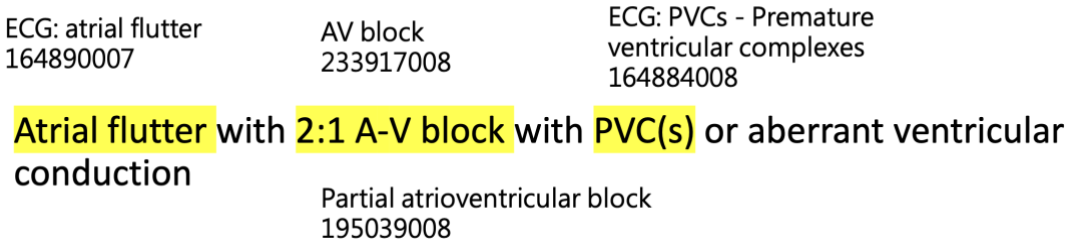

Fig. A2. Examples of machine-generated ECG reports mapped to SNOMED CT clinical terms. Highlighted terms indicate structured mappings such as atrial flutter, atrioventricular (AV) block, and premature ventricular complexes (PVCs), which facilitate interpretable representation of diagnostic findings.

TABLE A1
DEFINITIONS OF CVD CATEGORY AND MEDICAL HISTORY LABEL MAPPINGS

CVD Category	Medical History Label	ICD-9 Codes	ICD-10 Codes
CHD	CHD	410%, 411%, 4140%	I21%, I22%, I251%, I257%, I258%
CHD	CABG: Coronary artery bypass graft	361%, 362%	0210%, 0211%, 0212%, 0213%
CHD	PCI: Percutaneous coronary intervention	360%	0270%, 0271%, 0272%, 0273%
Stroke	Stroke	3466%, 433%, 434%, 436%, 4370%, 4371%	I63%, I65%, I66%
HF	HF	428%, 39891, 40201, 40211, 40291, 40401, 40403, 40411, 40413, 40491, 40493	I50%, I130%, I132%
	Hyperlipidemia	2720%, 2721%, 2722%, 2723%, 2724%	E780%, E781%, E782%, E783%, E784%, E785%
	Diabetes mellitus	250%	E10%, E11%, E13%

Atrial fibrillation	42731%	I480%, I482%
Hypertension	401%, 402%, 403%, 404%, 405%	I10%, I11%, I12%, I13%, I15%
Peripheral artery disease	4438%, 4439%	I738%, I739%

CABG and PCI were obtained from the ***procedures_icd*** table, while the other categories were derived according to ***diagnoses_icd*** table; "%" denotes wildcard matching of suffix chars.

II. TASK 2 & 3: MORTALITY AND LENGTH-OF-STAY (LOS) DATASET

Population selection

Fig. A3 shows the population selection process for the Mortality dataset. We included only the patient's first ICU admission and excluded records with ICU stays shorter than 24 hours. After applying these criteria, the dataset contained 40,167 records, of which 4,035 (10.04%) corresponded to in-hospital deaths.

The LOS dataset employed the same population as the Mortality dataset, with the length of ICU stay calculated for each patient. The average LOS was 4.05 days (SD = 5.19), and the distribution is shown in **Fig. A4**.

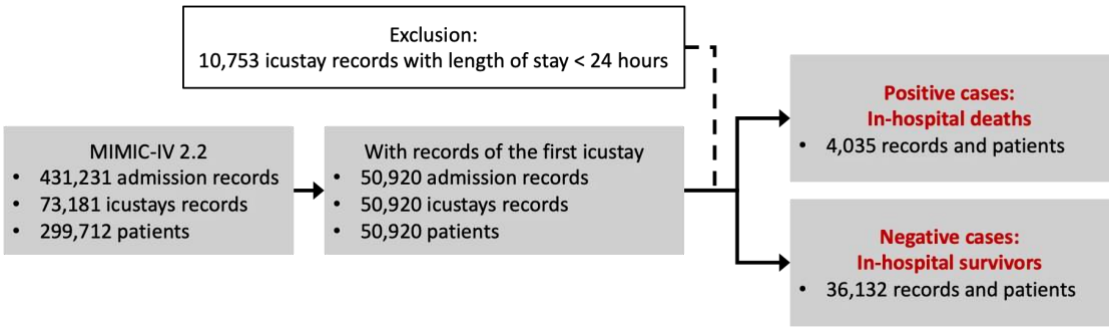

Fig. A3. Population selection process for Task 2 (in-hospital mortality). The final cohort included 40,167 first ICU admission records, with 4,035 mortality cases (10.04%).

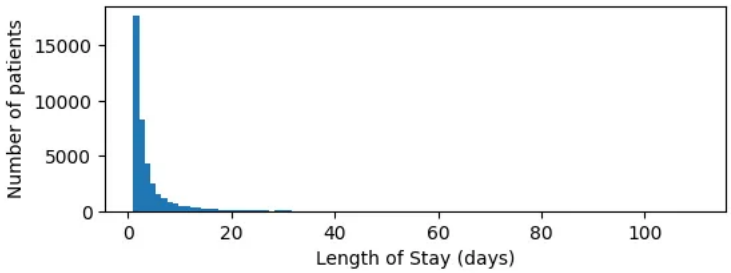

Fig. A4. Distribution of ICU length of stay (LOS) for Task 3. The mean LOS was 4.05 days with a standard deviation of 5.19 days.

Data description

1) **EHR Static Data:** Demographic information (gender, age) was extracted from the ***patient*** table, while admission details (admission type, admission location) were obtained from the ***admissions*** table.

Detailed variable summarize are available in *DataDescriptions/Mortality_Static.csv*, and the codebook is provided in *DataDescriptions/Codebook.md*.

2) **Vital Signs**: Hourly vital sign measurements were derived from the ***chartevents*** table. Only variables with less than 80% missingness were selected. Multiple ***itemid*** entries corresponding to the same variable (e.g., 220179, 220050, 224167, 227243 for systolic blood pressure) were merged into a single representative identifier (e.g., 220179). Detailed variable summaries are provided in *DataDescriptions/Mortality_Vitals.csv*.
3) **Laboratory Tests**: Laboratory test results were extracted from the ***chartevent*** and ***labevents*** tables. Only variables with less than 80% missingness were selected. Multiple ***itemid*** entries corresponding to the same variable were merged according to test item and specimen type (e.g., 50862 and 227456 for blood albumin were merged into 50862). Units were standardized before merging, and duplicate entries recorded at the same time were removed. Detailed variable summaries are provided in *DataDescriptions/Mortality_Labs.csv*.
4) **ECG Signals, Text, and Features**: The ECG preprocessing procedure was identical to Task 1. Detailed summaries of ECG features are provided in *DataDescriptions/Mortality_ECG.csv*.

B. Experimental Setup

I. Hyperparameters of MedM2T

Model Hyperparameters

Table B1 summarizes the model hyperparameters of MedM2T. When values differed across tasks, they are reported in the format **Task 1 / Task 2 / Task 3**.

Unimodal encoders:

1) **Static**: encoded by a multilayer perceptron (MLP).
2) **Labs**: encoded by the proposed sparse time-series encoder.
3) **Vitals**: integrating categorical and numerical vitals, using the same architecture as the multimodal encoder.
    a) Vitals (C): categorical vitals encoded with the sparse time-series encoder.
    b) Vitals (N): numerical vitals encoded with the hierarchical time-aware fusion model. This includes multi-scale high-frequency encoders (High-Freq 1 for a window size of 12, High-Freq 2 for a window size of 24) and a low-frequency encoder (Low-Freq).
4) **ECG**: encoded by the hierarchical time-aware fusion model, which integrates ECG-level representations across multiple time points.
    a) ECG (T): ECG text encoded with the embedding layer and MLP.
    b) ECG (S): ECG signals encoded with ResNet and MLP.
    c) ECG (F): ECG features encoded with MLP.
    d) ECG (Fusion): integration of ECG (T), ECG (S), and ECG (F) into ECG-level representations, using the same architecture as the multimodal encoder.

Multimodal encoder: composed of modality-specific encoders for each input type, a shared encoder, and the Bi-Modal attention modules.

Decoders: all implemented as a one-layer MLP.

Algorithm Hyperparameters

Table B2 lists the algorithm hyperparameters used during model training. An early stopping strategy was applied, with the maximum number of epochs set to 20.

TABLE B1
MODEL HYPERPARAMETERS OF MEDM2T

Component	Model	Parameter	Value
Static Encoder	MLP	Hidden Dim	[128, 128] / [128, 128] / [512, 512]
		Dropout	0.1
Static Decoder	MLP	Hidden Dim	[64] / [64] / [256]
		Dropout	0.5
Labs Encoder	Time Window Embedding Layer	Embed Dim	128 / 512 / 512
	Bi-LSTM	Hidden Dim	64 / 256 / 256
		Num Layers	2
	MLP	Hidden Dim	[128, 128] / [512, 512] / [512, 512]

		Dropout	0.05
Labs Decoder	MLP	Hidden Dim	[64] / [256] / [256]
		Dropout	0.1
Vitals Encoder	Shared Encoder (1-Layer MLP)	Embed Dim	512
		Dropout	0.05
	Bi-Modal Attention	Num Blocks	1
		Num Heads	64
		Dropout	0.1
Vitals Decoder	MLP	Hidden Dim	[512]
		Dropout	0.1
Vitals(C) Encoder	Time Window Embedding Layer	Embed Dim	256
	Bi-LSTM	Hidden Dim	128
		Num Layers	2
	MLP	Hidden Dim	[256, 256]
		Dropout	0.1
Vitals(C) Decoder	MLP	Hidden Dim	[128]
		Dropout	0.1
Vitals (N) Encoder	Self-Attention (High-Freq 1)	Embed Dim	8
		Num Heads	4
		Dropout	0.05
	ResNet (High-Freq 1)	Blocks Dim	[[16, 12], [32, 6]]
		Kernel Size	3
		Dropout	0.3
	MLP (High-Freq 1)	Hidden Dim	[128]
		Dropout	0.1
	Self-Attention (High-Freq 2)	Embed Dim	8
		Num Heads	4
		Dropout	0.05
	ResNet (High-Freq 2)	Blocks Dim	[[32, 24], [64, 12]]
		Kernel Size	3
		Dropout	0.3
	MLP (High-Freq 2)	Hidden Dim	[256]
		Dropout	0.1
	ResNet (Low-Freq)	Blocks Dim	[[384, 24], [192, 12]]
		Kernel Size	3
		Dropout	0.3
	MLP (Low-Freq)	Hidden Dim	[256]
		Dropout	0.1
Vitals (N) Decoder	MLP	Hidden Dim	[128]
		Dropout	0.25
ECG Encoder	ResNet	Blocks Dim	[[384, 5]]
		Kernel Size	3
		Dropout	0.3
	MLP	Hidden Dim	[256] / [256] / [512]
		Dropout	0.05
ECG Decoder	MLP	Hidden Dim	[128] / [128] / [256]
		Dropout	0.1
ECG (T) Encoder	Embedding Layer	Embed Dim	256
	MLP	Hidden Dim	[256, 256]
		Dropout	0.25
ECG (T) Decoder	MLP	Hidden Dim	[128]

		Dropout	0.25
ECG (S) Encoder	ResNet	Blocks Dim	[[64, 640], [128, 320], [196, 160], [256, 40], [320, 20]]
		Kernel Size	5
		Dropout	0.3
	MLP	Hidden Dim	[640]
		Dropout	0.1
ECG (S) Decoder	MLP	Hidden Dim	[320]
		Dropout	0.1
ECG (F) Encoder	MLP	Hidden Dim	[512] / [128] / [256]
		Dropout	0.05
ECG (F) Decoder	MLP	Hidden Dim	[256] / [64] / [128]
		Dropout	0.05
ECG (Fusion) Encoder	Shared Encoder (1-Layer MLP)	Embed Dim	256
		Dropout	0.05
	Bi-Modal Attention	Num Blocks	1
		Num Heads	64
		Dropout	0.1
ECG (Fusion) Decoder	MLP	Hidden Dim	[768]
		Dropout	0.1
Multimodal Encoder	Shared Encoder (1-Layer MLP)	Embed Dim	512
		Dropout	0.05
	Bi-Modal Attention	Num Blocks	1
		Num Heads	64
		Dropout	0.1
Multimodal Decoder	MLP	Hidden Dim	[1536] / [3840] / [3840]
		Dropout	0.1

When values differ across tasks, they are reported in the format Task 1 / Task 2 / Task 3; Vitals (C), Vitals (N) denote categorical and numerical vitals; ECG (T), ECG (S), ECG (F) denote ECG text, signals and features; ECG (Fusion) is a fusion of ECG text, signals and features.

TABLE B2

ALGORITHM HYPERPARAMETERS OF MEDM2T

Dataset	Parameter	Value
Static	Batch Size	128
	Optimizer	Adam
	Learning rate	0.0005 (core); 0.0001 (extended) / 0.0005 / 0.01
Labs	Batch Size	128
	Optimizer	Adam
	Learning rate	0.00005 / 0.0025 / 0.0025
Vitals	Batch Size	32
	Optimizer	Adam
	Learning rate	NA / 0.000001 / 0.00001
Vitals (C)	Batch Size	128
	Optimizer	Adam
	Learning rate	NA / 0.0005 / 0.001
Vitals (N)	Batch Size	128
	Optimizer	Adam
	Learning rate	NA / 0.001 / 0.0005
ECG	Batch Size	32
	Optimizer	Adam
	Learning rate	0.00005 / 0.00001 / 0.0005

ECG (T)	Batch Size	128
	Optimizer	Adam
	Learning rate	5E-05 / 0.0005 / 0.001
ECG (S)	Batch Size	128
	Optimizer	Adam
	Learning rate	0.005 / 0.001 / 0.0005
ECG (F)	Batch Size	128
	Optimizer	Adam
	Learning rate	0.0001 / 0.001 / 0.001
ECG (Fusion)	Batch Size	128
	Optimizer	SGD / Adam / Adam
	Learning rate	5E-05 / 0.001 / 5E-05
Multimodal	Batch Size	32 / 32 / 64
	Optimizer	Adam
	Learning rate	1E-06 (core); 5E-06 (extended) / 5E-05 / 1E-05

When values differ across tasks, they are reported in the format Task 1 / Task 2 / Task 3; Vitals (C), Vitals (N) denote categorical and numerical vitals; ECG (T), ECG (S), ECG (F) denote ECG text, signals and features; ECG (Fusion) is a fusion of ECG text, signals and features.

II. HYPERPARAMETERS OF COMPARED MODELS

We compared MedM2T against several state-of-the-art multimodal frameworks, including MultiBench, MultiModN, and HAIM. For MultiBench and MultiModN, most hyperparameters were adopted from their original configurations applied for the MIMIC datasets, with minor adjustments to hidden dimensions and learning rates. HAIM followed the hyperparameter tuning strategy recommended in the original work. Each modality's input type and encoder are summarized in **Table B3**, where time series (stats) refers to statistical feature extraction via the HAIM framework. MultiBench and MultiModN constructed multimodal learning frameworks using encoder–fusion–decoder architectures, whereas HAIM applied preprocessing followed by XGBoost for classification and regression.

For other compared models that did not explicitly provide hyperparameter settings, we used configurations consistent with MedM2T for similar data types and tasks, with additional tuning of learning rates.

TABLE B3

MODALITY TYPE AND ENCODER OF COMPARED MULTIMODAL FRAMEWORKS

Modality	**MultiBench**		**MultiModN**		**HAIM**	
	Type	**Encoder**	**Type**	**Encoder**	**Type**	**Model**
Static	Static	MLP	Static	MLP	Static	XGBoost
Labs	Time series	GRU	Time series (stats)		Time series (stats)	
Vitals (C)	Time series (stats)	MLP	Time series (stats)		Time series (stats)	
Vitals (N)	Time series	GRU	Time series (stats)		Time series (stats)	
ECG (T)	Static	MLP	Static		Static	
ECG (S)	Signal	ResNet	-		-	
ECG (F)	Static	MLP	Static		Static	

Time series (stats) denotes statistical features extracted from time series; Vitals (C), Vitals (N) denote categorical and numerical vitals; ECG (T), ECG (S), ECG (F) denote ECG text, signals, and features; ECG (Fusion) is a fusion of ECG text, signals, and features.

III. Sample Sizes of Tasks

For unimodal tasks, only the subset of samples with available data in the specific modality was used, whereas multimodal tasks utilized the full cohort. **Table B4** reports the number of samples for each task. For ECG-related modalities (ECG (T), ECG (S), ECG (F), and ECG (Fusion)), each ECG measurement record was treated as a separate training sample.

Task 1 is a multiclass classification problem. **Table B5** presents the distribution of samples across CVD categories for each modality.

ECG data were largely missing in Task 2 and Task 3, with only 46.7% sample having available ECG. In the ECG-available subset, only a few samples had over two ECG measurements in Task 2 and Task 3. **Table B6** summarizes the proportion of samples containing multiple ECG measurements in each task.

TABLE B4
SAMPLE SIZES OF TASKS

		Task 1	Task 2 & Task 3
Unimodal	Static	170777	40167
	Labs	167568	40039
	Vitals	-	40167
	Vitals (C)	-	40101
	Vitals (N)	-	40101
	ECG	170777	18750
	ECG (T)	294105	25213
	ECG (S)	294105	25213
	ECG (F)	294105	25213
	ECG (Fusion)	294105	25213
Multimodal		170777	40167

Vitals (C), Vitals (N) denote categorical and numerical vitals; ECG (T), ECG (S), ECG (F) denote ECG text, signals, and features; ECG (Fusion) is a fusion of ECG text, signals, and features.

TABLE B5
SAMPLE DISTRIBUTION ACROSS CVD CATEGORIES FOR TASK 1

	Static	Labs	ECG	Multimodal
non-CVD	125987	124353	125987	125987
CHD	18445	17385	18445	18445
Stroke	4927	4623	4927	4927
HF	21418	21207	21418	21418
SUM	170777	167568	170777	170777

TABLE B6
PROPORTION OF SAMPLES WITH MULTIPLE ECG MEASUREMENTS

k ECG	Task 1	Task 2 & Task 3
1	100%	100%
2	72%	26%

3	58%	6%
4	49%	2%
5	42%	1%
6	37%	0%
7	33%	0%
8	29%	0%
9	26%	0%
10	24%	0%

The percentage of samples containing more than k ECG records; the denominator is based on the ECG-available subset: Task 1 (N = 170,777) and Task 2/3 (N = 18,750).

C. Experimental Results

I. MedM2T Result

Table C1 presents the results of MedM2T under all unimodal and multimodal combinations across the three tasks, providing supplementary details corresponding to **Table IV** in the main manuscript.

The distinction between multimodal combinations highlights the diverse contributions of each modality. The exclusion of laboratory tests led to the largest performance drop in Task 1 and Task 2, whereas the exclusion of vital signs caused the greatest decline in Task 3. Notably, for Task 2, vitals achieved the best unimodal performance, whereas the exclusion of laboratory tests caused the largest degradation, suggesting that laboratory tests provide complementary and distinctive information when combined with other modalities.

TABLE C1

PERFORMANCE OF MEDM2T ACROSS UNIMODAL AND MULTIMODAL

		Task 1 (CVD)		**Task 2 (Mortality)**		**Task 3 (LOS)**	
		AUROC ↑	*AUPRC* ↑	*AUROC* ↑	*AUPRC* ↑	*MAE* ↓	*MSE* ↓
Unimodal	Static	0.717 / 0.846*	0.362 / 0.546*	0.678	0.170	2.89	26.53
	Labs	0.870	0.568	0.825	0.384	2.55	22.90
	Vitals	-	-	0.832	0.410	2.45	21.17
	Vitals (C)	-	-	0.657	0.172	2.85	26.34
	Vitals (N)	-	-	0.814	0.380	2.50	22.47
	ECG	0.846	0.540	0.734	0.212	2.77	25.95
	ECG (T)	0.700	0.400	0.667	0.166	2.85	26.70
	ECG (S)	0.764	0.461	0.720	0.199	2.77	24.47
	ECG (F)	0.683	0.380	0.648	0.161	2.85	26.16
	ECG (Fusion)	0.833	0.581	0.730	0.223	2.69	25.75
Multimodal	Static + Labs	0.922 / 0.946*	0.624 / 0.670*	0.885	0.505	2.50	21.21
	Static + Vitals	-	-	0.847	0.431	2.41	21.17
	Static + ECG	0.876 / 0.904*	0.582 / 0.627*	0.681	0.168	2.95	26.39
	Labs + Vitals	-	-	0.894	0.545	2.34	20.00
	Labs + ECG	0.931	0.668	0.877	0.496	2.51	21.31
	Vitals + ECG	-	-	0.831	0.407	2.43	21.37
	Static + Labs + Vitals	-	-	0.896	0.549	2.34	19.94
	Static + Labs + ECG	**0.940 / 0.947***	**0.686 / 0.706***	0.882	0.500	2.50	21.26
	Static + Vitals + ECG	-	-	0.847	0.428	2.42	21.16
	Labs + Vitals + ECG	-	-	0.892	0.543	2.34	20.08
	Static + Labs + Vitals + ECG	-	-	**0.901**	**0.558**	**2.31**	**19.98**

Asterisk (*) denotes using static core subset / extended subset; Vitals (C), Vitals (N) denote categorical and numerical vitals; ECG (T), ECG (S), ECG (F) denote ECG text, signals and features; ECG (Fusion) is a fusion of ECG text, signals and features. The number of samples for each modality is provided in Table B4.

Table C2 provides the detailed results of Task 1 (CVD prediction) across four categories: non-CVD, CHD, stroke, and HF, serving as supplementary details corresponding to **Fig. 5** in the main manuscript.

For unimodal, the static extended subset (including medical history and medications) achieved substantially better performance than the core subset across all classes. As more modalities were integrated, the performance gap between the core and extended subsets diminished, suggesting complementary information across modalities.

TABLE C2

PERFORMANCE OF MEDM2T FOR TASK 1 (CVD PREDICTION) ACROSS FOUR CATEGORIES

		non-CVD	CHD	Stroke	HF	Macro-AVG
		AUROC / AUPRC ↑				
Unimodal	S (c)	0.738 / 0.886	0.702 / 0.198	0.682 / 0.054	0.747 / 0.311	0.717 / 0.362
	S (e)	0.890 / 0.948	0.842 / 0.470	0.745 / 0.083	0.910 / 0.684	0.846 / 0.546
	L	0.909 / 0.959	0.889 / 0.571	0.784 / 0.124	0.896 / 0.619	0.870 / 0.568
	E	0.896 / 0.955	0.842 / 0.436	0.714 / 0.060	0.933 / 0.708	0.846 / 0.540
Multimodal	S (c) + L	0.958 / 0.984	0.938 / 0.664	0.868 / 0.178	0.924 / 0.669	0.922 / 0.624
	S (e) + L	0.980 / 0.992	0.952 / 0.717	0.893 / 0.206	0.957 / 0.765	0.946 / 0.670
	S (c) + E	0.918 / 0.966	0.874 / 0.510	0.761 / 0.078	0.950 / 0.774	0.876 / 0.582
	S (e) + E	0.944 / 0.976	0.911 / 0.616	0.802 / 0.110	0.959 / 0.804	0.904 / 0.627
	L + E	0.965 / 0.986	0.941 / 0.700	0.856 / 0.167	0.963 / 0.817	0.931 / 0.668
	S (c) + L + E	0.973 / 0.989	0.949 / 0.729	0.871 / 0.194	0.968 / 0.834	0.940 / 0.686
	S (e) + L + E	0.977 / 0.991	0.956 / 0.761	0.883 / 0.227	0.971 / 0.843	0.947 / 0.706

S (c) and S (e) denote using the core subset / extended subset; L denotes Labs; E denotes ECG.
The number of samples for each category is provided in Table B5.

II. BI-MODAL ATTENTION ABLATION STUDY IN TASK 1

Table C3 provides an ablation study of the Bi-Modal Attention module in Task 1 (CVD prediction). As reported in the main manuscript (**Table V**), Bi-Modal Attention enhances multimodal learning performance. Here, we further validate its effect across different multimodal combinations, a random subset of the multiclass task (N=50,000), and a binary classification task (CVD vs. non-CVD, N=50,000). In all cases, incorporating Bi-Modal Attention consistently improved both AUROC and AUPRC compared with models trained without it.

TABLE C3

ABLATION STUDY OF BI-MODAL ATTENTION IN TASK 1

		Bi-Modal Attention	AUROC ↑	AUPRC ↑
Full Dataset: Multiclass Task (N = 170777)	Static + Labs	w/	0.922	0.624
		w/o	0.894	0.608
	Static + ECG	w/	0.876	0.582
		w/o	0.870	0.581
	Labs + ECG	w/	0.931	0.668
		w/o	0.925	0.674
	Static + Labs +ECG	w/	0.940	0.686
		w/o	0.928	0.681
Random Subset: Multiclass Task (N = 50000)	Static + Labs + ECG	w/	0.936	0.681
		w/o		

Random Subset: Binary Task (N = 50000)	Static + Labs + ECG	w/	0.969	0.932
		w/o	0.961	0.923

Static is the core subset.

III. Performance of Comparative Frameworks

Table C4 compares the performance of MedM2T with other state-of-the-art multimodal frameworks (MultiBench, MultiModN, and HAIM) under unimodal and multimodal across the three tasks. This table provides supplementary details corresponding to **Table V** in the main manuscript. For MultiBench, MultiModN, and HAIM, we adopted the encoder configurations recommended in their original implementations. The input data types and encoders for each modality are summarized in **Table B3**.

Across all three tasks, MedM2T got the best results under multimodal integration, while HAIM showed competitive performance in unimodal settings with static or laboratory data.

TABLE C4

Comparison of MedM2T with other multimodal frameworks

		MedM2T	MultiBench	MultiModN	HAIM
			AUROC / AUPRC ↑		
Task 1 (CVD)	Static (core)	0.717 / 0.362	0.711 / 0.358	0.713 / 0.359	**0.718 / 0.364**
	Static (extended)	0.846 / 0.546	0.847 / 0.552	0.840 / 0.533	**0.853 / 0.557**
	Labs	**0.870** / 0.568	0.852 / 0.542	0.829 / 0.511	0.864 / **0.574**
	ECG	**0.846 / 0.540**	0.819 / 0.511	0.770 / 0.433	0.783 / 0.452
	Multimodal (core)	**0.940 / 0.686**	0.897 / 0.621	0.871 / 0.573	0.853 / 0.557
	Multimodal (extended)	**0.947 / 0.706**	0.916 / 0.649	0.889 / 0.593	0.899 / 0.633
			AUROC / AUPRC ↑		
Task 2 (Mortality)	Static	**0.678 / 0.170**	0.671 / 0.164	0.677 / 0.166	0.672 / 0.165
	Labs	0.825 / 0.384	0.812 / 0.384	0.831 / 0.409	**0.858 / 0.470**
	Vitals	**0.832 / 0.410**	0.781 / 0.345	0.774 / 0.314	0.822 / 0.399
	ECG	**0.734 / 0.212**	0.691 / 0.183	0.679 / 0.179	0.688 / 0.176
	Multimodal	**0.901 / 0.558**	0.833 / 0.418	0.856 / 0.409	0.890 / 0.540
			MAE / MSE ↓		
Task 3 (LOS)	Static	**2.89 / 26.53**	2.92 / 26.90	2.96 / 26.92	2.91 / 27.03
	Labs	2.55 / 22.90	2.96 / 26.91	2.95 / 26.99	**2.53 / 22.47**
	Vitals	**2.45 / 21.17**	2.92 / 26.56	2.96 / 26.92	2.51 / 22.67
	ECG	**2.77 / 25.95**	2.86 / 26.31	2.88 / 26.62	2.82 / 26.33
	Multimodal	**2.31 / 19.98**	2.93 / 26.82	2.95 / 26.92	2.43 / 21.89

Bold shows the best result; MultiBench reports the best multimodal results based on either LR or LRTF.
The number of samples for each modality is provided in Table B4.

REFERENCE