



Hyun Ryu¹, Cho Rong Bae^{1*}, Hyo Kyung Lee², Yongdoo Park³, You Ha Kwon¹, Kyung Cheon Seo¹, Jeong Eul Kwon², Jong Hoon Kim², Sejeong Jang², Bo Ryun Kim^{1*}

¹Department of Physical Medicine and Rehabilitation, Korea University College of Medicine, Seoul, Republic of Korea

²Department of Industrial Management Engineering, Korea University College of Engineering, Seoul, Republic of Korea

³Department of Biomedical Engineering, Korea University College of Medicine, Seoul, Republic of Korea

Introduction

- Cardiopulmonary exercise testing (CPET) is the gold standard for prescribing exercise intensity in cardiac rehabilitation (CR), but many patients—especially elderly or frail—cannot complete maximal tests.

Objective

- Develop and evaluate a machine learning model that combines early CPET patterns (shapelets) with clinical data to **predict individualized HR and MET ranges** for exercise prescription.

Method

I. Study Population

- Retrospective cohort: 336 patients, 826 CPET tests (2020–2024)
- After clinical data matching: 310 patients, 462 CPET tests analyzed

II. Data Used

- First 1.5–2 min of CPET (51 variables)

III. Study Workflow (3 Modules)

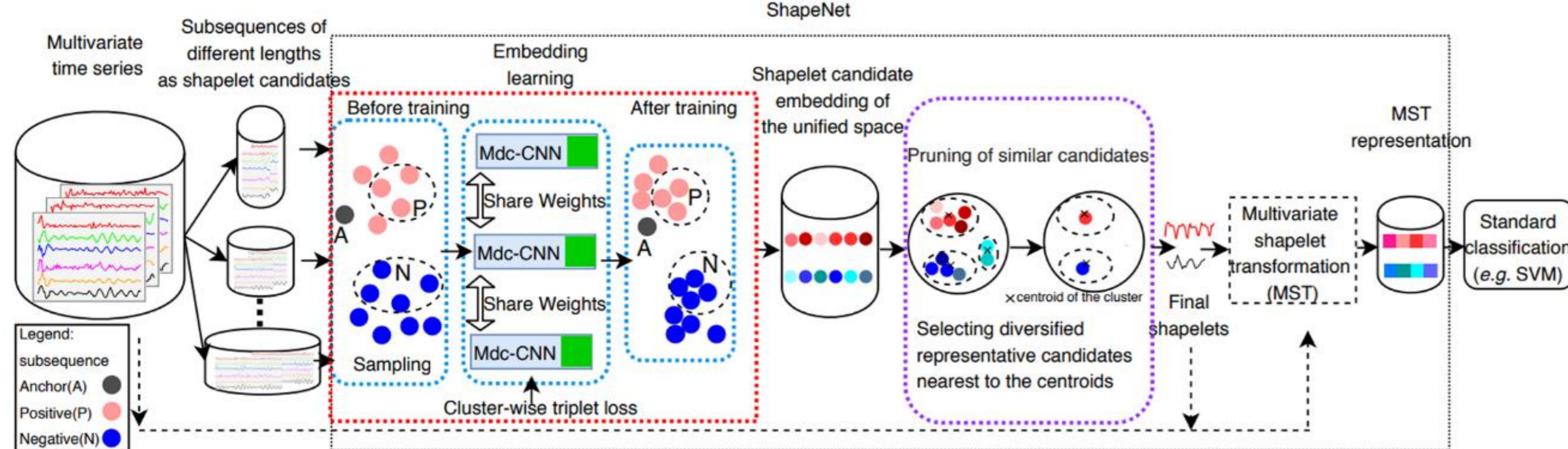


Figure 1. Study Framework for AI-Based Early CPET-Integrated Exercise Intensity Prediction

1. Class Definition (Physiological Phenotypes)

- Cluster patients using VO_{2max} , HR_{max} , MET_{max} , Fat_{max}
- Two classes:
 - Low-performance (below-average metrics)
 - High-performance (above-average metrics)

2. Shapelet Extraction & Selection

- Shapelets = distinctive short-term patterns in CPET time-series
- Extracted via ShapeNet representation learning
- Selected based on ability to distinguish classes
- Patient data transformed into feature vectors using minimum distance to shapelets

3. Exercise Intensity Prediction

- Combine shapelet features with clinical/demographic data
- Machine learning models: CatBoost, Random Forest, XGBoost, LightGBM, HistGB
- 3-fold cross-validation
- Model types: Shapelet-Only, Medical-Only, Integrated (Shapelet + Clinical)

IV. Outcomes & Metrics

- Predicted: HR upper/lower limits, MET upper/lower limits
- Performance metrics: MAE, RMSE, MAPE

Result

I. Optimal shapelets: 20 for best class discrimination

II. Prediction Performance

- Integrated model **outperformed** Shapelet-only and Medical-only models across all experiments

Table 1. Prediction Performance of Exercise Intensity Range in Integrated Model

	Upper Limit MAE	Lower Limit MAE
HR (bpm)	10-11	9-10
MET	0.7-0.9	0.6-0.8

Setting	Model	HR Upper			HR Lower			# of features
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	
Shapelet Only	CatBoost	12.74 ± 0.36	16.54 ± 0.48	10.96 ± 0.08	11.72 ± 0.07	15.92 ± 0.78	12.53 ± 1.35	20
	HistGB	13.71 ± 0.31	17.59 ± 0.22	11.76 ± 0.04	12.40 ± 0.45	16.78 ± 0.63	13.22 ± 1.31	
	LightGBM	13.46 ± 0.29	17.37 ± 0.23	11.55 ± 0.15	12.38 ± 0.29	16.68 ± 0.37	13.18 ± 1.14	
	RandomForest	12.15 ± 0.45	15.93 ± 0.69	10.48 ± 0.27	11.26 ± 0.16	15.44 ± 0.87	12.12 ± 1.36	
	XGBoost	13.20 ± 0.22	16.89 ± 0.25	11.36 ± 0.16	11.96 ± 0.22	16.11 ± 0.61	12.69 ± 1.28	
	Average	13.05 ± 0.33	16.86 ± 0.37	11.22 ± 0.14	11.94 ± 0.24	16.19 ± 0.65	12.75 ± 1.29	
Medical Only	CatBoost	11.59 ± 1.05	15.55 ± 1.46	9.90 ± 1.09	10.60 ± 0.85	15.01 ± 1.69	11.42 ± 2.08	170
	HistGB	12.05 ± 1.27	16.26 ± 1.63	10.29 ± 1.22	11.18 ± 0.71	15.72 ± 1.78	12.05 ± 2.04	
	LightGBM	11.80 ± 1.09	16.05 ± 1.40	10.06 ± 1.07	10.73 ± 0.73	15.34 ± 1.61	11.63 ± 2.06	
	RandomForest	11.38 ± 1.07	15.40 ± 1.51	9.81 ± 1.11	10.54 ± 0.86	14.88 ± 1.72	11.43 ± 2.10	
	XGBoost	11.82 ± 1.30	16.01 ± 1.66	10.10 ± 1.27	10.92 ± 0.95	15.42 ± 1.95	11.69 ± 2.16	
	Average	11.73 ± 1.16	15.86 ± 1.53	10.03 ± 1.15	10.80 ± 0.82	15.27 ± 1.75	11.65 ± 2.09	
Shapelet + Medical	CatBoost	11.15 ± 0.78	15.11 ± 1.34	9.58 ± 0.91	10.31 ± 0.51	14.50 ± 1.47	11.13 ± 1.73	190
	HistGB	11.16 ± 0.81	15.08 ± 1.40	9.54 ± 0.89	9.90 ± 0.52	14.33 ± 1.48	10.82 ± 1.76	
	LightGBM	10.98 ± 0.67	14.93 ± 1.23	9.38 ± 0.78	10.02 ± 0.51	14.38 ± 1.45	10.89 ± 1.75	
	RandomForest	10.73 ± 0.65	14.65 ± 1.19	9.21 ± 0.81	9.95 ± 0.47	14.30 ± 1.39	10.90 ± 1.70	
	XGBoost	10.82 ± 0.64	14.51 ± 1.12	9.23 ± 0.79	9.79 ± 0.53	14.16 ± 1.45	10.61 ± 1.70	
	Average	10.97 ± 0.71	14.86 ± 1.26	9.39 ± 0.84	9.99 ± 0.51	14.33 ± 1.45	10.87 ± 1.73	

Setting	Model	METs Upper			METs Lower			# of features
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	
Shapelet Only	CatBoost	0.96 ± 0.08	1.40 ± 0.27	19.72 ± 0.76	0.86 ± 0.06	1.25 ± 0.23	21.81 ± 0.25	20
	HistGB	1.01 ± 0.06	1.47 ± 0.22	21.04 ± 1.16	0.94 ± 0.07	1.33 ± 0.16	23.91 ± 1.21	
	LightGBM	0.99 ± 0.08	1.47 ± 0.21	20.57 ± 1.69	0.94 ± 0.08	1.37 ± 0.15	23.76 ± 1.93	
	RandomForest	0.96 ± 0.07	1.43 ± 0.34	20.25 ± 0.24	0.91 ± 0.08	1.40 ± 0.36	23.74 ± 0.51	
	XGBoost	1.00 ± 0.09	1.48 ± 0.35	21.10 ± 0.85	0.95 ± 0.07	1.46 ± 0.29	24.70 ± 0.37	
	Average	0.98 ± 0.08	1.45 ± 0.28	20.54 ± 0.94	0.92 ± 0.07	1.36 ± 0.24	23.58 ± 0.86	
Medical Only	CatBoost	0.86 ± 0.03	1.24 ± 0.26	18.62 ± 1.24	0.77 ± 0.03	1.11 ± 0.20	20.61 ± 2.01	170
	HistGB	0.88 ± 0.05	1.29 ± 0.30	19.08 ± 1.39	0.79 ± 0.04	1.18 ± 0.21	20.98 ± 0.90	
	LightGBM	0.90 ± 0.05	1.30 ± 0.29	19.43 ± 1.01	0.78 ± 0.03	1.18 ± 0.21	20.28 ± 0.79	
	RandomForest	0.90 ± 0.07	1.31 ± 0.34	19.51 ± 1.13	0.84 ± 0.07	1.22 ± 0.21	22.65 ± 3.27	
	XGBoost	0.88 ± 0.03	1.28 ± 0.27	18.90 ± 1.19	0.77 ± 0.05	1.16 ± 0.18	20.35 ± 2.97	
	Average	0.88 ± 0.05	1.28 ± 0.29	19.11 ± 1.19	0.79 ± 0.05	1.17 ± 0.20	20.98 ± 1.99	
Shapelet + Medical	CatBoost	0.70 ± 0.06	1.08 ± 0.33	14.39 ± 0.39	0.62 ± 0.05	0.95 ± 0.25	15.90 ± 1.44	190
	HistGB	0.69 ± 0.05	1.08 ± 0.30	14.29 ± 0.28	0.68 ± 0.04	1.03 ± 0.17	17.37 ± 0.97	
	LightGBM	0.70 ± 0.03	1.11 ± 0.29	14.44 ± 0.44	0.68 ± 0.03	1.03 ± 0.18	17.48 ± 1.22	
	RandomForest	0.76 ± 0.08	1.14 ± 0.32	15.84 ± 0.85	0.71 ± 0.07	1.04 ± 0.24	18.30 ± 1.38	
	XGBoost	0.70 ± 0.08	1.10 ± 0.35	14.31 ± 0.82	0.68 ± 0.04	1.14 ± 0.21	16.91 ± 1.31	
	Average	0.71 ± 0.06	1.10 ± 0.32	14.65 ± 0.56	0.67 ± 0.05	1.04 ± 0.21	17.19 ± 1.27	

Figure 2. Prediction Performance of Exercise Intensity Range

III. Feature Importance

- Early VO_2 dynamics
- HR trajectory
- VE/VO_2
- Fat oxidation-related variables

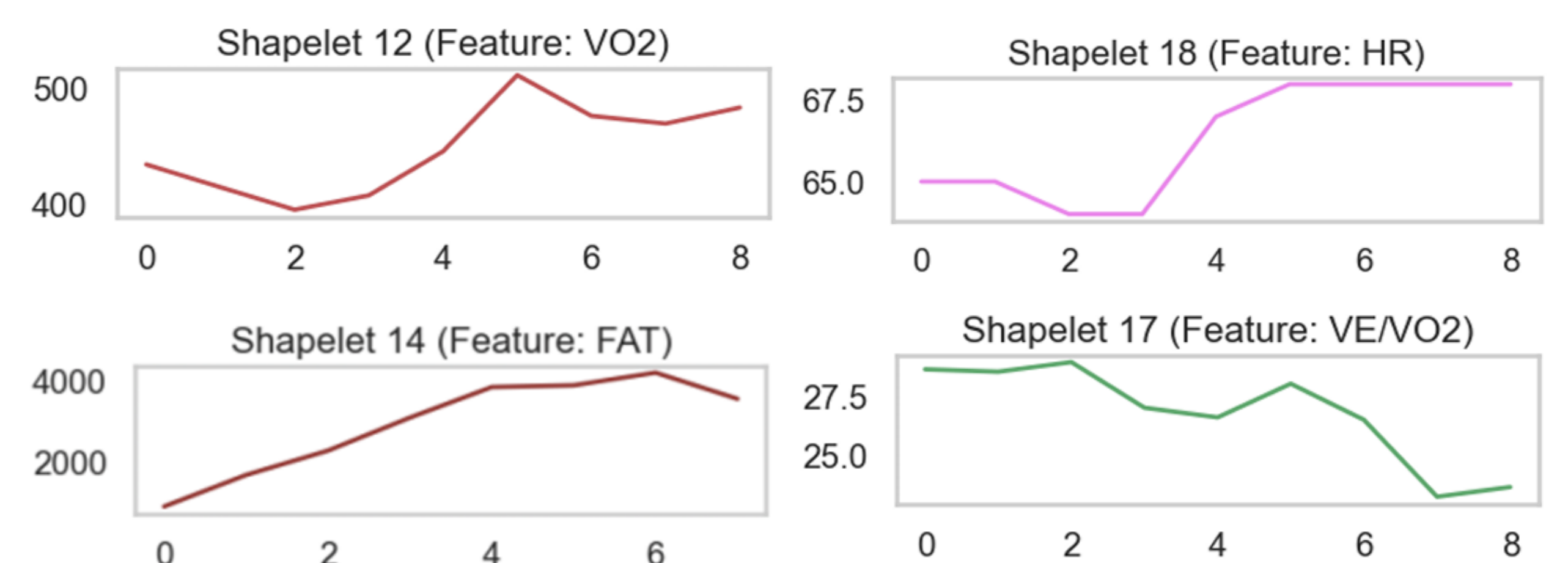


Figure 3. Representative Shapelets Derived from the Initial CPET Segment (VO_2 , HR, Fat Oxidation VE/VO_2)

Conclusion

- Physiological signatures derived from the **initial 1.5–2-minute segment of CPET** contain clinically meaningful information.
- Shapelet-based time-series analysis integrated with clinical data **enables accurate estimation of individualized exercise intensity ranges**, even in patients unable to complete maximal CPET.
- This approach may facilitate precision cardiac rehabilitation and development of real-time clinical decision-support system

