

Supplemental material for “Myocardial Infarction Detection with Incomplete Multi-View Data via Dual-Branch Gating Completion and Dirichlet Weighting”

Yadi Wang, Yulin Xie, Yi Xie, Lin Chen and Bingbing Jiang

APPENDIX I PROOF OF PROPOSITION 2

Notation. Before providing the proof, we briefly clarify the notations used below for completeness and reproducibility. $\hat{E}(f)$ denotes the empirical risk (average training loss) of the model f . $\mathcal{R}_v(f)$ represents the Rademacher complexity of the classifier corresponding to the v -th view, which measures model capacity. w^v is the fusion weight of the v -th view, and l^v is its corresponding loss value. $\mathbb{E}(\cdot)$ denotes the expectation operator. $\text{Cov}(\cdot, \cdot)$ represents the covariance, and $r(\cdot, \cdot)$ is the Pearson correlation coefficient.

For Eq. (16), the proof is as follows:

First, to prove Eq. (16), we can split it into two parts and compare them with static fusion respectively:

$$\sum_{v=1}^V \mathbb{E}(w_{\text{dynamic}}^v) \hat{E}(f_{\text{dynamic}}) \leq \sum_{v=1}^V w_{\text{static}}^v \hat{E}(f_{\text{static}}) \quad (1)$$

$$\sum_{v=1}^V \mathbb{E}(w_{\text{dynamic}}^v) \mathcal{R}_v(f_{\text{dynamic}}) \leq \sum_{v=1}^V w_{\text{static}}^v \mathcal{R}_v(f_{\text{static}}) \quad (2)$$

In theory, if the same type of function is effectively optimized, the same empirical risk should be obtained on the training data, and the intrinsic complexity of the single-view classifier $\mathcal{R}_v(f^v)$ is also unchanged, that is:

$$\hat{E}(f_{\text{dynamic}}) = \hat{E}(f_{\text{static}}) \quad (3)$$

$$\mathcal{R}_v(f_{\text{dynamic}}) = \mathcal{R}_v(f_{\text{static}}) \quad (4)$$

This work was supported by the Henan Province Science Foundation of Excellent Young Scholars (Grant 242300421171), the National Natural Science Foundation of China (Grant 62106066), the Postgraduate Education Reform and Quality Improvement Project of Henan Province (Grant YJS2024JD30, YJS2023JD28), the Henan Province University Science and Technology Innovation Team Support Plan (Grant 24IRT-STHN021) (*Corresponding authors: Yi Xie.*)

Yadi Wang, Yulin Xie, Yi Xie and Lin Chen are with the Henan Key Laboratory of Big Data Analysis and Processing, and also with the School of Computer and Information Engineering, Henan University, Kaifeng 475004, China (e-mail: yadiwang@henu.edu.cn; xieyulin@henu.edu.cn; xieyi@henu.edu.cn; LinChen@henu.edu.cn).

Bingbing Jiang is with the Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China, and also with the School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China (e-mail: jiangbb@hznu.edu.cn).

Then Eq. (1) and Eq. (2) can be simplified to prove only Eq. (5):

$$\mathbb{E}(w_{\text{dynamic}}^v) \leq w_{\text{static}}^v \quad (5)$$

From the above method, we can know that the weight construction method is $w^v = \lambda \frac{b^v b^{v\top}}{u^v + 1}$, and the weights of all views are multiplied by λ , but the weight ratio between views remains unchanged, that is, the weight of the view input with high confidence is still greater than the weight of the view input with low confidence, so the size of λ does not affect the experimental results. Therefore, there must be a λ^* , so that when $\lambda < \lambda^*$, $\mathbb{E}(w_{\text{dynamic}}^v) \leq w_{\text{static}}^v$, and Eq. (16) is proved. The proof of Eq. (17) is based on the definition of covariance, given by $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. Therefore, we only need to prove that $\text{Cov}(w_{\text{dynamic}}^v, l^v) < 0$. From the weight construction method above, we can see that the dynamic weight is negatively correlated with uncertainty. The greater the uncertainty, the greater the loss. Therefore, uncertainty and loss are positively correlated, which can be expressed as

$$r(u_v, \ell^v(x)) > 0 \quad (6)$$

It can be deduced that the dynamic weight and loss are negatively correlated, that is,

$$r(w_{\text{dynamic}}^v, \ell^v(x)) < 0 \quad (7)$$

where r is the Pearson correlation coefficient. Since the calculation formula of the Pearson correlation coefficient is $r(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$, where σ_x and σ_y are standard deviations and are both greater than 0, the covariance and the correlation coefficient have the same sign. When $r(w_{\text{dynamic}}^v, \ell^v(x)) < 0$, the covariance $\text{Cov}(w_{\text{dynamic}}^v, l^v) < 0$, and Eq. (17) is proved.

APPENDIX II SUPPLEMENTARY EXPERIMENTAL RESULTS

A. Statistical Significance Analysis of Model Performance

To further evaluate model performance, paired t -tests were conducted on the results of ten independent runs, as reported in Table I. The p -values indicate the statistical significance of pairwise comparisons, while $h = 1$ denotes a significant difference at $p < 0.05$. It should be noted that statistical

TABLE I: Pairwise paired t -test results of different algorithms across six evaluation metrics. Reported values are p -values, with statistically significant results highlighted in bold ($p < 0.05$).

Algorithm1	Algorithm2	ACC		F1		F2		Recall		Specificity		Precision	
		p -value	h(0/1)	p -value	h(0/1)	p -value	h(0/1)	p -value	h(0/1)	p -value	h(0/1)	p -value	h(0/1)
D2GCDW	DIMvLN	8.80e-05	1	0.00011	1	0.00011	1	0.00657	1	0.20833	0	0.00203	1
D2GCDW	LHGN	5.67e-06	1	1.50e-05	1	1.50e-05	1	8.13e-05	1	0.00549	1	0.00015	1
D2GCDW	DCP	1.17e-05	1	5.40e-12	1	5.40e-12	1	2.07e-10	1	0.00073	1	1.21e-09	1
D2GCDW	CPM-Nets	5.05e-07	1	5.27e-07	1	5.27e-07	1	0.19933	0	0.01085	1	6.46e-07	1
DIMvLN	LHGN	0.29232	0	0.50552	0	0.50552	0	8.11e-05	1	0.07479	0	0.10694	0
DIMvLN	DCP	0.00845	1	9.42e-09	1	9.42e-09	1	8.89e-09	1	0.00170	1	1.27e-09	1
DIMvLN	CPM-Nets	0.00955	1	0.15325	0	0.15325	0	0.02335	1	0.22916	0	0.38868	0
LHGN	DCP	0.28416	0	3.96e-09	1	3.96e-09	1	5.70e-08	1	0.00164	1	1.27e-07	1
LHGN	CPM-Nets	0.25495	0	0.25495	0	0.25495	0	0.12068	0	0.62222	0	0.24166	0
DCP	CPM-Nets	0.77562	0	4.17e-09	1	4.17e-09	1	3.64e-05	1	0.00201	1	2.24e-08	1

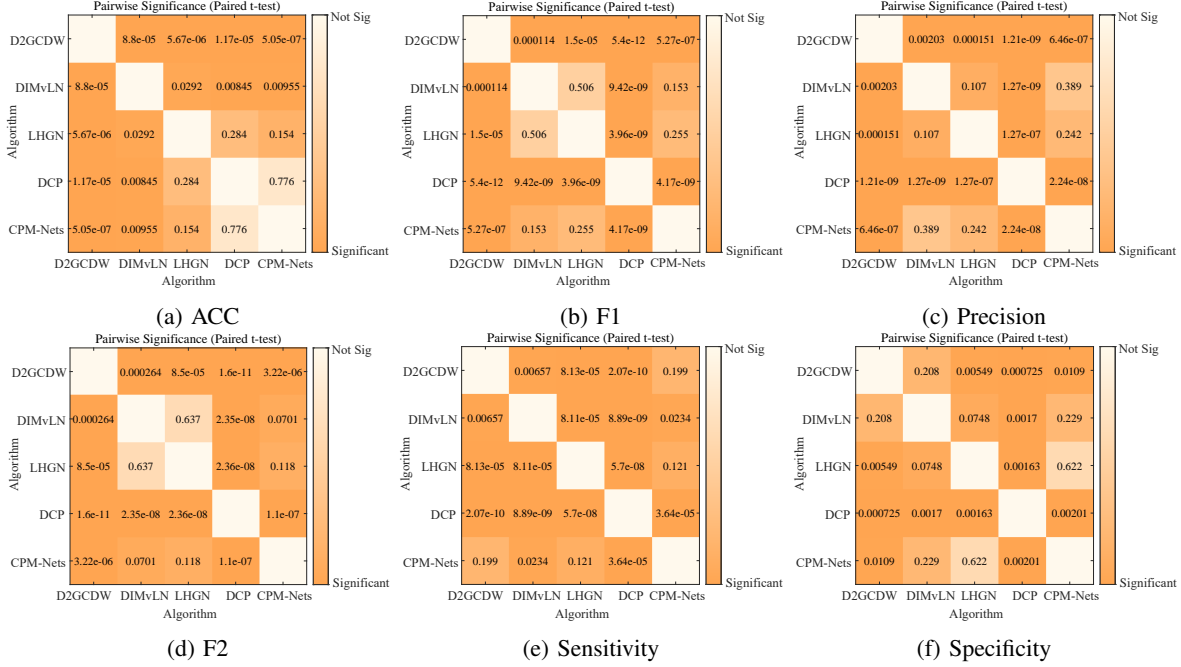


Fig. 1: Heatmaps of pairwise t -test results (p -values) for five models across six evaluation metrics.

significance alone does not imply superiority and should be interpreted together with the direction of the difference and the mean performance. As shown in Table I, D2GCDW achieves the highest mean values and shows statistically significant improvements over all baseline methods in terms of ACC, F1, F2, and Precision. In contrast, the inferior performance of DCP also leads to significant differences when compared with other methods, primarily reflecting its limitations. Overall, the superior performance of D2GCDW is supported by both higher mean values and statistical significance, indicating its robustness and practical effectiveness.

Figure 1 presents heatmaps of pairwise t -test p -values for five models across six evaluation metrics, where darker colors indicate stronger statistical significance. As shown, D2GCDW exhibits significant performance differences compared with baseline methods in most metrics, indicating its robust performance. In contrast, the significant differences associated with DCP primarily reflect its lower performance rather than an advantage.