

---

# Improving Evolutionary Multi-View Classification via Eliminating Individual Fitness Bias

---

Xinyan Liang<sup>1,2</sup>, Shuai Li<sup>1</sup>, Qian Guo<sup>3</sup>, Yuhua Qian<sup>1\*</sup>, Bingbing Jiang<sup>4</sup>,  
Tingjin Luo<sup>5</sup>, Liang Du<sup>1</sup>

<sup>1</sup> Institute of Big Data Science and Industry, Shanxi University

<sup>2</sup> State Key Laboratory of AI Safety, Beijing, 100086

<sup>3</sup> School of Computer Science and Technology, Taiyuan University of Science and Technology

<sup>4</sup> School of Information Science and Technology, Hangzhou Normal University

<sup>5</sup> College of Science, National University of Defense Technology

{liangxinyan48, LiShuai\_liuzhaona, czguoqian}@163.com, jinchengqyh@126.com  
jiangbb@hznu.edu.cn, tingjinluo@hotmail.com, duliang@sxu.edu.cn

## Abstract

Evolutionary multi-view classification (EMVC) methods have gained wide recognition due to their adaptive mechanisms. Fitness evaluation (FE), which aims to calculate the classification performance of each individual in the population and provide reliable performance ranking for subsequent operations, is a core step in such methods. Its accuracy directly determines the correctness of the evolutionary direction. That is, when FE fails to correctly reflect the superiority-inferiority relationship among individuals, it will lead to confusion in individual performance ranking, which in turn misleads the evolutionary direction and results in trapping into local optima. This paper is the first to identify the aforementioned issue in the field of EMVC and call it as fitness evaluation bias (FEB). FEB may be caused by a variety of factors, and this paper approaches the issue from the perspective of view information content: existing methods generally adopt joint training strategies, which restrict the exploration of key information in views with low information content. This makes it difficult for multi-view model (MVM) to achieve optimal performance during convergence, which in turn leads to FE failing to accurately reflect individual performance rankings and ultimately triggering FEB. To address this issue, we propose an evolutionary multi-view classification via eliminating individual fitness bias (EFB-EMVC) method, which alleviates the FEB issue by introducing evolutionary navigators for each MVM, thereby providing more accurate individual ranking. Experimental results fully verify the effectiveness of the proposed method in alleviating the FEB problem, and the EMVC method equipped with this strategy exhibits more superior performance compared with the original EMVC method. (The code is available at <https://github.com/LiShuailzn/Neurips-2025-EFB-EMVC>)

## 1 Introduction

Multi-view classification (MVC) aims to integrate data from multiple views to improve classification accuracy. A large number of methods have been proposed in this field [1–6]. Among them, evolutionary multi-view classification (EMVC) methods differs from traditional approaches where it dynamically fuses different views through the adaptive mechanism of evolutionary algorithms (EA) [7], rather than relying on static modeling. They have obtained important applications like protein

---

\*Corresponding Author.

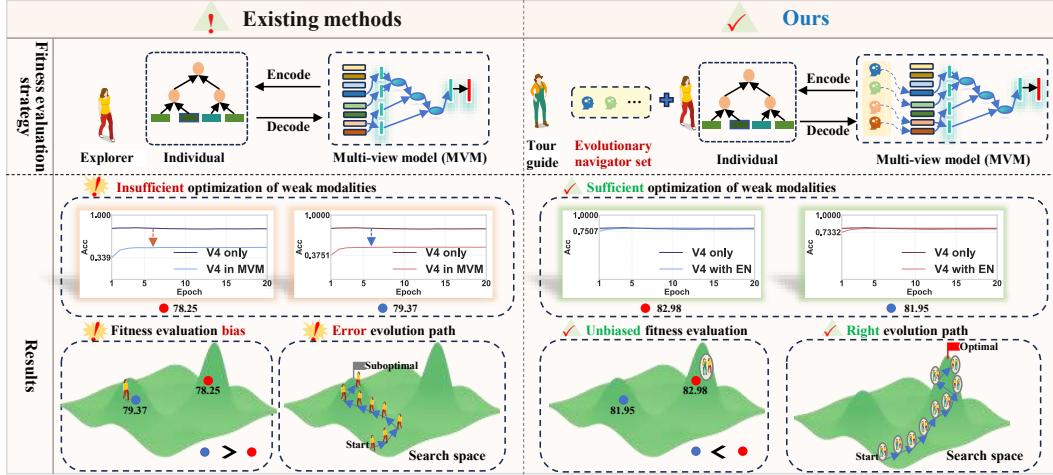


Figure 1: The issues of existing evolutionary multi-view classification and our solutions

secondary structure prediction [8]. Notably, fitness evaluation (FE), as a core step of such methods, runs through the entire process of individual selection, crossover, and mutation. It provides a reliable basis for performance ranking for subsequent operations, and thus its evaluation accuracy directly determines the correctness of the evolutionary direction.

However, when FE fails to correctly reflect the superiority-inferiority relationship among individuals, it will cause confusion in the ranking of individual performance, which in turn misleads the evolutionary direction and traps the process in local optima. This paper is the first to identify the above-mentioned issue in the field of EMVC and formally define it as fitness evaluation bias (FEB). FEB does not generally refer to any error in FE; instead, it specifically refers to the kind of bias that is sufficient to disrupt the true performance ranking of individuals. Given the core role of FE in the entire EMVC method, the occurrence of the FEB issue will in turn lead to systematic distortion in the whole evolutionary process. Specifically, in the selection phase, such biases will interfere with the screening of high-quality individuals, resulting in the abnormal loss of superior genes; in the crossover operation, due to the lack of accurate individual ranking relationships as guidance, it is difficult to form efficient gene combinations; mutation strategies, for lack of reliable guidance, degenerate into random perturbations. More seriously, the iterative optimization characteristic of EA will lead to the continuous accumulation of this bias, which in turn misleads the evolutionary direction and eventually causes the algorithm to fall into local optima (as shown in Fig. 1).

FEB may be caused by a variety of factors, and this paper approaches the issue from the perspective of view information content: in the multi-view model (MVM) decoded from individuals, there are significant differences in the information content among different views. Although the joint training strategy commonly adopted in existing methods can make full use of key information in views with high information content, it fails to fully explore key information in views with low information content. This leads to suboptimal utilization of multi-view data, which in turn causes MVM to fail to achieve optimal performance during the convergence process [9, 10]. Consequently, FE cannot accurately reflect the performance ranking of individuals, thereby triggering FEB. This phenomenon can be analogous to a mountaineer misjudging the highest point due to cognitive bias toward the terrain: if the mountaineer chooses a path based only on a local perspective (partial view information), they may stop prematurely at a suboptimal mountain top, thus missing the main peak that represents the global optimum. To our knowledge, none of the existing EMVC methods have considered this issue, which limits their performance.

To address the issue, we propose an evolutionary multi-view classification via eliminating individual fitness bias (EFB-EMVC) method, which alleviates the FEB problem by equipping each MVM with evolutionary navigators (ENs). EN realizes the directional optimization of view branches based on the knowledge distillation mechanism. Meanwhile, recent studies [11] have pointed out that cross-category information also plays a key role in the distillation process. Inspired by this, we further design EN to distill both corresponding category and cross-category information simultaneously,

so as to better guide the optimization of view branches. Specifically, we propose a wasserstein distance-oriented loss function (WDOLF) and construct a dual-path optimization mechanism to alleviate FEB: (1) MVM optimization ensures the consistency of the overall training objective; (2) the view distillation process fully explores the rich corresponding category and cross-category information in EN, thereby utilizing multi-view data more comprehensively. This mechanism can provide more accurate individual performance ranking, and based on this, calibrate the evolutionary direction in real time to guide the EA to iterate along a more correct path. This process is similar to an explorer moving forward under the guidance of a guide who is familiar with the terrain: with each step taken, the explorer can promptly obtain the correct route leading to the highest peak, thus avoiding falling into local optima (as shown in Fig. 1 for details). By virtue of the concise and efficient strategy of “global control + local in-depth exploration”, the potential of EMVC methods is fully unleashed. Notably, EN can be seamlessly integrated with other EMVC methods, demonstrating strong generality. The core contributions of this paper are summarized as follows:

1. In the field of evolutionary multi-view classification (EMVC), we are the first to identify and formally define the problem of fitness evaluation bias (FEB), and systematically analyze its impact on the entire evolutionary process.
2. To alleviate FEB, we equip multi-view model with evolutionary navigators (ENs) to fully explore the rich corresponding category and cross-category information therein, thereby achieving efficient utilization of multi-view data.
3. Experimental results fully verify the effectiveness of the proposed method in alleviating FEB. Moreover, EN can be seamlessly integrated into other EMVC methods, demonstrating strong compatibility.

## 2 Related Work

**Evolutionary Multi-View Classification (EMVC):** The EMVC method utilizes the adaptive mechanism of evolutionary algorithms (EA), enables continuous iterative optimization, and ultimately screens out satisfactory multi-view model (MVM) from the population. EDF [7] initializes the population in the search space composed of views and basic fusion operators, and gradually searches for satisfactory MVM through operations including fitness evaluation, selection, crossover, and mutation, pioneering research in this field. However, due to its inherent drawbacks such as long computation time, a series of improvement works have been initiated. For example, DC-NAS [12] accelerates the search process by adopting a divide-and-conquer strategy for data; CSG-NAS [13] and CoMO-NAS [14] further reduce complexity through the core structure search space, significantly improving algorithm efficiency; KS-NAS [15] reduces computational costs via introducing a dynamic knowledge base. As research advances, the credibility issue of MVM has gradually attracted attention. For instance, TEF [16] uses EA as a generator for high-quality pseudo-views, thereby further enhancing the credibility of MVM. Despite the fruitful achievements, the existing EMVC methods have not taken the problem of fitness evaluation bias (FEB) into account, which limits their performance. In contrast, the EFB-EMVC method proposed in this paper effectively alleviates the FEB problem by evolutionary navigators (ENs) into each MVM.

**Knowledge Distillation (KD):** The core idea of KD is to alleviate the challenges of complex models in deployment by transferring knowledge from the teacher model to the lightweight student model [17]. In this study, we draw on the basic idea of KD and introduce it into the EMVC framework, but its implementation relies on the EN. Through the EN, we can transfer knowledge in a targeted manner to the corresponding view branches of the MVM to assist in its training. Unlike traditional KD, the focus of this study is not on proposing a new distillation paradigm, but on using EN to alleviate the inherent FEB in the EMVC framework. Recent work [11] found that the cross-category information contained in the teacher model plays a key role in knowledge transfer. Inspired by this, the EN designed in this study not only provides corresponding category information, but also transmits cross-category information to better guide the optimization of each view branch. It is noted that the EN plays a core role in the process of alleviating FEB, while KD only serves as the implementation mechanism for knowledge transfer.

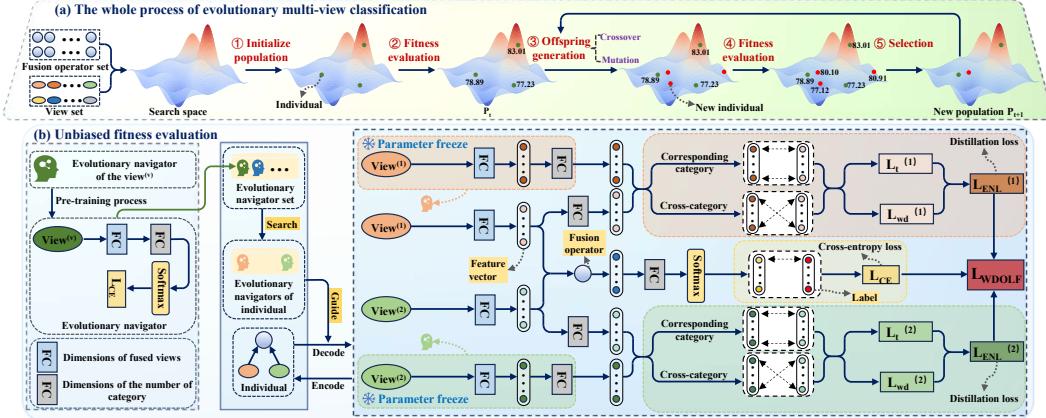


Figure 2: The overview of EFB-EMVC

### 3 Proposed Method

The proposed EFB-EMVC follows the process of the existing evolutionary multi-view classification (EMVC) method. The only difference is that EFB-EMVC alleviates fitness evaluation bias (FEB) by introducing evolutionary navigators (ENs) for each multi-view model (MVM) while existing methods do not. Our EFB-EMVC overview is illustrated in Fig. 2.

#### 3.1 The Alleviation of Fitness Evaluation Bias

In the field of EMVC, we are the first to identify and formally define the FEB issue, which exerts a systematic impact on the entire evolutionary process. To alleviate FEB, we equip each view branch in MVM with an EN and guide it by drawing on the knowledge distillation (KD) mechanism. However, traditional KD can only model corresponding category, while recent studies [11] have shown that rich cross-category information is crucial for the distillation process. Inspired by this, we design EN to distill both corresponding category and cross-category information simultaneously, thereby guiding the optimization of view branches more effectively.

Specifically, the EN is defined as a pre-trained teacher model. Taking the  $v$ -th view as an example: First, the EN corresponding to this view is decoded into a two-layer fully connected deep neural network (DNN), where the dimension of the first layer is the fused view dimension, the dimension of the second layer is the number of categories  $C$ , and a softmax layer is attached at the end of the network. Subsequently, the training data and test data are input into the DNN to complete the training and testing processes. After training, the optimal DNN parameters are frozen. For a given dataset, we pre-train ENs for all views and store the results in an EN set for subsequent retrieval. During the individual training process, according to the views corresponding to leaf nodes, the ENs required by the individual are retrieved from the EN set to serve as the knowledge source for that individual. The above process is shown in Fig. 2 (b) and is achieved via three steps.

**Step 1** aims to extract the category interrelations from the EN. Taking the  $v$ -th view as an example: **①** Obtain the predicted probability distribution outputs of all the training data by passing them into the corresponding EN; **②** Divide the training data into  $C$  category groups according to the resulting predicted probability distribution outputs; **③** For the  $c$ -th category divided by the EN, assuming it contains  $b_c$  training samples, we collect their logits, and then concatenate them into a feature matrix  $X_v^c \in \mathbb{R}^{C \times b_c}$ ; **④** Construct the kernel matrix  $K_v^c = X_v^c (X_v^c)^\top \in \mathbb{R}^{C \times C}$ . Notably, in some datasets, the EN corresponding to certain views may end up classifying fewer than  $C$  categories — this indicates that the EN judges there to be no samples belonging to some categories. For the sake of convenient description, we uniformly consider the number of its classified categories as still  $C$ . For the aforementioned categories with no corresponding samples, the elements in their kernel matrices are replaced with extremely small positive numbers; **⑤** After obtaining the kernel matrices  $\{K_v^c\}_{c=1}^C$  for all categories, we measure the category interrelations using the HSIC score, and then construct the mutual information matrix  $M_v$  for the  $v$ -th view as follows: for any two elements  $K_v^{c_1}$  and

$K_v^{c_2}$  from  $\{K_v^c\}_{c=1}^C$ , one computes  $\text{HSIC}(K_v^{c_1}, K_v^{c_2}) = \frac{1}{C^2} \text{tr} \left( \tilde{K}_v^{c_1} \tilde{K}_v^{c_2} \right)$ , where  $\tilde{K}_v^{c_1} = H K_v^{c_1} H$ ,  $\tilde{K}_v^{c_2} = H K_v^{c_2} H$ , and  $H = I - \frac{1}{C} \mathbf{1} \mathbf{1}^\top$ , with  $I$  as the identity matrix and  $\mathbf{1}$  a column vector of ones. The resulting mutual information matrix quantifies the statistical correlations between categories in the  $v$ -th view, where larger values indicate stronger associations between their feature distributions.

**Step 2** aims to effectively guide the training of each view branch by leveraging EN, thereby alleviating FEB. To this end, we have designed an optimization objective referred to as EN guidance loss (ENL). The core idea of ENL is as follows: during the training process, each view branch not only needs to learn information of their corresponding category from EN, but also should capture knowledge of cross-category relationships provided by EN—this enables the achievement of better optimization effects. The ENL loss function consists of two components: The first component is the cross-entropy loss between a view branch and EN with respect to the target category; The second component is the wasserstein distance (WD) between the output distributions of the two (i.e., a view branch and EN) with respect to non-target categories.

The loss function for the first part is  $L_t = -(p^T)_t \cdot \log(p^S)_t$ , where  $t$  denotes the target category,  $p^T = \text{softmax}(z^T)$  and  $p^S = \text{softmax}(z^S)$  represent the output probability distributions of the EN and the view branch.  $z^T$  and  $z^S$  denote the logits outputs of the EN and the view branch, respectively.

The second part is to calculate the WD between EN and view branches with respect to the probability distributions of non-target categories. For two discrete distributions  $P$  and  $Q$ , WD is defined as:

$$W(P, Q) = \min_{\gamma \in \Gamma} \sum_i \sum_j \gamma_{ij} \cdot c_{ij}, \quad (1)$$

where  $\sum_j \gamma_{ij} = p_i$ ,  $\sum_i \gamma_{ij} = q_j$ , and  $\gamma_{ij} \geq 0$ . Here,  $\gamma_{ij}$  denotes the transport plan from  $p_i$  to  $q_j$ ,  $c_{ij}$  is the transportation cost, and  $\Gamma$  is the set of feasible couplings that satisfy the marginal constraints.

Since both the EN and the view branch output discrete probability distributions over non-target categories, we adopt the discrete WD formulation. Specifically, we first mask the target category in the logits of both models to obtain the non-target feature vectors  $\hat{z}^T$  and  $\hat{z}^S$ , and apply the same masking to the mutual information matrix to derive the masked mutual information matrix  $\hat{M}_v$ . Then, we apply temperature-scaled softmax to  $\hat{z}^T$  and  $\hat{z}^S$ , yielding the non-target categories probability distributions  $\hat{p}^T$  and  $\hat{p}^S$ .

To efficiently approximate WD, we employ the entropy-regularized Sinkhorn algorithm[18]. The kernel matrix is defined as  $K = \exp(-\hat{M}_v/\epsilon)$ , where  $\epsilon = 0.05$  is the regularization parameter. The iteration matrix  $U$  is initialized as:  $U = \frac{1}{C-1} \cdot \mathbf{1}$ . Then, the updates are performed iteratively as:

$$V \leftarrow \frac{\hat{p}^T}{K^\top U}, \quad U \leftarrow \frac{\hat{p}^S}{KV}, \quad (2)$$

and after 10 iterations, the optimal transport plan is obtained as:

$$\gamma^* = \text{diag}(U) K \text{ diag}(V). \quad (3)$$

The WD is then computed as:

$$L_{wd} = \sum_i \sum_j \gamma_{ij}^* \cdot (\hat{M}_v)_{ij}. \quad (4)$$

Finally, the ENL loss is defined as:

$$L_{ENL} = \alpha L_t + \beta L_{wd}, \quad (5)$$

where  $\alpha$  and  $\beta$  are hyperparameters used to balance the contributions of the target-category cross-entropy loss and the non-target categories WD.

**Step 3** aims to combine the proposed ENL loss function with the traditional cross-entropy loss to jointly optimize the MVM and its individual view branches. This combined loss not only ensures the consistency of the overall training objective but also makes full use of the critical information from each view, thereby effectively alleviating the FEB problem.

When considering  $\mathcal{V}$  views, the whole loss function WDOLF is formulated as follows:

$$L_{WDOLF} = \gamma_1 L_{CE} + \gamma_2 \sum_{v=1}^{\mathcal{V}} L_{ENL}^{(v)} = \gamma_1 L_{CE} + \gamma_2 \alpha \sum_{v=1}^{\mathcal{V}} L_t^{(v)} + \gamma_2 \beta \sum_{v=1}^{\mathcal{V}} L_{wd}^{(v)}, \quad (6)$$

where  $L_{CE}$  represents the cross-entropy loss between the output of the MVM and the ground-truth labels, and  $L_{ENL}^{(v)}$  represents the ENL loss between the  $v$ -th view branch of the MVM and the corresponding EN.

### 3.2 The EMVC Method Driven by Unbiased Fitness Evaluation

The core steps of the EFB-EMVC method include population initialization, unbiased fitness evaluation, offspring generation, and selection.

**Population initialization:** First, in the search space composed of views and basic fusion operators, randomly initialize a population  $P$  consisting of  $k$  individuals encoded in the form of binary trees, where the leaf nodes consist of views and the branch nodes consist of basic fusion operators. In this paper, the basic fusion operators consists of concatenation, addition, multiplication, maximize and average [7]. For each individual, if the binary tree contains  $\mathcal{V}$  views, then it must contain  $\mathcal{V} - 1$  basic fusion operators. Each individual corresponds to MVM.

**Unbiased fitness evaluation:** First, according to the views corresponding to the leaf nodes of an binary tree, the ENs required by the individual are retrieved from the EN set of the current dataset. Next, the individual is decoded into an MVM. In this MVM, each view first passes through a fully connected layer (whose dimension is set to the dimension of the fused view). Subsequently, the data is processed through two streams: 1) one stream enters a fully connected layer (with the dimension equal to the number of categories) to generate a logits vector, and then the  $l_{ENL}$  distillation loss between this vector and the logits output by the EN corresponding to the view is calculated; 2) the other stream is fused with other views via a fusion operator, then passes through a fully connected layer (with the dimension equal to the number of categories) to generate the logits vector of the MVM, and finally enters the softmax layer to obtain a probability distribution, with which the cross-entropy loss is calculated against the true labels. Finally, all losses are combined into the  $l_{WDOLF}$  loss function, which is used to optimize the entire MVM and each view branch. Finally, the test accuracy at the optimal state during training is taken as the unbiased fitness value of the individual. The above process is detailed in Fig. 2 (b).

**Crossover and mutation:** The crossover rate is  $r = r1$ , and the mutation rate is  $r = r2$ . At this stage, following the principles of traditional evolutionary algorithms, the probability of the crossover operation is significantly higher than that of the mutation operation. For the crossover operation, a non-root node is first randomly selected from each of two randomly chosen binary trees to serve as the crossover point. The selected node along with its subtree is then pruned from the original tree and exchanged with the counterpart from the other tree. These subtrees are subsequently grafted onto the parent node of the original crossover point in the opposite tree, thereby generating two new individual structures. For the mutation operation, a node is randomly selected from the individual's tree structure. If the selected node is a fusion operation node, a new fusion operation is randomly sampled from a predefined list to replace the original operation. If the selected node is a view node, a new view is randomly chosen from a set of views to replace the original one.

---

#### Algorithm 1 EFB-EMVC Framework

---

- 1: **Input:**  $D = (X, Y)$ : training dataset;  $\hat{D} = (\hat{X}, \hat{Y})$ : test dataset;  $F$ : a set of basic fusion operators;  $T$ : iteration number; evolutionary navigator set;
- 2: **Output:** The satisfactory MVM and its corresponding accuracy;
- 3: **Population initialization:** Generate an initial population  $P_0$ ;
- 4: **Fitness evaluation:** Obtain the unbiased fitness values of all individuals in  $P_0$ ;
- 5: **for**  $t = 1$  to  $T$  **do**
- 6:   Generate offspring  $Q_t$  using the crossover operator;
- 7:   Conduct mutation on each individuals in  $Q_t$ ;
- 8:   Obtain the unbiased fitness values of all individuals in  $Q_t$ ;
- 9:   Select next generation population  $P_{t+1}$  from  $Q_t \cup P_t$  using a selection operator;
- 10: **end for**
- 11: **return** The satisfactory MVM and its corresponding accuracy.

---

**Selection:** Adopt the binary tournament selection method. Specifically, the selection operation has three application scenarios: (1) Crossover phase: Randomly select two individuals, compare their

fitness values, and retain the better one as Parent 1. Repeat this operation to select Parent 2. Then, perform the crossover operation on the two parent individuals; (2) Mutation phase: Randomly select two individuals and directly choose the one with the better fitness value for mutation; (3) Generating the next generation population: Merge the current population with the offspring population, randomly select two individuals, and retain the better one to be added to the next generation. Repeat this operation  $K$  times.

The algorithm framework of the EFB-EMVC method is shown in Algorithm 1.

## 4 Experiments

In this section, we aim to verify the effectiveness of EFB-EMVC method from four aspects: (1) Comparison with SOTA methods; (2) Impact analysis of the evolutionary navigator (EN) on EFB-EMVC; (3) Impact analysis of the distillation loss function ENL on EFB-EMVC; (4) Generality of evolutionary navigator. In addition, we conducted an in-depth analysis of the hyperparameters of EFB-EMVC and the experimental results are presented in Appendix A.4.

### 4.1 Experimental Setup

**Datasets.** In the experiments, nine multi-view datasets are used and they are MVoxCeleb [16], YoutubeFace [19], NUS-WIDE-128 (NUS) [20], Reuters [21], CB [7], MM-IMDB [22], NTU RGB-D [23], and EgoGesture [24]. For the Reuters dataset, two variants named Reuters5 and Reuters3 are generated by adding Gaussian noise [16]. Their detailed descriptions can be found in Appendix A.3.

**Evaluation metrics.** In the experimental process, to effectively avoid the randomness interference caused by data partitioning and network initialization operations, a 5-fold cross-validation strategy was adopted for the MVoxCeleb [16], YoutubeFace [19], NUS-WIDE-128 (NUS) [20], Reuters [16] and Reuters3 [16] datasets, dividing each dataset into a training set and a test set. The specific partitioning ratio is as follows: 80% of the samples are used for model training, and 20% are used for model testing. For the remaining datasets, since their original authors have already completed the partitioning of the training and test sets, the experiments were independently repeated five times. The final experimental results are presented in the form of average performance metrics and standard deviations to intuitively reflect the statistical stability of the model performance. For the MM-IMDB dataset, considering the existence of class imbalance, the weighted F1-score is selected as the evaluation metric; a higher value of this metric indicates better model performance. For the remaining datasets, accuracy is used as the evaluation metric.

### 4.2 Experimental Results

**Comparisons with SOTA methods.** In this section, we aim to validate the EFB-EMVC’s effectiveness by comparing with four kinds of SOTA methods. They are (1) fixed fusion operators-based methods including addition, average, max, multiplication, concatenation, MLB [25], MFB [26], TFN [27], LMF [28], and PTP [29]; (2) trustworthy multi-view classification methods including TMC [30], TMOA [31], ETMC [32], RCML [33]; (3) traditional adaptive multi-view classification methods including BV [7], SSV [7], MR [7], EmbraceNet [34], AWDR [35], RMAR [36]; (4) EMVC methods including EDF [7], DC-NAS [12], CoMO-NAS [14], CSG-NAS [13] and TEF [16].

The results are shown in Table 1. Based on these results, one can observe that: (1) Compared with the fixed fusion strategy, the EMVC methods demonstrate significant performance advantages on all datasets. This is attributed to its acquisition of a high-performance multi-view model (MVM) through adaptive selection. However, traditional multi-view classification methods, due to being restricted by manually designed features, sometimes have lower performance than that of the fixed fusion strategy. At the same time, the trusted multi-view classification method, because it adopts the way of late-stage desision fusion, results in insufficient feature interaction, and its performance sometimes cannot surpass that of the fixed fusion strategy. (2) Compared with existing EMVC methods, it can be observed that: although these methods outperform other types of methods in terms of performance metrics, the existence of fitness evaluation bias (FEB) causes such methods to fall into local optima, thereby preventing their performance potential from being fully unleashed. In contrast, EFB-EMVC effectively alleviates the FEB issue by introducing ENs into MVM, ensuring that the evolutionary algorithm (EA) can iterate and optimize along a relatively more correct direction to find a more

Table 1: Accuracy comparison results with SOTA methods (mean  $\pm$  standard deviation), where the best performance is highlighted in bold.

Methods	MVoxCeleb	YoutubeFace	NUS	Reuters5	Reuters3	CB
Add	87.53 $\pm$ 0.41	82.40 $\pm$ 0.23	72.81 $\pm$ 0.70	79.70 $\pm$ 0.25	83.46 $\pm$ 0.28	87.16 $\pm$ 0.17
Mul	72.31 $\pm$ 0.90	83.18 $\pm$ 0.14	64.58 $\pm$ 0.63	77.02 $\pm$ 0.38	81.89 $\pm$ 0.70	80.87 $\pm$ 1.16
Cat	87.98 $\pm$ 0.20	83.05 $\pm$ 0.56	72.32 $\pm$ 0.50	79.91 $\pm$ 0.28	83.66 $\pm$ 0.17	86.58 $\pm$ 0.11
Max	81.57 $\pm$ 0.41	81.49 $\pm$ 0.29	71.36 $\pm$ 0.47	80.02 $\pm$ 0.20	84.01 $\pm$ 0.28	84.21 $\pm$ 0.11
Avg	87.27 $\pm$ 0.33	82.23 $\pm$ 0.17	73.00 $\pm$ 0.51	79.69 $\pm$ 0.30	83.58 $\pm$ 0.28	87.05 $\pm$ 0.20
MLB (ICLR17)	87.11 $\pm$ 0.67	85.20 $\pm$ 0.28	70.60 $\pm$ 0.29	80.16 $\pm$ 0.15	83.80 $\pm$ 0.28	82.38 $\pm$ 0.32
MFB (TNNLS18)	85.23 $\pm$ 0.20	82.85 $\pm$ 0.17	71.34 $\pm$ 0.40	79.28 $\pm$ 0.21	83.25 $\pm$ 0.18	87.94 $\pm$ 0.32
TFN (EMNLP17)	57.53 $\pm$ 0.92	81.33 $\pm$ 0.19	63.66 $\pm$ 1.22	79.95 $\pm$ 0.30	83.73 $\pm$ 0.31	73.45 $\pm$ 0.30
LMF (ACL18)	89.92 $\pm$ 0.25	85.58 $\pm$ 0.22	71.74 $\pm$ 0.70	80.03 $\pm$ 0.15	83.75 $\pm$ 0.28	82.81 $\pm$ 0.18
PTP (NeurIPS19)	88.61 $\pm$ 0.36	85.18 $\pm$ 0.30	71.83 $\pm$ 0.50	80.10 $\pm$ 0.10	84.06 $\pm$ 0.20	85.08 $\pm$ 0.11
TMC (ICLR22)	73.13 $\pm$ 0.15	71.18 $\pm$ 2.27	72.73 $\pm$ 0.30	79.60 $\pm$ 0.56	84.23 $\pm$ 0.35	77.87 $\pm$ 0.22
TMOA (AAAI22)	84.72 $\pm$ 0.21	84.35 $\pm$ 0.25	72.60 $\pm$ 0.48	79.11 $\pm$ 0.43	84.19 $\pm$ 0.27	86.80 $\pm$ 0.10
ETMC (TPAMI23)	88.70 $\pm$ 0.15	79.63 $\pm$ 1.89	73.05 $\pm$ 0.67	79.80 $\pm$ 0.41	84.25 $\pm$ 0.42	--
RCML (AAAI24)	80.51 $\pm$ 0.41	81.95 $\pm$ 0.20	72.53 $\pm$ 0.55	81.39 $\pm$ 0.18	85.88 $\pm$ 0.29	--
BV (TEVC2021)	63.25 $\pm$ 0.14	82.01 $\pm$ 0.18	68.69 $\pm$ 0.59	80.61 $\pm$ 0.25	83.98 $\pm$ 0.14	77.08 $\pm$ 0.15
SSV (TEVC2021)	85.10 $\pm$ 0.23	84.43 $\pm$ 0.31	63.70 $\pm$ 0.64	79.51 $\pm$ 0.41	84.71 $\pm$ 0.22	87.02 $\pm$ 0.13
MR (TEVC2021)	79.92 $\pm$ 0.29	84.78 $\pm$ 0.21	64.39 $\pm$ 0.85	78.24 $\pm$ 0.45	84.17 $\pm$ 0.19	83.36 $\pm$ 0.21
EmbraceNet (IF19)	81.74 $\pm$ 0.34	80.90 $\pm$ 1.04	72.43 $\pm$ 0.38	80.07 $\pm$ 0.21	83.58 $\pm$ 0.25	85.85 $\pm$ 0.09
AWDR(PR19)	91.08 $\pm$ 0.09	85.11 $\pm$ 0.15	72.44 $\pm$ 0.66	79.69 $\pm$ 0.27	83.32 $\pm$ 0.32	86.66 $\pm$ 0.16
RMAR(INS22)	91.54 $\pm$ 0.11	85.21 $\pm$ 0.17	72.51 $\pm$ 0.67	79.84 $\pm$ 0.25	83.48 $\pm$ 0.25	85.36 $\pm$ 0.46
EDF (TEVC2021)	93.09 $\pm$ 0.20	85.83 $\pm$ 0.08	74.73 $\pm$ 0.45	81.12 $\pm$ 0.25	85.49 $\pm$ 0.21	88.55 $\pm$ 0.20
CoMO-NAS (ACMMM24)	--	--	74.24 $\pm$ 0.29	--	--	88.69 $\pm$ 0.38
CSG-NAS (IJCAI24)	--	--	74.52 $\pm$ 0.40	--	--	89.20 $\pm$ 0.06
DC-NAS (AAAI24)	92.19 $\pm$ 0.07	85.28 $\pm$ 0.14	74.35 $\pm$ 0.58	81.35 $\pm$ 0.28	85.86 $\pm$ 0.14	88.52 $\pm$ 0.13
TEF (ICLR25)	92.41 $\pm$ 0.12	86.02 $\pm$ 0.41	75.12 $\pm$ 0.57	82.26 $\pm$ 0.23	86.49 $\pm$ 0.10	--
<b>EFB-EMVC (ours)</b>	<b>94.82<math>\pm</math>0.12</b>	<b>87.67<math>\pm</math>0.17</b>	<b>75.79<math>\pm</math>0.76</b>	<b>82.66<math>\pm</math>0.24</b>	<b>86.51<math>\pm</math>0.12</b>	<b>89.67<math>\pm</math>0.18</b>

satisfactory solution. For instance, on the MVoxCeleb, YouTubeFace, and NUS datasets, the accuracy of EFB-EMVC is improved by approximately 1.73%, 1.65%, and 0.67% respectively compared with the suboptimal model. These results fully verify the effectiveness of this mechanism.

To further explore its performance, we carried out additional experiments on other datasets that are often used to validate the effectiveness of the EMVC methods. In this experiment, the SOTA methods compared with EFB-EMVC are roughly divided into two categories: (1) Single-view classification methods, such as Maxout MLP [37], VGG Transfer [38], Inflated ResNet-50 [39], Co-occurrence [40], ResNext-101 [41]; (2) Multi-view classification methods, such as Two-stream, GMU [22], CentralNet [42], MFAS [43], VGG-16 + LSTM [44], C3D + LSTM + RSTTM [45], I3D [46], MMTM [47], MTUT [48], 3D-CDC-NAS2 [49], BM-NAS [50], DC-NAS [12], CoMO-NAS [14], CSG-NAS [13], and MCTS-CSG [51] and HF-MNAS [52]. According to the results in Table 2, compared with single-view classification methods, multi-view classification methods generally achieve better performance. This is mainly due to their ability to fully utilize the rich information across different views, thereby achieving performance breakthroughs. Compared with SOTA methods, EFB-EMVC achieves the best performance on most datasets and also reaches performance comparable to SOTA methods on the EgoGesture dataset. This advantage stems from the EN mechanism of EFB-EMVC — this mechanism corrects the evolutionary direction in real time by alleviating the FEB issue.

**Impact analysis of the evolutionary navigator (EN) on EFB-EMVC.** In this experiment, we aim to analyze the impact of EN on EFB-EMVC (i.e., whether EN alleviates the problem of FEB?) by an ablation analysis on YouTubeFace dataset. To ensure the generality of this method, we took a population consisting of 15 randomly initialized individuals as the research object.

As shown in the left side of Fig. 3, the accuracy of all individuals after applying the EN exhibits a consistent improvement. It is noticed that its core value lies in the reconstruction of the performance order relationship on which the EMVC method relies. The selection mechanism of the EMVC method essentially depends on the performance order relationship to identify high-quality solutions. If this order relationship remains unchanged, it means that the EN only plays a role in enhancing performance and does not address the core defect of the EMVC method, that is, the FEB problem.

Table 2: Accuracy comparison results with SOTA methods (mean  $\pm$  standard deviation), where the best performance is highlighted in bold.

Method	MM-IMDB		NTU RGB-D		EgoGesture	
	Modality	F1-W (%)	Modality	Acc (%)	Modality	Acc (%)
Uni-view methods						
Modality 1	Text (T)	57.54	Video (V)	83.91	RGB (R)	93.75
Modality 2	Image (I)	49.21	Pose (P)	85.24	Depth (D)	94.03
Multi-view methods						
Two-stream (NeurIPS14)	I+T	60.81	V+P	88.60	--	--
GMU (ICLR17)	I+T	61.70	V+P	85.80	--	--
CentralNet (ECCV18)	I+T	62.23	V+P	89.36	--	--
MFAS (CVPR19)	I+T	62.50	V+P	$89.50 \pm 0.60$	--	--
MMTM (ICCV20)	--	--	V+P	88.92	R+D	93.51
MTUT (3DV19)	--	--	--	--	R+D	93.87
3D-CDC-NAS2 (TIP21)	--	--	--	--	R+D	94.38
BM-NAS (AAAI22)	I+T	$62.92 \pm 0.03$	V+P	$90.48 \pm 0.24$	R+D	$94.96 \pm 0.07$
DC-NAS (AAAI24)	I+T	$63.70 \pm 0.11$	V+P	$90.85 \pm 0.05$	R+D	$95.22 \pm 0.05$
CoMO-NAS (ACMMM24)	I+T	$63.84 \pm 0.16$	V+P	$90.94 \pm 0.02$	R+D	$95.25 \pm 0.03$
CSG-NAS (IJCAI24)	I+T	$64.12 \pm 0.12$	V+P	$91.12 \pm 0.03$	R+D	$95.25 \pm 0.04$
MCTS-CSG (IJMLC25)	--	--	V+P	$91.21 \pm 0.10$	R+D	$95.27 \pm 0.01$
HF-MNAS (TIP25)	I+T	64.17	V+P	91.15	R+D	<b>95.31</b>
<b>EFB-EMVC (ours)</b>	I+T	<b><math>64.53 \pm 0.05</math></b>	V+P	<b><math>91.30 \pm 0.03</math></b>	R+D	$95.30 \pm 0.03$

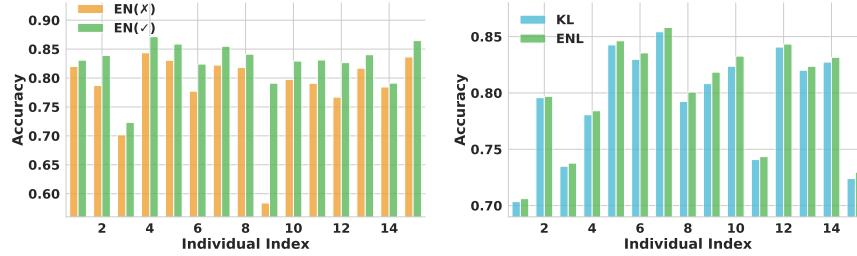


Figure 3: Impact of the EN and the ENL on EFB-EMVC. EN ( $\mathbf{X}$ ) represents that the final loss function only includes cross-entropy loss  $L_{CE}$  and there is no single-view distillation loss  $L_{ENL}$ .

Experimental results in Table 3 show that EN significantly changes the performance ranking relationships of 15 individuals. By calculating three correlation coefficients of the performance ranking relationships before and after applying EN, results indicate extremely low correlation between the two. Meanwhile, Table 4 systematically presents the distribution of individual pairs with changed performance ranking relationships before and after applying EN among the 15 individuals in the EN ablation experiment. Result shows that within the sample space consisting of 15 individuals, there are 11 pairs of individuals with significantly restructured performance ranking relationships.

The above experimental results fully demonstrate that the distortion effect of FEB on individual performance ranking relationships is extremely prominent, confirming that FEB is a key issue urgently needing to be addressed in the EMVC method. Meanwhile, the role of EN is not limited to performance improvement; instead, it fundamentally corrects the distorted individual performance ranking relationships, enabling EMVC methods to conduct effective optimization based on more accurate performance rankings. This reconstruction of ranking relationships breaks through the constraint bottleneck faced by traditional EMVC methods due to FEB, providing a new solution for algorithm design and theoretical development in this field.

Table 3: Rankings and Correlation Coefficients (PC: Pearson Correlation; SRC: Spearman’s Rank Correlation; KT: Kendall’s Tau)

EN	Sort the fitness values of 15 individuals	PC	SRC	KT
$\mathbf{X}$	[4, 15, 5, 7, 1, 8, 13, 10, 11, 2, 14, 6, 12, 3, 9]	0.2143	0.2143	0.1429
$\checkmark$	[4, 15, 5, 7, 8, 13, 2, 11, 1, 10, 12, 6, 14, 9, 3]			

Table 4: Pairs of individuals with changed fitness value relationships between EN ( $\times$ ) and EN ( $\checkmark$ )

Individual Pair	EN ( $\times$ )	EN ( $\checkmark$ )
[4,0,2,3,-0,-1, -0] vs [0,2,3,4,-2,-3, -0]	82.00 vs 78.70	83.09 vs 83.92
[4,0,2,3,-0,-1, -0] vs [1,0,4,2,-4,-2, -2]	82.00 vs 81.81	83.09 vs 84.12
[4,0,2,3,-0,-1, -0] vs [3,2,4,-3, -4]	82.00 vs 79.09	83.09 vs 83.11
[4,0,2,3,-0,-1, -0] vs [1,4,2,-4,-2]	82.00 vs 81.71	83.09 vs 84.02
[0,2,3,4,-2,-3, -0] vs [0,4,3,-4, -0]	78.70 vs 79.75	83.92 vs 82.93
[0,2,3,4,-2,-3, -0] vs [3,2,4,-3, -4]	78.70 vs 79.09	83.92 vs 83.11
[2,4,-4] vs [0,3,-4]	70.17 vs 58.39	72.33 vs 79.11
[3,4,-3] vs [3,4,0,2,-0,-2,-3]	77.73 vs 76.69	82.41 vs 82.66
[3,4,-3] vs [3,2,-1]	77.73 vs 78.45	82.41 vs 79.12
[0,4,3,-4,-0] vs [3,2,4,-3, -4]	79.75 vs 79.09	82.93 vs 83.11
[3,4,0,2,-0,-2,-3] vs [3,2,-1]	76.69 vs 78.45	82.66 vs 79.12

Table 5: Classification accuracy comparison of EDF and DC-NAS before and after applying EN

Methods	MVoxCeleb	YoutubeFace	NUS	Reuters5	Reuters3
EDF	93.09 $\pm$ 0.20	85.83 $\pm$ 0.08	74.73 $\pm$ 0.45	81.12 $\pm$ 0.25	85.49 $\pm$ 0.21
<b>EDF+EN</b>	<b>94.70<math>\pm</math>0.23</b>	<b>87.29<math>\pm</math>0.43</b>	<b>75.81<math>\pm</math>0.71</b>	<b>82.72<math>\pm</math>0.16</b>	<b>86.48<math>\pm</math>0.15</b>
DC-NAS	92.19 $\pm$ 0.07	85.28 $\pm$ 0.14	74.35 $\pm$ 0.58	81.35 $\pm$ 0.28	85.86 $\pm$ 0.14
<b>DC-NAS+EN</b>	<b>94.82<math>\pm</math>0.12</b>	<b>87.67<math>\pm</math>0.17</b>	<b>75.79<math>\pm</math>0.76</b>	<b>82.66<math>\pm</math>0.24</b>	<b>86.51<math>\pm</math>0.12</b>

**Impact analysis of ENL on EFB-EMVC.** In this part, we will conduct an in-depth analysis of the impacts of the KL divergence and the custom loss function ENL on EFB-EMVC to demonstrate that rich cross-category information is an important component of the KD process. To this end, we have selected the YouTubeFace dataset for the experiment. We first randomly initialize 15 individuals. Subsequently, we perform knowledge KD by applying the KL divergence and the ENL loss function respectively. Conclusions can be drawn from the results on the right side of Fig. 3. After applying the ENL loss function, the accuracy of individuals in the population has been improved to varying degrees. This also fully demonstrates the importance of the ENL loss function in leveraging rich cross-category information.

**Generality of the proposed evolutionary navigator.** In this experiment, we aim to verify whether EN can achieve a high degree of compatibility with other EMVC methods. To this end, we have selected two methods, EDF [7] and DC-NAS [12]. In EDF, individuals are encoded in the form of sequences, while in DC-NAS, individuals are encoded in the form of trees. As shown in Table 5, after integrating EN into these two methods, both methods have achieved significant performance improvements across all datasets. Taking the MVoxCeleb and YoutubeFace datasets as examples, the performance of EDF has been improved by 1.61% and 1.46%, and the performance of DC-NAS has been improved by 2.63% and 2.39%. This experimental result indicates that EN is capable of achieving a high degree of compatibility with various EMVC methods.

## 5 Conclusion

In the field of evolutionary multi-view classification (EMVC), this paper is the first to identify and formally define the fitness evaluation bias (FEB) issue, and systematically analyze its impact on the entire evolutionary process. On this basis, we propose an effective method to alleviate FEB from the perspective of view information content, namely EFB-EMVC. Specifically, EFB-EMVC introduces evolutionary navigators (ENs) into each multi-view model (MVM), and leverages the rich corresponding category and cross-category information contained in ENs to guide the optimization of each view, thereby effectively alleviating the FEB that is prevalent in existing EMVC methods. It is worth emphasizing that ENs can be seamlessly integrated into various multi-view learning methods, demonstrating strong compatibility and extensibility. In future work, there are still several important issues regarding FEB that require in-depth research. For instance: Beyond differences in information content, what other potential factors can cause FEB and what are the corresponding countermeasures? How to integrate efficient optimization techniques to reduce the training cost of EFB-EMVC? These issues will all be key directions for advancing the further development of EMVC.

## Acknowledgements

This work was supported by National Natural Science Foundation of China (Nos. 62306171, T2495251, 62406218, 62506219), the Science and Technology Major Project of Shanxi (No. 202201020101006), and Open Funding Programs of State Key Laboratory of AI Safety (No. 2025-12).

## References

- [1] Xinyan Liang, Yuhua Qian, Qian Guo, Honghong Cheng, and Jiye Liang. AF: An association-based fusion method for multi-modal classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9236–9254, 2022.
- [2] Jueqing Lu, Wray Buntine, Yuanyuan Qi, Joanna Dipnall, Belinda Gabbe, and Lan Du. Navigating conflicting views: Harnessing trust for learning. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025. URL <https://openreview.net/forum?id=uJ3JqtBYWk>.
- [3] Xinyan Liang, Shijie Wang, Yuhua Qian, Qian Guo, Liang Du, Bingbing Jiang, Tingjin Luo, and Feijiang Li. Trusted multi-view classification with expert knowledge constraints. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025. URL <https://openreview.net/forum?id=U64wEbM7NB>.
- [4] Yake Wei, Di Hu, Henghui Du, and Ji-Rong Wen. On-the-fly modulation for balanced multimodal learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(1):469–485, 2025.
- [5] Yang Yang, Hongpeng Pan, Qing-Yuan Jiang, Yi Xu, and Jinhui Tang. Learning to rebalance multi-modal optimization by adaptively masking subnetworks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4553–4566, 2025.
- [6] Gengyu Lyu, Zhen Yang, Xiang Deng, and Songhe Feng. L-VSM: Label-driven view-specific fusion for multiview multilabel classification. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4):6569–6583, 2025.
- [7] Xinyan Liang, Qian Guo, Yuhua Qian, Weiping Ding, and Qingfu Zhang. Evolutionary deep fusion method and its application in chemical structure recognition. *IEEE Transactions on Evolutionary Computation*, 25(5):883–893, 2021.
- [8] Yining Qian, Lijie Su, Meiling Xu, and Lixin Tang. Enhanced protein secondary structure prediction through multi-view multi-feature evolutionary deep fusion method. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 9(5):3352–3363, 2025.
- [9] Yang Yang, Hongpeng Pan, Qing-Yuan Jiang, Yi Xu, and Jinhui Tang. Learning to rebalance multi-modal optimization by adaptively masking subnetworks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4553–4566, 2025.
- [10] Xiaokang Peng, Yake Wei, Andong Deng, Dong Wang, and Di Hu. Balanced multimodal learning via on-the-fly gradient modulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8238–8247, 2022.
- [11] Jiaming Lv, Haoyuan Yang, and Peihua Li. Wasserstein distance rivals kullback-leibler divergence for knowledge distillation. *Proceedings of the Advances in Neural Information Processing Systems*, 37:65445–65475, 2024.
- [12] Xinyan Liang, Pinhan Fu, Qian Guo, Keyin Zheng, and Yuhua Qian. DC-NAS: Divide-and-conquer neural architecture search for multi-modal classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 13754–13762, 2024.
- [13] Pinhan Fu, Xinyan Liang, Tingjin Luo, Qian Guo, Yayu Zhang, and Yuhua Qian. Core-structures-guided multi-modal classification neural architecture search. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*, pages 3980–3988, 2024.

[14] Pinhan Fu, Xinyan Liang, Yuhua Qian, Qian Guo, Zhifang Wei, and Wen Li. CoMO-NAS: Core-structures-guided multi-objective neural architecture search for multi-modal classification. In *Proceedings of the 32nd ACM International Conference on Multimedia*, page 9126–9135, 2024.

[15] Zhihua Cui, Shiwu Sun, Qian Guo, Xinyan Liang, Yuhua Qian, and Zhixia Zhang. A fast neural architecture search method for multi-modal classification via knowledge sharing. In *Proceedings of the 34th International Joint Conference on Artificial Intelligence*, pages 5003–5011, 2025.

[16] Xinyan Liang, Pinhan Fu, Yuhua Qian, Qian Guo, and Guoqing Liu. Trusted multi-view classification via evolutionary multi-view fusion. In *Proceedings of the 13th International Conference on Learning Representations*, page 1–14, 2025.

[17] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[18] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

[19] Siwei Wang, Xinwang Liu, Li Liu, Wenxuan Tu, Xinzhong Zhu, Jiyuan Liu, Sihang Zhou, and En Zhu. Highly-efficient incomplete largescale multiview clustering with consensus bipartite graph. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9766–9775, 2022.

[20] Jinhui Tang, Xiangbo Shu, Guojun Qi, Zechao Li, Meng Wang, Shuicheng Yan, and Ramesh Jain. Tri-clustered tensor completion for social-aware image tag refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(8):1662–1674, 2017.

[21] MassihReza Amini, Nicolas Usunier, and Cyril Goutte. Learning from multiple partially observed views – an application to multilingual text categorization. In *Advances in Neural Information Processing Systems*, page 28–36, Vancouver, Canada, 2009.

[22] John Arevalo, Thamar Solorio, Manuel Montes-y Gómez, and Fabio A González. Gated multimodal units for information fusion. *arXiv preprint arXiv:1702.01992*, 2017.

[23] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1010–1019, 2016.

[24] Yifan Zhang, Congqi Cao, Jian Cheng, and Hanqing Lu. EgoGesture: A new dataset and benchmark for egocentric hand gesture recognition. *IEEE Transactions on Multimedia*, 20(5):1038–1050, 2018.

[25] Jin-Hwa Kim, Kyoung Woon On, Woosang Lim, Jeonghee Kim, Jung-Woo Ha, and Byoung-Tak Zhang. Hadamard product for low-rank bilinear pooling. In *Proceedings of the 5th International Conference on Learning Representations*, pages 1–10, 2017.

[26] Zhou Yu, Jun Yu, Chenchao Xiang, Jianping Fan, and Dacheng Tao. Beyond bilinear: Generalized multimodal factorized high-order pooling for visual question answering. *IEEE Transactions on Neural Networks and Learning Systems*, 29(12):5947–5959, 2018.

[27] Amir Zadeh, Minghai Chen, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. Tensor fusion network for multimodal sentiment analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1103–1114, 2017.

[28] Zhun Liu, Ying Shen, Varun Bharadhwaj Lakshminarasimhan, Paul Pu Liang, AmirAli Bagher Zadeh, and Louis-Philippe Morency. Efficient low-rank multimodal fusion with modality-specific factors. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2247–2256, 2018.

[29] Ming Hou, Jiajia Tang, Jianhai Zhang, Wanzeng Kong, and Qibin Zhao. Deep multimodal multilinear fusion with high-order polynomial pooling. *Proceedings of the 33rd Conference on Neural Information Processing Systems*, 32, 2019.

[30] Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification. In *Proceedings of the International Conference on Learning Representations*, pages 1–11, 2021.

[31] Wei Liu, Xiaodong Yue, Yufei Chen, and Thierry Denoeux. Trusted multi-view deep learning with opinion aggregation. *Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence*, 36(7):7585–7593, Jun. 2022.

[32] Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification with dynamic evidential fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2551–2566, 2023.

[33] Cai Xu, Jiajun Si, Ziyu Guan, Wei Zhao, Yue Wu, and Xiyue Gao. Reliable conflictive multi-view learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(14):16129–16137, 2024.

[34] JunHo Choi and JongSeok Lee. EmbraceNet: A robust deep learning architecture for multimodal classification. *Information Fusion*, 51:259–270, 2019.

[35] Muli Yang, Cheng Deng, and Feiping Nie. Adaptive-weighting discriminative regression for multi-view classification. *Pattern Recognition*, 88:236–245, 2019.

[36] Bingbing Jiang, Junhao Xiang, Xingyu Wu, Yadi Wang, Huanhuan Chen, Weiwei Cao, and Weiguo Sheng. Robust multi-view learning via adaptive regression. *Information Sciences*, 610:916–937, 2022.

[37] Ian Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. Maxout networks. In *Proceedings of the International Conference on Machine Learning*, pages 1319–1327. PMLR, 2013.

[38] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[39] Fabien Baradel, Christian Wolf, Julien Mille, and Graham W Taylor. Glimpse clouds: Human activity recognition from unstructured feature points. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 469–478, 2018.

[40] Chao Li, Qiaoyong Zhong, Di Xie, and Shiliang Pu. Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation. *arXiv preprint arXiv:1804.06055*, 2018.

[41] Okan Köpüklü, Ahmet Gunduz, Neslihan Kose, and Gerhard Rigoll. Real-time hand gesture detection and classification using convolutional neural networks. In *2019 14th IEEE International Conference on automatic Face & Gesture Recognition*, pages 1–8. IEEE, 2019.

[42] Valentin Vielzeuf, Alexis Lechervy, Stéphane Pateux, and Frédéric Jurie. Centralnet: a multi-layer approach for multimodal fusion. In *Proceedings of the European Conference on Computer Vision Workshops*, pages 0–0, 2018.

[43] Juan-Manuel Pérez-Rúa, Valentin Vielzeuf, Stéphane Pateux, Moez Baccouche, and Frédéric Jurie. MFAS: Multimodal fusion architecture search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6966–6975, 2019.

[44] Xiaodong Yang and YingLi Tian. Super normal vector for activity recognition using depth sequences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 804–811, 2014.

[45] Pavlo Molchanov, Xiaodong Yang, Shalini Gupta, Kihwan Kim, Stephen Tyree, and Jan Kautz. Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4207–4215, 2016.

[46] Joao Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6299–6308, 2017.

- [47] Hamid Reza Vaezi Joze, Amirreza Shaban, Michael L Iuzzolino, and Kazuhito Koishida. MMTM: Multimodal transfer module for cnn fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13289–13299, 2020.
- [48] Vikram Gupta, Sai Kumar Dwivedi, Rishabh Dabral, and Arjun Jain. Progression modelling for online and early gesture detection. In *2019 International Conference on 3D Vision*, pages 289–297. IEEE, 2019.
- [49] Zitong Yu, Benjia Zhou, Jun Wan, Pichao Wang, Haoyu Chen, Xin Liu, Stan Z Li, and Guoying Zhao. Searching multi-rate and multi-modal temporal enhanced networks for gesture recognition. *IEEE Transactions on Image Processing*, 30:5626–5640, 2021.
- [50] Yihang Yin, Siyu Huang, and Xiang Zhang. BM-NAS: Bilevel multimodal neural architecture search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8901–8909, 2022.
- [51] Guoqing Liu, Yuhua Qian, Xinyan Liang, and Pinhan Fu. Core structure-guided multi-modal classification via monte carlo tree search. *International Journal of Machine Learning and Cybernetics*, pages 1–11, 2025.
- [52] Kaifang Long, Guoyang Xie, Lianbo Ma, Qing Li, Min Huang, Jianhui Lv, and Zhichao Lu. Enhancing multimodal learning via hierarchical fusion architecture search with inconsistency mitigation. *IEEE Transactions on Image Processing*, 34:5458–5472, 2025. doi: 10.1109/TIP.2025.3599673.
- [53] Arsha Nagrani, Joon Son Chung, Weidi Xie, and Andrew Zisserman. Large-scale speaker verification in the wild. *Computer Speech & Language*, 60:101027, 2020.
- [54] Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yantao Zheng. Nus-wide: a real-world web image database from national university of singapore. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, 2009.

## A Appendix

In the supplemental material:

- **A.1.** Variables and Corresponding Meanings
- **A.2.** Experimental Settings
- **A.3.** Datasets
- **A.4.** Hyperparameter Analysis of EFB-EMVC
- **A.5.** Individuals Used in Ablation Experiment

### A.1 Variables and Corresponding Meanings

To help readers understand the content of this paper more clearly, this section provides a detailed organization and explanation of the variable symbols involved in the paper, with the specific corresponding relationships shown in Table 6.

Table 6: Variable names and their corresponding meanings

Variables	Meaning
$C$	The number of categories
$c$	The $c$ -th category
$v$	The $v$ -th view
$t$	The target category
$b_c$	The number of samples assigned to the $c$ -th category by the teacher model
$\mathcal{V}$	The number of views contained in the MVM
$X_v^c$	The feature matrix of the $c$ -th category in the $v$ -th view
$K_v^c$	The kernel matrix of the $c$ -th category in the $v$ -th view
$M_v$	The mutual information matrix of the $v$ -th view
$\hat{M}_v$	The mutual information matrix of the $v$ -th view after masking the target category
$K$	The kernel matrix in the Sinkhorn algorithm
$\text{HSIC}(K_v^{c_i}, K_v^{c_j})$	The HSIC score between the $c_i$ -th and $c_j$ -th category in the $v$ -th view
$z^T$	The logits output of the evolutionary navigator
$z^S$	The logits output of the view branch
$\hat{z}^T$	The logits of the evolutionary navigator after masking the target category
$\hat{z}^S$	The logits of the view branch after masking the target category
$p^T$	The probability distribution of the evolutionary navigator
$p^S$	The probability distribution of the view branch
$\hat{p}^T$	The probability distribution of the evolutionary navigator after masking the target category
$\hat{p}^S$	The probability distribution of the view branch after masking the target category
$\epsilon$	The regularization parameter
$U, V$	The iteration matrix
$\gamma^*$	The optimal transport plan

### A.2 Experimental Settings

In our experiments, all methods are implemented using TensorFlow 2.10.0. The computing environment includes Ubuntu 24.04.2 LTS as the operating system, equipped with an AMD EPYC processor with 160 physical cores (320 logical threads), 566 GB of DDR4 memory, and 8 NVIDIA GeForce RTX 5090 GPUs, each with 32 GB of VRAM. The experimental setup is based on Python 3.9.23 and CUDA 11.2.

**Parameter settings:**

**a) Training of multi-view model (MVM):** All MVM are trained using the Adam algorithm. The learning rate is set to 0.001, with a first-moment exponential decay rate of 0.9 and a second-moment

exponential decay rate of 0.999. Each network undergoes training for 200 epochs. To prevent overfitting, if the performance of a MVM does not improve after 10 epochs, the training process will be halted.

**b) EFB-EMVC Algorithm:** Inspired by the work of EDF[7], we set the population size to 28, the number of iteration rounds to 20, the crossover rate to 0.9, and the mutation rate to 0.2. For the MVoxCeleb [16] and CB [7] datasets, especially the CB dataset with up to 10,000 categories, directly applying our proposed  $\ell_{ENL}$  loss function will lead to an excessively large dimension of the mutual information matrix, resulting in the program being unable to run. In the ablation study section, we have fully verified the key role of cross-category information in the distillation process. To facilitate code reproduction, we use the traditional KL divergence loss instead of  $\ell_{ENL}$  on these two datasets to ensure the executability of the method and the stability of the experiment. In addition, for the MM-IMDB [22] dataset — a multi-label and multi-category dataset — the application of the  $\ell_{ENL}$  loss function may cause numerical instability during the optimization process due to potential dimension mismatch issues of the mutual information matrix. To address this issue, we still use the traditional KL divergence to replace the  $\ell_{ENL}$  loss function. To facilitate the experiment, only the following simplified settings are made for the weights of the loss function: The weight of the fusion output,  $\gamma_1$ , is set to  $1.0 \times \mathcal{V}$ , where  $\mathcal{V}$  represents the number of views of the current individual. For datasets using the  $\ell_{ENL}$  loss function, the output weight of each view,  $\gamma_2$ , is set to 1.0. For the cross-entropy loss between EN and the view branch on the target category, the parameter  $\alpha$  is set to 1.0; for the Wasserstein distance between their output distributions on non-target categories, the parameter  $\beta$  is set to 0.3. In addition, for datasets using KL divergence, the weight of each view is directly set to 1.0. The temperature coefficient is uniformly set to  $t=2$ .

### A.3 Datasets

All used datasets in our paper can be downloaded from the <https://github.com/LiShuailzn/Neurips-2025-EFB-EMVC>.

- **MVoxCeleb** [16] is a multi-view audio classification dataset that is constructed with Vox-Celeb dataset [53]. Each audio are extracted five view features and they are two deep feature including ecapa and resnet, and three traditional features including fbank, mfcc and spec. To study aim, Gaussian noise is added on ecapa and resnet mfcc. On this dataset, the specific parameter configuration of our EFB-EMVC method is as follows: the fusion view dimension to 128 and the reuse of view features is not allowed.
- **YouTube-Faces** [19]. The dataset includes 3,425 videos of 1,595 different people downloaded from YouTube. Similar to, we use a subset consists of 101,499 frames of 31 subjects and the same five features are extracted. The parameter configuration of the dataset is as follows: the fusion view dimension to 128, and the reuse of view features is not allowed.
- **NUS-WIDE-128 (NUS)** [20]. NUS dataset contains 43,800 single label images from 128 categories. For each image, six types of image features including color histogram (CH), color correlogram (CORR), edge direction histogram (EDH), wavelet texture (WT), block-wise color moments (CM) and bag of words based on SIFT descriptions (BoW), and one text feature are extracted. The dataset extended from the NUS-WIDE dataset [54]. In our experiments, we use its a subset consisting of 23,438 images from 10 category, including *animal*, *architecture*, *art*, *flowers*, *food*, *man*, *person*, *sky*, *toy*, and *water*. In this subset, each image is related to one label and each category includes at least 1,500 images. For the EFB-EMVC evolutionary algorithm parameters, the fusion view dimension to 128, and the reuse of view features is not allowed.
- **Reuters** [21]. Reuters is a multilingual multi-view dataset, each document is described by five different languages including English, French, German, Spanish and Italian. To make used model can work on this data, the dimensions of all views are reduced to 1,000 using PCA. Then, following, Gaussian noise is added to all views or 3 views for obtaining its two versions, named as Reuters5 [16] and Reuters3 [16], respectively. The parameter configuration of the dataset is as follows: the fusion view dimensions are set to 128, and the reuse of view features is not allowed.
- **CB** [7]. CB dataset designed for chemical structure image recognition in patent retrieval studies, which contains 100,000 chemical structure images distributed into 10,000 categories.

Among them, the specific experimental parameter settings are as follows: the fusion view dimension to 256, and the reuse of view features is not allowed.

- **MM-IMDB** [22]. MM-IMDB dataset for the multi-label film genre classification task, which contains a total of 23 categories. The dataset is divided into a training set of 15,552 films, a validation set of 2,608 films, and a test set of 7,799 films. To ensure fair comparison with other explicitly multi-view fusion approaches, we adopted the same neural network backbone models as BM-NAS and DC-NAS to extract various view features, using weighted F1 score as the evaluation metric. The specific parameter settings are as follows: fusion vector dimension  $FD$  is 128, and the reuse of view features is not allowed.
- **NTU** [23]. NTU dataset for multi-view action recognition task containing 60 categories. The training, validation and test sets include 23,760, 2,519 and 16,558 samples, respectively. To ensure the fairness of the experimental results, we followed the data preprocessing pipelines of BM-NAS and DC-NAS. Specifically, we used Inflated ResNet-50 and Co-occurrence as feature extractors for the skeleton and video views. For the EFB-EMVC evolutionary algorithm parameters, the fusion view dimension to 1024, and the reuse of view features is allowed.
- **Ego** [24]. Ego dataset for multi-view gesture recognition task containing 83 categories. The training set of this dataset includes 14,416 samples, the validation set includes 4,768 samples, and the test set includes 4,977 samples. We followed the methods of BM-NAS and DC-NAS, using ResNext-101 as the backbone network for RGB and depth video views. For the EFB-EMVC evolutionary algorithm parameters, the fusion view dimension to 512, and the reuse of view features is not allowed.

#### A.4 Hyperparameter Analysis of EFB-EMVC

**Impact Analysis of the fusion dimension on EFB-EMVC.** In this experiment, we selected the YoutubeFace and NUS datasets to investigate the impact of fusion view dimension on the performance of the EFB-EMVC method. Specifically, we first randomly selected 15 individuals as a fixed initial population. Subsequently, within the framework of this fixed initial population, we independently validated the performance of the model when the fusion view dimension was set to 32, 64, 128, 256, and 512 respectively. Experimental results were quantitatively presented using the average performance and maximum performance of the fixed initial population, as shown in Figure 4. It can be observed that on the YoutubeFace dataset, as the fusion dimension increases, the performance of the model shows a trend of continuous improvement and tends to stabilize at high dimensions; whereas on the NUS dataset, increasing the dimension does not bring significant benefits. The above results indicate that the impact of fusion dimension on the performance of the model exhibits dataset dependence: for datasets with higher representation complexity, higher dimensions can effectively enhance feature representation capability, while on datasets with more information redundancy, the benefits of blindly increasing dimensions are limited.

#### A.5 Individuals Used in Ablation Experiment

In this section, we will list in detail the individuals used in the ablation experiment part. Specifically, in Table 7, the left side shows the sample set of the EN ablation experiment, and the right side shows the sample set of the ENL ablation experiment.

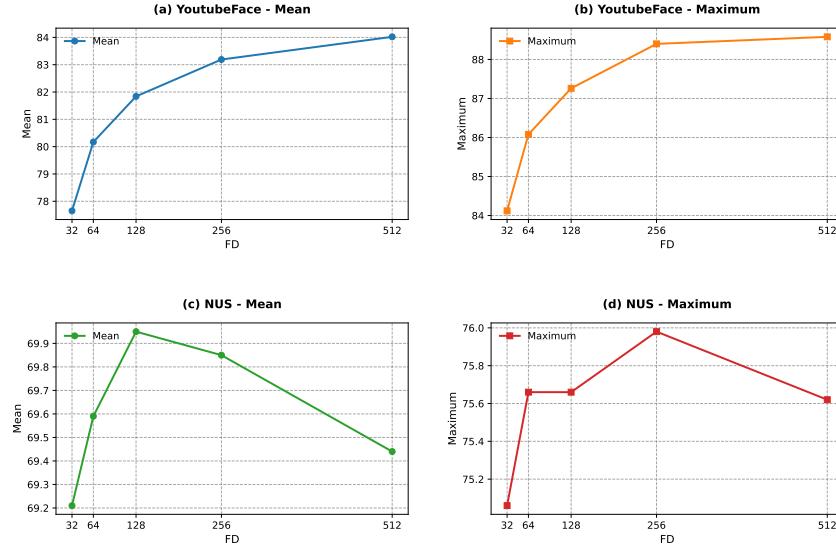


Figure 4: Mean and maximum value variations with integrate the view dimension (FD) on Youtube-Face and NUS datasets

Table 7: The set of individuals used in the ablation experiment

EN	ENL
[4,0,2,3,-0,-1, -0]	[4,0,-4]
[0,2,3,4,-2,-3,-0]	[1,0,-2]
[2,4,-4]	[4,1,1,-0,-2]
[4,3,0,1,2,-1,-1,-2,-4]	[0,0,-3,2,1,-3,-3]
[0,2,4,3,1,-1,-2,-3,-2]	[4,4,4,1,-0,-1,-1]
[3,4,-3]	[2,4,2,1,-1,-3,-0]
[1,4,3,0,2,-1,-4,-2,-3]	[3,4,4,4,1,-2,-4,-3,-0]
[1,0,4,2,-4,-2,-2]	[2,2,-4,0,3,0,-1,-0,-3]
[0,3,-4]	[0,1,1,-2,1,2,2,-1,-2]
[0,4,3,-4,-0]	[2,0,-0,2,3,3,-0,-0,-2]
[3,2,4,-3,-4]	[0,4,4,-4,-1,4,2,-0,-0]
[3,4,0,2,-0,-2,-3]	[4,4,2,2,-1,-0,-3]
[1,4,2,-4,-2]	[1,4,2,1,-0,-2,-3]
[3,2,-1]	[4,1,-3]
[3,2,1,4,-4,-1,-2]	[2,4,0,-0,-1]

## NeurIPS Paper Checklist

### 1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: Yes

Justification: The abstract and introduction have clearly identified the contribution of this paper as introducing evolutionary navigators for each multi-view model to address the fitness evaluation bias problem. Experimental results fully demonstrate the effectiveness of our proposed method, and evolutionary navigators can be seamlessly integrated into other evolutionary multi-view classification methods, demonstrating their strong compatibility. See section 4.2 for details.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

### 2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: Yes

Justification: This paper elaborates on the limitation of high training costs in the section 5, and mentions that optimization techniques will be considered in future work to reduce training costs.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

### 3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: Yes

Justification: This paper provides detailed explanations for all the formulas involved.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

### 4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: Yes

Justification: In Appendix A.2 and Appendix A.3, this paper elaborates on details such as the datasets used, the acquisition method for each dataset, the preprocessing workflow for each dataset, the specific values of all hyperparameters applied to each dataset, the network training approach, and the selection of optimizers. Additionally, section 4.1 provides a detailed description of the five-fold cross-validation method employed in the experiments, while section 4.2 thoroughly outlines the comparative methods involved. Through the above descriptions, the experimental results can be fully reproduced.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

## 5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: Yes

Justification: We will release the code and data with sufficient instructions.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

## 6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: Yes

Justification: All implementation details were provided in Appendix A.2, Appendix A.3 and section 4.1.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

## 7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: Yes

Justification: In this paper, to avoid the randomness caused by data partitioning and network initialization, we adopt a five-fold cross-validation strategy across all datasets. The experimental results are presented in the form of mean values and standard deviations. Details can be found in section 4.1.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

## 8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: No

Justification: This paper provides detailed explanations of the types of computing devices, memory, and storage information used in the experiments, see Appendix A.2, but does not include relevant computation time. This is because the focus of this paper is on addressing the fitness evaluation bias problem rather than accelerating the process.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

## 9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: Yes

Justification: We have checked the NeurIPS code of ethics for our paper.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

## 10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: Yes

Justification: The work in this paper enhances the accuracy and generalization of evolutionary multi-view classification methods in specific societal applications, including chemical structure recognition and medical image classification.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

## 11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: NA

Justification: Our paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

## 12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: Yes

Justification: The assets (datasets, code, models) used in the paper are open source, and our use follows the relevant protocols.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, [paperswithcode.com/datasets](http://paperswithcode.com/datasets) has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

### 13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: Yes

Justification: Documentation will be provided alongside the released code.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

### 14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: NA

Justification: Our paper doesn't involve the crowdsourcing experiments and research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

### 15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: NA

Justification: Our paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

#### 16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorosity, or originality of the research, declaration is not required.

Answer: NA

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.