Text-only Data Example in the Training Set

Query: What is the impact of using the same dataset for optimizing and measuring the performance of a model?

Text: Here, again, the unfair advantage of optimizing (selecting the models for the ensemble) and measuring performance on the same dataset appears.The advantage is small but systematic for the test split of ISIC (Fig. 5a); it is much more apparent for the challenging collection of clinical images of EDRA Atlas (Fig. 5b).

Answer 1: It can lead to an unfair advantage for the model.
Answer 2: Optimizing a model involves selecting certain parameters or features that improve its performance on a given dataset. If the same dataset is used to measure the model's performance, it may lead to an unfair advantage as the model has already been "tuned" to that specific dataset.

Text-image Data Example in the Training Set

Query: What is the best performing method for both detection and classification, according to the provided figure?

Text: 3) The last one is our proposed SFCN-OPI with both sibling branches and OPI (Ours in Table 1).
Figure: Table 1

Methods	Detection			Classification		
	P	R	F1	P	R	F1
FCN-5CLS	0.741	0.867	0.790	0.466	0.264	0.298
SFCN	0.784	0.844	0.807	0.450	0.561	0.496
SFCN-OPI-1	0.764	**0.890**	0.816	0.573	0.667	0.613
SFCN-OPI-2	0.788	0.885	0.828	0.674	0.759	0.711
SSAE	0.617	0.644	0.630	-	-	-
LIPSyM	0.725	0.517	0.604	-	-	-
CRImage	0.657	0.461	0.542	-	-	-
SR-CNN	0.783	0.804	0.793	-	-	0.683
SC-CNN	0.781	0.823	0.802	-	-	0.692
Ours	**0.819**	0.874	**0.834**	**0.718**	**0.774**	**0.742**

Note: the - means the results were not reported by that method.

Caption: Experimental results of ablation analysis, different training strategies of our method and comparison with other approaches. Note: the -means the results were not reported by that method.

Answer 1: Ours.
Answer 2: The best performing method for both detection and classification is Ours, as it achieves the highest F1 scores for both tasks. This can be seen in the "Ours" row of the table, where the F1 score for detection is 0.834 and the F1 score for classification is 0.742.