

SUPPLEMENTARY MATERIALS

A. Summary of Notational Conventions Used in the Taper

TABLE SI
THE SUMMARY OF SYMBOLS

Symbol	Meaning
T	The whole training set which contains V views
T_v^l	The training set of v -th view in l -th layer
K	The total number of classes.
N	The total number of samples in T
D_v	The dimensionality of the v -th view in T
Y_k	The k -th class in T
(x_i, y_i)	The sample with index i
F	The MVU-DF model
L^l	The l -th layer of MVUDF
$unit_v^l$	The v -th view learner in L^l
$view_v^l$	The v -th view in L^l , which contains $unit_v^l$ and A_v^l
$f_{v,s}^l$	The s -th forest in $view_v^l$
T_v^l	Training set of $view_v^l$
T^l	Training set of L^l
M	The number of trees in forest.
V	The number of views.
S	The number of forests in a view
p_k	The probability assigned to the k -th class for a sample by a single decision tree.
e_k	The evidence for the k -th class assigned to a sample by a single decision tree
E_k	The evidence for the k -th class assigned to a sample by a forest
b_k	The belief for the k -th class for a sample.
W	The non-prior information weight, default is $M + K$
$U, U^l, U_v^l, U_{v,s}^l$	Uncertainty of a sample, uncertainty of L^l , uncertainty of $view_v^l$ and uncertainty of $f_{v,s}^l$
$\mathcal{M}, \mathcal{M}^l, \mathcal{M}_v^l, \mathcal{M}_{v,s}^l$	Opinion of a sample, opinions of L^l , opinions of layer $view_v^l$, opinions of $f_{v,s}^l$
\mathcal{P}	The probability that converted from opinion \mathcal{M}
B	The number of bootstrap samples in gRIT
\mathcal{T}	The pruned decision rule set by gRIT algorithm
A_v^l, A^l	Intra-view interaction feature set of $view_v^l$ and inter-view interaction feature set of L^l
a	Feature
c	Feature set
\mathcal{C}	Feature pool generated using Cartesian operations
\mathbb{P}	The combinations pool

g_v^l , $g_{v,s}^l$	The sum of Gini importance of $unit_v^l$, the Gini importance of $f_{v,s}^l$
$T_{norm}^l, T_{easy}^l, T_{hard}^l, T_{lein}^l$	The sample set of normal type, easy type, hard type and less-information type in sample type assignment of l -th layer
w^l	The sample weights of L^l
τ	The hard degree

B. Comparison Models

- AdaBoost is an ensemble learning model that boosts the accuracy of a model by sequentially training multiple weak learners. It adjusts the weights of training samples in each step, placing more emphasis on previously misclassified samples to build a powerful classifier.
- Random Forest constructs multiple decision trees for classification or regression tasks. It reduces model variance and overfitting risk by randomly selecting data samples and feature subsets, and makes predictions through voting or averaging.
- Extra Trees(ET) is similar to Random Forest, but it further randomizes the construction of each decision tree. It randomly selects features for splitting at each node, increasing model diversity and reducing variance.
- Deep Forest demonstrates that deep learning frameworks can be implemented using non-differentiable modules. It achieves state-of-the-art performance in various classification and hybrid modeling tasks.
- GradientBoosting (GBDT) reduces the loss function by iteratively training a series of weak learners. In each training iteration, new weak learners enhance the overall predictive performance of the model by fitting the residuals of previous weak learners.
- XGBoost combines gradient boosting with regularization techniques. It performs well on large-scale datasets and high-dimensional features, offering faster training speed and better prediction performance.
- Catboost is designed specifically for handling data with categorical features. It automatically handles the encoding of categorical features, reducing the burden of feature engineering, and excels in handling class imbalance and missing data.
- LGBM is an improved version based on XGBoost, enhancing training speed through leaf-wise decision trees and histogram-based algorithms. It exhibits outstanding performance and lower memory usage.

C. Evaluation Metrics

AUROC is a good general-purpose metric for binary classification problems, particularly when both the positive and negative classes are of equal concern. It measures the classifier's ability to differentiate between the two classes, without being tied to a specific threshold. AUROC is relatively insensitive to class imbalance, making it a balanced measure for both skewed and balanced datasets.

The formulas for the True Positive Rate (TPR) and False Positive Rate (FPR) are:

$$TPR = \frac{TP}{TP + FN}, \quad (S1)$$

$$FPR = \frac{FP}{FP + TN}, \quad (S2)$$

In these formulas, TP , FN , FP , and TN denote the numbers of True Positives, False Negatives, False Positives, and True Negatives, respectively.

The formulas of AUROC is:

$$AUROC = \int_0^1 TPR d(FPR). \quad (S3)$$

AUPR is a highly useful metric when the focus is primarily on the positive class, which is often the case in imbalanced datasets or when False Negatives are particularly costly. Unlike AUROC, which considers both positive and negative classes, AUPR focuses more on the performance of the classifier concerning the positive class, making it more sensitive to False Negatives.

$$Precision = \frac{TP}{TP + FP}, \quad (S4)$$

$$Recall = \frac{TP}{TP + FN}. \quad (S5)$$

The formulas of AUPR is:

$$AUPR = \int_0^1 Precision d(Recall). \quad (S6)$$

Accuracy serves as a quick and intuitive metric, quantifying the proportion of correctly classified instances. However, it can be deceptive in imbalanced datasets or when different misclassifications have varied costs. Accuracy is best employed for problems where each class is equally important and where the class distribution is reasonably balanced. The formula for Accuracy is:

$$Accuracy = \frac{TP + TN}{N}, \quad (S7)$$

where N represent total number of samples.

D. Additional Experiment Results

a. Comparison between MVU-DF and Other Models on Multi-view Composite Data

TABLE SII. A

AUPR COMPARISON BETWEEN MVU-DF AND OTHER MODELS ON MULTI-VIEW COMPOSITE DATA

AUPR	AdaBoost	ExtraTrees	RandomForest	GBDT	CascadeForest	XGBoost	CatBoost	LGBM	MVU-DF
view1	0.6039±0.0453	0.5998±0.0338	0.6012±0.0343	0.6198±0.039	0.605±0.0335	0.6113±0.0322	0.5924±0.0325	0.6029±0.0378	0.7243±0.0376
view2	0.6936±0.0463	0.7743±0.0387	0.7745±0.0364	0.7474±0.0509	0.7679±0.0373	0.7616±0.0376	0.7633±0.0381	0.753±0.0364	0.7936±0.0357
view3	0.7012±0.0436	0.7764±0.037	0.7734±0.0381	0.7652±0.0409	0.7902±0.0296	0.778±0.033	0.7835±0.0321	0.7785±0.0263	0.7981±0.028
view4	0.7423±0.0266	0.7447±0.0381	0.7427±0.0391	0.745±0.0352	0.7616±0.0361	0.7267±0.0299	0.7815±0.0289	0.7271±0.0319	0.7955±0.0351
v1+v2	0.6373±0.0353	0.7507±0.0348	0.636±0.0436	0.6376±0.0456	0.6854±0.0399	0.6413±0.0366	0.6545±0.0376	0.6307±0.0339	0.7531±0.0366
v1+v3	0.6332±0.0384	0.7551±0.0317	0.6568±0.0391	0.6334±0.0372	0.6979±0.0393	0.6441±0.0356	0.6536±0.0346	0.6362±0.0343	0.7713±0.0316
v1+v4	0.6237±0.0482	0.7103±0.0449	0.6191±0.0344	0.63±0.0296	0.6347±0.0364	0.613±0.0324	0.6246±0.0408	0.629±0.0456	0.7567±0.0335
v2+v3	0.6824±0.0484	0.7862±0.0358	0.7858±0.0356	0.7546±0.0444	0.7911±0.0321	0.7642±0.0381	0.7753±0.0354	0.7745±0.0396	0.7953±0.0329
v2+v4	0.7228±0.0536	0.7943±0.0288	0.7929±0.03	0.7864±0.044	0.7925±0.0285	0.7845±0.0375	0.7897±0.0358	0.7711±0.0374	0.8193±0.0331
v3+v4	0.743±0.0401	0.8008±0.0293	0.8015±0.0301	0.7934±0.0291	0.8075±0.0288	0.7842±0.0376	0.7903±0.0299	0.7915±0.0275	0.8195±0.0287
v1+v2+v3	0.6518±0.0441	0.7678±0.0312	0.6711±0.0445	0.6382±0.0365	0.7201±0.0391	0.651±0.0347	0.6847±0.0438	0.6504±0.0362	0.7866±0.0318
v1+v2+v4	0.6437±0.032	0.7716±0.0335	0.6503±0.0449	0.6518±0.0441	0.7034±0.0432	0.6388±0.0409	0.6672±0.0459	0.6516±0.0332	0.795±0.0326
v1+v3+v4	0.6479±0.0399	0.774±0.0336	0.6669±0.0385	0.6485±0.0354	0.7103±0.0416	0.6505±0.0358	0.6727±0.0392	0.6554±0.038	0.8089±0.0292
v2+v3+v4	0.7214±0.0479	0.797±0.0311	0.7991±0.0316	0.785±0.033	0.7992±0.0251	0.7825±0.0274	0.7876±0.0315	0.793±0.0308	0.8212±0.0313
v1+v2+v3+v4	0.654±0.0356	0.7803±0.0296	0.6837±0.0472	0.658±0.0473	0.7339±0.0379	0.6555±0.0417	0.6902±0.0413	0.6684±0.0347	0.8164±0.0248
mean	0.6735±0.0417	0.7589±0.0341	0.7103±0.0378	0.6996±0.0395	0.7334±0.0352	0.6992±0.0354	0.7141±0.0365	0.7009±0.0349	0.7896±0.0322
win/tie/lose	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	-

TABLE SII. B

AUROC COMPARISON BETWEEN MVU-DF AND OTHER MODELS ON MULTI-VIEW COMPOSITE DATA

AUROC	AdaBoost	ExtraTrees	RandomForest	GBDT	CascadeForest	XGBoost	CatBoost	LGBM	MVU-DF
view1	0.4826±0.0516	0.4821±0.0474	0.4845±0.0428	0.4968±0.0493	0.4816±0.0409	0.4893±0.0387	0.4653±0.04	0.4859±0.0457	0.6434±0.0469
view2	0.622±0.0517	0.7096±0.0416	0.7087±0.0416	0.6747±0.0515	0.7066±0.0436	0.6932±0.0458	0.6948±0.0372	0.6886±0.0413	0.7292±0.0441
view3	0.6309±0.0475	0.7147±0.0417	0.7132±0.0416	0.7035±0.0412	0.7237±0.0333	0.7071±0.0351	0.7211±0.0363	0.7095±0.0345	0.7295±0.0402
view4	0.6556±0.0407	0.6901±0.0484	0.6889±0.0484	0.6756±0.0442	0.699±0.0452	0.6662±0.0417	0.7109±0.0373	0.6644±0.0423	0.7329±0.0433
v1+v2	0.5294±0.0394	0.682±0.0385	0.5274±0.0496	0.5192±0.0563	0.5902±0.0442	0.5252±0.0454	0.5528±0.0442	0.5158±0.0477	0.6862±0.0446
v1+v3	0.5182±0.0517	0.6861±0.0388	0.5606±0.0418	0.5206±0.0468	0.6134±0.0425	0.5362±0.0425	0.562±0.0403	0.5314±0.0422	0.7134±0.0412
v1+v4	0.5117±0.0598	0.606±0.0506	0.5019±0.0464	0.5073±0.0418	0.5234±0.046	0.4913±0.0365	0.5051±0.0459	0.509±0.0543	0.6896±0.0431
v2+v3	0.6096±0.0559	0.7265±0.0422	0.7248±0.0419	0.6895±0.0468	0.733±0.0388	0.7015±0.0393	0.7163±0.0359	0.7114±0.0381	0.7364±0.042
v2+v4	0.6583±0.0527	0.7377±0.0359	0.7368±0.0371	0.7216±0.0478	0.7368±0.0349	0.7202±0.0455	0.7223±0.0395	0.7066±0.0396	0.76±0.0422
v3+v4	0.6704±0.0428	0.7362±0.0368	0.7367±0.0398	0.7362±0.0339	0.7421±0.0357	0.7203±0.0372	0.7282±0.0349	0.7295±0.0354	0.7612±0.0404
v1+v2+v3	0.5473±0.0496	0.7067±0.0415	0.5797±0.0496	0.524±0.0511	0.65±0.043	0.5446±0.0473	0.5986±0.0464	0.5434±0.043	0.7278±0.0419
v1+v2+v4	0.5396±0.0469	0.7056±0.0422	0.5403±0.0525	0.5341±0.0537	0.6137±0.0489	0.5246±0.0473	0.557±0.0469	0.5418±0.0386	0.7392±0.0418
v1+v3+v4	0.5397±0.0496	0.7058±0.0409	0.5681±0.0436	0.5315±0.0491	0.6302±0.0433	0.547±0.0483	0.5827±0.0422	0.5493±0.0399	0.7486±0.0405
v2+v3+v4	0.6554±0.0482	0.7384±0.0393	0.7412±0.0385	0.7215±0.037	0.7406±0.0368	0.7198±0.034	0.7263±0.0367	0.7281±0.0334	0.7647±0.041
v1+v2+v3+v4	0.5525±0.0435	0.7213±0.0401	0.5936±0.0498	0.5349±0.0546	0.6651±0.0394	0.545±0.0507	0.6037±0.0438	0.5629±0.0478	0.7546±0.0368
mean	0.5815±0.0488	0.6899±0.0417	0.6271±0.0443	0.6061±0.047	0.6566±0.0411	0.6088±0.0424	0.6298±0.0405	0.6118±0.0416	0.7278±0.042
win/tie/lose	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	-

TABLE SII.C

ACCURACY COMPARISON BETWEEN MVU-DF AND OTHER MODELS ON MULTI-VIEW COMPOSITE DATA

Accuracy	AdaBoost	ExtraTrees	RandomForest	GBDT	CascadeForest	XGBoost	CatBoost	LGBM	MVU-DF
view1	0.5222±0.0424	0.5797±0.0284	0.5676±0.0292	0.5295±0.043	0.5835±0.0237	0.5429±0.0264	0.5765±0.0229	0.5511±0.0304	0.6295±0.0462
view2	0.6181±0.0346	0.6771±0.0392	0.6778±0.0372	0.6495±0.0349	0.6775±0.0367	0.6619±0.0424	0.6667±0.0321	0.666±0.0373	0.6883±0.0479
view3	0.6267±0.0411	0.6895±0.038	0.6895±0.0393	0.6819±0.0388	0.6997±0.0367	0.6663±0.0323	0.6917±0.0362	0.6705±0.0333	0.7013±0.0463
view4	0.6521±0.0404	0.6711±0.0417	0.6717±0.0414	0.6667±0.0416	0.6867±0.0418	0.6511±0.0397	0.6775±0.0345	0.6724±0.0311	0.6844±0.0419
v1+v2	0.554±0.04	0.6559±0.0257	0.5825±0.0252	0.5292±0.0389	0.6013±0.0192	0.5568±0.0336	0.5873±0.0184	0.5638±0.0294	0.6917±0.0457
v1+v3	0.5381±0.0446	0.6648±0.0286	0.5927±0.0275	0.5381±0.0459	0.621±0.0218	0.5698±0.0313	0.593±0.0228	0.5775±0.0329	0.7057±0.044
v1+v4	0.5352±0.0449	0.6102±0.0298	0.5787±0.0297	0.5152±0.0386	0.5952±0.0203	0.5432±0.0343	0.5895±0.0242	0.5517±0.0337	0.6892±0.0424
v2+v3	0.6013±0.0466	0.693±0.0417	0.6921±0.0445	0.6524±0.0418	0.6933±0.0371	0.673±0.034	0.6851±0.032	0.6787±0.0338	0.7165±0.0467
v2+v4	0.6429±0.0408	0.6902±0.0399	0.6902±0.0352	0.6819±0.0396	0.6943±0.0335	0.6816±0.0377	0.6876±0.0299	0.6686±0.0339	0.7203±0.0428
v3+v4	0.6524±0.0391	0.6924±0.0308	0.6895±0.0369	0.6937±0.0437	0.7029±0.0382	0.6683±0.0326	0.6832±0.0393	0.6857±0.0365	0.7216±0.0452
v1+v2+v3	0.5635±0.0316	0.6825±0.0304	0.6038±0.0267	0.5276±0.0452	0.6432±0.0258	0.573±0.0351	0.6067±0.028	0.581±0.0269	0.7076±0.0457
v1+v2+v4	0.5594±0.0368	0.6676±0.0296	0.5851±0.0258	0.5454±0.0456	0.6146±0.0215	0.5708±0.04	0.5946±0.0243	0.5721±0.0283	0.72±0.0477
v1+v3+v4	0.5575±0.04	0.6705±0.0368	0.599±0.0255	0.5346±0.0372	0.6241±0.0236	0.5775±0.043	0.6048±0.0228	0.5762±0.0257	0.7143±0.0482
v2+v3+v4	0.64±0.0398	0.7019±0.0409	0.7032±0.0398	0.6813±0.0381	0.699±0.0357	0.6702±0.0325	0.6857±0.0334	0.6784±0.0351	0.7267±0.0456
v1+v2+v3+v4	0.5616±0.029	0.6848±0.0212	0.6095±0.041	0.5381±0.0223	0.6524±0.036	0.5705±0.0266	0.6083±0.034	0.5762±0.0449	0.7127±0.0449
mean	0.5883±0.0394	0.6687±0.0335	0.6355±0.0337	0.5977±0.0397	0.6526±0.0301	0.6118±0.0348	0.6359±0.029	0.618±0.0329	0.702±0.0454
win/tie/lose	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	-

b. AS curve in Other Metrics

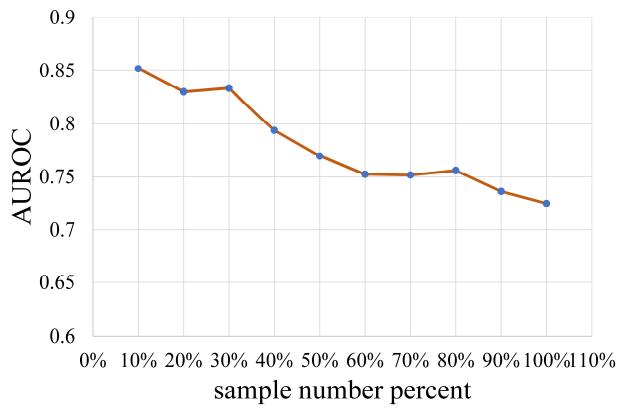


Figure S1 A. AS Curve in AUROC

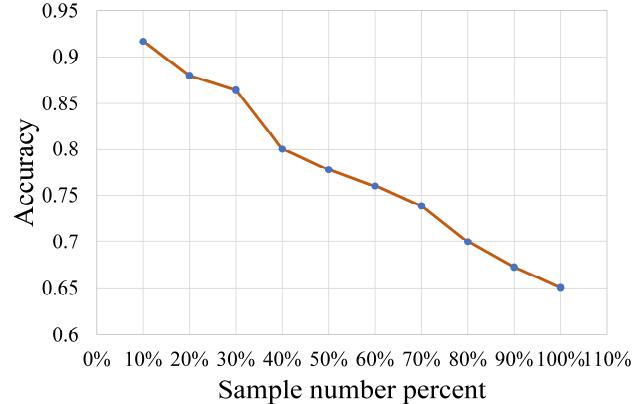


Figure S1 B. AS Curve in Accuracy

c. Feature Importance in Final Layer

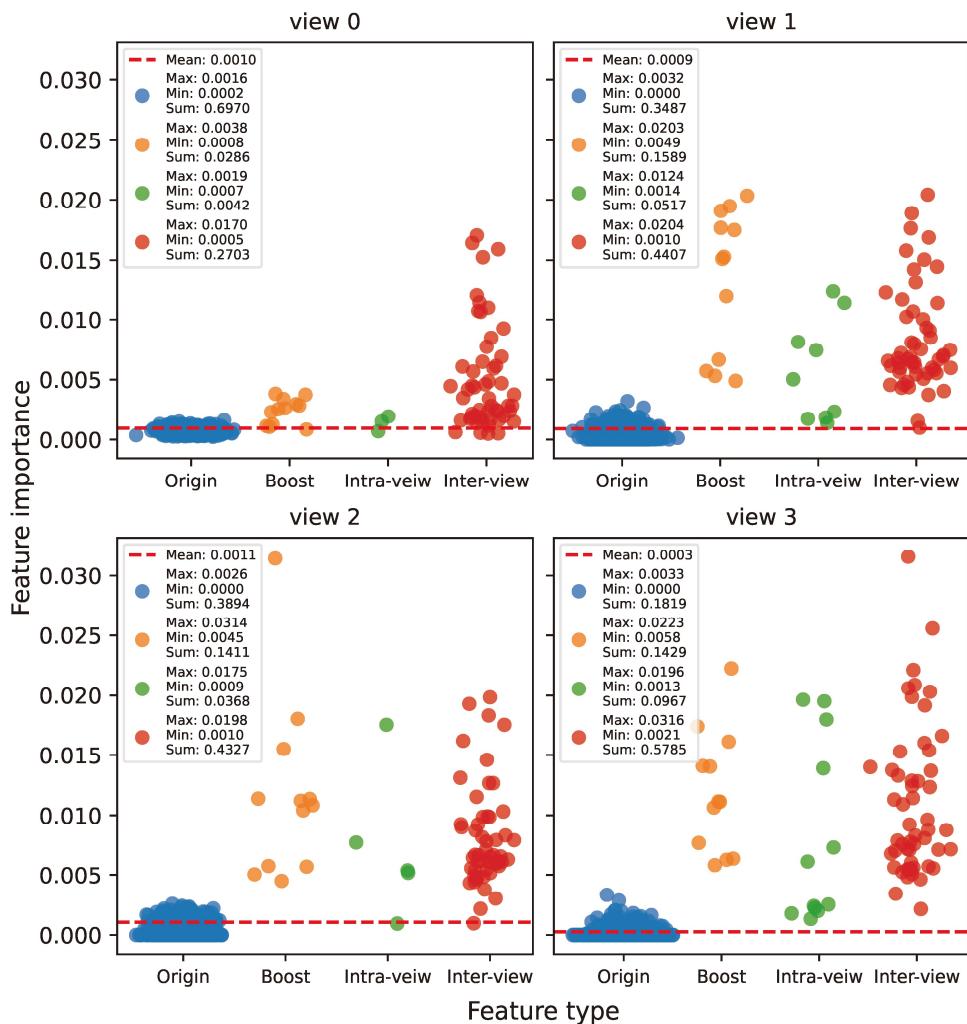


Figure S2 Feature importance in final layer for all views

TABLE SIII
THE NUMBER OF FEATURES FOR THE FOUR TYPES OF FEATURES.

View ID	# of origin	# of boost	# of intra-view	# of intra-view
1	978	12	3	53
2	1024	12	9	53
3	881	12	5	53
4	3705	12	12	53

d. Impact of Enabling or Disabling Inter-view Feature Interactions on View Performance

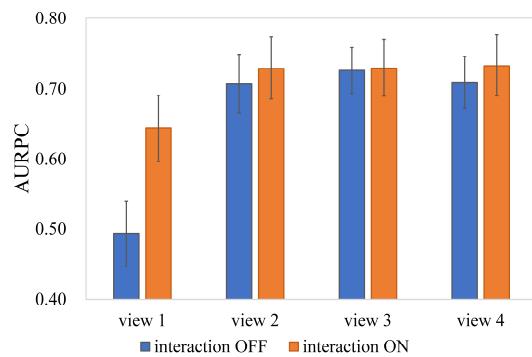


Figure S3 A Impact of enabling or disabling inter-view feature interactions on view performance with AUROC.

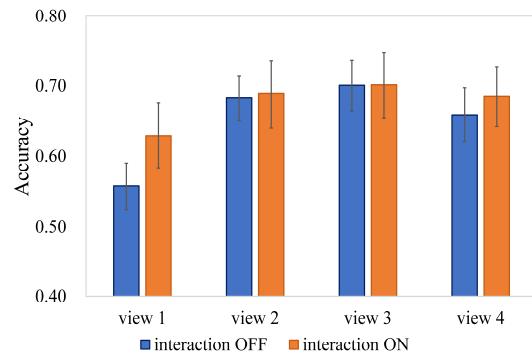
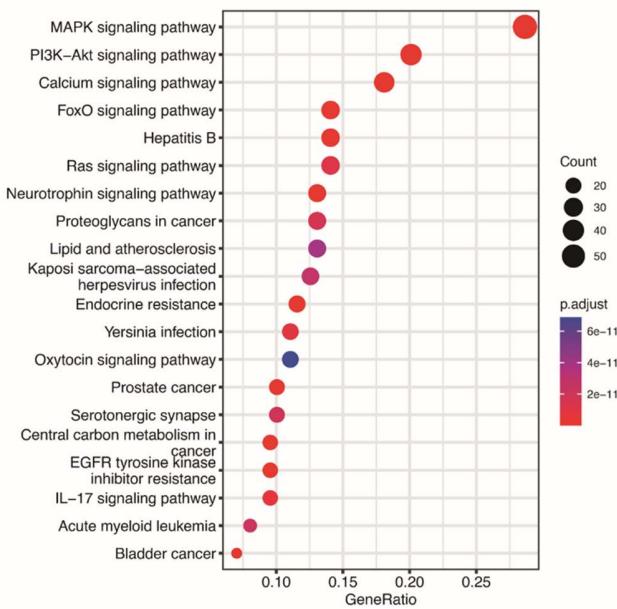


Figure S3 B Impact of enabling or disabling inter-view feature interactions on view performance with Accuracy.

E. Biological Pathways Enrichment of Two Uncertainty Sample Sets

A



B

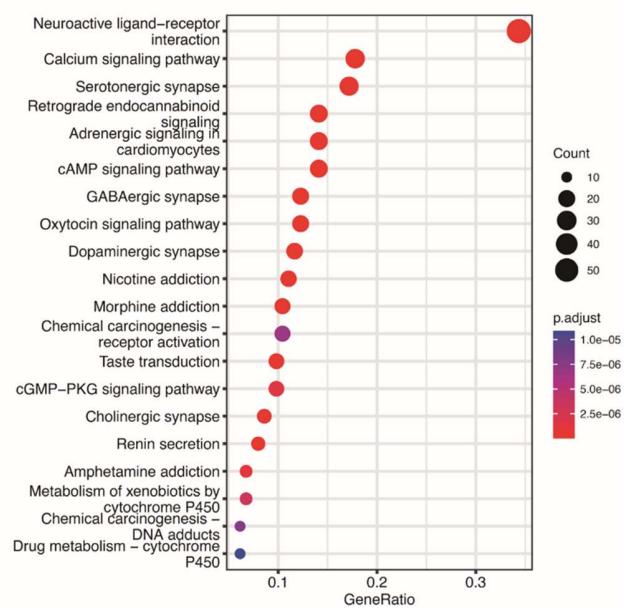


Figure S4. Biological pathways enrichment of two uncertainty sample sets

F. Data Complexity Criteria

We employed the following criteria to assess both subsets:

- 1) **Fraction of points on boundary (N1)** is a metric used in the context of machine learning and data analysis. It quantifies the proportion of data points that lie on or near the decision boundary of a classification or clustering algorithm. N1 is computed as the ratio of points located on the boundary to the total number of data points in the dataset. Boundary points are characterized as follows: For each sample, we closely examine the labels of its k nearest neighbors. If any of these neighboring labels differ from its own category, then we consider the sample to be on the decision boundary. It's worth noting that, when calculating the distance between samples, we use the Euclidean distance as our metric and set $k = 5$. This metric provides insights into the separability of data and the complexity of the decision boundaries of a given problem, which is valuable for evaluating the performance and generalization of machine learning models. A higher N1 value suggests a larger number of boundary points, indicating that the samples in the dataset are closer to the decision boundary, making the dataset potentially harder to be linearly separable.
- 2) **Non-parametric separability of classes (N2, N3)** are non-parametric metrics used to evaluate the separability of classes within a dataset without making strict assumptions about the data's underlying distribution. Next, we describe the computation methods for N2 and N3.

N2: Inter-class Nearest Neighbor Distance. For each sample in the dataset, we examine the labels of its $k = 5$ nearest neighbors, excluding the sample itself. If the label of the current sample differs from the label of its closest neighbor (the first nearest neighbor excluding itself), indicating they belong to different categories, we compute the Euclidean distance between the current sample and this neighbor. The N2 value represents the average of all these inter-class nearest neighbor distances.

N2 focuses on assessing class separability using methods that do not rely on predefined statistical models. N2 often employs distance-based techniques and nearest neighbor classifiers to determine the degree of separability between classes. This makes it particularly useful when dealing with datasets of varying complexities and unknown data distributions, as it offers flexibility in

measuring class separability based on the data's intrinsic characteristics. A smaller N2 value indicates less overlap between classes, making the samples easier to be distinguished based on their categories.

N3: Intra-class Nearest Neighbor Distance. For every sample in the dataset, we similarly observe the labels of its $k = 5$ nearest neighbors. If the label of the current sample matches the label of the farthest among its k nearest neighbors, indicating they are from the same category, we compute the average Euclidean distance between the current sample and all its k nearest neighbors. The (N3) value is the average of all these intra-class nearest neighbor distances.

N3 often explores density estimation, density-ratio estimation, and non-parametric clustering techniques to understand the underlying structure of the data. By doing so, it allows for a more flexible evaluation of class separability in complex datasets where traditional parametric models may not be suitable. A larger N3 value indicates greater separation between classes, implying an easier classification task.

- 3) **Fisher's discriminant ratio (F1):** Measures the tightness between two classes. A larger F1 value suggests that the difficulty of splitting two classes is lower.

$$F1 = \frac{(m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2}, \quad (S8)$$

where m_1 , m_2 , σ_1 and σ_2 are the means of the two classes and their standard deviation, respectively.

- 4) **Volume of overlap region (F2):**

$$F2 = \prod_i \frac{\min \max_i - \max \min_i}{\max \max_i - \min \min_i}, \quad (S9)$$

where $i = 1, \dots, D$ for a D-dimensional problem, and

$$\begin{aligned} \min \max_i &= \min\{\max(a_i, y_0), \max(a_i, y_1)\} \\ \max \min_i &= \max\{\min(a_i, y_0), \min(a_i, y_1)\} \\ \max \max_i &= \max\{\max(a_i, y_0), \max(a_i, y_1)\} \\ \min \min_i &= \{\min(a_i, y_0), \min(a_i, y_1)\} \end{aligned}$$

A smaller F2 value means less overlap between different classes, indicating potentially higher separability of the dataset.

