Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Trusted and Explainable Collective Learning for Conflictive Multi-View Decision-Making

Nengjun Zhu

zhu_nj@shu.edu.cn

Shanghai University

Zhiyu Zhang

Shanghai University

Shenghui Lan

The eighth people's hospital

Jian Cao

Shanghai Jiao Tong University

Siji Zhu

Ruijin Hospital

Research Article

Keywords: Multi-view learning, reliability modeling, interpretable model, multi-disciplinary decision-making, and conflictive opinion fusion

Posted Date: June 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8464754/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Knowledge and Information Systems on September 26th, 2026. See the published version at https://doi.org/10.1007/s10115-026-02884-1.

Trusted and Explainable Collective Learning for Conflictive Multi-View Decision-Making

Nengjun Zhu[1*], Zhiyu Zhang[1], Shenghui Lan[2], Jian Cao[3], Siji Zhu[4]

[1*]School of Computer Engineering and Science, Shanghai University, Shanghai, China.
[2]Department of Orthopaedics, The Eighth People's Hospital, Shanghai, China.
[3]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China.
[4]Department of General Surgery, Comprehensive Breast Health Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China.

*Corresponding author(s). E-mail(s): zhu_nj@shu.edu.cn;
Contributing authors: iri0116@shu.deu.cn;
d200978245@alumni.hust.edu.cn; cao-jian@sjtu.edu.cn;
zsj_mu@yeah.net;

Abstract

When processing multi-view data, conflicts may occur since different views have unique insights. Existing studies always consider conflict as a bad factor and thus adopt a negative operation, i.e., eliminating or minimizing conflicts. However, in a multi-view decision-making scenario, conflict can highlight view differences, inspire individuals to provide their own reasoning, and reveal the reliability of individual and collective decisions. To this end, we propose a novel trusted collective learning method (TECL) that can actively handle conflicts between individuals and obtain a collective opinion and its reliability interpretability by considering conflicts. Specifically, TECL first captures dual-concepts at both the feature and decision levels across different views. Then, it uses concepts to guide the assessment of individual opinion consistency. Subsequently, TECL learns view-specific evidence supporting individual opinions. To deal with conflicting opinions across views, TECL assembles a reliability-oriented collective learning

phase to determine which view owns a higher priority. It further combines priority and view-specific evidence to conclude the final collective opinion and its reliability. Experiments on a real-world multi-disciplinary consultation dataset demonstrate the superiority of our method and exhibit some interesting findings regarding conflictive multi-view decision-making. Our code is available at https://github.com/ifbettrer/TrustCL.

Keywords: Multi-view learning, reliability modeling, interpretable model, multi-disciplinary decision-making, and conflictive opinion fusion

1 Introduction

Multi-view data is widespread in real-world applications. Most existing solutions tend to integrate multi-view information with deep neural networks to achieve state-of-the-art performance in various application domains [1–6]. However, due to the property differences and inconsistency of multiple data sources, the inferred results by the multi-view deep learning method may be uncertain and unreliable since the traditional neural networks focus on the accuracy of the classifications but ignore the credibility and interpretability of the results, resulting in a significant limitation in various kinds of applications, especially in safety-critical applications (e.g., medical diagnosis [7] or autonomous driving [8]).

For reliability of results, an uncertainty-aware trusted multi-view classification (TMC) method [9] was proposed; however, it does not guarantee a decrease in the uncertainty of the result when integrating uncertain information extracted from the multi-view data. Most importantly, it spuriously *eliminates* conflicts by overlooking the "consistency" or "difference" between views in multi-view learning. Finally, it usually tends to choose informative views to make final decisions, making it hard to listen to the voice of the minority. Thus, for better modeling the conflicting views, an evidential conflictive multi-view learning (ECML) method [10] was proposed recently, which extracts evidence from different views and uses a conflicting opinion aggregation strategy to *minimize* their contradictory evidence. However, it simply treats evidence from each view equally and sums supporting evidence from different views. Still, it weakens the distinctive individual's contribution and underrates the value of conflicts in formulating a collective opinion. At the same time, neither of them takes into account the interpretability of the results and is unable to answer questions such as "Why did this view make such a decision?" For the interpretability of the results, several interpretable models [11, 12] have been proposed. Among them, the concept bottleneck model(CBM) [13] gave rise to a series of variants. However, the traditional CBM models pay more attention to why the final decision is made, while ignoring the concept of the feature level of individual views in multi-view data.

Nevertheless, real-world applications often require diverse perspectives and reasons for decisions, and conflicting views can foster deeper discussions and lead to better collective decisions. For instance, in a medical scenario involving a patient with a bone problem, a multidisciplinary discussion benefits from a holistic understanding

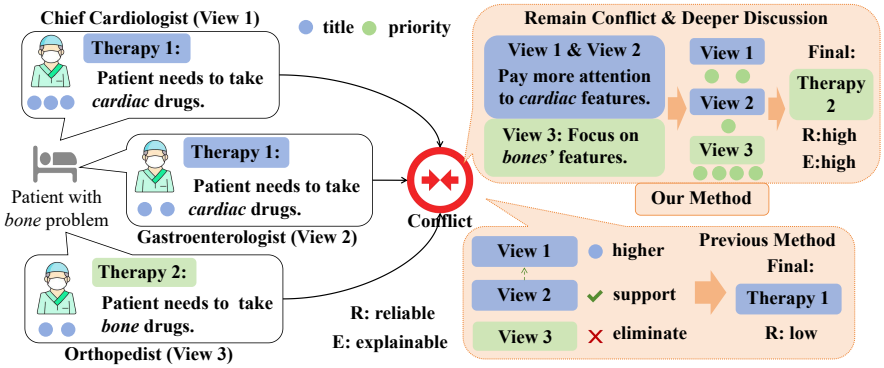


Fig. 1 Multidisciplinary discussion of situations.

of patient information. As shown in Figure 1, there is a *chief Doctor A*, who knows more about cardiology, and two *normal doctors*, B and C, with different expertise. They are in various positions and roles, which may lead to conflicting focuses when proposing therapies. A focuses on heart-related issues with B's support, while C pays more attention to the patient's bones because of his orthopedic expertise. In cases of conflict, previous studies often favored A's diagnosis due to its higher title (e.g., by modeling the static importance of views) or higher support (e.g., by considering the opinion of the majority), disregarding C's opinion. However, C's targeted expertise should be prioritized when addressing specific issues like orthopedic problems, as it can lead to more effective treatment. From another angle, because of A and B's disagreement, when taking C's opinion as the final therapy, the reliability of this decision will be reduced. Furthermore, existing research prominently focuses on the final therapy, failing to provide explanations for the individuals' underlying reasoning or the root causes of their conflicts. Consequently, the explainability of both individual and final decisions remains insufficient. In this case, we must consider each view's (e.g., doctor's) distinctive ability and decision in various instances (e.g., patients) and obtain them to serve as a basis for explaining the decision-making process. We can also evaluate the reliability of the conclusion by modeling the uncertainty of each view and the disagreement between their opinions.

To this end, we first propose a novel trusted collective learning method(TrustCL), which includes view-specific evidential learning, and reliability-oriented collective learning. To better explain the decision-making process, we then introduce dual-concepts learning and propose a trusted and explainable learning method(TECL). For dual-concepts learning, we use an attention mechanism to obtain the degree of attention each view pays to sample features, which serves as the feature concept. Subsequently, the enhanced features from each view, guided by their respective feature concepts, are input into decision concept extraction. It generates decisions for individual views, alongside their supporting attitudes toward the collective decision. Together, these outputs constitute the decision concepts. In the concepts guided consensus judgement stage, if the decisions and supporting attitudes of all views are consistent(i.e., without conflicts), a highly reliable collective decision can be directly obtained. If there

is no consensus, the process proceeds to the view-specific evidential learning stage to obtain a more feasible and reliable collective decision. For view-specific evidential learning, we first collect the supporting evidence using deep neural networks (DNNs), which can generate the probability of decision categories for a specific view. It enables a more accurate assessment of each view's trustworthiness and priority in subsequent individual modeling and conflict resolution. Different views may have conflicting opinions (i.e., decision categories), which we view as valuable information for reliability modeling. The conflicting individual opinions and the final collective opinion are all treated as learnable patterns of DNNs, so that DNNs can learn more about the focus of different views. Meanwhile, view-specific evidence is used to obtain the reliability of this view. For reliability-oriented collective learning, we merge individual opinions by considering their dynamic priority to obtain the collective opinion and its reliability. The dynamic priority is calculated based on the static importance of each view and the dynamic contribution of a view in different instances. Finally, we empirically compare TECL with multi-view learning baselines on real medical scenario datasets regarding multiple decision tasks and scenarios. Experiments show that TECL outperforms baselines from individual and collective perspectives on accuracy, Recall, and MRR metrics. The main contributions of our work are summarized as follows.

(1) We propose TrustedCL and TECL, which actively model conflicts among individuals, to highlight view distinctions and reveal the reliability of individual and collective decisions in multi-view decision-making.
(2) The proposed TECL can model concepts at both the feature level and individual decision-making level, providing reasons for individual and collective decisions within the model, thus enhancing its interpretability.
(3) The proposed methods can dynamically determine each view's priority and combine the priority and view-specific evidence to determine the final collective opinion and its reliability, improving the accuracy and credibility of decisions.
(4) Experiments on a real-world multidisciplinary consultation dataset demonstrate the superiority of TECL and TrustedCL and exhibit some interesting observations.

2 Related work

We briefly review related work on reliable multi-view learning, interpretable machine learning, and traditional collective learning, and discuss their relationship to the proposed trusted collective learning.

Reliable Multi-View Learning: Multi-view learning is an emerging direction that aims to improve generalization performance by leveraging multiple views or sources of data across various domains. By integrating information from diverse perspectives, multi-view learning methods effectively enhance accuracy, robustness, and generalization capabilities [14, 15]. Despite its potential advantages, ensuring the reliability of both the multi-view integration and the final decision is crucial, especially for noisy, corrupted, and conflicting data.

To handle conflicting data, the prevalent conflictive multi-view learning methods are mainly dedicated to eliminating the conflictive instance or minimizing conflicts. One line is based on partially view-aligned multi-view learning [16, 17]. Earlier

study [18] employed the differentiable agent of the Hungarian algorithm to establish alignment relationships for unaligned data. Along this line, researchers proposed noise-robust contrastive learning [19–21], the self-focused mechanism [22], to compute the alignment matrix. The seminal work, Trusted Multi-View Classification (TMC) [9], introduced the fusion of different views from an opinion perspective using the Dempster-Shafer Combination rule. Building upon TMC, [23] extended the approach by incorporating the pseudo-view, a concatenation of all other views, resulting in improved performance. Moreover, ECML [10] has been proposed and utilized to fuse conflicting opinions into a final solution to resolve conflicts. Concurrent research efforts, such as those by [24, 25], focus on multiview uncertainty estimation, enhancing the model's reliability.

However, these methods often operate negatively on conflicts, while each view in a real application has practical value, whether conflicting or not. Therefore, we propose a method to effectively merge opinions to reach a final decision while preserving the conflict.

Interpretable Machine Learning: Unlike post-hoc explanations of black-box models, interpretable models are designed to make their reasoning processes more understandable to humans, offering inherent ante-hoc interpretations for their decision-making [26]. In previous studies, the tree-regularization technique [27] and its variants [28] enable the decision boundaries of machine learning models to be approximated through a series of decision trees, thereby ensuring the interpretability of the target model. In cancer gene identification tasks, IMVRL-GCN [12] provides explanations by decomposing multi-view features into "shared features" and "specific features." VIB [29] filters out task-irrelevant noise through the information bottleneck principle, retaining only the most explanatory and crucial information flow.

In order to understand the decision made by models, CBMs [13, 30] and other variants [31] learn the expert-labeled concepts in the training data and utilize them for downstream tasks, which aim to make the AI system more interpretable and interactive. Among them, ante-hoc CBMs [13, 32, 33] learn the given concepts during the training process to obtain informative embedding. Some studies [34, 35] have pointed out that ante-hoc CBMs have serious label distortions in the absence of sufficiently high-level concepts. Our work adopts the method of constructing dual concepts and improves the label distortion by using low-level feature concepts to assist high-level decision concepts. In the medical field, further enhancing the interpretability of decisions helps doctors understand the decisions made by the model.

Traditional Collective Learning: Collective learning is a complex concept that is variously defined. It is generally conceptualized as a dynamic and cumulative process that produces knowledge [36]. Machine learning techniques associated with collective learning primarily involve proxy labeling operations on unsupervised data. Co-training [37, 38] leverages two (or more) views of the data, i.e., different sets of features representing the same data, to generate label information from each other. CL+GNN [39] proves that collective learning is a practical way to improve performance in the real world without changing the underlying architecture of GNN (graph neural network). It simulates label propagation in collective classification during training.

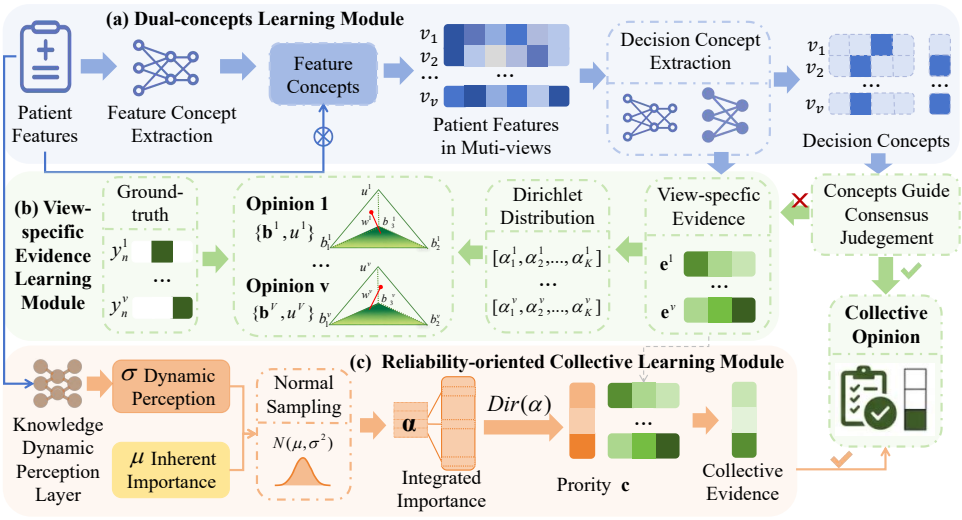


Fig. 2 Framework of TECL. The "red cross" symbol indicates that decisions have reached a consensus, while the "checkmark" symbol indicates that the decisions are in agreement or are sufficiently credible.

Trusted and Explainable Collective Learning: Although collective learning has been mentioned in previous approaches, its characteristics have been underutilized. For example, the existing approaches have not assigned specific roles to different individuals or inferred decision reliability by considering the conflicting opinions between individuals. Thus, inspired by the collective decision-making process in real-world scenarios, we propose a trusted and explainable collective learning method, which assesses individual opinions and formulates a final solution through the interaction of individuals, as a novel branch of reliable and explainable multi-view learning technologies.

3 The Definition and Method

3.1 Problem Definition

Suppose we are given a dataset with V views and N instances. We use $\boldsymbol{x}_n^v \in \mathbb{R}^{D_v} (v = 1, \ldots, V)$ to denote the feature vector for the v-th view of the n-th instance ($n = 1, ..., N$), where D_v is the dimensionality of the v-th view. The one-hot vector $\boldsymbol{y}_n^v \in \mathbb{R}^K$ and $\boldsymbol{y}_n \in \mathbb{R}^K$ denote the ground-truth category for the v-th view and collective opinion of the n-th instance, where K is the total number of categories. We also use $\hat{y}_n^v$ to present the ground-truth support attitude of v-th view for the n-th instance towards the collective opinion, with values of 0 or 1. In our problem, it allows $\boldsymbol{y}_n^v \neq \boldsymbol{y}_n$ and $\boldsymbol{y}_n^{v_i} \neq \boldsymbol{y}_n^{v_j}$. That is, the decision of one view might differ from the collective decision, and a conflict of decision exists between individuals (views).

Then, the goal of the proposed method is to accurately predict $\boldsymbol{y}_n^v$ and $\boldsymbol{y}_n$. Besides, it should provide the reliability of a specific view, i.e., R_n^v, and of collective opinion,

i.e., R_n. The priority of each view, i.e., c_n^v, is also calculated for conflictive multi-view decision-making.

3.2 Trusted and Explainable Collective Learning

Figure 2 shows the main frame of our proposed Trusted and Explainable Collective Learning (**TECL**), consisting of a **dual-concepts learning module**, a **view-specific evidence learning** module, and a **reliability-oriented collective learning** module.

In the first module, we use an attention mechanism to obtain the degree of attention an individual pays to sample features, which serves as the feature concepts $\boldsymbol{a}^v \in \mathbb{R}^{D_v}$. Then, we input the enhanced instance $\hat{\boldsymbol{x}}^v \in \mathbb{R}^{D_v}$ from each view, guided by their respective, into decision concepts extraction, which consists of an attitude capture network and DNN. It generates decisions of individual views, alongside their supporting attitudes toward the collective decision, which serve as the decision concepts $\boldsymbol{q}^v$. In the concepts guided consensus judgement stage, if the decisions and supporting attitudes of all views are consistent(i.e., without conflicts), a highly reliable collective decision can be directly obtained. If there is no consensus, the process proceeds to the second module to obtain a more feasible and reliable collective decision.

In the second module, we use evidential DNNs to collect each view's evidence $\boldsymbol{e}^v$ based on a supervised paradigm. After that, we can obtain the distribution of decision categories regarding each view, i.e., $\boldsymbol{\alpha}^v$, by parameterizing the view-specific evidence using a Dirichlet distribution. According to the distribution, we can calculate belief mass and uncertainty to represent the view's opinion $\boldsymbol{w}^v \in \mathbb{R}^{K+1}$. Specifically, we have fully considered the significance of individual viewpoints since we take $\boldsymbol{y}^v$ instead of $\boldsymbol{y}$ as a supervised signal. By doing so, we retain the conflicts and differences among different viewpoints. It enables DNNs to learn more informative and individual knowledge about specific viewpoints and obtain their high-order uncertainty values.

In the third module, we propose a reliability-oriented collective learning method. Specifically, we define priority distribution $\boldsymbol{c} \in \mathbb{R}^V$ as a weight to combine the evidence of each view. The priority distribution is learned through inherent importance $\boldsymbol{\mu} \in \mathbb{R}^V$ and dynamic perception importance $\boldsymbol{\sigma} \in \mathbb{R}^V$, where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are parameters of a normal distribution regarding the relations of different views when making decisions. Based on this, we fuse evidence, preserving characteristics and conflicts across views. Details will be elaborated as follows.

3.2.1 Dual-concepts Learning Module

This module aims to capture the dual concepts at both the feature and decision levels, acquiring enhanced features guided by feature concepts, and providing explanations for individual and collective opinions.

Given an input $\boldsymbol{x}_n^v$, which comprises feature information, the Feature Concepts Extraction based on attention mechanism captures the feature concepts $\boldsymbol{a}_n^v$, which consist of the varying degrees of instance-specific attention that different views pay to

various features:

$$\boldsymbol{a}_n^v = Attention(\boldsymbol{x}_n^v) \tag{1}$$

Then, through feature concepts guidance, we obtain the enhanced instances $\hat{\boldsymbol{x}}_n^v$:

$$\hat{\boldsymbol{x}}_n^v = \boldsymbol{x}_n^v \circ \boldsymbol{a}_n^v \tag{2}$$

Next, the enhanced instances are fed into an attitude capture network $h^v(.)$ and an evidential DNN $f^v(.)$ respectively to obtain the support attitude and view's evidence support. These two elements together form the decision concepts $\boldsymbol{q}_n^v$:

$$\boldsymbol{q}_n^v = \{f^v(\hat{\boldsymbol{x}}_n^v), h^v(\hat{\boldsymbol{x}}_n^v)\} \tag{3}$$

where support attitude $h^v(\hat{\boldsymbol{x}}_n^v)$ is a fixed value of either 0 or 1, indicating whether the current view supports the final decision. The view's evidence support will be discussed in detail in the next section. Finally, we initially obtain highly reliable collective opinions by determining whether the decision concepts are highly consistent (i.e., the opinions and attitudes are the same).

3.2.2 View-specific Evidence Learning Module

This module aims to get the evidence supporting and the reliability of each view. Thus, we need to calculate each view's evidence support $\boldsymbol{e}^v$ and reliability R^v. Inspired by the idea of subjective logic methods [40], we utilize evidential deep learning (EDL) to get the multinomial opinion and measure the reliability. It has two procedures shown as follows:

Procedure 1: Evidence acquisition. To facilitate the calculation of evidence, we first use subjective logic as follows: For K classification problems, a multinomial opinion of a specific view on an enhanced instance $\hat{\boldsymbol{x}}_n^v$ can be represented as a learnable vector $\boldsymbol{w} = \boldsymbol{b} \cup [u]$, where $\boldsymbol{b} = [b_1, \ldots, b_K]^\top$ assigns belief masses to categories of the instance based on the learned evidence, and the uncertainty mass u captures the degree of ambiguity or vacuity in the evidence. All elements of $\boldsymbol{w}$ are required to be non-negative, ranging from 0 to 1, and their sum should be equal to one as follows:

$$u + \sum_{k=1}^{K} b_k = 1 \tag{4}$$

To obtain b_k, we must get evidence support first, which is modeled by a view-specific evidential DNN $f^v(.)$:

$$\boldsymbol{e}^v = f^v(\hat{\boldsymbol{x}}_n^v \mid \boldsymbol{\theta}^v) \tag{5}$$

Among them, $\boldsymbol{e}^v = (e_1^v, \ldots, e_K^v)^\top$ and it is viewed as evidence support, and $\boldsymbol{\theta}^v$ are parameters of the v-th view-specific DNN. When we obtain $\boldsymbol{e}^v$, we add a Softplus layer to ensure that they are non-negative.

Procedure 2: Reliability calculation. Based on Equation (4)-(5), we can obtain the belief mass and uncertainty:

$$b_k^v = \frac{e_k^v}{S^v}, \quad u^v = \frac{K}{S^v} \tag{6}$$

where $S^v = \sum_{k=1}^{K}(e_k^v + 1)$ and it is a normalization term.

After the belief masses vector $\boldsymbol{b}^v$ and uncertainty mass u^v are obtained, we then use the opinion, i.e., $\boldsymbol{w^v} = \boldsymbol{b}^v \cup [u^v]$, to define the reliability R^v of the opinion as follows:

$$R^v = 1 - u^v = 1 - \frac{K}{S^v} = \sum_{k=1}^{K} b_k^v \tag{7}$$

According to the Dirichlet Probability Density Function (PDF) [10], the uncertainty in Dirichlet reflects the emptiness of evidence. If a class has a higher value of b_k^v or e_k^v, then its uncertainty will be lower, and it is more likely to be treated as the label of an instance because of its lower emptiness. Thus, the class probability d_k^v can be computed as $d_k^v = (e_k^v + 1)/S^v = \alpha^v/S^v$, where S^v and α^v are the Dirichlet strength and Dirichlet parameter in PDF, respectively.

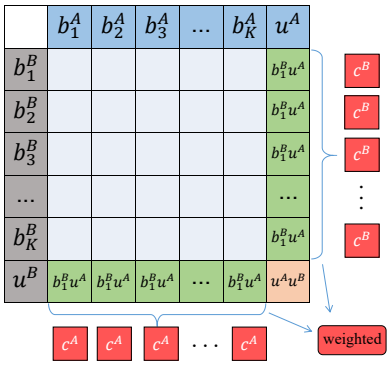


Fig. 3 Weighted projections.

3.2.3 Reliability-oriented Collective Learning Module

As mentioned, the knowledge from the multi-view data would show high conflict. These conflicts with view characteristics should be preserved, so we introduce reliability-oriented collective learning (RoCL) to model the conflict across views. RoCL pays attention to both the individual and the collective. It aggregates the explanation and conflicts of the individual to the whole, and hence the collective expressiveness. Then, we could capture unique knowledge and information through the RoCL's interaction mechanism.

Specifically, we first define view priorities $\boldsymbol{c} = (c^1, ..., c^V)^\top$, which focus on both the inherent and instance-specific importance of views. We integrate these two important

factors by modeling a Dirichlet distribution. Then, the priority vector $\boldsymbol{c}$ is applied to generate collective evidence and opinion by a weighted combination, and c^v is treated as a weight of view v. The detailed definitions are as follows.

Definition 1: Collective Opinion Generation. Let $\boldsymbol{w}^A = \boldsymbol{b}^A \cup [u^A]$ and $\boldsymbol{w}^B = \boldsymbol{b}^B \cup [u^B]$ be the opinions of view A and B in the same instance, respectively. Then, the collective opinion of these views, $\boldsymbol{w}^{A\underline{\Diamond}B} = \boldsymbol{w}^A \underline{\Diamond} \boldsymbol{w}^B = \boldsymbol{b}^{A\underline{\Diamond}B} \cup [u^{A\underline{\Diamond}B}]$, can be calculated as follows:

$$b_k^{A\underline{\Diamond}B} = \frac{c^A b_k^A u^B + c^B b_k^B u^A}{c^B u^A + c^A u^B} \tag{8}$$

$$u^{A\underline{\Diamond}B} = \frac{u^A u^B}{c^B u^A + c^A u^B} \tag{9}$$

where c^A and c^B represent the view priorities. The opinion $\boldsymbol{w}^{A\underline{\Diamond}B}$ represents the combination of the dependent opinions of A and B. As shown in the Figure 3, the molecule in Equation (8) is equivalent to, in the confidence space, using weights c^A and c^B to perform weighted projections on the vectors of the two views. The overlapping part (i.e., consensus confidence) $b_k^A u^B$ and $b_k^B u^A$ is enhanced. The size of the uncertainty region is inversely proportional to the degree of overlap of the confidence regions of the two views. $c^B u^A + c^A u^B$ normalizes and ensures that the total sum is a probability value. If the two views conflict, for example, both u^A and u^B are small, $c^B u^A + c^A u^B$ will decrease, which lead to the increase of $u^{A\underline{\Diamond}B}$.

Similarly, after obtaining the view priorities for more than three views, we can reach the collective opinion and evaluate its uncertainty according to the fusion rules:

$$\boldsymbol{w} = \boldsymbol{w}^1 \underline{\Diamond} \boldsymbol{w}^2 \underline{\Diamond} ... \boldsymbol{w}^V \tag{10}$$

A more detailed explanation is shown in the Discussion and Analysis. Next, we will introduce how to calculate the view priorities according to Definition 2.

Definition 2: The Priority of Individual View. A view's inherent/prior importance and instance-specific importance should determine the priorities simultaneously. The former is instance-free and is usually highly related to each view's historical position. For example, in real-world applications, when making collective decisions, the leader could guide the direction of discussions. Their opinions are more aligned with the final collective opinions and should therefore have higher priority. The latter requires that a view substantially impacts the decision of a specific instance, not due to its position, but because it may provide significant evidence.

Therefore, we employ a normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}_n^2)$ to model these two factors. The learnable mean vector $\boldsymbol{\mu} \in \mathbb{R}^V$ represents the prior distribution of view importance, whose initial value is the decision proportion of each view obtained through statistical normalization of the final decision situation in the dataset. And the variance $\boldsymbol{\sigma}_n \in \mathbb{R}^V$ represents the degree of priority shift for a specific instance $\boldsymbol{x}_n$. It is a dynamic perception that considers the relations between views and instances. Then, driven by the generation of evidence, we utilize the Dirichlet distribution to generate

the instance-specific priority distribution as follows:

$$\boldsymbol{c}_n \sim \text{Dirichlet}\left(\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}_n^2)\right) \tag{11}$$

where a sampled vector from the normal distribution is treated as a Dirichlet parameter that guides priority sampling. By doing this, we assign the generated pseudo-evidence to each view as weights.

Based on it, each element σ_n^v in $\boldsymbol{\sigma_n}$ is calculated as the output of a dynamic perception layer $g(\cdot)$ which captures the dynamic importance of a view:

$$\sigma_n^v = g(\hat{\boldsymbol{x}}_n^v) \tag{12}$$

Thus, we can obtain $\boldsymbol{\sigma_n} = [\sigma_n^1, \sigma_n^2, ..., \sigma_n^V]$ by training multiple view-specific perception layers. This is mainly to capture the changes in the instance-specific view importance caused by the fluctuating characteristics of different instances.

So far, we have obtained individual and collective opinions, and the dynamic priority of each view on a specific instance.

3.2.4 Loss Function

In this subsection, we introduce how to train view-specific evidential DNNs while considering historical individuals and collective opinions. For the details of Equation (5), we transfer traditional DNNs into evidential DNNs, in the context of reliability, by using $Softplus$ activation function as the output layer to ensure the output $\boldsymbol{e}^v$ of each view is non-negative.

As mentioned before, we must consider reliability and evidence in constructing loss functions. Traditional cross-entropy only ensures the correctness of the prediction. Thus, we employ the Dirichlet distribution to provide prior information on category probabilities and calculate the uncertainty under this condition. Specifically, this distribution's expectation can measure second-order uncertainty to ensure that the evidence of each view is more reliable, while the probability values in the softmax layer only capture first-order uncertainty.

The Probability Density Function (PDF) of the Dirichlet distribution is parameterized by $\boldsymbol{\alpha^v} = \left(\alpha_1^v, \ldots, \alpha_K^v\right)^\top$ and can get the second-order probability of each view and form the multinomial opinion $D(\boldsymbol{p}^v|\boldsymbol{\alpha}^v)$:

$$D(\boldsymbol{p}^v|\boldsymbol{\alpha}^v) = \begin{cases} \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^K p_k^{\alpha_k - 1}, & \textit{for } \boldsymbol{p}^v \in \mathcal{M}_K \\ 0, & \textit{otherwise} \end{cases} \tag{13}$$

where $B(\boldsymbol{\alpha})$ is the $Beta$ function, $\mathcal{M}_K$ is the K-dimensional unit simplex and $\boldsymbol{p^v} = (p_1^v, \ldots, p_K^v)^\top$ is the class assignment probabilities on the simplex $\mathcal{M}_K$, adhering to the following rules:

$$\sum_{k=1}^K p_k^v = 1 \tag{14}$$

The parameters of the Dirichlet distribution $\boldsymbol{\alpha}^v$ are defined as $\boldsymbol{\alpha}^v = \boldsymbol{e}^v + \mathbf{1}$ based on evidence supporting view-specific evidential DNNs.

Furthermore, utilizing the properties, e.g., $\boldsymbol{p}^v$, of the *Beta* and digamma function, the simplified loss calculation method of the specific view v is adjusted as follows:

$$\begin{aligned} L_{\text{ace}}(\boldsymbol{\alpha_n^v}, y^v) &= \int \left[\sum_{j=1}^{K} -y_{nj}^v \log p_{nj}^v \right] \frac{1}{B(\boldsymbol{\alpha_n^v})} \sum_{j=1}^{K} p_{nj}^{\alpha_{nj}^v - 1} \, dp_n^v \\ &= \sum_{j=1}^{K} y_{nj}^v \left(\psi(S_n^v) - \psi(\alpha_{nj}^v) \right) \end{aligned} \tag{15}$$

where the $\psi(.)$ is the digamma function, and S_n^v is defined in Equation (6). Equation (15) is the integral of the cross-entropy loss function on the simplex determined by $\boldsymbol{\alpha_n^v}$. It is more suitable for our evidence-based method, which can ensure that the correct label of each sample is higher than that of other categories.

However, it cannot guarantee that incorrect labels will generate less evidence. In order to predict that the evidence of incorrect labels will narrow down to 0, the following Kullback-Leibler (KL) divergence is employed, which can minimize the difference between the Dirichlet distributions after removing the non-misleading evidence, i.e., $D(\boldsymbol{p}^v|\tilde{\boldsymbol{\alpha}}^v)$, and the uniform Dirichlet distribution, i.e., $D(\boldsymbol{p}^v|\mathbf{1})$:

$$\begin{aligned} L_{KL}(\boldsymbol{\alpha}_n^v) &= KL[D(\boldsymbol{p}_n^v|\tilde{\boldsymbol{\alpha}}_n^v) \| D(\boldsymbol{p}_n^v|\mathbf{1})] \\ &= \log \left(\frac{\Gamma\left(\sum_{k=1}^{K} \alpha_{nk}^v\right)}{\Gamma\left(\prod_{k=1}^{K} \alpha_{nk}^v\right) \prod_{k=1}^{K} \Gamma(\tilde{\alpha}_{nk}^v)} \right) \\ &+ \sum_{k=1}^{K} (\tilde{\alpha}_{nk}^v - 1) \left[\psi(\tilde{\alpha}_{nk}^v) - \psi\left(\sum_{j=1}^{K} \tilde{\alpha}_{nj}^v \right) \right] \end{aligned} \tag{16}$$

where $\tilde{\boldsymbol{\alpha}}_n^v = (\alpha_n^v - 1) * (1 - y_n^v) + 1$ is the Dirichlet parameters after removal of the non-misleading evidence from predicted parameters α_n^v.

Meanwhile, guided by decision concepts, we use binary cross entropy with logits loss. The calculation method is as follows:

$$L_{bce}(q_n^v, \hat{y}^v) = -\left[\hat{y}_n^v \cdot \log(\sigma(q_n^v)) + (1 - \hat{y}_n^v) \cdot \log(1 - \sigma(q_n^v)) \right] \tag{17}$$

where $\hat{y}_n^v = 1 - (y_n^v \oplus y_n)$, indicates whether the current view's label aligns with collective opinion.

Based on these three parts, each instance $\boldsymbol{x}_n$ can get the loss $L_{\text{acc}}(\boldsymbol{\alpha}_n^v)$ of each view:

$$L_{\text{acc}}(\boldsymbol{\alpha}_n^v) = L_{\text{ace}}(\boldsymbol{\alpha}_n^v, y^v) + \lambda_t L_{KL}(\boldsymbol{\alpha}_n^v) + L_{bce}(q_n^v, \hat{y}^v) \tag{18}$$

where $\lambda_t = \min\left(1.0, \frac{t}{T}\right) \in [0, 1]$ is the annealing coefficient, t is the index of the current training epoch, and T is the annealing step. If the network pays too much attention to KL divergence in the early stage of training, it may lead to a lack of good exploration of the parameter space, resulting in a flat and uniform distribution of the network output. To prevent this situation from occurring, we can gradually increase the value of λ_t during training.

In summary, the view-specific loss is represented by $L_{\text{acc}}(\boldsymbol{\alpha}_n^v)$, while the collective loss is represented by $L_{\text{acc}}(\boldsymbol{\alpha}_n)$.

$$L_{\text{acc}}(\boldsymbol{\alpha}_n) = L_{\text{ace}}(\boldsymbol{\alpha}_n, y) + \lambda_t L_{KL}(\boldsymbol{\alpha}_n) \tag{19}$$

where $\boldsymbol{\alpha}_n = \boldsymbol{e}_n + \mathbf{1}$ and $\boldsymbol{e}_n$ is fused evidence support based on Equation (21). The overall loss function for a specific instance $\{\boldsymbol{x}_n^v\}_{v=1}^V$ can be calculated as:

$$L_{overrall} = L_{\text{acc}}(\boldsymbol{\alpha}_n) + \beta \sum_{v=1}^{V} L_{\text{acc}}(\boldsymbol{\alpha}_n^v) \tag{20}$$

By focusing on the correctness of individuals, the loss will penalize the result of each view and the overall prediction. Most existing deep learning methods aim to eliminate conflicts between multiple views. In contrast, our model does not prioritize the similarity of results among multinomial opinions. TrustCL regards different views as the whole and conflict as the motive.

3.3 Discussion and Analysis

Discussion. In most previous studies, final uncertainty is only related to the uncertainty of each view, ignoring its importance. Instead, we consider both the individual uncertainty and the priority of views when fusing evidence support. The collective opinion generation, i.e., $\boldsymbol{w}^{A\underline{\Diamond}B} = \boldsymbol{w}^A \underline{\Diamond} \boldsymbol{w}^B$, is actually equivalent to aggregating view-specific evidence using a weighted mechanism, i.e., $\boldsymbol{e}^{A\underline{\Diamond}B} = c^A * \boldsymbol{e}^A + c^B * \boldsymbol{e}^B$.

Regarding the scenario where the number of views is larger than two, the fusion rule is extended as follows:

$$\begin{aligned} \boldsymbol{e} &= \boldsymbol{e}^1 \underline{\Diamond} \boldsymbol{e}^2 \underline{\Diamond} ... \boldsymbol{e}^V \\ &= c^1 * \boldsymbol{e}^1 + c^2 * \boldsymbol{e}^2 + ... + c^V * \boldsymbol{e}^V \end{aligned} \tag{21}$$

where $\sum_{v=1}^V c^v = 1$. Through this rule, if the evidence of all views is consistent, the fusion evidence will be enhanced. If there is a conflict in the views, the evidence will be reconciled by the weights.

During the opinion aggregation stage, priority distribution $\boldsymbol{c}$ may enhance the influence of low-reliability views on the acquisition of collective opinion. This is because in the process of obtaining $\boldsymbol{c}$, we perceived the importance of dynamic views based on the different emphasis features of specific instances, which is different from the relative inherent importance of multiple views. This can enable the collective to break the

absoluteness of the inherent status for the final opinion and focus on the viewpoints of a few views with lower inherent importance but more relevant to specific instances.

Our fusion rule also leads to the scaling of the integrated evidence at the same time, resulting in a decline in the reliability of the collective opinion. This is normal because the situation where the minority is noticed is often accompanied by the existence of conflicts, and the reliability of the view itself is low. Naturally, the reliability decreases after the fusion of multiple views. However, when all views support the basic facts, the model will pay more attention to the views with higher inherent importance and higher reliability, which also makes the reliability of the collective opinion relatively high. Unlike this, if only simple average pooling operations are used like that in the work [10], it will not particularly enhance the influence of specific instance views on the final decision. In the case of consistent view decisions, it may also magnify the influence of views with lower reliability, resulting in a significant reduction in the reliability of the final decision.

Time complexity analysis. To analyze the time complexity of this algorithm, we need to consider the computational cost of each part in the algorithm. Our model is a multi-view classification algorithm based on deep learning, and it consists of the following main parts:

Data preprocessing: It includes filling in missing values, normalizing continuous variables, and uniquely coding categorical variables. The time complexity of these operations is usually proportional to the size of the dataset, i.e., $O(n)$, where n is the number of samples in the dataset.

Model training: Model training is the most computationally expensive part of the algorithm. In each training epoch, the algorithm needs to perform forward propagation and backward propagation on the entire training set. The time complexity of forward propagation and backward propagation depends on the structure of the model and the size of the dataset. For this algorithm, the model consists of V EvidenceCollector modules (corresponding to V views), a global MLP module, and V Attention and DCollector modules. Assume that the total number of parameters for each view-related module (including its EvidenceCollector, Attention, and DCollector) is p_v. Also, assume that the number of parameters for the global MLP module is p_m. For a dataset containing n samples, the forward propagation complexity for all views is $O(n \cdot V \cdot p_v)$, and the complexity of the global module is $O(n \cdot p_m)$. Combining these, the total number of parameters in the model is $P = (V \cdot p_v + p_m)$, and the time complexity of a single forward pass is $O(nP)$. Due to the symmetry between gradient calculation and forward propagation, the time complexity of backpropagation is also $O(nP)$. The training process requires e epochs. Therefore, the total time complexity of the model training is $O(2nPe)$.

Model Evaluation: In the model evaluation stage, the algorithm needs to propagate the test set forward and calculate the performance indicators. The time complexity of forward propagation is proportional to the structure of the model and the size of the test set, that is, $O(n_{test}p)$, where n_{test} is the number of samples in the test set.

To sum up, the total time complexity of this algorithm is $O(n+2nPe+n_{test}p)$. In practical applications, in order to reduce the time complexity, methods such as using more efficient model structures, reducing the number of training epochs, and using faster hardware can be considered.

4 Experiment

We evaluate our model using a real-world dataset, taking the Multi-Disciplinary Medical Recommendation (MDMR) as an example. Then, we analyze the relationship between the view priority, reliability, dual-concepts, and prediction performance.

4.1 Experimental Setup

Dataset. BCDB [41] is a real-world multi-disciplinary consultation dataset, recording the physical indicators of patients with breast cancer and 9,015 instances of doctors' decision data. The doctors' decisions include six doctors' individual opinions and final recommendations for each patient. Notice that the individual decisions might be different. We choose different properties and doctors' individual opinions (the ground-truth categories) for a patient as different views to model how a doctor would diagnose the patient. Then, we consider the final collective recommendation as a result of multidisciplinary discussions among all doctors, i.e., the result of merging multiple views. Besides, we transform patient-specific information into more manageable numerical data.

Compared Methods. We compare our method with several previous methods on two tasks: One is predicting individual decisions, namely single-view-oriented learning, and the other is predicting collective decisions, namely multi-view-oriented learning.

For single-view-oriented learning methods, we choose: 1) **MLPsingle**: The traditional Multilayer Perceptron method is used for a single view, and information only flows in one direction in the network, which learns each view separately. 2) **TMC**: Trusted Multi-view Classification [9] is the pioneer uncertainty-aware method, which focuses more on the uncertainty estimation problem and produces a reliable classification result for individual views.

For multi-view-oriented learning methods, we choose: 1) **MLPshare**: Utilizes a Multilayer Perceptron method, all views are regarded as a complete network, and no longer any form of method is used. 2) **ECML**: Evidential Conflictive Multi-view Learning [10] is an uncertainty perception method based on the view of conflict between views, aiming to minimize the degree of conflict. 3) **ECMLboth**: Based on evidential conflictive multi-view learning, the view and the final view are used to supervise the learning of the model jointly, and the conflict between views is retained for uncertainty perception.

Evaluation Metrics. We select three commonly used metrics, i.e., **accuracy (A)**, **Recall (R)**, and **Mean Reciprocal Rank (MRR)**. According to real-world requirements, we use the top-1 indicator to calculate such metrics' values. As aforementioned, we have two prediction tasks, and thus we have two different ways to calculate the values. For single-view-oriented learning methods, the three metrics all indicate the average performance of predictions for each view, marked as **IA, IMRR, and IR**. For multi-view-oriented learning methods, the metrics are utilized to evaluate the performance of predicting collective decisions, marked as **CA**, **CMRR**, and **CR**.

Implement details. For the proposed TrustCL and TECL, we separate the dataset into the feature matrix $\boldsymbol{X}$ and the label matrix $\boldsymbol{Y}$, where $\boldsymbol{Y}$ contains the individual decisions of six views and the final collective decisions. They are all utilized as

supervision information. The Adam optimizer [42] is used for neural network training, with a learning rate of 0.003, and the L_2-norm regularization setting is $1e-5$. The data is divided into the training set (80%) and the test set (20%). During the training period, we can adjust the KL divergence regularization of λ by adjusting the epoch to achieve a better training effect, and set the annealing coefficient to 50. We run each method five times and report the average value and standard deviation in the table. This model is implemented in PyTorch and runs on an Intel(R) Core(TM) i7-9700 CPU with 40GB graphics memory.

Table 1 Comparison of predicted results about individual and collective opinions.

	Individual			Collective		
model	IA	IMRR	IR	CA	CMRR	CR
MLPsingle	70.6 ± 1.8	81.2 ± 1.1	71.0 ± 1.8	74.1 ± 1.9	77.6 ± 1.5	74.5 ± 1.7
MLPshare	0	0	0	72.9 ± 1.2	82.1 ± 0.9	73.4 ± 1.7
ECML	73.5 ± 0.9	81.7 ± 0.3	73.4 ± 0.6	77.7 ± 0.3	84.0 ± 0.2	77.3 ± 0.2
ECMLboth	71.8 ± 0.7	79.2 ± 0.2	71.5 ± 0.4	77.9 ± 0.3	84.5 ± 0.2	77.3 ± 0.3
TMC	70.6 ± 1.1	80.6 ± 1.0	70.8 ± 1.6	74.5 ± 0.2	82.9 ± 0.5	74.8 ± 0.7
Ours(TrustCL)	74.3 ± 0.7	82.5 ± 0.3	74.1 ± 0.5	**78.4±0.4**	85.1 ± 0.3	78.0 ± 0.3
Ours(TECL)	**74.5±0.5**	**83.0±0.6**	**74.2±0.3**	78.0 ± 0.9	**85.3±0.3**	**78.5±0.2**

Note: Bold value indicates the best performance and the underlined value indicates the second-best performance in each column. The colored area represents the performance of our model.

4.2 Experimental Results

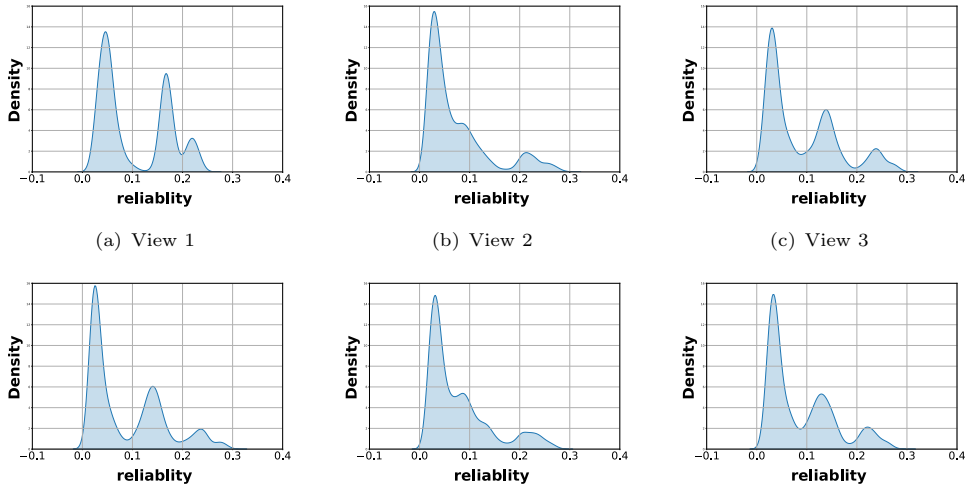
Performance Comparison. Table 1 compares the predicted results about individual and collective decisions, respectively. It can be found that MLP focuses entirely on the independent learning of single views, thus achieving relatively good performance on some metrics in the single-view dimension. TMC introduces a trust mechanism, which can take into account the influence of different perspectives in the single-view dimension and has better performance compared to other baselines. Meanwhile, ECML introduces the concept of minimizing conflicts, which is beneficial for alignment among multiple views and has better performance in obtaining collective opinions. On all accounts, we can see that our models perform best compared to other models. Furthermore, TECL performs slightly better than TrustCL across most metrics, especially at the individual level. This is because (1) TrustCL and TECL emphasize and capture the global knowledge and information generated. (2) they combine all views and learns collectively, which performs better than considering unique views (i.e., MLPsingle). And (3) compared to TrustCL, TECL strengthens the features that different individuals focus on at the feature concept level, leading to a better capture of individual decision-making preferences.

Ablation Experiment. To evaluate the effect of our knowledge aggregation strategy, we conduct three variations of TrustCL: 1) TrustCL-$\boldsymbol{\mu}$, which uses $\boldsymbol{\mu}$ instead of

Table 2 Ablation comparison on individual and collective opinion predictions.

	Individual			Collective		
model	IA	IMRR	IR	CA	CMRR	CR
TrustCL-$\boldsymbol{\mu}$	72.9 ± 1.0	81.6 ± 0.6	72.6 ± 0.9	77.9 ± 0.2	84.1 ± 0.3	77.3 ± 0.2
TrustCL-$\boldsymbol{\sigma}$	63.5 ± 2.2	73.1 ± 1.2	63.2 ± 2.1	77.3 ± 0.2	83.9 ± 0.4	77.0 ± 0.4
TrustCL-$\mathcal{N}$	71.9 ± 0.9	79.4 ± 0.3	71.6 ± 0.6	77.9 ± 0.3	84.3 ± 0.3	77.6 ± 0.3
TrustCL	74.3 ± 0.7	82.5 ± 0.3	74.1 ± 0.5	**78.4±0.4**	85.1 ± 0.3	78.0 ± 0.3
TECL	**74.5±0.5**	**83.0±0.6**	**74.2±0.3**	78.0 ± 0.9	**85.3±0.3**	**78.5±0.2**

Note: Bold value indicates the best performance and the underlined value indicates the second-best performance in each column.



(a) View 1 (b) View 2 (c) View 3

(d) View 4 (e) View 5 (f) Collective reliability

Fig. 4 Density of single-view and collective reliability.

$\boldsymbol{c}$ in Equation (11) as weights when merging views, focusing solely on inherent importance for the perception of reliability; 2) TrustCL-$\boldsymbol{\sigma}$, replaces $\boldsymbol{c}$ with $\boldsymbol{\sigma}$ as views' weights, emphasizing synergy between views; 3) TrustCL-$\mathcal{N}$, which only uses normal distribution to calculate view priority for TrustCL, ignoring the use of Dirichlet distribution.

The results of the ablation experiment on individual opinion predictions and collective opinion predictions are shown in Table 2, respectively. Among them, the model that does not consider the instance-specific importance performs the worst in each indicator, which can better prove the necessity of dynamically adjusting the importance of each view in response to the differences in the characteristics of different instances during the process of priority acquisition. We cannot merely consider the inherent importance resulting from the historical status of the view, but should take into account multiple dimensions, e.g., the characteristics of the instance, the features

that the view focuses on, and its historical status, to obtain the final collective opinion. Guided by dual-concepts, the model can also focus more effectively on its own "specialized" tasks. So, it can be seen that TrustCL and TECL perform best among these models due to their comprehensive approach.

Collective Importance Visualization. Figure 4(a)-4(e) and Figure 4(f) illustrate the density distribution of single-view reliability and the reliability distribution after fusion, respectively. We can observe that View 1, which has a higher title, exhibits a greater density of high reliability. In General, the fused reliability distribution should tend to align more closely with that of View 1. However, the reliability density after fusion is similar to that of View 3 and View 4. This indicates that the status of the fused view has changed, and the priority c enhances the importance of other views.

Moreover, we can see that the fused reliability is skewed towards the medium range, resulting in a relative decrease in both low and high reliability. This might be because the information conflict among reliable viewpoints reduces the overall reliability of the final decision. Among them, the addition of priority c, in the case of view conflicts, the evidence will also be reconciled by weights, resulting in the scaling of reliability. This is in line with our thinking in the discussion and analysis of section 3.

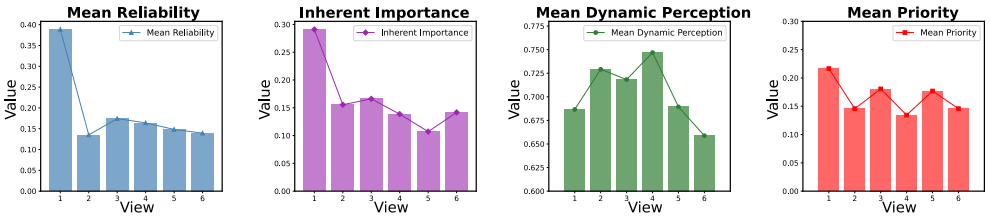
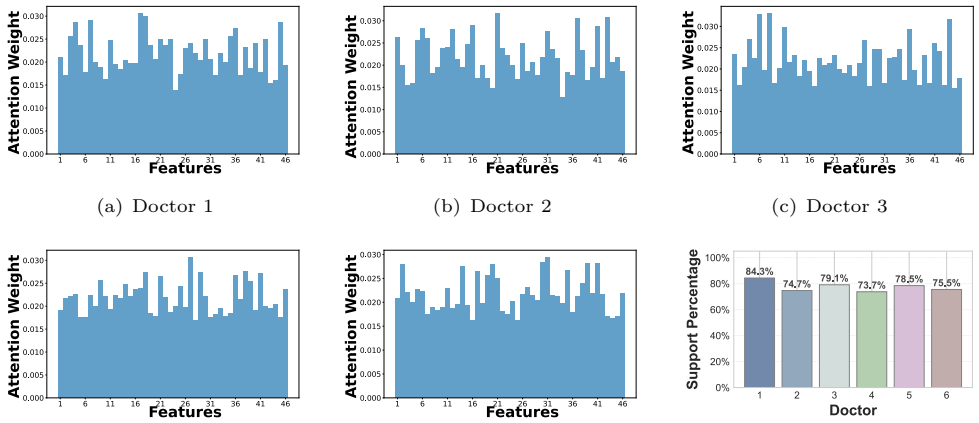


Fig. 5 Mean reliability, inherent importance, dynamic perception, and priority for all instances.

Figure 5 shows the whole results of the mean reliability, inherent importance, priority, and dynamic perception of each view. We can find that mean reliability and inherent importance have more similar trends, indicating that reliability effectively reflects the known importance of views. In contrast, the mean dynamic perception is quite different from the others, especially for view 1. Obviously, dynamic perception pays more attention to the features of the instance, resulting in a significant change in trend. This indicates that, under the combined effect of inherent importance and dynamic perception, the priority of different views will change dynamically as well.

MDMR Estimation. First, we obtain the overall attention and support attitude of different doctors (i.e., views) to different features. As shown in Figure 6(a) to Figure 6(e), we can find that the allocation of attention to features by different doctors is different, and the focus is different. This might be related to the different specialties of doctors. Differences in specialties may lead to conflicts among doctors in diagnostic decision-making for specific patients, as well as variations in doctors' support attitudes towards collective decision-making, as shown in Figure 6(f).

Therefore, we simulate multidisciplinary discussions, evaluating the effectiveness of TrustCL and TECL from three aspects: feature concepts, support attitudes, and



(a) Doctor 1 (b) Doctor 2 (c) Doctor 3

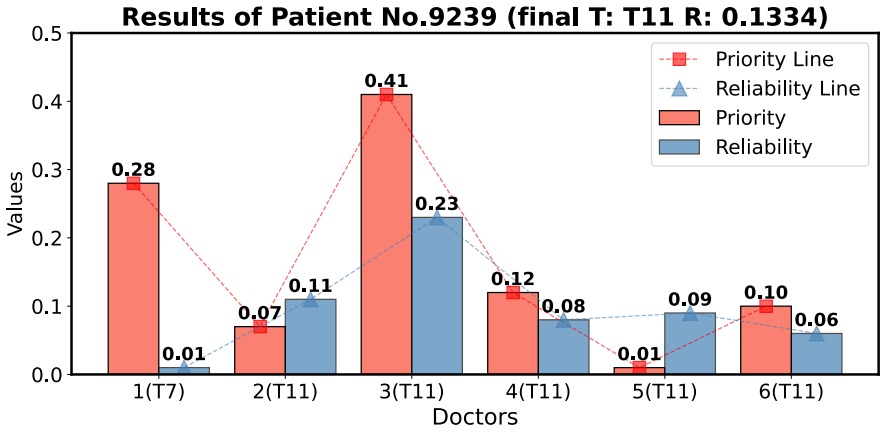
(d) Doctor 4 (e) Doctor 5 (f) Support attitude per doctor

Fig. 6 (a) to (e) are overall feature attention weights of different doctors. (f) is the overall support attitude of doctors for collective decisions.

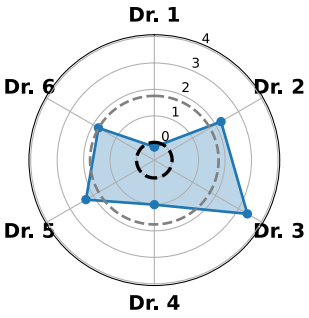
uncertainty, in order to assess their interpretability and reliability. The whole results of patient No.9239 are presented in Figure 7. From the results, Doctor 1 no longer holds a dominant position in terms of priority and reliability, and has an opposing attitude towards the collective decision; while Doctor 3 is more suitable for the current patient's treatment, achieving the highest priority and reliability, and showing strong support for the collective decision. This may be because Doctor 3 pays more attention to the specific characteristics of the patient's condition. As shown in Figure 7(c), we can see that he focuses more on the 9th attribute "rs_21_gene", which other doctors do not pay attention to. Meanwhile, the 0th attribute "is_ct" is given special attention by Doctor 1, while Doctor 3 doesn't focus on it much. Perhaps it may be less important to the patient's condition. There is a conceptual conflict between the two doctors, leading to the consequence that the collective reliability is midway between the lowest and the highest. Other doctors made the same decisions as Doctor 3, showing weak support for the collective decision, and their focus on attributes overlapped with Doctor 3's, rather than focusing on the 0th attribute.

In contrast, we select another patient, No.5827, whose doctors' decisions and support attitudes are consistent, resulting in a correspondingly high reliability of the collective decision. The whole results of patient N0.9239 are presented in Figure 8. It can be seen that the attributes that doctors focus on are concentrated in a few specific areas. Doctors 2 and 5 have higher priority, indicating that they are better suited to the current patient's treatment, surpassing the influence of their inherent status. Both also strongly support the collective decision. While other doctors hold a weaker supportive attitude, especially Doctor 1.

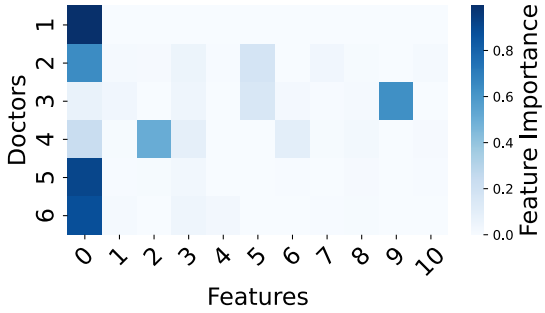
In summary, we can ensure that: 1) Our models fully respect the distinctions of individuals and notice the voice of the minority, regardless of titles. Moreover, it effectively evaluates the reliability of individual opinions. 2) Our TECL can effectively capture the attributes that different perspectives focus on regarding a specific sample



(a) Decisions and their reliability



(b) Doctor support scores



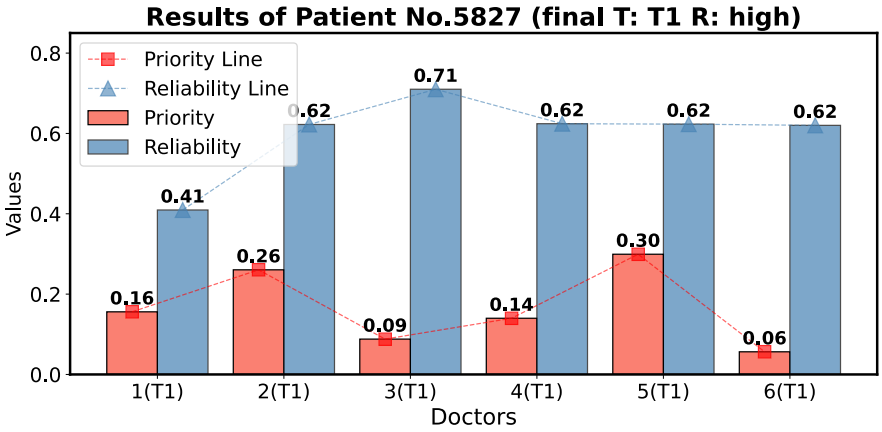
(c) Feature concepts heatmap

Fig. 7 Whole results of patient No.9239. In 7(b), the black dashed circle represents the zero mark, and the gray dashed circle represents the average support level. Below the zero mark indicates opposition, between the zero mark and the average support line indicates weak support, and above the average support level indicates strong support.

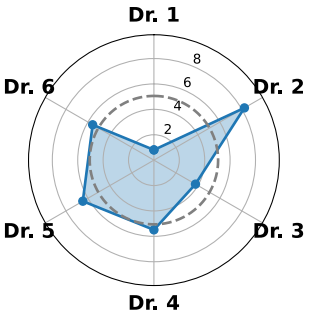
and their attitudes towards collective opinions, providing explanations for individual opinions and the conflicts that arise between individuals. 3) Our models emphasize individual expression in collective opinions. The conflict of views results in a less reliable final result. TrustCL and TECL outperform others in experiments and combine view-specific evidence with priority during the fusion process, so the final collective decisions made are feasible.

5 Conclusion

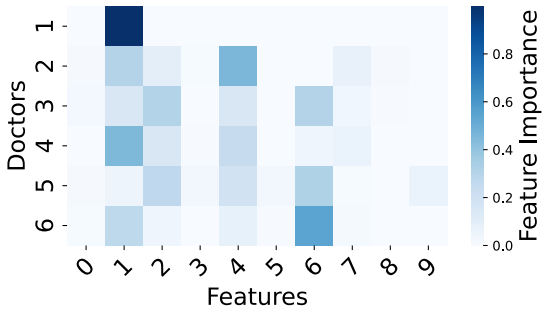
In this paper, we propose the Trusted and Explainable Collective Learning (TECL) framework to address opinion conflicts in multi-view decision-making. Unlike traditional conflict-reduction methods, TECL actively models these conflicts to highlight



(a) Decisions and their reliability



(b) Doctor support scores



(c) Feature concepts heatmap

Fig. 8 Whole results of patient No.5827. In 8(b), there is no black dashed circle because all the doctors hold a supportive attitude.

inherent view distinctions and evaluate the reliability of both individual and collective decisions. The framework employs a dual-concept learning mechanism at both the feature and decision levels, which significantly enhances model interpretability while capturing nuanced cross-view interactions. Furthermore, by learning view-specific evidence and incorporating a reliability-oriented learning phase to determine the priority of each view, it synthesizes conflicting opinions to generate a collective decision along with its confidence level. Extensive experiments on a real-world breast cancer multidisciplinary consultation dataset demonstrate that TECL consistently outperforms state-of-the-art baselines, validating its robustness in complex, conflict-heavy scenarios. Future work will explore the integration of Large Language Models (LLMs) to further advance decision intelligence in large-scale semantic contexts.

Acknowledgment

This work was supported by the Seed Program of Shanghai Municipal Health Commission for Medical Innovation and Transformation (2025ZZ1009), the National Natural Science Foundation of China (62202282, 82573726), the Intelligent Healthcare Program of Shanghai Municipal Health Commission (2025ZHYL003), and the Interdisciplinary Program of Shanghai Jiao Tong University (YG2024QNB05).

References

[1] Lu, M., Xu, Y., Chu, W., Gu, J., Gao, M.: Contrastive multi-view clustering based on multi-head attention mechanisms and three-way decision. Knowledge and Information Systems **67**(7), 6115–6138 (2025) https://doi.org/10.1007/s10115-025-02403-8

[2] Bachman, P., Hjelm, R.D., Buchwalter, W.: Learning representations by maximizing mutual information across views. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. NIPS'19. Curran Associates Inc., Red Hook, NY, USA (2019)

[3] Tian, Y., Krishnan, D., Isola, P.: Contrastive multiview coding. In: Proceedings of the Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI. ECCV'20, pp. 776–794 (2020)

[4] Sun, S., Dong, W., Liu, Q.: Multi-view representation learning with deep gaussian processes. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(12), 4453–4468 (2021)

[5] Zhang, C., Cui, Y., Han, Z., Zhou, J.T., Fu, H., Hu, Q.: Deep partial multi-view learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(5), 2402–2415 (2022) https://doi.org/10.1109/TPAMI.2020.3037734

[6] Noori, A., Balafar, M.A., Bouyer, A., Salmani, K.: A novel contrastive multi-view framework for heterogeneous graph embedding. Knowledge and Information Systems **67**(5), 4373–4399 (2025) https://doi.org/10.1007/s10115-025-02346-0

[7] Zhu, N., Huang, J., Cao, J., Hu, L., Yuan, Z., Gao, H.: R2v-mif: Rule-to-vector contrastive learning and multi-channel information fusion for therapy recommendation. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. IJCAI'24, pp. 2634–2641 (2024)

[8] Yang, Z., Chen, J., Miao, Z., Li, W., Zhu, X., Zhang, L.: Deepinteraction: 3d object detection via modality interaction. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. NIPS '22 (2022)

[9] Han, Z., Zhang, C., Fu, H., Zhou, J.T.: Trusted multi-view classification. In: Proceedings of the International Conference on Learning Representations. ICLR'21 (2021)

[10] Xu, C., Si, J., Guan, Z., Zhao, W., Wu, Y., Gao, X.: Reliable conflictive multi-view learning. Proceedings of the AAAI Conference on Artificial Intelligence **38**(14), 16129–16137 (2024)

[11] Wu, B., Hong, S., Lin, R.: Explainable prediction for business process activity with transformer neural networks. Knowledge and Information Systems **67**(10), 9221–9252 (2025) https://doi.org/10.1007/s10115-025-02501-7

[12] Yang, J., Fu, H., Xue, F., Li, M., Wu, Y., Yu, Z., Luo, H., Gong, J., Niu, X., Zhang, W.: Multiview representation learning for identification of novel cancer genes and their causative biological mechanisms. Briefings in Bioinformatics **25** (2024) https://doi.org/10.1093/bib/bbae418

[13] Koh, P.W., Nguyen, T., Tang, Y.S., Mussmann, S., Pierson, E., Kim, B., Liang, P.: Concept Bottleneck Models (2020). https://arxiv.org/abs/2007.04612

[14] Gao, Z., Liu, Y., Wu, F., Shi, N., Shi, Y., Zhuang, X.: A reliable and interpretable framework of multi-view learning for liver fibrosis staging. In: Medical Image Computing and Computer Assisted Intervention. MICCAI'23 (2023)

[15] YU, Z., DONG, Z., YU, C., YANG, K., FAN, Z., CHEN, C.L.P.: A review on multi-view learning. Frontiers of Computer Science **19**(7), 197334 (2025)

[16] Wen, Y., Wang, S., Liao, Q., Liang, W., Liang, K., Wan, X., Liu, X.: Unpaired multi-view graph clustering with cross-view structure matching. IEEE Transactions on Neural Networks and Learning Systems, 1–15 (2023)

[17] Zhang, T., Liu, X., Gong, L., Wang, S., Niu, X., Shen, L.: Late fusion multiple kernel clustering with local kernel alignment maximization. IEEE Transactions on Multimedia (2021)

[18] Huang, Z., Hu, P., Zhou, J.T., Lv, J., Peng, X.: Partially view-aligned clustering. In: Advances in Neural Information Processing Systems. NIPS'20, vol. 33, pp. 2892–2902 (2020)

[19] Yang, M., Li, Y., Huang, Z., Liu, Z., Hu, P., Peng, X.: Partially view-aligned representation learning with noise-robust contrastive loss. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). CVPR'21, pp. 1134–1143 (2021)

[20] Wang, X., Wang, Y., Wang, Y., Huang, A., Liu, J.: Trusted semi-supervised multi-view classification with contrastive learning. IEEE Transactions on Multimedia **26**, 8268–8278 (2024)

[21] Xu, C., Zhang, Y., Guan, Z., Zhao, W.: Trusted multi-view learning with label noise. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24 (2024)

[22] Zhang, X., Chen, M., Mu, J., Zong, L.: Adaptive view-aligned and feature augmentation network for partially view-aligned clustering. In: Proceedings of the Advances in Knowledge Discovery and Data Mining. KDD'23, pp. 223–235 (2023)

[23] Han, Z., Zhang, C., Fu, H., Zhou, J.T.: Trusted multi-view classification with dynamic evidential fusion. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(2), 2551–2566 (2023)

[24] Jung, M.C., Zhao, H., Dipnall, J., Gabbe, B., Du, L.: Uncertainty estimation for multi-view data: The power of seeing the whole picture. In: Advances in Neural Information Processing Systems. NIPS '22, pp. 6517–6530 (2022)

[25] Jung, M.C., Zhao, H., Dipnall, J., Du, L.: Beyond unimodal: generalising neural processes for multimodal uncertainty estimation. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. NIPS '23 (2023)

[26] Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., Zhong, C.: Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges (2021)

[27] Wu, M., Hughes, M.C., Parbhoo, S., Zazzi, M., Roth, V., Doshi-Velez, F.: Beyond sparsity: tree regularization of deep models for interpretability. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. AAAI'18 (2018)

[28] Li, J., Wang, C., Chen, J., Zhang, H., Dai, Y., Wang, L., Wang, L., Nandi, A.K.: Explainable cnn with fuzzy tree regularization for respiratory sound analysis. IEEE Transactions on Fuzzy Systems **30**(6), 1516–1528 (2022) https://doi.org/10.1109/TFUZZ.2022.3144448

[29] Alemi, A.A., Fischer, I., Dillon, J.V., Murphy, K.: Deep Variational Information Bottleneck (2019). https://arxiv.org/abs/1612.00410

[30] Losch, M., Fritz, M., Schiele, B.: Semantic bottlenecks: Quantifying and improving inspectability of deep representations **129**(11), 3136–3153 (2021) https://doi.org/10.1007/s11263-021-01498-0

[31] Kazhdan, D., Dimanov, B., Jamnik, M., Liò, P., Weller, A.: Now You See Me (CME): Concept-based Model Extraction (2020)

[32] Kim, E., Jung, D., Park, S., Kim, S., Yoon, S.: Probabilistic concept bottleneck models. In: Proceedings of the 40th International Conference on Machine Learning. ICML'23. JMLR.org, Honolulu, Hawaii, USA (2023)

[33] Sheth, I., Kahou, S.E.: Auxiliary losses for learning generalizable concept-based models. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. NIPS '23. Curran Associates Inc., Red Hook, NY, USA (2023)

[34] Yuksekgonul, M., Wang, M., Zou, J.: Post-hoc concept bottleneck models. In: The Eleventh International Conference on Learning Representations. ICLR '23 (2023)

[35] Kazhdan, D., Dimanov, B., Andrés-Terré, H., Jamnik, M., Lio', P., Weller, A.: Is disentanglement all you need? comparing concept-based & disentanglement approaches. ArXiv **abs/2104.06917** (2021)

[36] Garavan, T.N., Carbery, R.: Collective Learning, pp. 646–649. Springer, New York, NY (2012)

[37] Nigam, K., Ghani, R.: Analyzing the effectiveness and applicability of co-training. In: Proceedings of the Ninth International Conference on Information and Knowledge Management. CIKM '00, pp. 86–93 (2000)

[38] Chen, M., Weinberger, K.Q., Blitzer, J.C.: Co-training for domain adaptation. In: Proceedings of the 24th International Conference on Neural Information Processing Systems. NIPS'11, pp. 2456–2464. Curran Associates Inc., Granada, Spain (2011)

[39] Hang, M., Neville, J., Ribeiro, B.: A collective learning framework to boost gnn expressiveness for node classification. In: Proceedings of the 38th International Conference on Machine Learning. PMLR'21 (2021)

[40] Sensoy, M., Kaplan, L., Kandemir, M.: Evidential deep learning to quantify classification uncertainty. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. NIPS'18, pp. 3183–3193 (2018)

[41] Zhu, N., Cao, J., Shen, K., Chen, X., Zhu, S.: A decision support system with intelligent recommendation for multi-disciplinary medical treatment. ACM Trans. Multimedia Comput. Commun. Appl. **16**(1s) (2020)

[42] Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: Proceedings of the International Conference on Learning Representations. ICLR '14 (2014)