Appendix

A DEFINITION OF DIRICHLET DISTRIBUTION

Definition 2 (Dirichlet Distribution). *The Dirichlet distribution for categorical distributions $\boldsymbol{p_i}$ is built as:*

$$Dir\left(\boldsymbol{p_i} \mid \boldsymbol{e}_i + 1\right) = \begin{cases} \frac{1}{B(\boldsymbol{e}_i+1)} \prod_{k=1}^{K} p_{ik}^{e_{ik}} & \text{for } \boldsymbol{p_i} \in S_K, \\ 0 & \text{otherwise,} \end{cases}$$

where $B\left(\boldsymbol{e_i}+1\right)$ denotes the $K-$dimensional multinomial beta function, $\boldsymbol{e}_i^v + 1$ represents the Dirichlet parameters with a uniform uncertainty prior on each class, and S_K is the $K-$dimensional unit simplex. Once the parameters of distribution are calculated, the expected probability $\boldsymbol{p_i}$ is the mean of the $Dir\left(\boldsymbol{p_i} \mid \boldsymbol{e}_i + 1\right)$.

B ALGORITHM FOR CONFORMALIZED MULTI-VIEW DEEP CLASSIFICATION

In this section, we will give an overall algorithm framework of Conformalized Multi-view Deep Classification to show the optimization process of our method in details.

Algorithm 1: Algorithm for Conformalized Multi-view Deep Classification (CMDC)

/***Training***/
Input: $\mathcal{X}_{train} = \left\{\left\{x_i^v\right\}_{v=1}^{V}, y_i\right\}_{i=1}^{N}$, $\mathcal{X}_{cal} = \left\{\left\{x_i^v\right\}_{v=1}^{V}, y_i\right\}_{i=1}^{M}$.
Initialize: Initialize the parameters of the neural network.
while *not converged* **do**
  for $i = 1 : N$ **do**
    for $v = 1 : V$ **do**
      $\boldsymbol{e}_i^v \leftarrow$ non-negative ENN output;
      $O_i^v \leftarrow$ using the Eq.(1);
    end
    Obtain $O_i = \left\{\boldsymbol{b}_i = \{b_{ik}\}_{k=1}^{K+1}\right\}$ with Eq.(3);
    Obtain $\boldsymbol{e_i}$, u_i with Eq.(1);
  end
  Obtain $\hat{p}$ and $\hat{q}$ using the Algorithm 2;
  Obtain the overall loss using Eq.(7);
  Update the networks with gradient descent using Eq.(7);
end
Output: parameters of neural networks.
/***Test***/
Input: $\mathcal{X}_{cal} = \left\{\left\{x_i^v\right\}_{v=1}^{V}, y_i\right\}_{i=1}^{M}$, $\left\{\left\{x_{test}^v\right\}_{v=1}^{V}, y_{test}\right\}$.
Initialize: parameterized by the training network.
Obtain $\hat{p}, \hat{q}$ using the Algorithm 2;
Obtain $C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)$ using the Algorithm 3;
Output: the prediction sets $C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)$.

C PROOFS OF THE PROPOSITIONS

In this part, we will give the detailed proofs of the following propositions shown in Section 4. **We have noticed a few spelling errors in the original paper regarding Proposition 2, especially the Eq.(15). In the subsequent proofs, we rectify these errors for accuracy.**

Algorithm 2: Algorithm for Calculation of $\hat{p}$ and $\hat{q}$

Input: $\mathcal{X}_{cal} = \left\{\left\{x_i^v\right\}_{v=1}^{V}, y_i\right\}_{i=1}^{M}$.
$E_p = \{\}$, $E_q = \{\}$
for $i = 1 : M$ **do**
  for $v = 1 : V$ **do**
    $\boldsymbol{e}_i^v \leftarrow$ output from parameterized model;
    $O_i^v \leftarrow$ using the Eq.(1);
  end
  Obtain $O_i = \left\{\boldsymbol{b}_i = \{b_{ik}\}_{k=1}^{K+1}\right\}$ with Eq.(3);
  $E_\theta\left(\left\{x_i^v\right\}_{v=1}^{V}, y_i\right) \leftarrow$ using the Eq.(4);
  $E_\theta\left(\left\{x_i^v\right\}_{v=1}^{V}, K+1\right) \leftarrow$ using the Eq.(4);
  $E_p \leftarrow E_p \cup E_\theta\left(\left\{x_i^v\right\}_{v=1}^{V}, y_i\right)$;
  $E_q \leftarrow E_q \cup E_\theta\left(\left\{x_i^v\right\}_{v=1}^{V}, K+1\right)$;
end
$E_p \leftarrow$ Sort E_p from the smallest to the largest.
$E_q \leftarrow$ Sort E_q from the smallest to the largest.
$\hat{p} \leftarrow Quantile\left(E_p; \left\lceil \frac{(1-\tau)(1+M)}{M} \right\rceil\right)$;
$\hat{q} \leftarrow Quantile\left(E_q; \left\lceil \frac{(1-\tau)(1+M)}{M} \right\rceil\right)$;
Output: $\hat{p}$ and $\hat{q}$.

Algorithm 3: Algorithm for Construction of Prediction Set $C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)$

Input: $\left\{\left\{x_{test}^v\right\}_{v=1}^{V}, y_{test}\right\}$, $\hat{p}$ and $\hat{q}$.
$C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right) = \{\}$;
for $k = 1 : K+1$ **do**
  $E_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}, k\right) \leftarrow$ using the Eq.(4);
  if $k \neq K+1$ **then**
    $C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right) \cup k \leftarrow \left\{E_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}, k\right) \leqslant \hat{p}\right\}$;
  end
  else
    $C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right) \cup K+1 \leftarrow \left\{E_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}, k\right) > \hat{q}\right\}$;
  end
end
Output: $C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)$.

Proposition (Classification Performance Improvement). *Let $\widehat{R}\left(h\left(\left\{x_i^v\right\}_{v=1}^{\cdot}; \theta\right)\right) = \frac{1}{N}\sum_{i=1}^{N} \ell_{acc}\left(\theta\right)_i$ be the empirical risk for the data with any subset of views. Under the regularization of confidence* *calibration, the inequality*

$$\widehat{R}\left(h\left(\left\{x_i^v\right\}_{v=1}^{V};\theta\right)\right) \leq \widehat{R}\left(h\left(\left\{x_i^v\right\}_{v=1}^{V'};\theta\right)\right) \tag{19}$$

is satisfied with $V' \in \mathbb{N}^* \leq V$.

Proof: *For convenience, we use* θ *and* θ' *to replace* $h\left(\left\{x_i^v\right\}_{v=1}^{V};\theta\right)$ *and* $h\left(\left\{x_i^v\right\}_{v=1}^{V'};\theta\right)$. *Let* $\boldsymbol{\alpha} = \mathbf{e} + 1 = [\alpha_{i1}, \ldots, \alpha_{iK}]$, $\boldsymbol{\alpha}' = \mathbf{e}' + 1 = [\alpha'_{i1}, \ldots, \alpha'_{iK}]$, *we have:*

$$\widehat{R}(\theta) - \widehat{R}(\theta') \Leftrightarrow \ell_{acc}(\theta)_i - \ell_{acc}(\theta')_i$$

$$\underset{\{y_{ik}=1\}}{\overset{Eq.(8)}{\Leftrightarrow}} log(S_i) - log(\alpha_{ik}) - log(S'_i) + log\left(\alpha'_{ik}\right) \Leftrightarrow \log\left(\frac{S_i \cdot \alpha'_{ik}}{\alpha_{ik} \cdot S'_i}\right).$$

Under the regularization of confidence calibration, which shrinks the evidence of incorrect classes to zero, we have $S = S' + e''_{ik}$, $\alpha_{ik} = \alpha'_{ik} + e''_{ik}$ *using the fusion rule defined in Definition 1, where* e''_{ik} *is the joint evidence for the correct label* k *integrated from the* V'^{th} *view to the* V^{th} *view. Then we have:*

$$\log\left(\frac{S_i \cdot \alpha'_{ik}}{\alpha_{ik} \cdot S'_i}\right) \leqslant 0 \Leftrightarrow \frac{S_i \cdot \alpha'_{ik}}{\alpha_{ik} \cdot S'_i} \leqslant 1$$

$$\Leftrightarrow \frac{\left(S'_i + e''_{ik}\right) \cdot \alpha'_{ik}}{\left(\alpha'_{ik} + e''_{ik}\right) \cdot S'_i} \leqslant 1 \Leftrightarrow \frac{\alpha'_{ik} + \frac{e''_{ik} \cdot \alpha'_{ik}}{S'_i}}{\alpha'_{ik} + e''_{ik}} \leqslant 1.$$

The equality holds since $\frac{\alpha'_{ik}}{S'_i} \leqslant 1$. □

Proposition (Conformal Risk Control Guarantee). *Given any test data point* $\left\{x_{test}^v\right\}_{v=1}^{V}$, *suppose the Conformal Risk of our model* $h\left(\left\{x_{test}^v\right\}_{v=1}^{V};\theta\right)$ *is defined as* $CR(\theta)$, *we have:*

$$CR(\theta) = CR_{mis}(\theta) + CR_{outlier}(\theta) \leq 2\tau. \tag{20}$$

where $CR_{mis}(\theta)$ *indicates the risk of miscoverage that true label is not contained in the prediction set, and* $CR_{outlier}(\theta)$ *is the risk of points are wrongly identified as outliers, defined as:*

$$CR_{mis}(\theta) = \mathbb{E}\left(\mathbb{P}\left(\left\{y_{test}^{true} \in \{1, \ldots, K\}\right\} \notin C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)\right)\right). \tag{21}$$

$$CR_{outlier}(\theta) = \mathbb{E}\left(\mathbb{P}\left(\{\hat{y}_{test} = \{K+1\}\} \in C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)\right)\right). \tag{22}$$

Notably, the above guarantee also ensures that:

$$\mathbb{P}\left(y_{test}^{true} \in C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)\right) \geq 1 - \tau. \tag{23}$$

Proof: *Here, we first consider the case of* $k \in \{1, \ldots, K\}$. *Suppose we have calibration dataset* $\mathcal{X}_{cal} = \left\{\left\{X_i^v\right\}_{v=1}^{V}, y_i\right\}_{i=1}^{M}$ *and any test sample* $\left\{\left\{x_{test}^v\right\}_{v=1}^{V}, y_{test}\right\}$. *Then the threshold* $\hat{p}$ *can be defined as:*

$$\hat{p} = \inf\left\{p : \frac{\left|\left\{i : E_\theta\left(\left\{x_i^v\right\}_{v=1}^{V}, y_i\right) \leq p\right\}\right|}{M} \geq \frac{\lceil (M+1)(1-\tau) \rceil}{M}\right\} \tag{24}$$

We have the resulting prediction sets:

$$C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right) = \left\{k : E_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}, k\right) \leq \hat{p}\right\}, \tag{25}$$

where $k \in \{1, \ldots, K\}$.

Here we assume the conformal scores calculated on $\mathcal{X}_{cal}$ *for* $i = 1, \ldots, M$ *are sorted so that*

$$E_\theta^{cal}\left(\left\{x_1^v\right\}_{v=1}^{V}, y_1\right) < \ldots < E_\theta^{cal}\left(\left\{x_M^v\right\}_{v=1}^{V}, y_M\right). \tag{26}$$

Let $E_1^{cal} < \ldots < E_M^{cal}$ *to replace for convenience, we have that* $\hat{p} = E^{cal}_{\lceil (M+1)(1-\tau) \rceil}$ *when* $\tau \geq \frac{1}{M+1}$ *and* $\hat{p} = \infty$ *otherwise, which indicates the* $\left\lceil \frac{(1-\tau)(1+M)}{M} \right\rceil$*-quantile of the ordered* $\left\{E_\theta^{cal}\left(\left\{x_M^v\right\}_{v=1}^{V}, y_M\right)\right\}_{i=1}^{M}$ *with true label* $y_i \in [1, \ldots, K]$.

Note that in the case $\hat{p} = \infty$, $C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right) = \mathcal{Y}$, *so the coverage property is trivially satisfied; thus we only have to handle the case when* $\tau \geq \frac{1}{M+1}$. *We proceed by noticing the equality of the two events:*

$$\left\{k \in C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)\right\} = \left\{E_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}, k\right) \leq \hat{p}\right\}. \tag{27}$$

Combing this with the definition of $\hat{p}$ *yields*

$$\left\{k \in C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)\right\} = \left\{E_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}, k\right) \leq E^{cal}_{\lceil (M+1)(1-\tau) \rceil}\right\}. \tag{28}$$

By exchangeability of the variables $\mathcal{X}_{cal} = \left\{\left\{X_i^v\right\}_{v=1}^{V}, y_i\right\}_{i=1}^{M}$ *and* $\left\{x_{test}^v\right\}_{v=1}^{V}$, *we have*

$$\mathbb{P}\left(E_{test} \leq E_z^{cal}\right) = \frac{z}{M+1} \tag{29}$$

for any integer z. *In other words,* E_{test} *is equally likely to fall in anywhere between the calibration points* $E_1^{cal}, \ldots, E_M^{cal}$. *Note that above, the randomness is over all variables* $E_1, \ldots, E_M, E_{test}$. *From here, we conclude*

$$\mathbb{P}\left(E_{test} \leq E^{cal}_{\lceil (M+1)(1-\tau) \rceil}\right) = \left\lceil \frac{(1-\tau)(1+M)}{M+1} \right\rceil \geq 1 - \tau, \tag{30}$$

which means, for any test samples:

$$\mathbb{P}\left(y_{test}^{true} \in C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)\right) \geq 1 - \tau. \tag{31}$$

In contrast,

$$\mathbb{P}\left(y_{test}^{true} \notin C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)\right) < \tau. \tag{32}$$

Similarly, the predicted $\hat{y}_{test}$ *can derive the following property according the above proof:*

$$\mathbb{P}\left(\{\hat{y}_{test} = \{K+1\}\} \notin C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)\right) \geq 1 - \tau, \tag{33}$$

which means, for any test samples

$$\mathbb{P}\left(\{\hat{y}_{test} = \{K+1\}\} \in C_\theta\left(\left\{x_{test}^v\right\}_{v=1}^{V}\right)\right) < \tau. \tag{34}$$

Then we have:

$$CR(\theta) = CR_{mis}(\theta) + CR_{outlier}(\theta) \leq 2\tau. \tag{35}$$

□

Table 4: Comparison with popular multi-view learning methods (trained on 70% of data) based on Accuracy (ACC, %).

Method	HAND	SCENE	ANIMAL	CAL	CUB	HMDB	MRNet
DCCA [2]	94.01±1.20	51.21±1.67	81.10±1.22	80.00±0.77	78.12±1.00	44.90±1.37	87.21±2.79
DCCAE [48]	97.01±0.23	52.12±0.31	84.97±0.21	88.11±0.79	80.42±0.88	48.00±0.80	88.23±0.11
DTCCA [50]	96.01±0.10	58.32±0.20	82.77±0.10	90.01±0.26	80.31±0.22	55.01±0.87	84.01±0.32
CPM-Nets [53]	94.00±1.10	65.23±0.02	85.21±0.20	89.11±1.88	85.67±0.02	66.56±1.10	88.03±0.00
DUA-Nets [15]	98.00±0.21	65.23±0.11	87.81±1.44	93.01±0.22	79.83±1.50	62.12±0.87	89.61±1.00
MVTCAE [19]	97.00±0.23	64.22±0.01	86.21±0.12	91.01±0.44	90.00±0.12	72.89±2.01	93.01±1.22
TMC [17]	98.23±0.09	66.57±0.03	88.21±0.41	93.01±0.10	90.04±1.01	74.10±1.22	91.01±0.80
TMDOA [29]	98.30±0.21	70.10±0.12	90.00±0.01	93.38±0.04	91.78±1.22	85.49±0.57	94.01±1.22
ETMC [18]	98.20±0.23	66.01±0.07	88.90±0.25	92.77±0.37	90.04±1.21	74.22±0.88	92.02±1.21
SMDC [28]	98.89±0.10	72.06±0.22	93.46±0.00	96.40±0.10	92.06±0.16	90.00±0.20	92.08±1.00
CMDC (Ours)	**99.17±0.52**	**75.18±1.11**	**94.57±0.55**	**97.72±0.97**	**93.05±0.44**	**91.56±0.40**	**94.29±0.79**

Proposition (Generalization). *Let X be a set of N samples with label Y, $\theta \in \mathbb{B}^d$ be the parameter of loss function in a finite d-dimensional unit ball. Define generalization risk as:*

$$R(\theta) = \mathbb{E}_{(X,Y)}\left[\ell_{acc}(\theta)_i\right]. \tag{36}$$

Let $\theta^* = \operatorname{argmax}_{\theta \in \mathbb{B}^d} R(\theta)$ be the optimal parameter in the unit ball, $\widehat{\theta} = \operatorname{argmax}_{\theta \in A} \widehat{R}(\theta)$ be the optimal parameter of empirical risk among a candidate set A. With probability at least $1-\delta$ we have,

$$R(\theta^*) \le R\left(\widehat{\theta}\right) + \frac{3 + \sqrt{2\varepsilon^2 d \ln(N) + 4\varepsilon^2 \ln(2/\delta)}}{\sqrt{N}}. \tag{37}$$

Proof: *We assume the $\ell_{acc}(\theta)_i$ is strictly bounded by intervals $[0, \varepsilon]$. Let $\epsilon = \frac{3}{\sqrt{N}}$ and $\Delta = \frac{\sqrt{\varepsilon^2 d \ln(3/\epsilon) + \varepsilon^2 \ln(2/\delta)}}{\sqrt{2N}}$. For any fixed θ, according to Hoeffding's inequality, we have,*

$$\Pr\left\{\left|\widehat{R}(\theta) - R(\theta)\right| > \Delta\right\} \leqslant 2\exp\left(-\frac{2N^2\Delta^2}{\sum_{i=1}^{N}(\varepsilon-0)^2}\right) = \frac{\delta}{(3/\epsilon)^d}.$$

With probability no less than $1-\delta$ we have

$$\forall\theta : \left|\widehat{R}(\theta) - R(\theta)\right| \leqslant \frac{\sqrt{\varepsilon^2 d \ln(3/\epsilon) + \varepsilon^2 \ln(2/\delta)}}{\sqrt{2N}}. \tag{38}$$

Let A be an ϵ-cover of $\mathbb{B}^d$ (i.e. $\forall\theta \in \mathbb{B}^d \exists \theta'' \in A : \|\theta - \theta''\| \le \epsilon$), $\widehat{\theta} = \operatorname{argmax}_{\theta \in A} \widehat{R}(\theta)$ be the optimal parameter of empirical risk, we can obtain

$$\begin{aligned}
\forall\theta'' \in A : R\left(\widehat{\theta}\right) &\ge \widehat{R}\left(\widehat{\theta}\right) - \frac{\sqrt{\varepsilon^2 d \ln(3/\epsilon) + \varepsilon^2 \ln(2/\delta)}}{\sqrt{2N}} \\
&\ge \widehat{R}(\theta'') - \frac{\sqrt{\varepsilon^2 d \ln(3/\epsilon) + \varepsilon^2 \ln(2/\delta)}}{\sqrt{2N}} \\
&\ge R(\theta'') - 2\frac{\sqrt{\varepsilon^2 d \ln(3/\epsilon) + \varepsilon^2 \ln(2/\delta)}}{\sqrt{N}}.
\end{aligned}$$

The first and third inequality hold since Eq. (38), the second inequality holds since $\widehat{\theta} = \operatorname{argmax}_{\theta \in A} \widehat{R}(\theta)$.

According to the Lipschitz-continuity of $\ell_{acc}(\theta)_i$ w.r.t. to θ, $\forall\theta \in \mathbb{B}^d$ for classification, we have

$$\begin{aligned}
R(\theta) &\le R\left(\widehat{\theta}\right) + \epsilon + 2\frac{\sqrt{\varepsilon^2 d \ln(3/\epsilon) + \varepsilon^2 \ln(2/\delta)}}{\sqrt{N}} && (39) \\
&\le R\left(\widehat{\theta}\right) + \frac{3 + \sqrt{2\varepsilon^2 d \ln(N) + 4\varepsilon^2 \ln(2/\delta)}}{\sqrt{N}}. && (40)
\end{aligned}$$

□

D ADDITIONAL EXPERIMENTS

Here, we extensively evaluate the proposed method on real-world multi-view datasets, considering metrics such as Accuracy (ACC), Expected Calibration Error (ECE) [16], Inefficiency (Ineff), and Coverage (Cov), where Ineff measures the size of prediction sets and Cov measures the coverage of true label contained in the prediction sets. All experimental settings are consistent with those described in the main body.

D.1 Supplementary Comparison with Popular Methods

To thoroughly investigate the effectiveness of our model, we conducted detailed comparisons with several popular models for multi-view deep classification trained on 70% of the data. The comprehensive experimental results are presented in Table 4. It is evident that our approach demonstrates state-of-the-art performance across all multi-view datasets with the same training data, as indicated by the accuracy (ACC) metric. For instance, considering the CUB dataset, our model achieves an approximate 1% improvement in ACC. This result suggests that our proposed CMDC model achieves the best classification performance compared to other methods trained on the same dataset.

D.2 Analysis of Parameter τ

In this subsection, we delve into the effect of τ, a parameter crucial to our method. Here, we set $\tau = 0.05$, ensuring that the true label is included in the prediction set with a probability of 95%. Table 5 showcases the performance of CMDC with $1-\tau = 95\%$, com-

Table 5: Comparison with popular multi-view learning methods (trained on 70% of data) based on Accuracy (ACC, %) with $1 - \tau = 95\%$.

Method	HAND	SCENE	ANIMAL	CAL	CUB	HMDB	MRNet
DCCA [2]	94.01±1.20	51.21±1.67	81.10±1.22	80.00±0.77	78.12±1.00	44.90±1.37	87.21±2.79
DCCAE [48]	97.01±0.23	52.12±0.31	84.97±0.21	88.11±0.79	80.42±0.88	48.00±0.80	88.23±0.11
DTCCA [50]	96.01±0.10	58.32±0.20	82.77±0.10	90.01±0.26	80.31±0.22	55.01±0.87	84.01±0.32
CPM-Nets [53]	94.00±1.10	65.23±0.02	85.21±0.20	89.11±1.88	85.67±0.02	66.56±1.10	88.03±0.00
DUA-Nets [15]	98.00±0.21	65.23±0.11	87.81±1.44	93.01±0.22	79.83±1.50	62.12±0.87	89.61±1.00
MVTCAE [19]	97.00±0.23	64.22±0.01	86.21±0.12	91.01±0.44	90.00±0.12	72.89±2.01	93.01±1.22
TMC [17]	98.23±0.09	66.57±0.03	88.21±0.41	93.01±0.10	90.04±1.01	74.10±1.22	91.01±0.80
TMDOA [29]	98.30±0.21	70.10±0.12	90.00±0.01	93.38±0.04	91.78±1.22	85.49±0.57	94.01±1.22
ETMC [18]	98.20±0.23	66.01±0.07	88.90±0.25	92.77±0.37	90.04±1.21	74.22±0.88	92.02±1.21
SMDC [28]	98.89±0.10	72.06±0.22	93.46±0.00	96.40±0.10	92.06±0.16	90.00±0.20	92.08±1.00
CMDC (Ours)	**99.17±0.22**	**74.56±0.92**	**94.24±0.28**	**97.53±0.00**	**93.00±0.00**	**90.89±0.55**	**94.01±0.67**

Table 6: Analysis of CMDC with $\tau = 0.05$ on SCENE.

Method	ECE ↓	Ineff ↓	Cov ↑
CMDC ($\tau = 0.1$)	**0.0119 ± 0.00**	**2.9901 ± 0.02**	**0.9063 ± 0.00**
CMDC ($\tau = 0.05$)	**0.0091 ± 0.00**	**3.8674 ± 0.01**	**0.9524 ± 0.00**

pared with other state-of-the-art multi-view deep learning methods, demonstrating superior accuracy (ACC).

Furthermore, we extend our analysis to the SCENE dataset, comprising 15 classes, by evaluating the Expected Calibration Error (ECE), Inefficiency (Ineff), and Coverage (Cov) metrics to assess the impact of different τ values. The results presented in Table 6 illustrate our method's ability to consistently achieve an appropriate coverage rate $(1 - \tau)$ across various τ values. Comparing the results of Top-1 accuracy (ACC) shown in Table 5, which is 74.56%, with the Top-k ACC (which indicates the coverage rate) of 95.24% shown in Table 6, demonstrates the effectiveness of our method. This high coverage rate allows our method to provide precise prediction sets, leading to more reliable results for decision-makers. Moreover, the analysis reveals a positive correlation between the probability of coverage rate $1 - \tau$ and the Cov metric, while observing a decrease in the Ineff metric, signifying an increase in the prediction set size.

E SOURCE CODE

The code for this work can be found in the Supplementary Materials / source-code.zip file.