

# Quantifying Classification Uncertainty using Regularized Evidential Neural Networks

**Xujiang Zhao, Yuzhe Ou**

UT Dallas, TX 75080  
{xujiang.zhao, yxo170030}  
@utdallas.edu

**Lance Kaplan**

US Army Research Lab  
Adelphi, MD 20783, USA  
lkaplan@ieee.org

**Feng Chen**

UT Dallas, TX 75080  
fchen5@albany.edu

**Jin-Hee Cho**

Virginia Tech  
Falls Church, VA 22043  
jicho@vt.edu

## Abstract

Traditional deep neural nets (NNs) have shown the state-of-the-art performance in the task of classification in various applications. However, NNs have not considered any types of uncertainty associated with the class probabilities to minimize risk due to misclassification under uncertainty in real life. Unlike Bayesian neural nets indirectly inferring uncertainty through weight uncertainties, evidential neural networks (ENNs) have been recently proposed to support explicit modeling of the uncertainty of class probabilities. It treats predictions of an NN as subjective opinions and learns the function by collecting the evidence leading to these opinions by a deterministic NN from data. However, an ENN is trained as a black box without explicitly considering different types of inherent data uncertainty, such as vacuity (uncertainty due to a lack of evidence) or dissonance (uncertainty due to conflicting evidence). This paper presents a new approach, called a *regularized ENN*, that learns an ENN based on regularizations related to different characteristics of inherent data uncertainty. Via the experiments with both synthetic and real-world datasets, we demonstrate that the proposed regularized ENN can better learn of an ENN modeling different types of uncertainty in the class probabilities for classification tasks.

## Introduction

Inherent data uncertainty due to different root causes has been serious hurdles in providing effective solutions for real world problems. Critical safety issues often may occur because different types of uncertainty have not been sufficiently considered and even some perceived uncertainty tends to be improperly interpreted because of lack of knowledge on the cause of the uncertainty. For example, when an autonomous vehicle misclassifies objects, this may lead to highly risky situations. Deep learning (DL) models have recently gained tremendous attention in the data science community (Kipf and Welling 2016; Veličković et al. 2018). Despite their superior performance in various tasks (e.g., classification or regression), they are limited in dealing with different types of data uncertainty. Predictive uncertainty estimation using Bayesian NNs (BNNs) has been explored for classification prediction or regression in the computer vi-

sion applications (Malinin and Gales 2018). They considered well-known uncertainty types, such as aleatoric uncertainty and epistemic uncertainty, where aleatoric uncertainty only considers data uncertainty caused by statistical randomness (e.g., observation noises) while epistemic uncertainty refers to model uncertainty introduced by limited knowledge or ignorance in collected data. On the other hand, in the belief / evidence theory, Subjective Logic (SL) (Jøsang, Cho, and Chen 2018) considered vacuity (or a lack of evidence) as the key dimension of uncertainty. A recent work on SL has defined other aspects of uncertainty, such as dissonance, consonance, vagueness, and monosonance (Jøsang, Cho, and Chen 2018).

Traditional deep neural nets (NNs) have been popularly applied in solving classification tasks. However, NNs have not been used to deal with uncertainties associated with the class probabilities to minimize risk due to misclassification under uncertainty in real life applications. Methods using BNNs have been proposed to indirectly predict uncertainty through weight uncertainties. Recently, techniques using evidential neural networks (ENNs) have been proposed to support explicit modeling of the uncertainty of class probabilities. It treats predictions of a neural net as subjective opinions and learns the function that collects the evidence leading to these opinions by a deterministic NN from data. However, an ENN is still trained as a black box without explicitly considering different types of data uncertainty (e.g., vacuity or dissonance). This paper presents a new approach that learns an ENN based on regularizations related to different types of inherent data uncertainty, which we call it *regularized ENNs*.

Our **key contributions** are summarized as follows:

1. Our proposed regularized ENNs-based method is the first that explicitly considers different types of inherent data uncertainty in learning an ENN for quantifying uncertainty in classification tasks.
2. We validated the performance of our regularized ENNs-based method via extensive simulation experiments using both synthetic and real-world datasets. We conducted the performance comparison of predicting different types of uncertainty when applying classification methods based on the regularized ENNs, ENNs (without regularization), and NNs.

## Motivational Scenarios

This paper focuses on scenarios where decision making is modeled as a classification problem with the aim of quantifying uncertainty of decision making in terms of classification prediction. Such cases are common in the government and public sectors with the following examples:

- *Self-driving cars safety*: An autonomous car may deploy a deep learning system that fails to distinguish the white side of a truck against bright sky because such type of the truck has not been known before. In this case, vacuity (i.e., a lack of evidence) can represent a measure of the extent the detected object belongs to a class of cars that has never been known before (i.e., not in a considered domain). Further, if the truck is not detected as a single class type, it increases uncertainty due to vagueness. Or if the truck is detected with either one type or the other due to conflicting evidence, it also increases uncertainty because of conflicting evidence.
- *Public health*: The outbreak of an unknown disease occurs and we observe that the correlated symptoms and the spread patterns cannot be reliably predicted. As described in other examples, a lack of evidence (i.e., vacuity), conflicting evidence supporting either one or the other (i.e., dissonance among evidence) or vagueness by failing in identifying as a single object may significantly lead to increasing uncertainty which introduces high challenging in decision making.
- *Security in network infrastructure*: A power network system can be targeted by a cyberattack, which can break down the system which can introduce tremendous damage because this attack is not timely detected. Uncertainty introduced by vacuity, dissonance, or vagueness (as each term is discussed above) can delay decisions that can immediately take actions to minimize the damage.

To focus subsequent discussion on a single concrete example, we will consider image classification, a critical component in autonomous cars. We will quantify the uncertainty of an image classifier in recognizing objects of a known or unknown class in images using neural network models.

## Related Work

**Uncertainty Quantification in Bayesian Deep Learning (BDL)**: Machine/deep learning (ML/DL) researchers considered *aleatoric* uncertainty (AU) and *epistemic* uncertainty (EU) based on BNNs for computer vision applications. AU consists of homoscedastic uncertainty (i.e., constant errors for different inputs) and heteroscedastic uncertainty (i.e., different errors for different inputs) (Gal 2016). A BDL framework was presented to estimate both AU and EU simultaneously in regression settings (e.g., depth regression) and classification settings (e.g., semantic segmentation) (Kendall and Gal 2017). Later *distributional uncertainty* is defined based on distributional mismatch between the test and training data distributions (Malinin and Gales 2018). *Dropout variational inference* (Gal and Ghahramani 2016) was proposed as one of key approximate inference techniques in BNNs. Other methods (Eswaran, Günnemann,

and Faloutsos 2017; Zhang et al. 2018) measured overall uncertainty in node classification but were not focused on uncertainty decomposition and GNNs.

**Uncertainty Quantification in Belief/Evidence Theory**: In belief/evidence theory domain, uncertainty reasoning has been substantially explored in Fuzzy Logic (De Silva 2018), Dempster-Shafer Theory (DST) (Sentz, Ferson, and others 2002), or Subjective Logic (SL) (Jøsang 2016). Unlike the efforts in ML/DL above, belief/evidence theory focused on reasoning of inherent uncertainty in information resulting from unreliable, incomplete, deceptive, and/or conflicting evidence. SL considered uncertainty in subjective opinions in terms of *vacuity* (i.e., a lack of evidence) and *vagueness* (i.e., failure of discriminating a belief state) (Jøsang 2016). Recently, other dimensions of uncertainty have been studied, such as *dissonance* (due to conflicting evidence) and *consonance* (due to evidence about composite subsets of state values) (Jøsang, Cho, and Chen 2018).

## Background

This section provides the background knowledge to understand this work including: (1) subjective opinions in SL; (2) uncertainty dimensions in a subjective opinion; and (3) evidential NNs to predict subjective opinions.

**Subjective Opinions in SL**: A multinomial opinion in a given proposition  $x$  is represented by  $\omega_Y = (\mathbf{b}_Y, u_Y, \mathbf{\alpha}_Y)$  where a domain is  $\mathbb{Y} \equiv \{1, \dots, K\}$  and the additivity requirement of  $\omega_y$  is given as  $\sum_{y \in \mathbb{Y}} \mathbf{b}_Y(y) + u_Y = 1$ . To be specific, each parameter indicates,

- $\mathbf{b}_Y$ : *belief mass distribution* over  $\mathbb{Y}$ ;
- $u_Y$ : *uncertainty mass* representing *vacuity of evidence*;
- $\mathbf{\alpha}_Y$ : *base rate distribution* over  $\mathbb{Y}$ .

The projected probability distribution of a multinomial opinion is given by:

$$\mathbf{P}_Y(y) = \mathbf{b}_Y(y) + \mathbf{\alpha}_Y(y)u_Y, \quad \forall y \in \mathbb{Y} \quad (1)$$

Multinomial probability density over a domain of cardinality  $K$  is represented by the  $K$ -dimensional Dirichlet PDF where the special case with  $K = 2$  is the Beta PDF as a binomial opinion. Denote a domain of  $K$  mutually disjoint elements in  $\mathbb{Y}$  and  $\mathbf{\alpha}_Y$  the strength vector over  $y \in \mathbb{Y}$  and  $\mathbf{p}_Y$  the probability distribution over  $\mathbb{Y}$ . Dirichlet PDF with  $\mathbf{p}_Y$  as  $K$ -dimensional variables is defined by:

$$\text{Dir}(\mathbf{p}_Y; \mathbf{\alpha}_Y) = \frac{1}{B(\mathbf{\alpha}_Y)} \prod_{y \in \mathbb{Y}} \mathbf{p}_Y(y)^{\mathbf{\alpha}_Y(y)-1} \quad (2)$$

where  $\frac{1}{B(\mathbf{\alpha}_Y)} = \frac{\Gamma(\sum_{y \in \mathbb{Y}} \mathbf{\alpha}_Y(y))}{\prod_{y \in \mathbb{Y}} \Gamma(\mathbf{\alpha}_Y(y))}$ ,  $\mathbf{\alpha}_Y(y) \geq 0$ , and  $\mathbf{p}_Y(y) \neq 0$  if  $\mathbf{\alpha}_Y(y) < 1$ .

We term *evidence* as a measure of the amount of supporting observations collected from data in favor of a sample to be classified into a certain class. Let  $\mathbf{r}_Y(y) \geq 0$  be the evidence derived for the singleton  $y \in \mathbb{Y}$ . The total strength  $\mathbf{\alpha}_Y(y)$  for the belief of each singleton  $y \in \mathbb{Y}$  is given by:

$$\mathbf{\alpha}_Y(y) = \mathbf{r}_Y(y) + \mathbf{\alpha}_Y(y)W, \quad (3)$$

$$\text{where } \mathbf{r}_Y(y) \geq 0, \quad \forall y \in \mathbb{Y}$$

where  $W$  is a non-informative weight representing the amount of uncertain evidence and  $\alpha_Y(y)$  is the base rate distribution. Given the Dirichlet PDF, the expected probability distribution over  $\mathbb{Y}$  is:

$$\mathbb{E}_Y(y) = \frac{\alpha_Y(y)}{\sum_{y_i \in \mathbb{Y}} \alpha_Y(y_i)} = \frac{r_Y(y) + \alpha_Y(y)W}{W + \sum_{y_i \in \mathbb{Y}} r_Y(y_i)} \quad (4)$$

The observed evidence in the Dirichlet PDF can be mapped to the multinomial opinions by:

$$\mathbf{b}_Y(y) = \frac{r(y)}{S}, \quad u_x = \frac{W}{S}, \quad (5)$$

where  $S = \sum_{y_i \in \mathbb{Y}} \alpha_Y(y_i)$ . With loss of generality, we set  $\alpha_Y(y) = \frac{1}{K}$  and the non-informative prior weight (i.e.,  $W = K$ ), and hence  $\alpha_Y(y) \cdot W = 1$  for each  $y \in \mathbb{Y}$ .

**Uncertainty Characteristics of Subjective Opinions:** In (Jøsang, Cho, and Chen 2018), the multidimensional uncertainty dimensions of a subjective opinion based on the formalism of SL are discussed, such as singularity, vagueness, vacuity, dissonance, consonance and monosonance. These different dimensions of uncertainty can be observed from binomial, multinomial, or hyper opinions depending on their characteristics (e.g., vagueness is only observed in hyper opinions when a composite belief exists). As we deal with a multinomial opinion in this work, we discuss two main types of uncertainty dimensions that can be estimated in a multinomial opinion, which are *vacuity* and *dissonance*.

The main cause of vacuity is due to a lack of evidence or knowledge, which corresponds to uncertainty mass,  $u$ , of an opinion in SL as:

$$\text{Vac}(\alpha_Y) = \frac{W}{S}. \quad (6)$$

This type of uncertainty refers to uncertainty because the analyst does not have sufficient information or knowledge to analyze the uncertainty.

The *dissonance* of an opinion can be caused by the same amount of conflicting evidence and is estimated by the difference between singleton belief masses (e.g., class labels), leading to ‘inconclusiveness’ in decision making situations. For example, given a four-state multinomial opinion with  $(b_1, b_2, b_3, b_4, u, a) = (0.25, 0.25, 0.25, 0.25, 0.0, a)$  based on Eq. (5), although  $u$  (vacuity) is zero, one cannot make a decision if there are the same amounts of beliefs supporting respective beliefs. Given a multinomial opinion with non-zero belief masses, the measure of dissonance can be obtained by:

$$\text{Diss}(\alpha_Y) = \sum_{y_i \in \mathbb{Y}} \left( \frac{\mathbf{b}_Y(y_i) \sum_{y_j \in \mathbb{Y} \setminus y_i} \mathbf{b}_Y(y_j) \text{Bal}(y_j, y_i)}{\sum_{y_j \in \mathbb{Y} \setminus y_i} \mathbf{b}_Y(y_j)} \right), \quad (7)$$

where the relative mass balance between a pair of belief masses  $\mathbf{b}_Y(y_j)$  and  $\mathbf{b}_Y(y_i)$  is expressed by:

$$\text{Bal}(y_j, y_i) = 1 - \frac{|\mathbf{b}_Y(y_j) - \mathbf{b}_Y(y_i)|}{\mathbf{b}_Y(y_j) + \mathbf{b}_Y(y_i)}, \quad (8)$$

Recall that we measure the dissonance only when the belief mass is non-zero. If all belief masses are zero with vacuity

being 1 (i.e.,  $u_Y = 1$ ), the dissonance (or balance) will be set to zero.

The relative mass balance has its maximum at 1 when  $\mathbf{b}_Y(y_j) = \mathbf{b}_Y(y_i)$ . The relative mass balance has the minimum at 0 when one of the belief masses equals 0. In case of the zero sum of belief masses in both the nominators and the denominators of Eq. (7) and Eq. (8), it must be assumed that the fraction of the sums of belief masses equals 1. But notice that even with high vacuity (a lack of evidence), high dissonance can be observed.

The above two uncertainty measures (i.e., vacuity and dissonance) can be interpreted using class-level evidence measures of subjective opinions. As in Figure 1, given three classes (1, 2, and 3), we have three subjective opinions  $\{\alpha_1, \alpha_2, \alpha_3\}$ , represented by the three-class evidence measures as:  $\alpha_1 = (6, 1, 1)$ ,  $\alpha_2 = (20, 20, 20)$ ,  $\alpha_3 = (20, 20, 50)$ . The vacuity and dissonance measures of the three subjective opinions can be calculated via Eq. (6) and Eq. (7), respectively.

When the evidence measures of the three classes of a subjective opinion decrease, the corresponding vacuity increases. When the evidence measures of the three classes of a subjective opinion become closer (and therefore are more conflicting with each other), the corresponding dissonance increases.

**Evidential Neural Networks (ENNs):** The gold standard for DNNs is to use the softmax operator to convert the continuous activations of the output layer to class probabilities. Although the softmax function provides a point estimate for the class probabilities of a sample, it does not provide the associated uncertainty. ENNs are similar to classical NNs, except that the softmax layer is replaced by an activation layer (e.g., ReLU) to ascertain non-negative output, which is taken as the evidence vector for the predicted Dirichlet distribution.

Given a sample  $i$ , let  $f(\mathbf{x}_i|\Theta)$  represent the evidence vector predicted by the network for the classification, where  $\mathbf{x}_i \in \mathbb{R}^L$  is the observation of a feature in this sample and  $\Theta$  is a set of network parameters. Then, the corresponding Dirichlet distribution has parameters  $\alpha_i = f(\mathbf{x}_i|\Theta) + 1$ . Once the parameters of this distribution are calculated, its mean (i.e.,  $\alpha_i/S_i$ ) can be taken as an estimate of the class probabilities.

Let  $\mathbf{y}_i$  be a dummy vector encoding the ground-truth class of observation  $\mathbf{x}_i$  with  $y_{ij} = 1$  and  $y_{ik} = 0$  for all  $k \neq j$ , and  $\alpha_i$  be the parameters of the Dirichlet density on the predictors. The Dirichlet density  $\text{Dir}(\mathbf{p}_i; \alpha)$  is the prior on the Multinomial distribution  $\text{Multi}(\mathbf{y}_i|\mathbf{p}_i)$ . The following sum of squared loss (SSL) is used to estimate the parameters  $\alpha_i$  based on the sample  $i$ :

$$\begin{aligned} \mathcal{L}(f(\mathbf{x}_i|\Theta), \mathbf{y}_i) &= \int \frac{\|\mathbf{y}_i - \mathbf{p}_i\|_2^2}{B(\alpha_i)} \prod_{j=1}^K p_{ij}^{(\alpha_{ij}-1)} d\mathbf{p}_i \\ &= \sum_{j=1}^K \left( y_{ij}^2 - 2y_{ij}\mathbb{E}[p_{ij}] + \mathbb{E}[p_{ij}^2] \right). \end{aligned} \quad (9)$$

In the current design of ENNs, an ENN is trained as a black box using the above squared error function as the loss

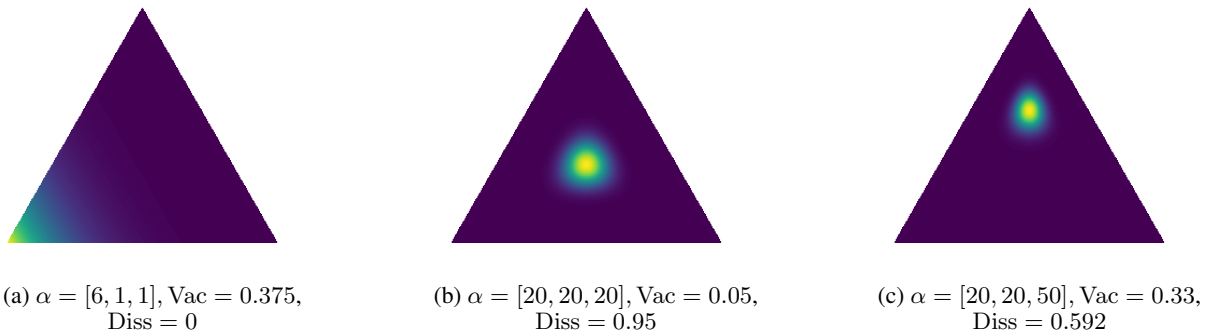


Figure 1: Illustration of different vacuity and dissonance of subjective opinions based on their evidence measures.

function with a regularization term to promote epistemic uncertainty

$$\mathcal{L}(\Theta) = \mathbb{E}_{(\mathbf{x}_i, \mathbf{y}_i) \sim \mathcal{D}} [\mathcal{L}(\alpha_i, \mathbf{y}_i) + \lambda_t KL [\text{Dir}(\mathbf{p}_i; \alpha_i^{-\mathbf{y}_i}) || \text{Dir}(\mathbf{p}_i; \mathbf{1})]], \quad (10)$$

where  $KL[\cdot]$  represent the Kullback-Leibler divergence. Note that  $\alpha_i^{-\mathbf{y}_i}$  is the evidence predicted by the network except that the ground truth class evidence is set to zero. The regularization term is necessary to bias the network to generate high epistemic uncertainty where it is likely to make errors such as decision boundaries. However, the training of ENN does not necessarily bias it to return high epistemic far from the training data at points also far from decision boundaries.

Given the various characteristics of uncertainty that subjective opinions can describe, we propose a regularized ENN in the section below to learn an NN that is more effective to quantify uncertainty for classification tasks through the regularization on multidimensional uncertainty, such as vacuity and dissonance.

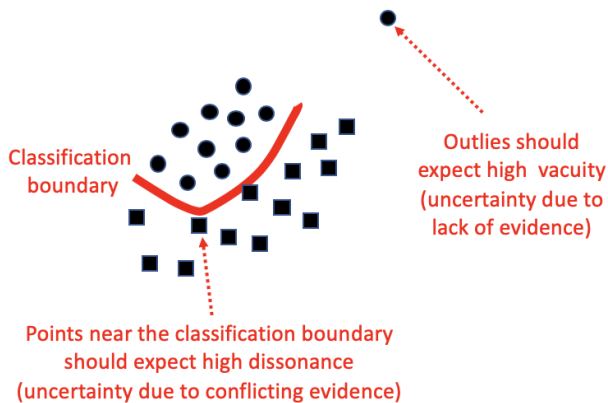


Figure 2: Illustration of boundary points with high dissonance and outliers (or OOD samples) with high vacuity.

**Regularized ENNs:** Given a set of samples  $\mathcal{D} = \{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_N, \mathbf{y}_N)\}$ , we can identify some samples that should expect high vacuity or dissonance using heuristic ways. As shown in Figure 2, the samples with conflicting labels in their neighbors should expect high dissonance.

These samples are often close to the classification boundary. In addition, outliers (or out-of-distribution samples) should expect high vacuity.

Let  $\mathcal{D}_{\text{OOD}}$  be the set of out-of-distribution samples in  $\mathcal{D}$  and  $\mathcal{D}_{\text{BOD}}$  be the set of samples whose neighboring samples have conflicting evidence, where the set of samples in in-distribution ( $\mathcal{D}_{\text{IN}}$ ) is set to  $\mathcal{D}_{\text{IN}} = \mathcal{D} - \mathcal{D}_{\text{OOD}} - \mathcal{D}_{\text{BOD}}$ . We then propose a training method using a regularized ENN that minimizes the following loss function:

$$\mathcal{L}(\Theta) = \mathbb{E}_{(\mathbf{x}_i, \mathbf{y}_i) \sim \mathcal{D}} [\mathcal{L}(f(\mathbf{x}_i | \Theta), \mathbf{y}_i)] - \lambda_1 \mathbb{E}_{(\mathbf{x}_i, \mathbf{y}_i) \sim \mathcal{D}_{\text{OOD}}} [\text{Vac}(f(\mathbf{x}_i | \Theta))] - \lambda_2 \mathbb{E}_{(\mathbf{x}_i, \mathbf{y}_i) \sim \mathcal{D}_{\text{BOD}}} [\text{Diss}(f(\mathbf{x}_i | \Theta))], \quad (11)$$

over the parameters  $\Theta$  of  $f$ , where  $\lambda_1$  and  $\lambda_2$  are the tradeoff parameters that decide the importance of the expected uncertainty dimensions associated with vacuity and dissonance.

## Experiments & Results

We evaluate our proposed method following the experimental setup in (Louizos and Welling 2017). We use the standard LeNet with ReLU non-linearities as the neural network architecture. All experiments are implemented in TensorFlow (Abadi et al. 2015) where the Adam optimizer (Kingma and Ba 2014) is used with default settings for training.

### Experimental Setup

**Dataset:** The CIFAR10 dataset consists of 60,000  $32 \times 32$  colour images in 10 classes, with 6,000 images per class. The original classification setting used 50,000 training images and 10,000 test images.

**Out-of-Distribution (OOD) Setting:** We used a subset of the classes in the CIFAR10 dataset for training and the rest of the subset for testing OOD uncertainty. To be specific, for training, we use the samples from the first five categories {airplane, automobile, bird, cat, deer} in the training set from the CIFAR10 dataset as  $\mathcal{D}_{\text{IN}}$  (i.e., in-distribution), another two categories {ship, truck} as  $\mathcal{D}_{\text{OD}}$  (i.e., out-of-distribution), and the left three categories {dog, frog, horse} for testing the OOD. For  $\mathcal{D}_{\text{BOD}}$  (i.e., boundary-out-of-distribution) dataset, we select the boundary sample by taking the following steps: (1) calculate the distance between each training image; (2) for each image, choose top 10 similar images as observed evidence; (3) calculate dissonance

based the 10 pieces of observed evidence for each training image; and (4) choose the 500 images with the highest dissonance. Here we measure the similarity based on the cosine similarity metric (Bishop 2006). by:

$$\text{Sim}(A, B) = \frac{A \cdot B}{\|A\|_2 \cdot \|B\|_2}, \quad (12)$$

where  $A$  and  $B$  are the feature vectors of two images and the raw image pixels are treated as the features for the cosine similarity.

**Synthetic Dataset:** We generate a dataset of three classes (1, 2, and 3) in a two dimensional space. We sample 1,000 points of each class (1, 2, and 3) from  $\mathcal{N}\left(\begin{pmatrix} -2 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$ ,  $\mathcal{N}\left(\begin{pmatrix} 0 \\ 1.464 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$ , and  $\mathcal{N}\left(\begin{pmatrix} 2 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$ , respectively. In addition, we sample 100 OOD points from  $\mathcal{N}\left(\begin{pmatrix} -8 \\ -8 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$  and  $\mathcal{N}\left(\begin{pmatrix} 8 \\ -8 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$  each. Therefore, we have 3,000 in-distribution points and 200 out-of-distribution points in total. We expect that the testing points near the boundary of the three classes should expect high dissonance measures and the testing points that are far away from the 3,000 training points should expect high vacuity measures.

**Comparison Schemes:** We compared the following approaches:

- **L2** corresponds to the standard deterministic neural nets with softmax output and weight decay;
- **ENN** refers to evidential deep learning model (Sensoy, Kaplan, and Kandemir 2018);
- **ENN-Vac** uses ENN with the vacuity regularization only;
- **ENN-Diss** uses ENN with the dissonance regularization only;
- **ENN-Vac-Diss** uses ENN with both the vacuity and dissonance regularization.

**Parameter Tuning:** We trained the same model LeNet architecture using 20 and 50 filters with size  $5 \times 5$  at the first and second convolutional layers, respectively; and 500 hidden units are used for the fully connected layer, same as in the ENN model (Sensoy, Kaplan, and Kandemir 2018). For all comparison methods, we set `batch_size` = 1000, `learning_rate` = 0.01, `dropout_rate` = 0.9 and `weight_decay_coefficient` = 0.005. For vacuity and dissonance regularization parameters, we set  $\lambda_1 = 0.01$ ,  $\lambda_2 = 0.01$ . The values of the parameters used are obtained by tuning them to show the optimal effect of using regularization for synthetic data.

**Metrics:** Our approach directly quantifies vacuity and dissonance based on Eqs. (6) and (7), respectively. However, some approaches use entropy to measure prediction uncertainty as in (Louizos and Welling 2017), i.e., uncertainty in predicted probabilities increases as their entropy increases. To be fair, we use the same metric for evaluating prediction

uncertainty for OOD in the rest of the paper and for class probabilities. We use the empirical CDFs over the range of possible entropy  $[0, \log 5]$  for all models trained with the CIFAR10 dataset. The curves closer to the bottom right corner of the plot are considered desirable, indicating maximum entropy in all predictions, leading to high OOD detection (Louizos and Welling 2017).

## Experimental Results

**Results with Synthetic Datasets:** In this experiment, we trained the model based on the in-distribution class (i.e., yellow, red and white nodes) and tested the rest of nodes except that ENN-Vac and ENN-Vac-Diss also use outlier nodes (pink nodes at both bottom left and bottom right corners) in the training, ranging in  $[-10, 10] \times [-10, 10]$ , where each coordinate represents the values of node features in the given synthetic dataset. In Figure 3, the yellow, red and white nodes are training nodes where the contour map indicates the different level of strengths from purple (weak, low uncertainty) to green (strong, high uncertainty) for a given uncertainty (i.e., vacuity, dissonance). Here, we evaluate the performance of the all four schemes (i.e., ENN, ENN-Vac, ENN-Diss, and ENN-Vac-Diss) in terms of classification predictions and how different uncertainty types are measured under different schemes. Figures 3a and 3c show vacuity from ENN and ENN-Vac, respectively. The observed high vacuity in Figure 3c (ENN-Vac) is more reasonable as the opinions of nodes are far away from the training data. That is, since OOD class nodes are expected to have lack of evidence with high vacuity. Region influenced by our OOD training nodes extend to the whole bottom area which gives more advantages of OOD detection. Figures 3b and 3d show the observed dissonance of nodes’ subjective opinions. As expected, Figure 3d shows a more reasonable level of dissonance than Figure 3b, exhibiting stronger dissonance in the boundary areas among three classes. Figure 3 also demonstrates that the choice of OOD samples is critical for Vacuity to be high far from the training data. Ideally, the vacuity should be high at the upper portions and sides of Figure 3c as well. It indicates that the selection of appropriate OOD samples affects the accurate estimation of vacuity.

**Results with Cifar Datasets:** Figure 4 shows the empirical CDFs for all models trained with the CIFAR10 dataset. The curves closer to the bottom right corner of the plot are desirable, indicating maximum entropy in all predictions (Louizos and Welling 2017). It is clear that the uncertainty estimates of our model ENN-Vac is significantly better than those of the baseline methods (ENN and Softmax), which shows the benefit of using vacuity as an indicator of uncertainty to improve OOD detection. ENN-Diss is comparable with ENN, showing that dissonance regularization does not help significantly for the OOD detection. One interesting pattern shows that ENN-Vac-Diss outperformed among all other methods, demonstrating that considering both vacuity and dissonance regularized can significantly improve the OOD detection.

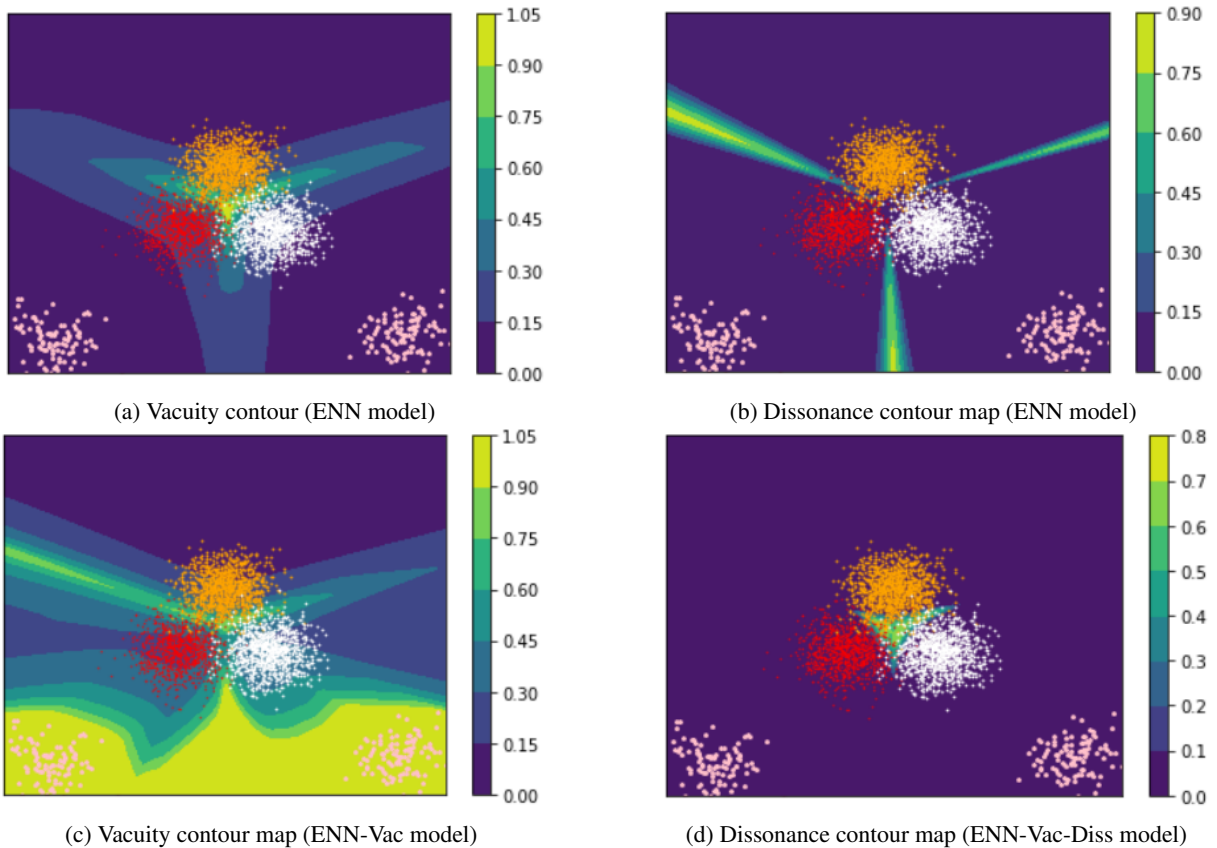


Figure 3: Comparison between the vacuity and dissonance of the predictions of ENN ((a) and (b)) and regularized ENN ((c) and (d)) where each node’s feature value is represented by a coordinate with the range in  $[-10, 10] \times [-10, 10]$ .

## Conclusion & Future Work

In this work, we developed the regularized ENNs that learn an ENN based on the regularization associated with two different uncertainty dimensions, vacuity and dissonance. Via extensive experiments based on both synthetic and real datasets, we demonstrated more reasonable levels of estimated vacuity and dissonance in classification prediction and out-of-distribution detection. From our study, the following **key findings** are obtained:

- Overall our proposed ENN considering both vacuity and dissonance (i.e., ENN-Vac-Diss) outperformed among all in terms of the performance in the out-of-distribution task. In addition, our proposed ENN with vacuity considered (ENN-Vac) outperformed among all baselines as the vacuity is regularized in an ENN.
- Our proposed ENN-Vac and ENN using dissonance (ENN-Diss) showed more desirable uncertainty measurement for vacuity and dissonance, respectively. To be specific, ENN-Vac showed high vacuity in more outliers (out-of-distribution) samples. In addition, when ENN-Diss is used, higher dissonance is observed particularly at the boundary areas between different classes. This provides great promise for this regularized ENNs considering different uncertainty dimensions to be applicable in vari-

ous types of classification prediction and OOD detection tasks.

In our future work, we plan to: (1) investigate the clear reasons of unstable performance of ENN considering multiple dimensions of uncertainty; (2) investigate how OOD samples can be generated to more effectively estimate vacuity; (3) validate the performance of our *regularized* ENNs approach based on more real-world datasets (e.g., object detection datasets); and (4) extend our proposed work with more uncertainty regularized that can deal with vagueness in hyper opinions in Subjective Logic.

## References

- Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; and Zheng, X. 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

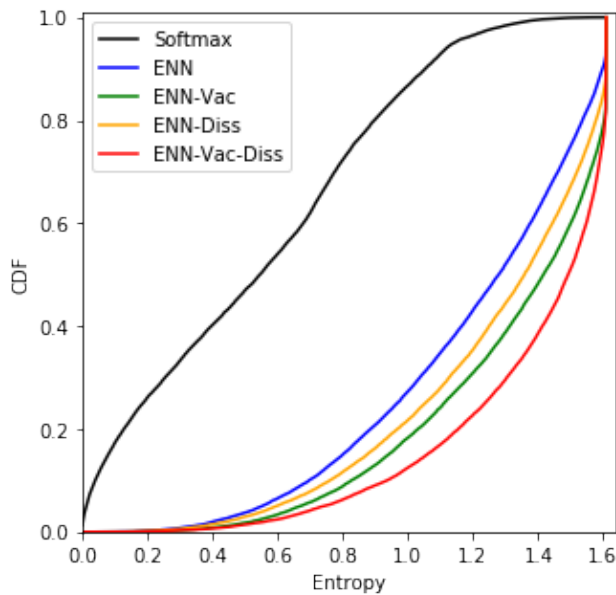


Figure 4: Empirical CDF for the entropy of the predictive distributions on the subset of the classes of the CIFAR10 dataset.

Bishop, C. 2006. *Pattern recognition and machine learning*. Springer-Verlag New York.

De Silva, C. W. 2018. *Intelligent control: fuzzy logic applications*. CRC press.

Eswaran, D.; Günnemann, S.; and Faloutsos, C. 2017. The power of certainty: A dirichlet-multinomial model for belief propagation. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, 144–152. SIAM.

Gal, Y., and Ghahramani, Z. 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *ICML*, 1050–1059.

Gal, Y. 2016. *Uncertainty in deep learning*. University of Cambridge.

Jøsang, A.; Cho, J.-H.; and Chen, F. 2018. Uncertainty characteristics of subjective opinions. In *Fusion*, 1998–2005. IEEE.

Jøsang, A. 2016. *Subjective logic*. Springer.

Kendall, A., and Gal, Y. 2017. What uncertainties do we need in Bayesian deep learning for computer vision? In *NIPS*, 5574–5584.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kipf, T. N., and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Louizos, C., and Welling, M. 2017. Multiplicative normalizing flows for variational bayesian neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, 2218–2227. JMLR. org.

Malinin, A., and Gales, M. 2018. Predictive uncertainty estimation via prior networks. *arXiv preprint arXiv:1802.10501*.

Sensoy, M.; Kaplan, L.; and Kandemir, M. 2018. Evidential deep learning to quantify classification uncertainty. In *NIPS*, 3183–3193.

Sentz, K.; Ferson, S.; et al. 2002. *Combination of evidence in Dempster-Shafer Theory*, volume 4015. Citeseer.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lió, P.; and Bengio, Y. 2018. Graph attention networks. *ICLR*.

Zhang, Y.; Pal, S.; Coates, M.; and Üstebay, D. 2018. Bayesian graph convolutional neural networks for semi-supervised classification. *arXiv preprint arXiv:1811.11103*.