

# An Extension of the Evidential Neural Network Classifier based on Contextual Discounting

Serigne Mamadou Diène  
UR 3926 LGI2A  
Univ. Artois, Béthune  
F-62400, France.

Sébastien Ramel  
UR 3926 LGI2A  
Univ. Artois, Béthune  
F-62400, France.

Frédéric Pichon  
UR 3926 LGI2A  
Univ. Artois, Béthune  
F-62400, France.

David Mercier  
UR 3926 LGI2A  
Univ. Artois, Béthune  
F-62400, France.

smamadou.diene@univ-artois.fr sebastien.ramel@univ-artois.fr frederic.pichon@univ-artois.fr david.mercier@univ-artois.fr

**Abstract**—In this article, an extension of the Evidential Neural Network (ENN) introduced by Denœux is proposed by replacing the classical discounting operation with a refinement of the latter, called the contextual discounting (CD) operation, which allows for class-specific reliability degrees and yields a more expressive representation of uncertainty. This extension, named CD-ENN, is trained using the cross-entropy of the normalized contour function, making it efficient to learn its additional parameters. Experimental results on UCI datasets show that CD-ENN outperforms the original ENN in terms of predictive performance, as measured by the recently introduced generalized negative log-likelihood criterion.

**Index Terms**—Dempster-Shafer Theory, Evidential Neural Network, Contextual Discounting

## I. INTRODUCTION

Evidential classifiers quantify classification uncertainty using Dempster-Shafer belief functions [2], [13]. They permit the distinction between aleatory uncertainty and epistemic uncertainty [6], which is important in various situations [9]. The *Evidential K-Nearest Neighbors* (EKNN) algorithm, introduced by Denœux [3], is arguably the most well-known evidential classifier. In this method, each of the  $K$  nearest neighbors of an instance  $\mathbf{x}$  to be classified, provides a belief function or, equivalently, a so-called mass function, expressing a belief about the (unknown) true class of  $\mathbf{x}$ . This mass function is obtained by discounting the neighbor's class information according to a reliability degree that decreases as the distance between the neighbor and  $\mathbf{x}$  increases. Finally, the mass functions established from the  $K$  nearest neighbors are combined using Dempster's rule, yielding a final mass function that reflects the overall belief regarding  $\mathbf{x}$ 's class membership.

An adaptive extension of this algorithm was later proposed by Denœux [4] in the form of a neural network known as the *Evidential Neural Network* (ENN). In this extension, the (final) mass function representing the class of an instance  $\mathbf{x}$  is computed using a limited number of prototypes (instead of nearest neighbors), enabling faster classification and significantly reducing memory requirements. More precisely, each prototype induces a mass function obtained by discounting the prototype's class membership information, and then the mass

functions from the prototypes are merged using Dempster's rule to obtain the final mass function about the class of  $\mathbf{x}$ . An additional key advantage of ENN is that it can be used as a replacement for the traditional *softmax* layer in deep neural networks, transforming them into evidential variants offering a finer quantification of predictive uncertainty [14].

Recently, Denœux et al. [7] introduced the *Contextual Discounting Evidential K-Nearest Neighbors* (CD-EKNN) algorithm, an extension of EKNN that uses contextual discounting [11] instead of discounting. The contextual discounting is a refinement of discounting, which allows assigning class-specific reliability degrees. Furthermore, CD-EKNN differs also from EKNN (and ENN) in the error function used for training: the cross-entropy of the normalized contour function is used rather than the mean squared error of the pignistic transform. This error function can be computed in linear time as a function of the number of classes.

The objective of this work is to explore similar modifications to the original ENN algorithm, specifically integrating contextual discounting and using the cross-entropy of the normalized contour function for training. The resulting model, referred hereafter as *Contextual Discounting Evidential Neural Network* (CD-ENN), is more flexible than ENN, yet its training computational complexity is identical.

The remainder of this paper is organized as follows. Necessary notions of the theory of belief functions are provided in Section II. In Section III, the ENN algorithm is first recalled and then its extension CD-ENN is proposed. The interest of CD-ENN is shown experimentally in Section IV. Finally, Section V concludes the paper.

## II. BACKGROUND

This section provides an overview of the key concepts necessary for the remainder of the paper. Further details can be found, for example, in [13], [7]. The fundamental principles of Dempster-Shafer (DS) theory are introduced in Section II-A, followed by a presentation of the classical and contextual discounting operations in Section II-B.

### A. Basic Concepts

When a question  $Q$  has an answer belonging to a finite set  $\Omega = \{\omega_1, \dots, \omega_c\}$ , the uncertainty regarding this answer can

Serigne Diène's PhD work is funded by the Hauts-de-France region and by Artois University.

be modeled using a *mass function* (MF)  $m$ , defined from the power set of  $\Omega$ , denoted  $2^\Omega$ , to  $[0, 1]$  s.t.  $\sum_{A \subseteq \Omega} m(A) = 1$ . The value  $m(A)$  represents the amount of belief allocated to the hypothesis that the correct answer lies within the subset  $A$  of  $\Omega$ . Any subset  $A \subseteq \Omega$  for which  $m(A) > 0$  is referred to as a *focal element* of  $m$ . A *Bayesian MF* is a MF having only singletons as focal elements (it corresponds then to the probability distribution  $p(\omega) = m(\{\omega\})$  for all  $\omega \in \Omega$ ).

A MF  $m$  is in one-to-one correspondence with a *belief function*  $Bel$  and a *plausibility function*  $Pl$ , respectively defined for any  $A \subseteq \Omega$  by  $Bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B)$  and  $Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$ .

The *contour function*  $pl$  is the restriction of  $Pl$  to the singletons of  $\Omega$ . That is, for each  $\omega \in \Omega$ , we have:  $pl(\omega) = Pl(\{\omega\})$ .

A MF can be approximated by a probability distribution, using the plausibility transform [1], denoted hereafter by  $plp$  and which consists in normalizing the contour function:

$$plp(\omega) = \frac{pl(\omega)}{\sum_{\omega' \in \Omega} pl(\omega')}, \quad \forall \omega \in \Omega. \quad (1)$$

When two independent and reliable MFs  $m_1$  and  $m_2$  are defined on the same set  $\Omega$ , they can be combined using *Dempster's rule* which gives the combined MF denoted  $m_{1 \oplus 2} = m_1 \oplus m_2$  and defined for all non-empty subsets  $A \subseteq \Omega$ , as follows:

$$(m_1 \oplus m_2)(A) = \frac{1}{1 - \kappa} \sum_{B \cap C = A} m_1(B)m_2(C), \quad (2)$$

where the conflict coefficient  $\kappa$  is given by  $\kappa = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$ , and the mass assigned to the empty set is defined as  $m_{1 \oplus 2}(\emptyset) = 0$ . Dempster's rule is commutative and associative. Furthermore, it satisfies

$$plp_{1 \oplus 2}(\omega) = \frac{pl_1(\omega)pl_2(\omega)}{\sum_{\omega' \in \Omega} pl_1(\omega')pl_2(\omega')}, \quad \forall \omega \in \Omega, \quad (3)$$

which means that the probabilistic approximation of the combination by Dempster's rule of two MFs can be computed directly from their contour functions.

### B. Discounting

To account for the reliability of a MF, Shafer [13] introduced the concept of discounting, where a MF  $m$  is corrected given a reliability degree  $\beta \in [0, 1]$ , into a discounted MF denoted by  ${}^\beta m$  and defined by:

$${}^\beta m = \beta m + (1 - \beta) m_\Omega, \quad (4)$$

where  $m_\Omega$  is the vacuous MF s.t.  $m_\Omega(\Omega) = 1$ , and  $1 - \beta$  is called the discount rate.

The contour function associated with  ${}^\beta m$  is given by [7]:

$${}^\beta pl(\omega) = 1 - \beta + pl(\omega)\beta, \quad \forall \omega \in \Omega. \quad (5)$$

This classical discounting operation inspired the development of the *contextual discounting* (CD) defined with a vector of reliability degrees  $\beta = (\beta_1, \dots, \beta_c)$ , where  $\beta_q \in [0, 1]$  for  $q = 1, \dots, c$ , is the reliability degree given that  $\omega_q \in \Omega$  is the

true state. Its formal derivation [11] results into a contextually discounted MF denoted by  ${}^\beta m$  and defined as, for all  $A \subseteq \Omega$ :

$${}^\beta m(A) = \sum_{B \subseteq A} m(B) \prod_{\omega_q \in A \setminus B} (1 - \beta_q) \prod_{\omega_q \in \bar{A}} \beta_q, \quad (6)$$

and its associated contour function is given by:

$${}^\beta pl(\omega_q) = 1 - \beta_q + pl(\omega_q)\beta_q, \quad q = 1, \dots, c. \quad (7)$$

Equation (5) is retrieved from (7) when the vector of reliability degrees  $\beta$  contains only one value: a reliability degree  $\beta$ .

The discounting and contextual discounting operations are at play, as mentioned in the Introduction, in E(K)NN and CD-EKNN, respectively. They have been used also in some other classification schemes, e.g., to correct a classifier based on its confusion matrix [8].

## III. CONTEXTUAL-DISCOUNTING EVIDENTIAL NEURAL NETWORK CLASSIFIER

In this section, we first recall the ENN classifier before introducing the proposed CD-ENN classifier based on the contextual discounting.

### A. Evidential Neural Network classifier

The *Evidential Neural Network* (ENN) classifier [4] is a prototype-based classifier. It is defined using  $\ell$  prototypes, where each prototype  $i$ , for  $i = 1, \dots, \ell$ , is characterized by a vector  $\mathbf{p}^{(i)} \in \mathbb{R}^p$  and is equipped with two parameters  $\alpha^{(i)} \in (0, 1)$  and  $\gamma^{(i)} > 0$ , together with a class proportion vector  $\mathbf{u}^{(i)} = (u_1^{(i)}, \dots, u_c^{(i)})$ , such that  $u_q^{(i)} \in [0, 1]$  and  $\sum_{q=1}^c u_q^{(i)} = 1$ . We denote hereafter by  $\mathbf{p} = (\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(\ell)})$ , the  $(p \times \ell)$ -dimensional matrix of prototype vectors and by  $\mathbf{u} = (\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(\ell)})$ , the  $(c \times \ell)$ -dimensional matrix of prototype class proportions. In addition let  $\boldsymbol{\alpha} = (\alpha^{(1)}, \dots, \alpha^{(\ell)})$  and  $\boldsymbol{\gamma} = (\gamma^{(1)}, \dots, \gamma^{(\ell)})$  be the vectors storing the parameters  $\alpha^{(i)}$  and  $\gamma^{(i)}$  associated to all the prototypes.

For a given instance  $\mathbf{x} \in \mathbb{R}^p$  to be classified, the prediction of ENN relies on evaluating the similarity  $s^{(i)}$  between  $\mathbf{x}$  and each prototype  $i$  defined by:

$$s^{(i)} = \alpha^{(i)} \exp(-\gamma^{(i)} \|\mathbf{x} - \mathbf{p}^{(i)}\|^2). \quad (8)$$

Each similarity  $s^{(i)}$  is interpreted as the reliability degree<sup>1</sup> of the Bayesian MF  $m^{(i)}$  defined by  $m^{(i)}(\{\omega_q\}) = u_q^{(i)}$ , for  $q = 1, \dots, c$ . In other words, the class proportion information  $\mathbf{u}^{(i)}$ , associated with prototype  $i$ , is discounted according to  $s^{(i)}$ , which depends on the distance between the instance to be classified  $\mathbf{x}$  and the prototype vector  $\mathbf{p}^{(i)}$ . Using (4), each prototype  $i$  induces then a MF  $\hat{m}^{(i)}$ , defined as:

$$\begin{aligned} \hat{m}^{(i)} &= (\hat{m}^{(i)}(\{\omega_1\}), \dots, \hat{m}^{(i)}(\{\omega_c\}), \hat{m}^{(i)}(\Omega)) \\ &= (u_1^{(i)} s^{(i)}, \dots, u_c^{(i)} s^{(i)}, 1 - s^{(i)}). \end{aligned} \quad (9)$$

<sup>1</sup> $s^{(i)}$  is formally identical to the activation of the  $i$ -th unit of the hidden layer of a radial basis function network with exponential activation function.

Finally, the output MF  $\hat{m}$  of ENN about the class of instance  $\mathbf{x}$  is obtained by combining MFs  $\hat{m}^{(i)}$ , over all the prototypes, using Dempster's rule of combination (2). It is given by:

$$\hat{m} = \oplus_{i=1}^{\ell} \hat{m}^{(i)}. \quad (10)$$

Let  $\mathcal{D} = \{(\mathbf{x}^{[t]}, y^{[t]})\}_{t=1}^n$ , be some training instances with feature values  $\mathbf{x}^{[t]} \in \mathbb{R}^p$  and class labels  $y^{[t]} \in \Omega$ . The evidential classifier ENN can be trained in several ways. In [4], the mean squared error as well as the cross-entropy of the so-called pignistic transform were considered; these error criteria can be computed in linear time as a function of the number of classes. In this paper, following [7], we use as loss function, the cross-entropy of the plausibility transform, given by

$$\ell_{\mathcal{D}}(\mathbf{p}, \mathbf{u}, \gamma, \alpha) = - \sum_{t=1}^n \ln(plp^{[t]}(y^{[t]})), \quad (11)$$

where  $plp^{[t]}$  denotes the plausibility transform (1) of the output MF  $\hat{m}^{[t]}$  (10) computed by ENN for the training instance  $\mathbf{x}^{[t]}$  of  $\mathcal{D}$ . We remark that, thanks to the property (3), we have

$$plp^{[t]}(\omega) = \frac{\prod_{i=1}^{\ell} \hat{pl}^{[t],(i)}(\omega)}{\sum_{\omega' \in \Omega} \prod_{i=1}^{\ell} \hat{pl}^{[t],(i)}(\omega')}, \quad \forall \omega \in \Omega, \quad (12)$$

where  $\hat{pl}^{[t],(i)}(\omega_q) = 1 - s_q^{[t],(i)} + u_q^{(i)} s_q^{[t],(i)}$ ,  $q = 1, \dots, c$ , is the contour function of MF  $\hat{m}^{[t],(i)}$  (9) obtained from prototype  $i$  for instance  $\mathbf{x}^{[t]}$ . Therefore, computing the error function (11) is also linear in the number of classes.

### B. Contextual-Discounting ENN classifier

In the formulation of the ENN classifier recalled previously, the MF  $\hat{m}^{(i)}$ , induced by prototype  $i$  for a given instance  $\mathbf{x}$ , is based on the use of the classical discounting. This amounts to assuming that the prototype  $i$  has reliability  $s^{(i)}$ , when it is only known that the actual class of the instance  $\mathbf{x}$  is in  $\Omega$  [11]. Inspired by the contextual discounting-based extension of EKNN called CD-EKNN [7], and the fact that ENN is itself an (prototype-based) extension of EKNN, we propose to replace the classical discounting in ENN by the contextual discounting. Specifically, we introduce the reliability degree  $s_q^{(i)}$  that the prototype  $i$  provides reliable information about instance  $\mathbf{x}$ , given that the true class of the instance is  $\omega_q$ . We define  $s_q^{(i)}$ , for all  $i = 1, \dots, \ell$  and all  $q = 1, \dots, c$ , as

$$s_q^{(i)} = \alpha^{(i)} \exp(-\gamma_q^{(i)} \|\mathbf{x} - \mathbf{p}^{(i)}\|^2), \quad (13)$$

with  $\gamma_q^{(i)} > 0$ . Accordingly, from (6), the evidence from prototype  $i$ , given its class information represented by Bayesian MF  $m^{(i)}$  s.t.  $m^{(i)}(\{\omega_q\}) = u_q^{(i)}$ , for all  $q = 1, \dots, c$ , becomes:

$$\hat{m}^{(i)}(A) = \sum_{\omega_k \in A} m^{(i)}(\{\omega_k\}) \left( \prod_{\omega_q \in A \setminus \{\omega_k\}} (1 - s_q^{(i)}) \prod_{\omega_\ell \in \bar{A}} s_\ell^{(i)} \right) \quad (14)$$

The output MF  $\hat{m}$  of CD-ENN about the class of instance  $\mathbf{x}$  is then obtained in a similar way as that of ENN, i.e., by

combining by Dempster's rule the MFs  $\hat{m}^{(i)}$  defined by (14) over all the prototypes.

The CD-ENN model shares then common parameters with the ENN model: the prototype vectors  $\mathbf{p}$ , the class proportion vectors  $\mathbf{u}$ , and the alpha parameters  $\alpha$ . However, they differ in the  $\gamma$  parameters: ENN has  $\ell$  parameters  $\gamma^{(i)}$ ,  $i = 1, \dots, \ell$  (i.e., one parameter per prototype), whereas CD-ENN has  $c \times \ell$  parameters  $\gamma_q^{(i)}$ ,  $i = 1, \dots, \ell$ ,  $q = 1, \dots, c$  (i.e., one parameter per prototype and class).

Another important difference with ENN is that CD-ENN can produce an output MF that has as focal sets all possible subsets of  $\Omega$ , whereas ENN only produces an output MF whose focal sets are the singletons of  $\Omega$  and  $\Omega$  itself. Nonetheless, if the cross-entropy of the plausibility transform is used as error criterion, then training CD-ENN has the same complexity as training ENN, due to the contour function  $\hat{pl}^{[t],(i)}$  of MF  $\hat{m}^{[t],(i)}$  (14) obtained from prototype  $i$  for instance  $\mathbf{x}^{[t]}$ , satisfying  $\hat{pl}^{[t],(i)}(\omega_q) = 1 - s_q^{[t],(i)} + u_q^{(i)} s_q^{[t],(i)}$ ,  $q = 1, \dots, c$ .

### C. Illustrative Example

To illustrate the difference between ENN and CD-ENN, let us borrow the 3-class synthetic dataset used in [7, Section 3.2] to demonstrate CD-EKNN. Figures 1 and 2 show the contour plots of output MFs computed respectively by the ENN and the CD-ENN models<sup>2</sup> for this dataset (to optimize space, the output mass given to  $\Omega$  by the two models is provided in Figure 3). It can be seen that CD-ENN allocates masses to all subsets of  $\Omega$ , whereas ENN only allocates masses to singletons and to  $\Omega$ . Importantly, the masses allocated by CD-ENN to the non-singleton subsets seem relevant. (Additionally, by comparing Figure 2 to [7, Figure 1], it may be remarked that the output MF of CD-ENN looks like a smooth version of that of CD-EKNN, similarly as the output MF of ENN appears to be a smooth version of that of EKNN, which is seen by comparing Figure 1 to [7, Figure 2].)

## IV. NUMERICAL EXPERIMENTS

In this section, the ENN and CD-ENN models are compared experimentally using 10 real-world classification datasets provided by the UCI [10] and summarized in Table I<sup>3</sup>. The experimental protocol we followed is presented in Section IV-A and the obtained results are provided in Section IV-B.

### A. Experimental protocol

For both models, and for a given dataset with  $c$  classes, we considered  $\ell = c$  prototypes, i.e., as many prototypes as there are classes. Furthermore, each dataset was split (using stratified random sampling) into training and test sets, with respectively 80% and 20% of the instances. The prototype vectors  $\mathbf{p}$  were initialized as the centroids of the  $k$ -means clusters (with, then,  $k = \ell = c$ ) of the training instances and their class membership proportion vectors  $\mathbf{u}$  were initialized as the class

<sup>2</sup>The parameters of the two models were determined using the procedure described in the first paragraph of Section IV-A.

<sup>3</sup>For the Ecoli dataset, we only used quantitative attributes 2, 3, 6, 7, and the four most frequent classes 'im', 'pp', 'imU' and 'cp'.

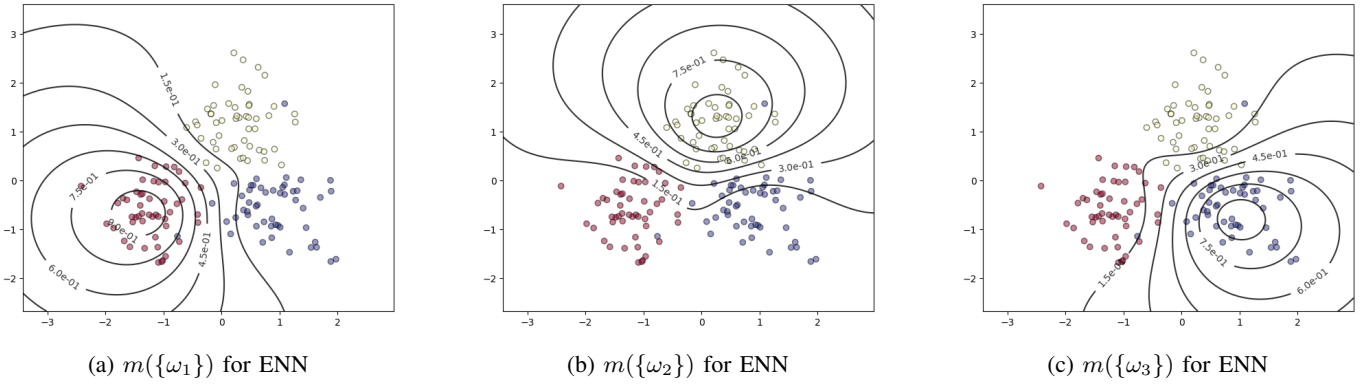


Fig. 1: Contour plots of the output MF obtained by ENN on the 3-class dataset from [7, Section 3.2]. Data from class 1 ( $\omega_1$ ) appear in red, those from class 2 ( $\omega_2$ ) in yellow, and those from class 3 ( $\omega_3$ ) in blue.

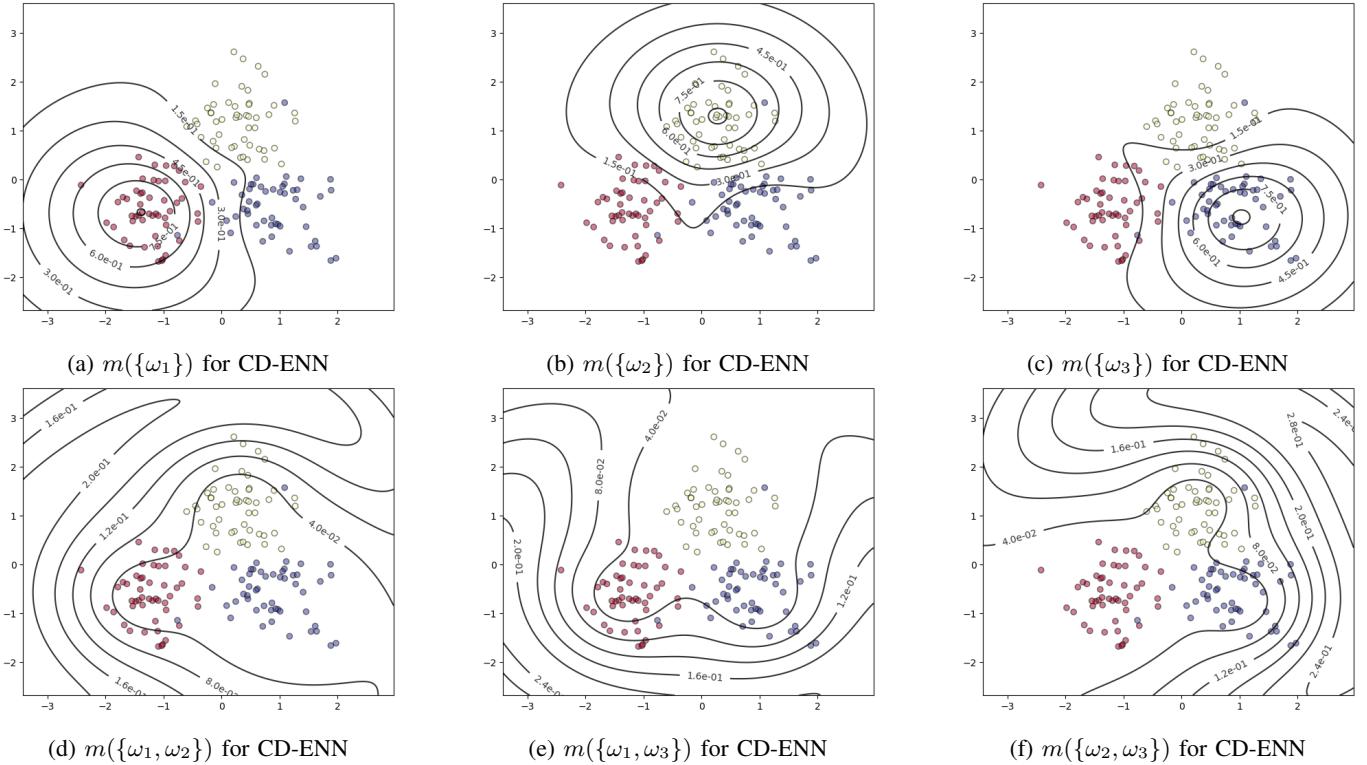


Fig. 2: Contour plots of the output MF obtained by CD-ENN on the 3-class dataset from [7, Section 3.2].

TABLE I: UCI Datasets used

Datasets	# Classes	# Features	# Instances
Pima	2	8	768
Wine red	6	11	1599
Wine white	7	11	4898
Dry Bean	7	16	13611
Ecoli	4	4	307
Glass	6	9	214
Heart	2	8	462
Ionosphere	2	34	351
Vertebral	3	6	310
Sonar	2	60	208

proportions in the associated clusters. Parameters  $\alpha^{(i)}$ ,  $\gamma^{(i)}$  and  $\gamma_q^{(i)}$  were initialized as  $\alpha^{(i)} = 0.5$  and  $\gamma^{(i)} = \gamma_q^{(i)} = 0.1$ , for all  $i = 1, \dots, \ell$  and all  $q = 1, \dots, c$ . Then, all parameters (including  $\mathbf{p}$  and  $\mathbf{u}$ ) were learned by minimizing the error function (11) over the learning set, using the ADAM optimizer with a learning rate set to 0.01, a validation split of 20% and an early stopping criterion with a patience of 5.

To evaluate and compare the performance on the test set of a given dataset, of the learned ENN and CD-ENN models, we need a performance measure, which takes into account the fact that they are evidential classifiers, i.e., classifiers outputting a MF for a given instance, which is a richer piece

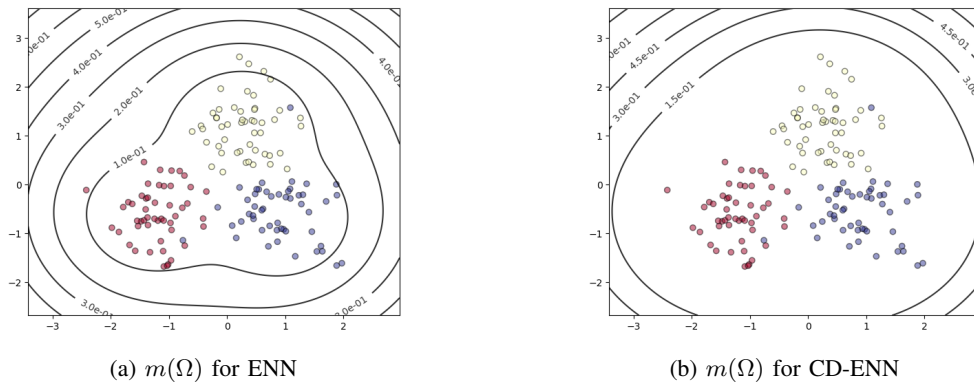


Fig. 3: Output mass for  $\Omega$  obtained by ENN and CD-ENN on the 3-class dataset from [7, Section 3.2].

of information than a probability measure. By considering an instance  $\mathbf{x}$  whose true class is  $y \in \Omega$  and for which we have a prediction regarding its class in the form of a MF  $m$ , the quality of this prediction can be evaluated appropriately using the *Generalized Negative Log-Likelihood* (GNLL), proposed originally by Denœux [5] for regression and adapted recently to classification in [12]. It is defined as:

$$\mathcal{L}(y, m) = -\frac{1}{2} \ln Bel(\{y\}) - \frac{1}{2} \ln Pl(\{y\}), \quad (15)$$

where  $Bel$  and  $Pl$  are the belief and plausibility functions, respectively, associated to  $m$ . The loss (15) is minimized when the degrees of belief and of plausibility assigned to the true class are high. Moreover, if  $m$  is Bayesian, i.e., it corresponds to a probability distribution  $p$ , we have  $Bel(\{y\}) = Pl(\{y\}) = p(y)$  and the GNLL (15) reduces to the standard cross-entropy, i.e.,  $\mathcal{L}(y, m) = -\ln p(y)$ . The performance of ENN and CD-ENN on the test set of a given dataset was measured using the average GNLL over this test set.

The entire procedure described in this section was repeated 20 times for each dataset.

### B. Results

For each dataset and for both the ENN and CD-ENN models, we computed the mean and standard deviation (std), over the 20 repetitions, of their average test GNLL; these are reported in Table II.

TABLE II: Mean ( $\pm$ std) test GNLL for ENN and CD-ENN

Datasets	ENN( $\downarrow$ )	CD-ENN( $\downarrow$ )
Pima	$3.06 \pm 0.47$	<b>0.593</b> $\pm 0.13$
Wine red	$1.04 \pm 0.03$	<b>1.005</b> $\pm 0.02$
Wine white	$1.14 \pm 0.01$	<b>1.12</b> $\pm 0.01$
Dry Bean	$0.64 \pm 0.03$	<b>0.63</b> $\pm 0.02$
Ecoli	<b>0.57</b> $\pm 0.08$	$0.66 \pm 0.06$
Glass	$1.41 \pm 0.17$	<b>1.27</b> $\pm 0.06$
Heart	$1.01 \pm 0.32$	<b>0.66</b> $\pm 0.04$
Ionosphere	$3.1 \pm 0.52$	<b>2.97</b> $\pm 0.99$
Vertebral	<b>0.68</b> $\pm 0.06$	$0.73 \pm 0.07$
Sonar	$1.06 \pm 0.38$	<b>0.85</b> $\pm 0.07$

We can see that CD-ENN outperforms ENN in terms of the mean test GNLL on a majority of the 10 considered UCI

datasets. Moreover, the differences are statistically significant at level 0.05, according to paired  $t$ -tests, except for the *Dry Bean* and *Ionosphere* datasets.

### V. CONCLUSION

In this paper, we have shown how contextual discounting can be integrated into the ENN classifier as a replacement for the discounting operation, resulting in a more expressive evidential classifier called CD-ENN. Comparisons on real UCI data have shown that CD-ENN can lead to better predictive performance, as measured by the generalized negative log-likelihood. In addition, CD-ENN maintains the same training computational complexity as ENN, making it a promising potential alternative. Future developments will focus on testing with other datasets, studying the impact of the number of prototypes, and using CD-ENN as a replacement for the softmax layer in a deep neural network.

### REFERENCES

- [1] B. R. Cobb and P. P. Shenoy, "On the plausibility transformation method for translating belief function models to probability models," *Int. J. Approx. Reason.*, vol. 41, no. 3, pp. 314–330, 2006.
- [2] A. P. Dempster, "Upper and lower probabilities induced by a multiple valued mapping," *Ann. Math. Stat.*, vol. 38, pp. 325–339, 1967.
- [3] T. Denœux, "A k-nearest neighbor classification rule based on dempster-shafer theory," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 5, pp. 804–813, 1995.
- [4] —, "A neural network classifier based on dempster-shafer theory," *IEEE Trans. Syst., Man, Cybern. A*, vol. 30, no. 2, pp. 131–150, 2000.
- [5] —, "Quantifying prediction uncertainty in regression using random fuzzy sets: the ennreg model," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 10, pp. 3690–3699, 2023.
- [6] —, "Uncertainty quantification in logistic regression using random fuzzy sets and belief functions," *Int. J. of Approx. Reason.*, vol. 168, p. 109159, 2024.
- [7] T. Denœux, O. Kanjanatarakul, and S. Sriboonchitta, "A new evidential K-nearest neighbor rule based on contextual discounting with partially supervised learning," *Int. J. of Approx. Reason.*, vol. 113, pp. 287–302, 2019.
- [8] Z. Elouedi, E. Lefevre, and D. Mercier, "Discountings of a belief function using a confusion matrix," in *22nd IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, 2010, pp. 287–294.
- [9] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods," *Mach. Learn.*, vol. 110, pp. 457–506, 2021.
- [10] M. Kelly, R. Longjohn, and K. Nottingham, "The uci machine learning repository," <https://archive.ics.uci.edu>.

- [11] D. Mercier, B. Quost, and T. Denœux, "Refined modeling of sensor reliability in the belief function framework using contextual discounting," *Inf. Fusion*, vol. 9, no. 2, pp. 246–258, 2008.
- [12] F. Pichon, S. Diène, T. Denœux, S. Ramel, and D. Mercier, "r-ERBFN: an extension of the evidential rbf accounting for the dependence between positive and negative evidence," in *Scalable Uncertainty Management*, ser. LNAI, S. Destercke, M. V. Martinez, and G. Sanfilippo, Eds., vol. 15350. Palermo, Italy: Springer, 2024.
- [13] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, N.J.: Princeton University Press, 1976.
- [14] Z. Tong, P. Xu, and T. Denœux, "An evidential classifier based on Dempster-Shafer theory and deep learning," *Neurocomputing*, vol. 450, pp. 275–293, 2021.